Steering Machine Learning Ecosystems of Interacting Agents

Meena Jagadeesan

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2025-62
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-62.html

May 14, 2025

Steering Machine Learning Ecosystems of Interacting Agents

by

Meena Jagadeesan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael I. Jordan, Co-chair
Assistant Professor Jacob Steinhardt, Co-chair
Assistant Professor Nika Haghtalab
Associate Professor Anca D. Dragan
Professor Federico Echenique

Spring 2025

Steering Machine Learning Ecosystems of Interacting Agents

Abstract

Steering Machine Learning Ecosystems of Interacting Agents

by

Meena Jagadeesan

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Michael I. Jordan, Co-chair

Assistant Professor Jacob Steinhardt, Co-chair

When machine learning models such as large language models (LLMs) and recommender systems are deployed into human-facing applications, these models interact with humans, companies, and other models within a broader ecosystem. However, the resulting multi-agent interactions often induce unintended ecosystem-level outcomes, including clickbait in classical content recommendation ecosystems, and more recently, safety violations and market concentration in nascent LLM ecosystems. The core issue is that ML models are classically analyzed as a single agent operating in isolation, so standard evaluation approaches in machine learning fail to capture ecosystem-level outcomes at the society-level, market-level, and algorithm-level.

This thesis investigates how to characterize and steer ecosystem-level outcomes, focusing on LLM ecosystems and content recommendation ecosystems. To tackle this, we augment the typical algorithmic perspective on machine learning with an economic and statistical perspective. The key idea is to trace ecosystem-level outcomes back to the incentives of interacting agents (i.e., ML models, humans, and companies) and back to the ML pipeline for training models.

In the first part, we investigate how competition between model-providers influences ecosystem-level performance trends and market outcomes. We demonstrate that scaling trends are fundamentally altered, and we develop technical tools to evaluate proposed AI policy. In the second part, we investigate how ML models deployed in content recommendation ecosystems influence content creation. We characterize how recommendation models shape the content supply via creator incentives, and how generative models shape which types of users produce content. In the third part, we investigate repeated interactions between a human and a ML model. We develop evaluation metrics which account for competing preferences, and design near-optimal incentive-aware algorithms.

More broadly, this thesis takes a step towards a vision of machine learning ecosystems where the interactions between ML models, humans, and companies are steered towards the desired ecosystem-level outcomes.

To my family and friends, for their love, support, and laughter.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I'd like to thank my advisors Mike Jordan and Jacob Steinhardt. Both of them have invested so much in me and have always encouraged me to explore my own interests. I really admire how Mike finds deep connections between different fields and builds new bridges between them. I've learned so much from him about how to approach interdisciplinary research, and he's inspired me to always keep broadening my knowledge of mathematical tools. Mike has also been so supportive as an advisor, and I've left every meeting with him feeling very excited about my work. I really admire how Jacob identifies deeply important problems in the world that very few other people are even aware of, and how he develops techniques to solve those problems from the ground up. As an advisor, Jacob has patiently helped me build research skills like paper writing and brainstorming research projects, even when it would have been so much faster for him to just do these tasks himself.

I've also been incredibly fortunate to be mentored by several other faculty members at Berkeley. I'd like to thank Nika Haghtalab for being on my committee and for her generous mentorship throughout my PhD. Over the course of several joint projects, I've learned so much from Nika about research in learning in games and how to navigate an interdisciplinary research career. Furthermore, I'd like to thank Anca Dragan and Federico Echenique for being on my committee and for their valuable feedback during my quals. I'm additionally grateful to Anca for supporting a recommender systems meeting that I co-organized, and to Federico for helping connect me with the econ community. Finally, I'd like to thank Moritz Hardt, who was at Berkeley during the initial stretch of my PhD. I was very fortunate to be mentored by Moritz on some of my early projects, and I learned so much from him about how to distill clean conceptual insights.

During my PhD, I spent two wonderful and very productive summers at Microsoft Research New England in Boston, where I was mentored by Nicole Immorlica and Brendan Lucier. I collaborated with Nicole and Brendan on several projects over the course of those two summers. I'd like to thank Nicole and Brendan for teaching me so much about how to write a good applied modelling paper and about research in EconCS more generally. They've also been very generous about supporting my career, including by setting up collaborations with other senior researchers in the EconCS community (Nageeb Ali and Alex Slivkins) and other MSR interns (Kate Donahue).

The work done during my PhD was joint with a number of fantastic collaborators: Nageeb Ali, Eshwar Ram Arunachaleswaran, Natalie Collina, Jessica Dai, Sarah Dean, Evan Dong, Kate Donahue, Bailey Flanigan, Nikhil Garg, Nika Haghtalab, Moritz Hardt, Xinyan Hu, Nicole Immorlica, Mike Jordan, Erik Jones, Leqi Liu, Brendan Lucier, Celestine Mendler-Dünner, Alex Pan, Chara Podimata, Alex Slivkins, Jacob Steinhardt, Yixin Wang, Alex Wei, and Tijana Zrnic. They're all amazing researchers, and I've learned so much from all of them. I'd especially like to thank the more senior researchers for mentoring me at different stages in my PhD, and the more junior researchers for being patient with me as I learned how to be an effective mentor.

I'd like to thank the incredible staff at Berkeley—including Louise Verkin, Naomi Yamasaki, Susanne Kauer, Jean Nguyen, Carissa Cloud, Angie Abbatecola, Ami Katagiri, and Roxana Infante. A special thanks to Louise and Naomi for helping me navigate complicated logistical tasks throughout my PhD, and to all of the BAIR staff for making BWW an incredible place to do research.

I've really enjoyed being part of Mike's group and Jacob's group. I'm very grateful for all of the fun conversations in the office and at group socials, and I've learned so much from everyone. Thanks to Jacob's group—including Collin Burns, Frances Ding, Jiahai Feng, Dan Hendrycks, Xinyan Hu, Erik Jones, Gabe Mukobi, Alex Pan, Adam Sealfon, Alex Wei, Kayo Yin, and Ruiqi Zhong—for getting me to start thinking about language models really early on in my PhD. Thanks to Mike's group—including Nivasini Ananthakrishnan, Liviu Aolaritei, Stephen Bates, Linda Cai, Tatjana Chavdarova, Tiffany Ding, Alireza Fallah, Clara Fannjiang-Wong, Paula Gradu, Wenshuo Guo, Xinyan Hu, Jivat Kaur, Kirthi Kandasamy, Praneeth Karimireddy, Karl Krauth, Tianyi Lin, Eric Mazumdar, Drew Nguyen, Ezinne Nwankwo, Reese Pathak, Esther Rolf, Nilesh Tripuraneni, Annie Ulichney, Francisca Vasconcelos, Ian Waudby-Smith, Serena Wang, Yixin Wang, Alex Wei, Xuelin Yang, Yaodong Yu, Eric Zhao, Lydia Zakynthinou, Manolis Zampetakis, and Tijana Zrnic—for exposing me to a broad set of topics across statistics and machine learning theory.

Beyond Mike's group and Jacob's group, I benefited from interacting with the broader machine learning theory community at Berkeley—including Kush Bhatia, Micah Carroll, Irene Chen, Lijie Chen, Mihaela Curmei, Yuval Dagan, Jessica Dai, Sarah Dean, Nikhil Garg, Cassidy Laidlaw, Celestine Mendler-Dünner, John Miller, Chara Podimata, Abhishek Shetty, and Kunhe Yang. I'd especially like to thank Nikhil Garg for his mentorship and career advice throughout my PhD. Furthermore, I'd like to thank Celestine Mendler-Dünner for her mentorship during my initial projects during my PhD, and Kush Bhatia for being such a generous BAIR mentor during my first year. Finally, I'd like to thank Irene Chen and Lijie Chen for giving me detailed feedback on my job talk this semester.

Outside of Berkeley, I've enjoyed interacting with many amazing faculty members and students in the broader research community. I'd especially like to thank Jon Kleinberg for his encouragement during the later stretch of my PhD and for his generous support of my job market. I'd also thank everyone who gave me advice about how to navigate different stages of the job market.

My PhD research was generously supported by an Open Philanthropy AI Fellowship, a Paul and Daisy Soros Fellowship, and a Berkeley fellowship. These fellowships helped give me the flexibility to pursue my own research interests and to collaborate with a broad set of researchers across institutions.

Before coming to Berkeley, I spent four wonderful years at Harvard. I'd like to thank my undergrad advisor Jelani Nelson for investing so much in me, for taking a chance on me when I was very inexperienced, and for his generosity. I'm also grateful to other researchers—including Shuchi Chawla, Cynthia Dwork, Suriya Gunasekar, Stratos Idreos, Scott Kominers, and Ilya Razenshteyn—who mentored me when I was an undergrad and very patiently helped me build research skills.

I'd like to thank all of my incredible friends, who have brought so much happiness, laughter, and love into my life. To my friends in the Bay area: I'll always remember the Napa trips, the long walks through the Berkeley hills, and the bougie brunches in San Francisco. A special thank you to Katie and Neerja for being my first friends in grad school, and for their friendship through the PhD. To my friends outside the Bay area: even though we've lived on opposite coasts for the past five years, I'm very grateful that we've managed to see each often surprisingly often, between my two summers in Boston and so many trips to New York. A special thank you to Katherine H. and Katherine D. for so many years of friendship, for all the bougie dinners in New York and in San Francisco, and for helping me stay connected to the world outside of academia.

Most of all, I'd like to thank my family for everything that they've done for me. I would not have made it here without them. Thank you to my wonderful grandparents for always believing in me and instilling in me a love of learning. I'm deeply grateful to my parents for inspiring me to follow in their footsteps and become a computer scientist, my brother Ravi for making Exeter, then Cambridge, and now the Bay Area feel like home, and my fiancé Erik for his love and patience through all of the highs and lows. Thank you for supporting me through so much and for helping me chase my dreams.

Part I

Introduction

Chapter 1

Overview

Modern machine learning models assist us on cognitive tasks, curate the digital content that we consume, and automate physical tasks such as driving. When ML models are deployed into human-facing applications, these models interact with humans, companies, and other models within a broader ecosystem. For instance, a large language model (LLM) interacts with consumers who query the model, developers who fine-tune the model for specific downstream tasks, companies that train the model, companies that train competing models, and other LLM agents collaborating on a shared task. As another example, a content recommender system interacts with users who consume digital content, creators who produce the content, other competing platforms, and generative models which are increasingly used for content creation.

However, interactions between these agents (i.e., models, humans, and companies) often lead to *unintended ecosystem-level outcomes*. For example, content recommendation platforms such as YouTube routinely become saturated with content with misleading titles or flashy thumbnails (Meyerson, 2012). YouTube has battled *clickbait* for the past decade, and clickbait even continues to be a global issue for YouTube today (The YouTube Team, 2024). Taking a closer look, a key driver of clickbait is the interactions between the recommender system and content creators. Specifically, content creators are typically rewarded based on winning recommendations, and recommender systems typically optimize for engagement metrics. This incentivizes content creators to artificially "game" these engagement metrics by using misleading titles and flashy thumbnails.

Shifting our focus to nascent LLM ecosystems, the interactions between ML models, humans, and companies are starting to lead to unintended ecosystem-level outcomes in these ecosystems as well. For example, when Microsoft deployed an LLM-based assistant in 2023, a consumer managed to extract sensitive information about the LLM-based assistant and posted it on Twitter; in a later interaction, the LLM-based assistant retrieved that Twitter post and started threatening the consumer (Perrigo, 2023). This example illustrates how *safety violations* emerge from the interactions between LLM-based assistants, consumers, and the internet. As another example, policymakers have raised concerns about market-level outcomes such as *market concentration* where a small handful of LLM companies attract most of the consumers (Vipra and Korinek, 2023).

The fundamental issue is that standard evaluation approaches in machine learning fail to capture ecosystem-level outcomes, since machine learning models are classically analyzed as a single agent operating in isolation. For example, performance is classically measured in terms of the loss of a single model over a distribution; this perspective underpins evaluation approaches such as *benchmarks* as well as phenomena such as *scaling laws* (Kaplan et al., 2020). However, this perspective does not capture how consumers in LLM ecosystems choose between different fine-tuned models based on their individual preferences, and how recommendation models shape the incentives of content creators. Classical paradigms in machine learning—such as empirical risk minimization—further assume that each model-provider optimizes for the loss of its ML model over a distribution. However, this assumption is violated when multiple companies compete for consumers and thus strategically train their models to attract consumers away from other models.

1.1 Our contributions

In this thesis, we take an interdisciplinary approach on machine learning ecosystems, where we augment the typical algorithmic perspective on machine learning with an economic and statistical perspective. The key idea is to trace ecosystem-level outcomes back to the *incentives of interacting agents* and back to the *ML pipeline for training models*. Incentives emerge from each agent (i.e., ML model, human, or company) optimizing for their own objective, different agents having competing objectives, and one agent's behaviors influencing the behavior of other agents; economic models provide a useful toolkit to formalize these incentives. The ML pipeline for training models captures empirical details such as the data that the model is trained on, the metrics used to evaluate the model, and how the model is pretrained and fine-tuned; statistical frameworks provide a useful toolkit to formalize these empirical details. Bringing these toolkits together introduces rich technical challenges, such as analyzing incentives in data-driven environments, with high-dimensional actions, and with multi-objective learning.

We specifically adopt this interdisciplinary approach in order to *characterize* and *steer* ecosystem-level outcomes in LLM ecosystems and content recommendation ecosystems (Table 1.1). To characterize ecosystem-level outcomes, for nascent LLM ecosystems, we make forecasts about future ecosystem-level performance trends, and we provide mathematical and empirical support for these forecasts. For content recommendation systems, we trace empirically observed outcomes for the content supply back to interactions between ML models and content creators. To steer ecosystem-level outcomes towards societal objectives, we design and evaluate interventions for policymakers and companies. We develop technical tools to evaluate policy interventions, and we design incentive-aware learning algorithms which aim to help companies account for ecosystem-level impacts.

This thesis is organized by the type of interaction between agents.

	Characterize ecosystem-level outcomes	**Steer ecosystem-level outcomes**
Large language model ecosystems	*Ecosystem-level performance trends (Parts II, IV)*	*Technical tools to assess policy interventions (Part II)*
Recommendation ecosystems	*Impact of ML models on the content supply (Part III)*	*Design of algorithm-level interventions (Parts III, IV)*

Table 1.1: This thesis investigates how to characterize (left) and steer (right) ecosystem-level outcomes in ML ecosystems. The focus is on large language model ecosystems (top) and content recommendation ecosystems (bottom).

- **Part II: Model-Provider Competition.** This part investigates how competition between model-providers impacts ecosystem-level performance trends and market outcomes. Specifically, model-providers strategically train their models to attract consumers away from other models, to steer consumers towards monetizable behaviors, or to comply with regulation. We show that these competitive pressures can fundamentally alter scaling trends (Jagadeesan et al., 2023b; 2024). Moreover, we develop technical tools to predict the market outcomes of proposed AI policy targeting safety violations (Jagadeesan et al., 2024) and mandating data sharing (Jagadeesan et al., 2023c), and to quantify the market power of a recommendation platform (Hardt et al., 2022).

- **Part III: Incentives for Digital Content Creation.** This part investigates how ML models deployed in content recommendation ecosystems implicitly shape incentives for digital content creation. Specifically, recommendation models and generative models both shape how content creators are incentivized to design their content. We characterize how recommendation models shape the resulting content supply on the platform (Jagadeesan et al., 2023a; Immorlica et al., 2024), and how generative models shape whether digital content is produced by content creators or directly by consumers themselves (Ali et al., 2025).

- **Part IV: Repeated Human-AI Interactions.** This part investigates repeated interactions between a human and a ML model, focusing on the case of competing preferences. Specifically, the preferences of the model-provider—or the learned preferences of the ML model itself—are often misaligned with the preferences of humans who interact with the model. We develop evaluation metrics that account for these competing preferences. We then design incentive-aware algorithms that perform near-optimally against these evaluation metrics (Jagadeesan et al., 2023d; 2022; Donahue et al., 2024b).

Within each part, we also briefly discuss other joint works (Hu et al., 2023; Dai et al., 2024; Pan et al., 2024; Arunachaleswaran et al., 2025) which contribute to the themes of each part.

Part II

Model-Provider Competition

Chapter 2

Overview

Model-providers often compete with each other for consumer usage. For example, in LLM ecosystems, developers that fine-tune LLMs compete with each other to attract consumers, and companies that train LLMs also compete with each other. In content recommendation ecosystems, platforms compete with each other to attract users and generate profit. In both types of ecosystems, the nature of competition is further shaped by policymakers and regulators, who have raised concerns about market concentration and about safety violations, and have started to design policy-level interventions in response (Stigler Committee, 2019; European Union, 2022a; Vipra and Korinek, 2023; European Union, 2024).

The resulting competitive forces influence how model-providers design their models and thus also affect how ecosystem-level outcomes should be evaluated. Specifically, in LLM ecosystems, developers or companies may strategically train or fine-tune their models to attract consumers away from other model-providers or to comply with regulatory requirements. In recommendation ecosystems, a platform's power in the surrounding market affects how much the platform's ML model can steer consumers towards profitable patterns. In both cases, since competition shapes how model-providers design their models, this means that performance should be evaluated at an ecosystem-level, rather than at the level of individual model-providers. Moreover, since proposed AI policy shapes the incentives of model-providers, these policies should also be evaluated at an ecosystem-level.

2.1 Our contributions

This part investigates ecosystem-level performance trends and market outcomes under competing model-providers. We first focus on the impact on *scaling laws* (Kaplan et al., 2020), showing how competitive forces can fundamentally alter how increases to scale (i.e., data, compute, parameters) affect performance. We also develop technical tools to predict market outcomes when AI policy interacts with the ML pipeline for training models.

- In Chapter 3, we investigate scaling laws under competition between developers who fine-tune a pretrained model. We theoretically and empirically show that scaling laws can become non-monotone: that is, training the pretrained model with more resources can counterintuitively lead to worse ecosystem-level accuracy for consumers.

- In Chapter 4, we investigate when a new LLM company is able to enter the market with less data than incumbent LLM companies. We characterize how reputational damage from safety violations—which is shaped by regulation—affects these data-driven barriers to entry for new companies. En route, we derive data scaling laws in multi-objective environments, which illustrate how regulatory pressure can reduce the data efficiency of learning.
- In Chapter 5, we investigate how competition between recommendation platforms affects user utility at the ecosystem-level. We show that competition does not necessarily align market outcomes with user utility, contradicting typical economic intuition. Furthermore, under policy interventions which force platforms to share data with each other, this misalignment still persists.
- In Chapter 6, we illustrate a connection between a recommendation platform's power in the market and the platform's ability to steer users with its ML model. We then use this to define a measure of power that captures the maximum possible causal impact that any ML model deployed by the platform can have on users. This measure of power provides technical tools that could inform the ongoing debate about antitrust enforcement for recommendation platforms.

2.2 Methodological theme

In this part, a common methodological theme is to leverage ideas from a subfield of economics called *industrial organization*, but with an eye towards the details of the ML pipeline for training models.

In some works, we build on the theory of industrial organization, and view the ML model deployed by a company as its product. We generalize standard economic models of product selection (e.g., (Hotelling, 1929)) and price competition (e.g., Baye and Kovenock (2008)) to capture the rich space of fine-tuned models (Chapter 3) and the interdependence between data, user choices, and recommendation model performance (Chapter 5).

In other works, we build on conceptual ideas from industrial organization about market entry and market power. We develop quantitative formalizations which are tailored to data-driven, multi-objective ML pipelines (Chapter 4) and to the distribution shifts triggered by ML models (Chapter 6).

Chapter 3

Developers Fine-tuning a Pretrained Model

This chapter is based on *"Improved Bayes Risk Can Yield Reduced Social Welfare Under Competition"* (Jagadeesan et al., 2023b), which is joint work with Michael I. Jordan, Jacob Steinhardt, and Nika Haghtalab.

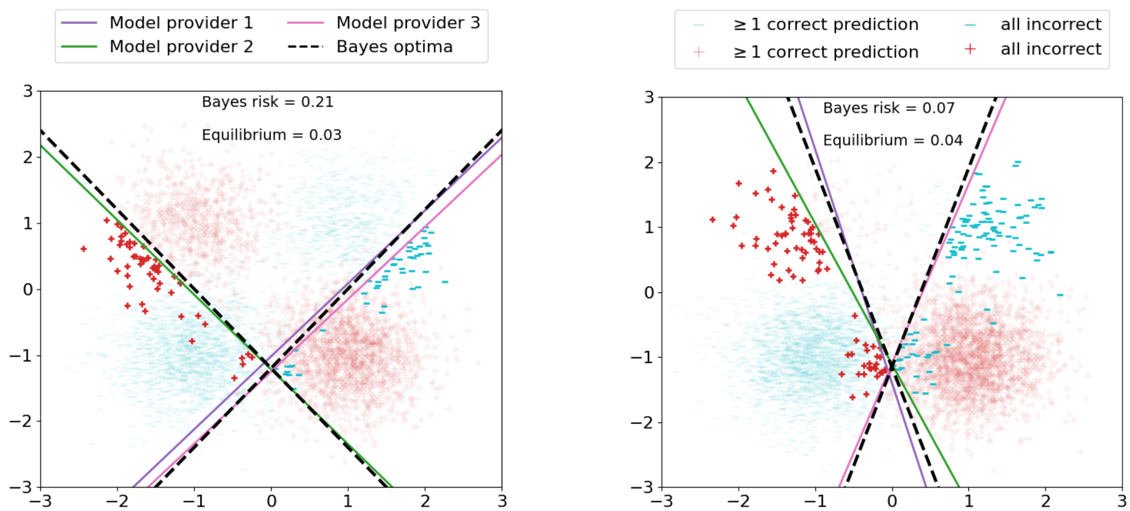
3.1 Introduction

Scaling trends in machine learning suggest that increasing the scale of a system consistently improves predictive accuracy. For example, scaling laws illustrate that increasing the number of model parameters (Kaplan et al., 2020; Sharma and Kaplan, 2020; Bahri et al., 2024) and amount of data (Hoffmann et al., 2022) can reliably improve model performance, leading to better representations and thus better predictions for downstream tasks (Hernandez et al., 2021).

However, these scaling laws typically take the perspective of a single model-provider in isolation, when in reality, model-providers often compete with each other for users. In emerging marketplaces built on a *foundation model* (Bommasani et al., 2021), different model-providers *fine-tune* or *prompt* the foundation model in different ways to attract users (Example 1). For example, in the recently released GPT store[1], model-developers create specialized versions of ChatGPT and compete for user usage.

A distinguishing feature of competing model-providers is that users can switch between model-providers and select a model-provider that offers them the highest predictive accuracy for their specific requests. This breaks the direct connection between predictive accuracy of a single model-provider in isolation and social welfare across competing model-providers, and raises the question: *what happens to scaling laws when model-providers compete with each other?*

[1]See `https://openai.com/blog/introducing-the-gpt-store`.



(a) (b)

Figure 3.1: Comparison of equilibrium loss on two data distributions, one with high Bayes risk (left) and one with lower Bayes risk (right). Each plot shows the linear predictors chosen at equilibrium under competition between three model-providers (solid lines), along with two approximately Bayes-optimal predictors (dashed lines). The equilibrium social loss is lower in the left plot than the right plot, even though the Bayes risk is much higher. The intuition is that approximate Bayes optima disagree on more data points in the left plot than in the right plot; thus, users have a greater likelihood of at least one predictor offering them a correct prediction, which increases the overall predictive accuracy for users (i.e., the social welfare).

In this chapter, we show that the typical intuition about scaling laws can fundamentally break down under competition. Surprisingly, even monotonicity can be violated: increasing scale can *decrease* the overall predictive accuracy (social welfare) for users. We study increases to scale through the lens of *data representations* (i.e., learned features), motivated by how increasing scale generally improves representation quality (Bengio et al., 2013). We exhibit several multi-class classification tasks where better data representations (as measured by Bayes risk) *decrease* the overall predictive accuracy (social welfare) for users, when varying data representations along several different axes.

From a conceptual perspective, the lens of data representations offers a clean formalization of emerging marketplaces built on a foundation model (Example 1). In such marketplaces, the foundation model is pretrained on a large amount of data, and different model-providers *fine-tune* the foundation model in different ways. We conceptualize pretraining as *learning data representations (e.g., features)* and fine-tuning as *learning a predictor from these representations*. In this formalization, increasing the *scale* of the foundation model improves the data representations accessible to model-providers during finetuning. We defer the details of the model to Chapter 3.2, and we make the connection to finetuning explicit in image classification experiments in Chapters 3.4.3-3.4.4.

The basic intuition for how the overall predictive accuracy can be non-monotonic in the data representation quality (i.e., Bayes risk) is illustrated in Figure 3.1. When data representations are low quality, any predictor will be incorrect on a large fraction of users, and near-optimal predictors may disagree on large subpopulations of users. Model providers are thus incentivized to choose complementary predictors that cater to different subpopulations (market segments), thus improving the overall predictive accuracy for users. In contrast, when representations are high quality, each optimal predictor is incorrect on only a small fraction of users, and near-optimal predictors likely agree with each other on most data points. As a result, model-providers are incentivized to select similar predictors, which decreases the overall predictive accuracy for users.

To study when non-monotonicity can occur, we first focus on a stylized setup that permits closed-form calculations of the social welfare at equilibrium (Chapter 3.3). Using this characterization, in three concrete binary classification setups, we show that the equilibrium social welfare can be non-monotonic in Bayes risk. In particular, we vary representations along three axes—the per-representation Bayes risks, the noise level of representations, and the dimension of the data representations—and exhibit non-monotonicity in each case (Figure 3.2).

Going beyond the stylized setup of Chapter 3.3, in Chapter 3.4 we consider linear function classes and demonstrate empirically that the social welfare can be non-monotonic in the data representation quality. We consider binary and 10-class image classification tasks on CIFAR-10 where data representations are obtained from the last-layer representations of AlexNet, VGG16, ResNet18, ResNet34, and ResNet50, pretrained on ImageNet. Better representations (as measured by Bayes risk) can again perform worse under competition (Figures 3.4 and 3.5). We also consider synthetic data where we can vary representation quality more systematically, again finding ubiquitous non-monotonicities.

Altogether, our results demonstrate that the classical setting of a single model-provider can be a poor proxy for understanding multiple competing model-providers. This suggest that caution is needed when inferring that increased social welfare necessarily follows from the continuing trend towards improvements in predictive accuracy in machine learning models. Machine learning researchers and regulators should evaluate methods in environments with competing model-providers in order to reasonably assess the implications of raw performance improvements for social welfare.

3.1.1 Related work

Our work connects to research threads on the *welfare implications of algorithmic decisions* and *competition between data-driven platforms*.

Welfare implications of algorithmic decisions. Recent work investigates *algorithmic monoculture* (Kleinberg and Raghavan, 2021; Bommasani et al., 2022), a setting in which multiple model-providers use the same predictor. In these works, monoculture is intrinsic to the decision-making pipeline: model-providers are given access to a shared algorithmic ranking (Kleinberg and Raghavan, 2021) or shared components in the training pipeline (Bommasani et al., 2022). In contrast, in our work, monoculture may arise endogenously from competition, as a result of scaling trends. Model-providers are always given access to the same function classes and data, but whether or not monoculture arises depends on the quality of data representations and its impact on the incentives of model-providers. Our work thus offers a new perspective on algorithmic monoculture, suggesting that it may arise naturally in competitive settings as a side effect of improvements in data representation quality.

More broadly, researchers have identified several sources of mismatch between predictive accuracy and downstream welfare metrics. This includes *narrowing* of a classifier under repeated interactions with users (Hashimoto et al., 2018), *preference shaping* of users induced by a recommendation algorithm (Carroll et al., 2022; Dean and Morgenstern, 2022; Curmei et al., 2022), *strategic adaptation* by users under a classifier (Brückner et al., 2012; Hardt et al., 2016), and the *long-term impact of algorithmic decisions* (Liu et al., 2018; 2020b).

Competition between data-driven platforms. Our work is also related to the literature on competing predictors. The model in our paper shares similarities with the work of Ben-Porat and Tennenholtz (2017; 2019), who studied equilibria between competing predictors. Ben-Porat and Tennenholtz (2017; 2019) show that empirical risk minimization is not an optimal strategy for a model-provider under competition and design algorithms that compute the best-responses; in contrast, our focus is on the equilibrium social welfare and how it changes with data representation quality. The specifics of our model also slightly differ from the specifics of Ben-Porat and Tennenholtz (2017; 2019). In their model, each user has an accuracy target that they wish to achieve and randomly chooses between model-providers that meet that accuracy target; in contrast, in our model, each user noisily chooses the model-provider that minimizes their loss and model-providers can have asymmetric market reputations.

Our work also relates to *bias-variance games* (Feng et al., 2022) between competing model-providers. However, Feng et al. (2022) focus on the the equilibrium strategies for the model-provider, but do not consider equilibrium social welfare for users; in contrast, our work focuses on the equilibrium social welfare. The model of Feng et al. (2022) also differs from the model in our work. In Feng et al. (2022), a model-provider action is modeled as choosing an error *distribution* for each user, where the randomness in the error is intended to capture randomness in the training data samples and in the predictor; moreover, the action set includes error distributions with a range of different variances. In contrast, in our population-level setup with deterministic predictors, the error distribution for every user is always a point mass (variance 0). Thus, the equilibrium characterization of Feng et al. (2022) does not translate to our model. The specifics of the model-provider utility in the work of Feng et al. (2022) differs slightly from our model as well.

Other aspects studied in this research thread include competition between model-providers using *out-of-box* learning algorithms that do not directly optimize for market share (Ginart et al., 2021; Kwon et al., 2022; Dean et al., 2024a), competition between model-providers selecting *regularization parameters* that tune model complexity (Iyer and Ke, 2022), competition between *bandit algorithms* where data directly comes from users (Aridor et al., 2025; Jagadeesan et al., 2023c), and competition between *algorithms dueling* for a user (Immorlica et al., 2011). Our work also relates to *classical economic models of product differentiation* such as Hotelling's model (Hotelling, 1929; d'Aspremont et al., 1979) (see Anderson et al. (1992) for a textbook treatment), as well as the emerging area of *platform competition* (see, e.g., Jullien and Sand-Zantman, 2021; Calvano and Polo, 2021a).

3.2 Model

We focus on a multi-class classification setup with input space $X \subseteq \mathbb{R}^d$ and output space $Y = \{0, 1, 2, \ldots, K-1\}$. Each user has an input x and a corresponding true output y, drawn from a distribution $\mathcal{D}$ over $X \times Y$. Model providers choose predictors f from some model family $\mathcal{F} \subseteq (\Delta(Y))^X$ where $\Delta(Y)$ is the set of distributions over Y. A user's loss given predictor f is $\ell(f(x), y) = \mathbb{P}[y \neq f(x)]$. (In Chapter 3.3, we take $\mathcal{F} = \{0, 1, 2, \ldots, K-1\}^X$ to be all deterministic functions mapping inputs to classes, while in Chapter 3.4 we consider linear predictors of the form $f(x) = \text{softmax}(Wx + b)$.)

We study competition between $m \geq 2$ model-providers for users, building on the model of Ben-Porat and Tennenholtz (2017; 2019). We index the model-providers by $[m] := \{1, 2, \ldots, m\}$, and let f_j denote the predictor chosen by model provider j. After the model-providers choose predictors $f_1, \ldots, f_m$, each user then chooses one of the m model-providers to link to, based on prediction accuracy. Model-providers aim to optimize the number of users that they win. (We note that this model is stylized and will make several simplifying assumptions; we defer a detailed discussion of the implications of these assumptions to Chapter 3.5.)

As an illustrative example, these model components can be mapped onto marketplaces where model-providers each fine-tune a shared foundation model as follows.

Example 1 (Model-providers finetuning a shared foundation model)**.** *Consider an emerging marketplace where different model-providers fine-tune or prompt a foundation model in different ways. (A foundation model is a large model such as GPT-4 that is pretrained on a large amount of data and can be adapted to large range of downstream tasks.*[2]*) A real-world example of such a marketplace is the recently released GPT store*[3] *where model-providers create specialized versions of chatGPT via prompting and compete for user usage. As another example, our experiments in Chapters 3.4.3-3.4.4 operate on a simulated image classification marketplace where model-providers fine-tune a large model (e.g., ResNet34) pretrained on ImageNet.*

[2] See Bommasani et al. (2021) for an introduction to the foundation model paradigm.
[3] See `https://openai.com/blog/introducing-the-gpt-store`.

As a simplified and stylized model for such marketplaces, the input x captures the final-layer representation learned by the foundation (pretrained) model. Let the function class $\mathcal{F}$ be linear functions. The model provider thus selects a linear function $f \in \mathcal{F}$ operating on the last-layer representations x, which captures finetuning while freezing all of the model parameters except for those in the last layer. Each model-provider optimizes for the number of users that they win, which captures that the finetuning objective is a proxy for market share. In this setup, increasing the scale of the pretrained model (e.g., by increasing the number of parameters or the amount of data) leads to improvements in data representations x accessible to model-providers during finetuning.

With this example in place, we now formalize user decisions, model-provider incentives, and the quality of the market outcome for users.

User decisions. Users noisily pick the model-provider offering the best predictions for them. That is, a user with representation x and true label y chooses a model-provider $j^*(x, y)$ such that the loss $\ell(f_{j^*(x,y)}(x), y)$ is the smallest across all model-providers $j \in [m]$, subject to noise in user decisions. More formally, we model user noise with the logit model (Train, 2009), also known as the Boltzmann rationality model:

$$\mathbb{P}[j^*(x, y) = j] = \frac{e^{-\ell(f_j(x),y)/c}}{\sum_{j'=1}^{m} e^{-\ell(f_{j'}(x),y)/c}}, \tag{3.1}$$

where $c > 0$ denotes a noise parameter. We extend this model to account for uneven market reputations across decisions in Chapter 3.3.5.

Model provider incentives. A model-provider's utility is captured by the market share that they win. That is, model-provider j's utility is

$$u(f_j; \mathbf{f}_{-j}) := \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}} \left[\mathbb{P}[j^*(x, y) = j]\right],$$

where $\mathbf{f}_{-j}$ denotes the predictors chosen by the other model-providers and where the expectation is over (x, y) drawn from $\mathcal{D}$. Since the market shares always sum to one, this is a constant-sum game.

Each model-provider chooses a *best response* to the predictors of other model-providers. That is, model-provider j chooses a predictor f_j^* such that

$$f_j^* \in \arg\max_{f_j \in \mathcal{F}} u(f_j; \mathbf{f}_{-j}).$$

The best-response captures that model-providers optimize for market share. In practice, model-providers may do so via A/B testing to steer towards predictors that maximize profit, or by actively collecting data on market segments where competitors are performing poorly.

We study market outcomes $\mathbf{f} = (f_1^*, f_2^*, \ldots, f_m^*)$ that form a Nash equilibrium. Recall that $(f_1^*, f_2^*, \ldots, f_m^*)$ is a *pure strategy Nash equilibrium* if for every $j \in [m]$, model-provider j's predictor is a best-response to $\mathbf{f}^*_{-j}$: that is, $f_j^* \in \arg\max_{f_j \in \mathcal{F}} u(f_j; \mathbf{f}^*_{-j})$. In well-behaved instances, pure-strategy equilibria exist (see theoretical results in Chapter 3.3 and simulation results in Chapter 3.4). However, for our results in Chapter 3.3.5, we must turn to mixed strategy equilibria where model-providers instead choose distributions μ_j over $\mathcal{F}$.

Quality of market outcome for users. We are interested in studying the quality of a market outcome $\mathbf{f} = (f_1, f_2, \ldots, f_m)$ in terms of user utility. The quality of $\mathbf{f}$ is determined by the overall *social loss* that it induces on the user population, after users choose between model-providers:

$$\texttt{SL}(f_1, \ldots f_m) := \mathbb{E}[\ell(f_{j^*(x,y)}(x), y)]. \tag{3.2}$$

When $f_1^*, \ldots, f_m^*$ is a Nash equilibrium, we refer to $\texttt{SL}(f_1^*, \ldots f_m^*)$ as the *equilibrium social loss.*

Our goal is to study how the equilibrium social loss changes when the representation quality (i.e., the quality of the input representations X) improves. We formalize representation quality as the minimum risk $\texttt{OPT}_{\text{single}}$ that a single model-provider could have achieved on the distribution $\mathcal{D}$ with the model family $\mathcal{F}$. This means that $\texttt{OPT}_{\text{single}}$ is equal to the Bayes risk:

$$\texttt{OPT}_{\text{single}} := \min_{f \in \mathcal{F}} \mathbb{E}\left[\ell(f(x), y)\right].$$

In the following sections, we show that the equilibrium social loss $\texttt{SL}(f_1^*, \ldots f_m^*)$ can be non-monotonic in the representation quality (as measured by $\texttt{OPT}_{\text{single}}$), when representations are varied along a variety of axes.

3.3 Non-monotonicity of Equilibrium Social Loss in a Stylized Setup

To understand when non-monotonicity can occur, we first consider a stylized setup (described below) that permits closed-form calculations of the social loss. We first show a simple mathematical example that illustrates non-monotonicity (Chapter 3.3.1). We characterize the equilibrium social loss in this setup for binary classification (Chapter 3.3.2), and apply this characterization to three concrete setups that vary representation quality along different axes (Chapter 3.3.3): we show that the equilibrium social loss can be non-monotonic in Bayes risk in all of these setups (Figures 3.2b-3.2c). Finally, we extend our theoretical characterization from Chapter 3.3.2 to setups with more than 2 classes (Chapter 3.3.4), and we extend our model and results to model-providers with unequal market reputations (Chapter 3.3.5).

Specification of stylized setup. Assume the input space X is finite and let $\mathcal{F} = \mathcal{F}_{\text{all}}^{\text{multi-class}}$ contain all deterministic functions from X to $\{0, 1, \ldots, K-1\}$. For simplicity, we also assume that users make noiseless decisions (i.e., we take $c \to 0$), so a user's choice of model-provider $j^*(x, y)$ is specified as follows:

$$\mathbb{P}[j^*(x,y) = j] = \begin{cases} 0 & \text{if } j \notin \arg\min_{j' \in [m]} \mathbb{1}[y \neq f_{j'}(x)] \\ \frac{1}{\left|\arg\min_{j' \in [m]} \mathbb{1}[y \neq f_{j'}(x)]\right|} & \text{if } j \in \arg\min_{j' \in [m]} \mathbb{1}[y \neq f_{j'}(x)]. \end{cases} \tag{3.3}$$

In other words, users pick the model-provider with minimum loss, choosing randomly in case of ties. We show that pure strategy equilibria are guaranteed to exist in this setup.

Proposition 1. *Let X be a finite set of representations, let there be $K \geq 2$ classes, let $\mathcal{F} = \mathcal{F}_{all}^{multi\text{-}class}$, and let $\mathcal{D}$ be the distribution over (X, Y). Suppose that user decisions are noiseless (i.e., user decisions are given by (3.3)). For any $m \geq 2$, there exists a pure strategy equilibrium.*

3.3.1 Simple mathematical example of non-monotonicity

We show a simple example where improving data representation quality (i.e. Bayes risk) reduces the equilibrium social welfare. Consider a distribution over binary labels given by $\mathbb{P}[Y = 1] = 0.6$ and $\mathbb{P}[Y = 0] = 0.4$, and suppose that there are $m = 3$ model-providers. We consider two different sets of representations X_1 and X_2, which give rise to two different distributions $\mathcal{D}_1$ over $X_1 \times Y$ and $\mathcal{D}_2$ over $X_2 \times Y$ satisfying $\mathbb{P}[Y = 1] = 0.6$ and $\mathbb{P}[Y = 0] = 0.4$.

Suppose that $X_1 = \{x_0\}$ consists of the trivial representation which provides no information about users. The distribution $\mathcal{D}_1$ is specified by $\mathbb{P}_{\mathcal{D}_1}[Y = 1 \mid X_1 = x_0] = 0.6$ and $\mathbb{P}_{\mathcal{D}_1}[Y = 0 \mid X_1 = x_0] = 0.4$. In this case, the Bayes risk is 0.4. Moreover, it is not difficult to see that $f_1^*(x_0) = f_2^*(x_0) = 1$ and $f_3^*(x_0) = 0$ is an equilibrium. (The reason that $f_1(x_0) = f_2(x_0) = f_3(x_0) = 1$ is not an equilibrium is that model provider 3 would deviate to $f_3^*(x_0) = 0$ and *increase* their utility from 1/3 to 0.4.) Since the model-providers collectively offer both labels for the representation x_0, each user has the option to choose either label, so the equilibrium social loss $\texttt{SL}(f_1^*, f_2^*, f_3^*) = 0$.

Next, suppose that $X_2 = \{x_1, x_2\}$ consists of binary representations that provide some nontrivial information about users. In particular, the distribution $\mathcal{D}_2$ is specified by equally likely representations $\mathbb{P}_{\mathcal{D}_2}[X_2 = x_1] = \mathbb{P}_{\mathcal{D}_2}[X_2 = x_2] = 0.5$. The conditional distribution $Y \mid X_2$ is specified by $\mathbb{P}_{\mathcal{D}_2}[Y = 1 \mid X_2 = x_1] = 0.4$, $\mathbb{P}_{\mathcal{D}_2}[Y = 0 \mid X_2 = x_1] = 0.6$, $\mathbb{P}_{\mathcal{D}_2}[Y = 1 \mid X_2 = x_2] = 0.8$, and $\mathbb{P}_{\mathcal{D}_2}[Y = 0 \mid X_2 = x_2] = 0.2$. In this case, the Bayes risk goes down to 0.3. Moreover, it is not difficult to see that $f_1^*(x_1) = f_2^*(x_1) = 0$, $f_3^*(x_1) = 1$, and $f_1^*(x_2) = f_2^*(x_2) = f_3^*(x_2) = 1$ is an equilibrium. (Intuitively, the reason that $f_1^*(x_2) = f_2^*(x_2) = f_3^*(x_2) = 1$ occurs at equilibrium in this setup is that no model provider $i \in [m] = \{1, 2, 3\}$ wants to deviate to $f_i(x_2) = 0$, since this would *decrease* their utility on $X_2 = x_2$ from 1/3 to 0.2.) Since users with representation x_2 no longer have the option to choose the label of 0, the equilibrium social loss is $\texttt{SL}(f_1^*, f_2^*, f_3^*) = 0.1$.

As a result, even though the Bayes risk is lower for representations in the second setup than for the representations in the first setup, the equilibrium social loss is higher. This instantation thus provides a simple mathematical example where non-monotonicity occurs. In the remaining sections, we consider more general setups that elucidate what factors drive non-monotonicity.

3.3.2 Characterization of the equilibrium social loss for binary classification

To generalize the above example, we analyze general instantations of the stylized setup, focusing first on binary classification. Let $\mathcal{F}_{\text{all}}^{\text{binary}}$ denote the function class $\mathcal{F}_{\text{all}}^{\text{multi-class}}$ in the special case of $K = 2$ classes. Since $\mathcal{F}_{\text{all}}^{\text{binary}}$ lets model-providers make independent predictions about each representation x, the only source of error is noise in individual data points. To capture this, we define the *per-representation Bayes risk* $\alpha(x)$ to be:

$$\alpha(x) := \min(\mathbb{P}(y = 1 \mid x), \mathbb{P}(y = 0 \mid x)). \tag{3.4}$$

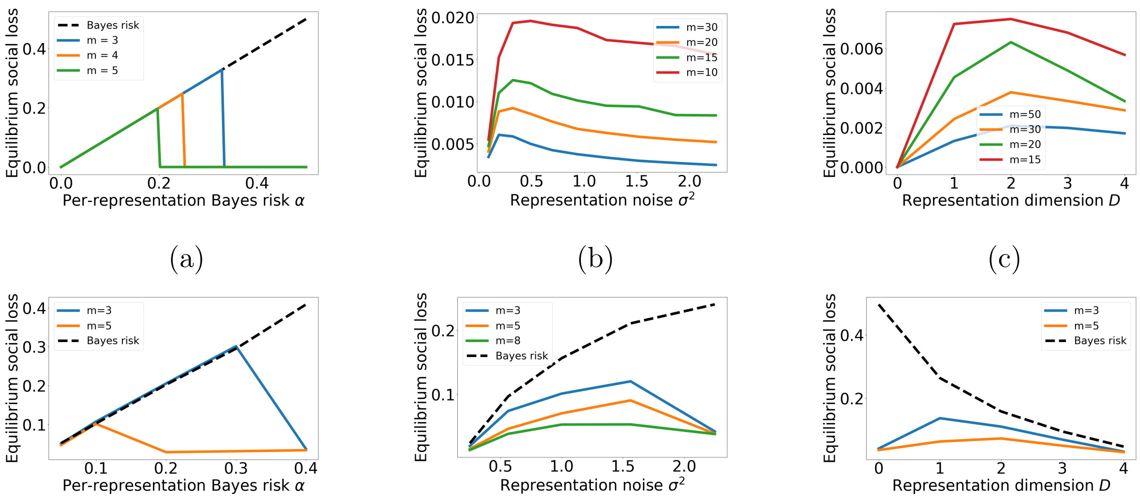
The value $\alpha(x)$ measures how random the label y is for a given representation x. As a result, $\alpha(x)$ is the minimum error that a model-provider can hope to achieve on the given representation x. Increasing $\alpha(x)$ increases the Bayes risk $\texttt{OPT}_{\text{single}}$: in particular, $\texttt{OPT}_{\text{single}}$ is equal to the average value $\mathbb{E}[\alpha(x)]$ across the population. The equilibrium social loss, however, depends on other aspects of $\alpha(x)$.

We characterize the equilibrium social loss in terms of the per-representation Bayes risks in the following proposition. Our characterization focuses on pure-strategy equilibria, which are guaranteed to exist in this setup (see Proposition 1).

Proposition 2. *Let X be a finite set, let $K = 2$, and let $\mathcal{F} = \mathcal{F}_{all}^{binary}$. Suppose that user decisions are noiseless (i.e., user decisions are given by (3.3)). Suppose also that $\alpha(x) \neq 1/m$ for all $x \in X$.*[4] *At any pure strategy Nash equilibrium $f_1^*, \ldots, f_m^*$, the social loss $\mathit{SL}(f_1^*, \ldots, f_m^*)$ is equal to:*

$$\mathit{SL}(f_1^*, \ldots, f_m^*) = \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}} \left[\alpha(x) \cdot \mathbb{1}[\alpha(x) < 1/m]\right]. \tag{3.5}$$

The primary driver of Proposition 2 is that as the per-representation Bayes risk $\alpha(x)$ decreases, the equilibrium predictions for x go from *heterogeneous* (different model-providers offer different predictions for x) to *homogenous* (all model-providers offer the same prediction for x). In particular, if $\alpha(x)$ is below $1/m$, then all model-providers choose the Bayes optimal label $y^* = \arg\max_{y'} \mathbb{P}[y' \mid x]$, so predictions are homogeneous; on the other hand, if $\alpha(x)$ is above $1/m$, then at least one model-provider will choose $1 - y^*$, so predictions are heterogeneous. When predictions are heterogeneous, each user is offered perfect predictive accuracy by some model-provider, which results in zero social loss. On the other hand, if predictions are homogeneous and all model-providers choose the Bayes optimal label, the social loss on x is the per-representation Bayes risk $\alpha(x)$. Putting this all together, the equilibrium social loss takes the value in (3.5). We defer a proof of Proposition 2 to Chapter A.2.



(a) (b) (c)

(d) (e) (f)

Figure 3.2: Equilibrium social loss (y-axis) versus data representation quality (x-axis) given m model-providers, for different function classes $\mathcal{F}$ (rows) and when representations are varied along different aspects (columns). Top row: $\mathcal{F} = \mathcal{F}_{\text{all}}^{\text{binary}}$, with closed-form formula from Proposition 2. Bottom row: linear functions, computed via simulation (Chapter 3.4). We vary representations with respect to per-representation Bayes risk (a,d), noise level (b,e), and dimension (c,f). The dashed line indicates the Bayes risk (omitted if it is too high to fit on the axis). The Bayes risk is monotone, but the equilibrium social loss is non-monotone.

3.3.3 Non-monotonicity along several axes of varying representations

Using Proposition 2, we next vary representations along several axes and compute the equilibrium social loss, observing non-monotonicity in each case.

Setting 1: Varying the per-representation Bayes risks. Consider a population with a single value of x that has Bayes risk $\alpha(x) = \alpha$. We vary representation quality by varying α from 0 to 0.5. Figure 3.2a depicts the result: by Proposition 2, the equilibrium social loss is zero if $\alpha > 1/m$ and is α if $\alpha < 1/m$, leading to non-monotonicity at $\alpha = 1/m$. When there are $m \geq 3$ model-providers, the equilibrium social loss is thus non-monotonic in α. (For $m = 2$, where $\alpha = 1/2$ is the maximum possible per-representation Bayes risk, the equilibrium social loss is monotone in α.) As the number of model-providers increases, the non-monotonicity occurs at a higher data representation quality (a lower Bayes risk).

Setting 2: Varying the representation noise. Consider a one-dimensional population given by a mixture of two Gaussians (one for each class), where each Gaussian has variance

[4]When $\alpha(x) = 1/m$, there turn out to be multiple pure-strategy equilibria with different social losses.

σ^2 (see Chapter A.1 for the details of the setup). We vary the parameter σ to change the representation quality. Intuitively, a lower value of σ makes the Gaussians more well-separated, which improves representation quality (Bayes risk). By Proposition 2, the equilibrium social loss is $\mathbb{E}[\alpha(x) \cdot \mathbb{1}[\alpha(x) < 1/m]]$. For each value of σ, we estimate the equilibrium social loss by sampling representations x from the population and taking an average.[5] Figure 3.2b depicts the result: the equilibrium social loss is non-monotonic in σ (and thus the Bayes risk). Again, as the number of model-providers increases, the non-monotonicity occurs at a higher representation quality (a lower Bayes risk).

Setting 3: Varying the representation dimension. We consider a four-dimensional population (X^{all}, Y), and let the representation X consist of the first D coordinates of X^{all}, for D varying from 0 to 4 (see Chapter A.1 for full details). Intuitively, a higher dimension D makes the representations more informative, thus improving representation quality (Bayes risk). As before, for each value of D, we estimate the equilibrium social loss by sampling representations x from the population and taking an average. Figure 3.2c depicts the result: the equilibrium social loss is once again non-monotonic in the representation dimension D (and thus the Bayes risk).

Discussion. Settings 1-3 illustrate that equilibrium social loss can be non-monotonic in Bayes risk when representations are improved along many qualitatively different axes. The intuition is that varying representations along these axes can increase the values of $\alpha(x)$ for inputs x; by Proposition 2, these changes to $\alpha(x)$ can lead to non-monotonicity in the equilibrium social loss. We will revisit Settings 1-3 for richer market structures (Chapter 3.3.5) and for linear predictors and noisy user decisions (Chapter 3.4.2).

3.3.4 Generalization to more than 2 classes

While our analysis has thus far focused on classification with $K = 2$ classes, the number of classes K can be much larger in practice. As a motivating example, consider content recommendation tasks where each class represents a different genre of content; since the content landscape can be quite diverse, we would expect K to be fairly large.[6] This motivates us to extend our theoretical characterization in Proposition 2 to classification with $K \geq 2$ classes.

For the case of $K \geq 2$ classes, the appropriate analogue of the per-representation Bayes risk is the per-class-per-representation Bayes risk, defined to be:

$$\alpha^i(x) := \mathbb{P}(y = i \mid x) \tag{3.6}$$

[5] Strictly speaking, we can't directly apply Proposition 2 to this setup since X is infinite. We circumvent this issue by applying Proposition 2 on a sample of the representations.

[6] When K is large, even if users can "search" for and "consume" content on their own without relying on model-provider predictions, we expect that our measure of social loss would still be a good proxy for the loss experienced by users. In particular, it would be prohibitively expensive for users to try out all K classes, so classes that are not suggested to the user by any model-provider's predictions might be effectively inaccessible to the user.

for each $x \in X$ and $i \in \{0, 1, \ldots, K-1\}$. Observe that $1 - \max_{0 \le i \le K-1} \alpha^i(x)$ is the minimum error that a single model-provider can hope to achieve on x, and $\texttt{OPT}_{\text{single}}$ is equal to the average value $\mathbb{E}[1 - \max_{0 \le i \le K-1} \alpha^i(x)]$ across the population. The equilibrium social loss, however, depends on other aspects of the $\alpha^i(x)$ values.

We characterize the equilibrium social loss in terms of the per-class-per-representation Bayes risks in the following proposition. Our characterization again focuses on pure-strategy equilibria, which are guaranteed to exist in this setup by Proposition 1.

Proposition 3. *Let X be a finite set, let there be $K \ge 2$ classes, let $\mathcal{F} = \mathcal{F}_{all}^{multi\text{-}class}$. Suppose that user decisions are noiseless (i.e., user decisions are given by (3.3)). Let $c = \min_{x \in X} \max_{0 \le i \le K-1} \alpha^i(x)$. Then, at any pure strategy Nash equilibrium $f_1^*, \ldots, f_m^*$, the social loss $\mathit{SL}(f_1^*, \ldots, f_m^*)$ is bounded as*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\sum_{i=1}^{K} \alpha^i(x) \cdot \mathbb{1}\left[\alpha^i(x) < \frac{c}{m}\right]\right] \le \mathit{SL}(f_1^*, \ldots, f_m^*) \le \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}\left[\sum_{i=1}^{K} \alpha^i(x) \cdot \mathbb{1}\left[\alpha^i(x) \le \frac{1}{m}\right]\right]. \tag{3.7}$$

The high-level intuition for Proposition 3 is similar to the intuition for Proposition 2, except that each class needs to be considered separately. In particular, when class i occurs sufficiently frequently for the representation x (i.e., when $\alpha^i(x)$ is not too small), then some model-provider will label x as i; on the other hand, if the class i occurs very infrequently for x, then no model-provider will label x as i. We defer a proof of Proposition 3 to Chapter A.2.

While Proposition 3 is conceptually a generalization of Proposition 2, the details of Proposition 3 slightly differ. In particular, Proposition 3 does not completely pin down the equilibrium social loss, and there is a factor of c slack in the constraint on each $\alpha^i(x)$ in (3.7) between the upper and lower bounds. Nonetheless, since the value $c = \min_{x \in X} \max_{0 \le i \le K-1} \alpha^i(x)$ measures the minimum accuracy of the Bayes optimal predictor across all inputs x, we expect that "reasonable" representations (i.e., representations which are sufficiently informative) would have c equal to a constant. When c is a constant, there is at most a constant factor slack in the $\alpha^i(x)$ constraints in (3.7) between upper and lower bound.

For similar reasons to Proposition 2, Proposition 3 implies that the equilibrium social loss can be non-monotonic in the representation quality (i.e., the Bayes risk). As a concrete example, consider the following adaptation of Setting 1 in Chapter 3.3.3: let there be a population with a single value of x where $\alpha_0(x) = 1 - 2\alpha$, $\alpha_1(x) = \alpha$, and $\alpha_2(x) = \alpha$ for some $\alpha < 1/4$. In this setup, we see that $c \ge 1/2$. By Proposition 3, the equilibrium social loss is 2α if $\alpha < 1/(2m)$, and the equilibrium social loss is 0 if $\alpha > 1/m$; on the other hand, the Bayes risk is equal to 2α for any $\alpha < 1/4$. This illustrates that the equilibrium social loss is non-monotonic in the Bayes risk. We expect that other setups similar to those in Chapter 3.3.3 will also lead to non-monotonicity for multi-class tasks.

3.3.5 Generalization to unequal market reputations

While we assumed above that users evenly break ties between model-providers, in reality, users might be more likely to choose model-providers with a higher market reputation (e.g., established, popular model-providers). This motivates us to incorporate market reputations into user decisions.

Formally, we assign to each model-provider j a *market reputation* w_j, and we replace the logit model in (3.1) with a weighted logit variant. When $c \to 0$, rather than breaking ties uniformly, they are instead broken proportionally to w_j:

$$\mathbb{P}[j^*(x,y) = j] = \begin{cases} 0 & \text{if } j \notin \arg\min_{j' \in [m]} \mathbb{1}[y \neq f_{j'}(x)] \\ \frac{w_j}{\sum_{j'' \in [m]} w_{j''} \cdot \mathbb{1}[j'' \in \arg\min_{j' \in [m]} \mathbb{1}[y \neq f_{j'}(x)]]} & \text{if } j \in \arg\min_{j' \in [m]} \mathbb{1}[y \neq f_{j'}(x)]. \end{cases} \tag{3.8}$$

See Chapter 3.5 for further discussion of this model. For simplicity, we assume that market reputations are normalized to sum to one.

Similarly to Proposition 2, we derive a closed-form formula for the equilibrium social loss, focusing on the case of binary classification with $m = 2$ model-providers for analytic tractability. We observe non-monotonicity as before, but with a more complex functional form.

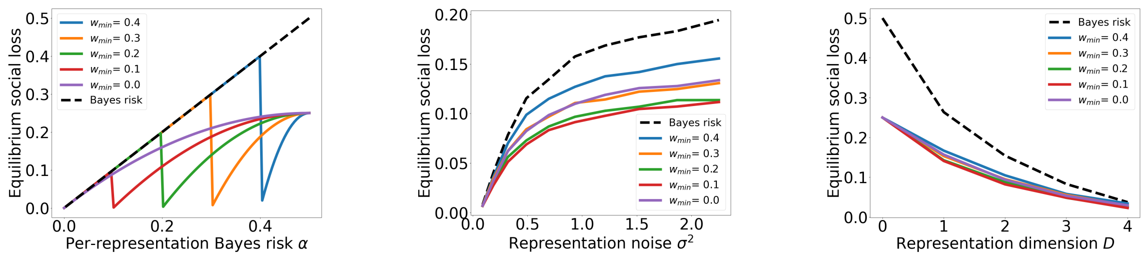
Proposition 4. *Let X be a finite set, let $K = 2$, and let $\mathcal{F} = \mathcal{F}_{all}^{binary}$. Suppose there are $m = 2$ model-providers with market reputations w_{min} and w_{max}, where $w_{max} \geq w_{min}$ and $w_{max} + w_{min} = 1$. Suppose that user decisions are given by (3.8), and that $\alpha(x) \neq w_{min}$ for all $x \in X$.*[7] *At any (mixed) Nash equilibrium (μ_1, μ_2), the expected social loss $\mathbb{E}_{\substack{f_1 \sim \mu_1 \\ f_2 \sim \mu_2}}[\mathtt{SL}(f_1, f_2)]$ is equal to:*

$$\mathbb{E}_{(x,y) \sim \mathcal{D}} \left[\underbrace{\frac{(\alpha(x) - w_{min}) \cdot (w_{max} - \alpha(x))}{(1 - 2 \cdot w_{min})^2}}_{(A)} \cdot \mathbb{1}[\alpha(x) > w_{min}] + \underbrace{\alpha(x)}_{(B)} \cdot \mathbb{1}[\alpha(x) < w_{min}] \right]. \tag{3.9}$$

The high-level intuition for Proposition 4, like for Proposition 2, is that the equilibrium predictions go from heterogeneous to homogenous as $\alpha(x)$ decreases. Term (A), which is realized for large $\alpha(x)$, captures the equilibrium social loss for heterogeneous predictions. Term (B), which is realized for small $\alpha(x)$, captures the equilibrium social loss for homogeneous predictions. We defer the proof of Proposition 4 to Chapter A.2.

The details of Proposition 4 differ from Proposition 2 in several ways. First, the transition point from heterogeneous to homogeneous predictions occurs at $\alpha(x) = w_{\min}$ as opposed to $\alpha(x) = 1/2$. In particular, the transition point depends on the market reputations rather than only the number of model-providers. Second, the equilibria have *mixed strategies*

[7]As with Proposition 2, when $\alpha(x)$ is equal to $w_{\min}$ for some value of x, there are multiple equilibria.



(a) (b) (c)

Figure 3.3: Equilibrium social loss (y-axis) versus data representation quality (x-axis) given two model-providers with market reputations $[1 - w_{\min}, w_{\min}]$ when representations are varied along different aspects (columns). The equilibrium social loss is computed via the closed-form formula from Proposition 4. We vary representations with respect to per-representation Bayes risk (a), noise level (b), and dimension (c). The dashed line indicates the Bayes risk. The Bayes risk is monotone for all 3 axes of varying representations; on the other hand, the equilibrium social loss is non-monotone in the per-representation Bayes risk and monotone in noise level and dimension.

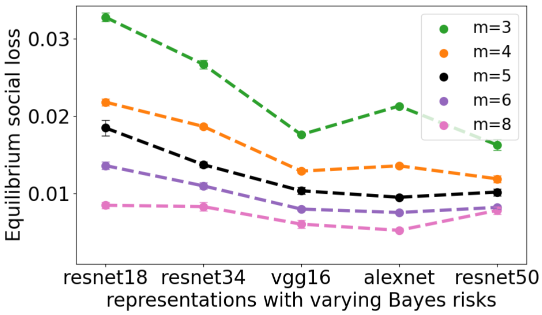
rather than pure strategies, because pure-strategy equilibria do not necessarily exist when market reputations are unequal (see Lemma 133 in Chapter A.2). Third, the social loss at a representation x is no longer equal to zero for heterogeneous predictions—in particular, term (A) is now positive for all $\alpha(x) > w_{\min}$ and increasing in $\alpha(x)$.

To better understand the implications of Proposition 4, we revisit Settings 1-3 from Chapter 3.3.3, considering the same three axes of varying representations with the same distributions over (x, y). In contrast to Chapter 3.3.3, we consider 2 competing model-providers with unequal market positions rather than m competing model providers with equal market positions. Our results, described below, are depicted in Figure 3.3.

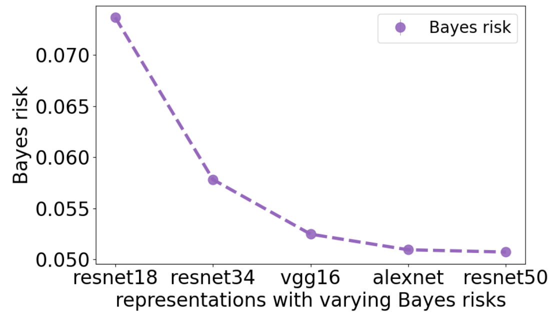
Setting 1: Varying the per-representation Bayes risks. Consider the same setup as Setting 1 in Chapter 3.3.3. Figure 3.3a depicts the non-monotonicity of the equilibrium social loss in the per-representation Bayes risk α across different settings of market reputations for 2 competing model-providers. The discontinuity occurs at the smaller market reputation $w_{\min}$. Thus, as the market reputations of the 2 model-providers become closer together, the non-monotonicity occurs at a lower data representation quality (higher Bayes risk).

Settings 2-3: Varying the representation noise or representation dimension. Consider the setups from Settings 2-3 in Chapter 3.3.3. Figures 3.3b-3.3c depicts that the equilibrium social loss is *monotone* in data representation quality (Bayes risk) across different settings of market reputations for 2 competing model-providers.

Discussion. To interpret these results, observe that for 2 model-providers with equal market reputations ($w_{\min} = 0.5$), the equilibrium social loss is always equal to the Bayes risk by



(a)



(b)

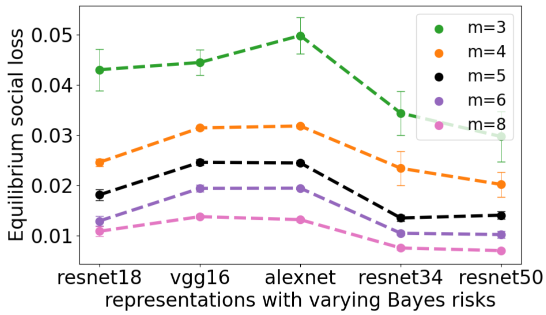
Figure 3.4: Equilibrium social loss (left) and Bayes risk (right) on a binary classification task on CIFAR-10 (Chapter 3.4.3). Representations are generated from different networks pre-trained on ImageNet. The points show the equilibrium social loss when m model-providers compete with each other (left) and the Bayes risk of a single model-provider in isolation (right). While Bayes risk is decreasing in this representation ordering, the equilibrium social loss is non-decreasing in this ordering. The equilibrium social loss is thus non-monotonic in representation quality as measured by Bayes risk. Error bars are 1 standard error.

Propositions 2-4, which trivially implies monotonicity. In contrast, Figure 3.3 shows that for unequal market positions ($w_{\min} < 0.5$), the equilibrium social loss is non-monotonic in Bayes risk for Setting 1, though it is still monotonic in Bayes risk for Settings 2 and 3. (For comparison, recall from Figures 3.2a-3.2c that for $m \gg 2$ model-providers with equal market reputations, non-monotonicity was exhibited for all three settings.) An interesting open question is identify other axes of varying representations, beyond Setting 1, which lead to non-monotonicity for 2 model-providers with unequal market reputations.

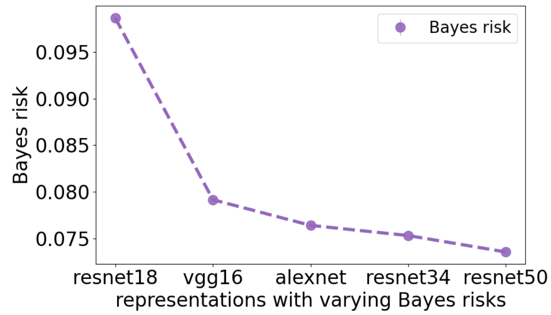
3.4 Empirical Analysis of Non-monotonicity for Linear Predictors

We next turn to linear predictors and demonstrate empirically that the social welfare can be non-monotonic in data representation quality in this setup as well.[8] We take $X = \mathbb{R}^D$ and we let the model parameters be ϕ. For binary classification, we let $\mathcal{F}_{\text{linear}}^{\text{binary}}$ be the family of linear predictors $f_{w,b} = \text{sigmoid}(\langle w, x\rangle + b)$ where $w \in \mathbb{R}^D$, $b \in \mathbb{R}$, and $\phi = [w, b]$. Similarly, for classification with more than 2 classes, we let $\mathcal{F}_{\text{linear}}^{\text{multi-class}}$ be the family of linear predictors $f_{W,b} = \text{softmax}(Wx + b)$ where $w \in \mathbb{R}^{|Y| \times D}$, $b \in \mathbb{R}^{|Y|}$, and $\phi = [W, b]$. Since this setting no longer admits closed-form formulae, we numerically estimate the equilibria using

[8]The code can be found at `https://github.com/mjagadeesan/competition-nonmonotonicity`.



(a)



(b)

Figure 3.5: Equilibrium social loss (left) and Bayes risk (right) on a 10-class classification task on CIFAR-10 (Chapter 3.4.4). Representations are generated from different networks pre-trained on ImageNet. The points show the equilibrium social loss when m model-providers compete with each other (left) and the Bayes risk of a single model-provider in isolation (right). While Bayes risk is decreasing in this representation ordering, the equilibrium social loss is non-decreasing in this ordering. The equilibrium social loss is thus non-monotonic in representation quality as measured by Bayes risk. Error bars are 1 standard error.

a variant of *best-response dynamics*, where model-providers repeatedly best-respond to the other predictors.

We first show on low-dimensional synthetic data on a binary classification task that the insights from Chapter 3.3.3 readily generalize to linear predictors (see Figures 3.2d-3.2f). We then turn to natural data, considering binary and 10-class image classification tasks for CIFAR-10 and using pretrained networks—AlexNet, VGG16, and various ResNets—to generate high-dimensional representations (ranging from 512 to 4096). In this setting we again find that the equilibrium social loss can be non-monotonic in the Bayes risk (see Figure 3.4 and Figure 3.5).

3.4.1 Best-response dynamics implementation

To enable efficient computation, we assume the distribution $\mathcal{D}$ corresponds to a finite dataset with N data points. We calculate equilibria using an approximation of best-response dynamics. Model-providers (players) iteratively (and approximately) best-respond to the other players' actions. We implement the approximate best-response as running several steps of gradient descent.

In more detail, for each $j \in [m]$, we initialize the model parameters ϕ as mean zero Gaussians with standard deviation σ. Our algorithm then proceeds in stages. At a given stage, we iterate through the model-providers in the order $1, \ldots, m$. When j is chosen, first we decide whether to reinitialize: if the risk $\mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f_\phi(x), y)]$ exceeds a threshold ρ, we re-initialize w_j and b_j (sampling from mean zero Gaussians as before); otherwise, we do not reinitialize. Then we run gradient descent on $u(\cdot; \mathbf{f}_{-j})$ (computing the gradient on the full dataset of N points) with learning rate η for I iterations, updating the parameters ϕ. We run this gradient descent step up to 2 more times if the risk $\mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f_\phi(x), y)]$ exceeds a threshold ρ'. At the end of a stage, the stopping condition is that for every $j \in [m]$, model-provider j's utility $u(f_j, \mathbf{f}_{-j})$ has changed by at most ε relative to the previous stage. If the stopping condition is not met, we proceed to the next stage.

3.4.2 Simulations on synthetic data

We first revisit Settings 1-3 from Chapter 3.3.3, considering the same three axes of varying representations with the same distributions over (x, y). In contrast to Chapter 3.3.3, we restrict the model family to linear predictors $\mathcal{F}^{\text{binary}}_{\text{linear}}$ instead of allowing all predictors $\mathcal{F}^{\text{binary}}_{\text{all}}$. We also set the noise parameter c in user decisions (3.1) to 0.3. Our goal is to examine if the findings from Chapter 3.3 generalize to this new setting.

We compute the equilibria for each of the following (continuous) distributions as follows. First, we let $\mathcal{D}$ be the empirical distribution over $N = 10,000$ samples from the continuous distribution. Then we run the best-response dynamics described in Chapter 3.4.1 with $\rho = 0.3$, $I = 5000$, $\eta = 0.1$, $\varepsilon = 0.01$, and $\sigma = 0.1$. We then compute the equilibrium social loss according to (3.2). We also compute the Bayes optimal predictor with gradient descent. See Chapter A.1 for full details.

Our results, described below, are depicted in Figures 3.2d-3.2f (row 2). We compare these results with Figures 3.2a-3.2c (row 1), which shows the analogous results for $\mathcal{F}^{\text{binary}}_{\text{all}}$ from Chapter 3.3.3.

Setting 1: Varying the per-representation Bayes risks. Consider the same single x setup as in Setting 1 in Chapter 3.3.3. The only parameter of the predictor is the bias $b \in \mathbb{R}$ (i.e., we treat x as zero-dimensional). Figure 3.2d shows that the equilibrium social loss is non-monotonic in α, which mirrors the non-monotonicity in Figure 3.2a.

Setting 2: Varying the representation noise. Consider the same one-dimensional mixture-of-Gaussians distribution as in Setting 2 in Chapter 3.3.3. (The weight w is one-dimensional.) We again vary the noise σ to change the representation quality. Figure 3.2e shows that the equilibrium social loss is non-monotonic in the noise σ, which again mirrors the non-monotonicity in Figure 3.2b.

Setting 3: Varying the representation dimension. Consider the same four-dimensional population as in Setting 3 in Chapter 3.3.3. We vary the representation dimension D from 0 to 4 to change the representation quality. Figure 3.2f shows that the equilibrium social loss is non-monotonic in the dimension D, which once again mirrors the non-monotonicity in Figure 3.2c.

Discussion. In summary, in Figure 3.2, rows 1 and 2 exhibit similar non-monotonicities. This illustrates that the insights from Chapter 3.3.2 translate to linear predictors and noisy user decisions.

3.4.3 Simulations on CIFAR-10 for binary classification

We next turn to experiments with natural data. While we have directly varied the informativeness of data representations thus far, representations in practice are frequently generated by pretrained models (Example 1). The choice of the pretrained model implicitly influences representation quality, as measured by Bayes risk on the downstream task. In this section, we consider how the equilibrium social loss changes with representations generated from pretrained models of varying quality. We restrict the model family to linear predictors $\mathcal{F}_{\text{linear}}^{\text{binary}}$ and set the noise parameter c in user decisions (3.1) to 0.1.

We consider a binary image classification task on CIFAR-10 (Krizhevsky, 2009) with 50,000 images. Class 0 is defined to be $\{\text{airplane}, \text{bird}, \text{automobile}, \text{ship}, \text{horse}, \text{truck}\}$ and the class 1 is defined to be $\{\text{cat}, \text{deer}, \text{dog}, \text{frog}\}$. We treat the set of 50,000 images and labels as the population of users, meaning that it is both the training set and the validation set.[9] Representations are generated from five models—AlexNet (Krizhevsky et al., 2012), VGG16 (Simonyan and Zisserman, 2015), ResNet18, ResNet34, and ResNet50 (He et al., 2016)—pretrained on ImageNet (Deng et al., 2009). The representation dimension is 4096 for AlexNet and VGG16, 512 for ResNet18 and ResNet34, and 2048 for ResNet50.

We compute the equilibria as follows. First, we let $\mathcal{D}$ be the distribution described above with $N = 50,000$ data points. Then we run the best-response dynamics described in Chapter 3.4.1 for $m \in \{3, 4, 5, 6, 8\}$ model-providers with $\rho = \rho' = 0.3$, $I = 2000$, $\varepsilon = 0.001$, $\sigma = 0.5$, and a learning rate schedule that starts at $\eta = 1.0$. We then compute the equilibrium social loss according to (3.2). We also compute the Bayes risk using gradient descent. For full experimental details, see Chapter A.1.

Figure 3.4 shows that the equilibrium social loss can be non-monotone in the Bayes risk. For example, for $m = 3$, VGG16 outperforms AlexNet, even though the Bayes risk of VGG16 is substantially higher than the Bayes risk of AlexNet. Interestingly, the location of the non-monotonicity differs across different values of m. For example, for $m = 5$ and $m = 8$, AlexNet outperforms ResNet50 despite having a higher Bayes risk, but ResNet50 outperforms AlexNet for $m = 3$ and $m = 4$.

3.4.4 Simulations on CIFAR-10 for 10-class classification

While our empirical analysis has thus far focused on binary classification, we now turn to classification with more than 2 classes. In particular, we consider a ten class CIFAR-10 (Krizhevsky, 2009) task with 50,000 images. The labels are specified by the CIFAR-10 classes in the original dataset. We treat the set of 50,000 images and labels as the population of users, meaning that it is both the training set and the validation set. Representations are

[9] We make this choice to be consistent with the rest of the paper, where we focus on population-level behavior and thus do not consider generalization error.

generated from the same five models—AlexNet (Krizhevsky et al., 2012), VGG16 (Simonyan and Zisserman, 2015), ResNet18, ResNet34, and ResNet50 (He et al., 2016)—pretrained on ImageNet (Deng et al., 2009). We restrict the model family to linear predictors $\mathcal{F}_{\text{linear}}^{\text{multi-class}}$ and again set the noise parameter c in user decisions (3.1) to 0.1.

We compute the equilibria as follows. First, we let $\mathcal{D}$ be the distribution described above with $N = 50,000$ data points. Then we run the best-response dynamics described in Chapter 3.4.1 for $m \in \{3, 4, 5, 6, 8\}$ model-providers with $\rho = 0.7$, $\rho' = 1.0$, $I = 2000$, $\varepsilon = 0.001$, $\sigma = 0.5$, and a learning rate schedule that starts at $\eta = 1.0$. As before, we compute the equilibrium social loss according to (3.2), and we also compute the Bayes risk using gradient descent. For full experimental details, see Chapter A.1.

Figure 3.5 shows that the equilibrium social loss can be non-monotone in the Bayes risk. For example, across all five values of m, ResNet18 outperforms VGG16, even though the Bayes risk of ResNet is substantially higher than the Bayes risk of VGG16. Furthermore, for $m = 3$, VGG16 outperforms AlexNet despite having a larger Bayes risk. Interestingly, the shape of the equilibrium social loss curve for each value of m (Figure 3.5a) appears qualitatively different than the analogous equilibrium social loss curve for binary classification (Figure 3.4a).

3.5 Discussion of Model Assumptions

We highlight and discuss several assumptions that we make in our stylized model.

3.5.1 Assumptions on user decisions

Our primary model for user decisions given by (3.1) is the standard logit model for discrete choice decisions (Train, 2009) which is also known as the Boltzmann rationality model. In the limit as $c \to 0$, a user with representation x and label y select from the set of model-providers $\arg\min_{j \in [m]} \ell(f_j(x), y)$ that achieve the minimum loss; in particular, the user chooses a model-provider from this set with probability proportional to the model-provider's market reputation. For $c > 0$, the specification in equation (3.1) captures that users evaluate a model-provider based on a noisy perception of the loss.

While this model implicitly assumes that a user's choice of platform is fully specified by the platforms' choices of predictor (i.e. platforms are ex-ante homogeneous), we extend this model in Chapter 3.3.5 to account for uneven market reputations across decisions. These market reputations are modeled as global weights in the logit model for discrete choice. Given market reputations $w_1, \ldots, w_m$, users choose a predictor according to:

$$\mathbb{P}[j^*(x, y) = j] = \frac{w_j \cdot e^{-\ell(f_j(x), y)/c}}{\sum_{j'=1}^{m} w_{j'} \cdot e^{-\ell(f_{j'}(x), y)/c}}. \tag{3.10}$$

When the market reputations are all equal ($w_1 = \ldots = w_m$), equation (3.10) exactly corresponds to (3.1). When the market reputations w_j are not equal, equation (3.10) captures that users place a higher weight on model-providers with a higher market reputation. This captures that users are more likely to choose a popular model-provider than a very small model-provider without much reputation. However, this formalization does assume that market reputations are global across users and that market reputations surface as tie-breaking weight in the noiseless limit.

Implicit in this model is asymmetric information between the model-providers and users. While the only information that a model-provider has about users is their representations, a user can make decisions based on noisy perceptions of their own loss (which can depend on their label). This captures that, even if users are unlikely to know their own labels, users can experiment with multiple model-providers to (noisily) determine which one maximizes their utility. The inclusion of market reputations reflects that users are more likely to experiment with and ultimately choose popular model-providers than less popular model-providers.

3.5.2 Assumption of global data representations

Our results assume that all model-providers share the same representations x for each user and thus improvements in representations x are experienced by all model-providers. This assumption is motivated by emerging marketplaces where different model-providers utilize the same foundation model, but *fine-tune* the model in different ways (Example 1). An interesting direction for future work would be to incorporate heterogeneity or local improvements in the data representations.

3.5.3 Assumption on model-provider action space

We make the simplifying assumption that the only action taken by model-providers is to choose a classifier from a pre-specified class. This formalization does not capture other actions (such as data collection and price setting) that may be taken by the platform. Incorporating other model-provider decisions would be an interesting avenue for future work.

3.6 Discussion

We showed that the monotonicity of scaling trends can be violated under competition. In particular, we demonstrated that when multiple model-providers compete for users, improving data representation quality (as measured by Bayes risk) can *increase* the overall loss at equilibrium. We exhibited the non-monotonicity of the equilibrium social loss in the Bayes risk when representations are varied along several axes (per-representation Bayes risk, noise, dimension, and pre-trained model used to generate the representations).

An interesting direction for future work is to further characterize the regimes when the equilibrium social loss is monotonic versus non-monotonic in data representation quality as measured by Bayes risk. For example, an interesting open question is to generalize our theoretical results from Chapter 3.3 to more general function classes and distributions of market reputations. Moreover, another interesting direction would be generalize our empirical findings from Chapter 3.4 to other axes of varying data representations and to non-linear classes of predictors. Finally, while we have focused on classification tasks, it would be interesting to generalize our findings to regression tasks with continuous outputs or to generative AI tasks with text-based or image-based outputs.

More broadly, the non-monotonicity of equilibrium social welfare in scale under competition establishes a disconnect between scaling trends in the single model-provider setting and in the competitive setting. In particular, typical scaling trends (e.g. (Kaplan et al., 2020; Sharma and Kaplan, 2020; Bahri et al., 2024; Hoffmann et al., 2022; Hernandez et al., 2021))—which show increasing scale reliably increases predictive accuracy for a single model-provider in isolation—may not translate to competitive settings such as digital marketplaces. Thus, understanding the downstream impact of scale on user welfare in digital marketplaces will likely require understanding how scaling trends behave under competition. We hope that our work serves as a starting point for analyzing and eventually characterizing the scaling trends of learning systems in competitive settings.

Chapter 4

Companies Training Language Models

This chapter is based on *"Safety vs. Performance: How Multi-Objective Learning Reduces Barriers to Market Entry"* (Jagadeesan et al., 2024), which is joint work with Michael I. Jordan and Jacob Steinhardt.

4.1 Introduction

Large language models and other large-scale machine learning (ML) models have led to an important shift in the information technology landscape, one which has significant economic consequences. Whereas earlier generations of ML models provided the underpinnings for platforms and services, new models—such as language models—are themselves the service. This has led to new markets where companies offer language models as their service and compete for user usage. As in other markets, it is important to reason about market competitiveness: in particular, to what extent there are barriers to entry for new companies.

A widespread concern about these markets is that new companies face insurmountable *barriers to entry* that drive market concentration (Vipra and Korinek, 2023). The typical argument is that incumbent companies with high market share can purchase or capture significant amounts of data and compute,[1] and then invest these resources into the training of models that achieve even higher performance (Kaplan et al., 2020). This suggests that the company's market share would further increase, and that the scale and scope of this phenomenon would place incumbent companies beyond the reach of new companies trying to enter the market. The scale is in fact massive—language assistants such as ChatGPT and Gemini each have hundreds of millions of users (Cook, 2024). In light of the concerns raised by policymakers (Vipra and Korinek, 2023) and regulators (The White House, 2023; European Union, 2022b) regarding market concentration, it is important to investigate the underlying economic and algorithmic mechanisms at play.

[1] Large companies can afford these resources since the marketplace is an economy of scale (i.e., fixed costs of training significantly exceed per-query inference costs). They also generate high volumes of data from user interactions.

While standard arguments assume that market share is determined by model performance, the reality is that the incumbent company risks reputational damage if their model violates safety-oriented objectives. For example, incumbent companies face public and regulatory scrutiny for their model's safety violations—such as threatening behavior (Perrigo, 2023), jailbreaks (Wei et al., 2023), and releasing dangerous information (The White House, 2023)—even when the model performs well in terms of helpfulness and usefulness to users. In contrast, new companies face less regulatory scrutiny since compliance requirements often prioritize models trained with more resources (The White House, 2023; California Legislature, 2024), and new companies also may face less public scrutiny given their smaller user bases.

In this chapter, we use a multi-objective learning framework to show that the threat of reputational damage faced by the incumbent company can reduce barriers to entry. For the incumbent, the possibility of reputational damage creates pressure to align with safety objectives in addition to optimizing for performance. Safety and performance are not fully aligned, so improving safety can reduce performance as a side effect. Meanwhile, the new company faces less of a risk of reputational damage from safety violations. The new company can thus enter the marketplace with significantly less data than the incumbent company, a phenomenon that our model and results formalize.

Model and results. We analyze a stylized marketplace based on multi-objective linear regression (Chapter 4.2). The performance-optimal output and the safety-optimal output are specified by two different linear functions of the input x. The marketplace consists of two companies: an incumbent company and a new company attempting to enter the market. Each company receives their own unlabelled training dataset, decides what fraction of training data points to label according to the performance-optimal vs. safety-optimal outputs, and then runs ridge regression. The new company requires a less stringent level of safety to avoid reputational damage than the incumbent company. We characterize the *market-entry threshold* N_E^* (Definition 1) which captures how much data the new company needs to outperform the incumbent company.

First, as a warmup, we characterize N_E^* when the new company faces no safety constraint and the incumbent company has infinitely many data points (Chapter 4.3). Our key finding is that the new company can enter the market with finite data, even when the incumbent company has infinite data (Theorem 5; Figure 4.1). Specifically, we show that the threshold N_E^* is finite; moreover, it is increasing in the correlation (i.e., the alignment) between performance and safety, and it is decreasing in a problem-specific scaling law exponent.

Next, we turn to more general environments where the incumbent has finite data $N_I < \infty$ (Chapter 4.4.2). We find that the threshold N_E^* scales sublinearly with the incumbent's dataset size N_I, as long as N_I is sufficiently large. In fact, the threshold N_E^* scales at a slower rate as N_I increases: that is, $N_E^* = \Theta(N_I^c)$ where the exponent c is decreasing in N_I (Theorem 8; Figure 4.3). For example, for concrete parameter settings motivated by language models (Hoffmann et al., 2022), the exponent c decreases from 1 to 0.75 to 0 as N_I increases. In general, the exponent c takes on up to three different values depending on N_I, and is strictly smaller than 1 as long as N_I is sufficiently large.

Finally, we turn to environments where the new company also faces a nontrivial safety constraint, assuming for simplicity that the incumbent company again has infinite data (Chapter 4.4.3). We find that N_E^* is finite as long as the new company faces a strictly weaker safety constraint than the incumbent. When the two safety thresholds are closer together, the new company needs more data and in fact needs to scale up their dataset size at a faster rate: that is, $N_E^* = \Theta(D^{-c})$, where D measures the difference between the safety thresholds and where the exponent c increases as D decreases (Theorem 9; Figure 4.4). For the parameter settings in (Hoffmann et al., 2022), the exponent c changes from -2.94 to -3.94 to an even larger value as D decreases. In general, the exponent c takes on up to three different values.

Technical tool: Scaling laws. To prove our results, we derive scaling laws for *multi-objective* high-dimensional linear regression, which could be of independent interest (Chapter 4.4.1; Figure 4.2). We study optimally-regularized ridge regression where some of the training data is labelled according to the primary linear objective (capturing performance) and the rest is labelled according to an alternate linear objective (capturing safety).

We characterize data-scaling laws for both the loss along the primary objective and the excess loss along the primary objective relative to an infinite-data ridgeless regression. Our scaling laws quantify the rate at which the loss (Theorem 6; Figure 4.2a) and the excess loss (Theorem 7; Figure 4.2b) decay with the dataset size N, and how this rate is affected by the fraction of data labelled according to each objective and other problem-specific quantities. Our analysis improves upon recent works on scaling in multi-objective environments (e.g., Jain et al., 2024; Song et al., 2024) by allowing for non-identity covariances and problem-specific regularization, which leads to new insights about scaling laws as we describe below.

Our results reveal that the scaling rate becomes slower as the dataset size increases, illustrating that multi-objective scaling laws behave qualitatively differently from classical single-objective environments. While a typical scaling exponent in a single-objective environment takes on a single value across all settings of N, the scaling exponent for multi-objective environments decreases as N increases. In particular, the scaling exponent takes on *three different values* depending on the size of N relative to problem-specific parameters. The intuition is that the regularizer must be carefully tuned to N in order to avoid overfitting to training data labelled according to the alternate objective, which in turn results in the scaling exponent being dependent on N (Chapter 4.5).

Discussion. Altogether, our work highlights the importance of looking beyond model performance when evaluating market entry in machine learning marketplaces. Our results highlight a disconnect between market entry in single-objective environments versus more realistic multi-objective environments. More broadly, a company's susceptibility to reputational damage affects how they train their model to balance between different objectives. As we discuss in Chapter 4.6, these insights have nuanced implications for regulators who wish to promote both market competitiveness and safety compliance, and also generalize beyond language models to online platforms.

4.1.1 Related work

Our work connects to research threads on *competition between companies* as well as *scaling laws and high-dimensional linear regression.*

Competition between model-providers. Our work contributes to an emerging line of work studying how competing companies (i.e., model-providers) strategically design their machine learning pipelines to attract users. Company actions range from choosing a function from a model class (Ben-Porat and Tennenholtz, 2017; 2019; Jagadeesan et al., 2023b), to selecting a regularization parameter (Iyer and Ke, 2022), to choosing an error distribution over user losses (Feng et al., 2022), to making data purchase decisions (Dong et al., 2019; Kwon et al., 2022), to deciding whether to share data (Gradwohl and Tennenholtz, 2023), to selecting a bandit algorithm (Aridor et al., 2025; Jagadeesan et al., 2023c). While these works assume that companies win users solely by maximizing (individual-level or population-level) accuracy, our framework incorporates the role of *safety violations* in impacting user retention implicitly via reputational damage. Moreover, our focus is on quantifying the barriers to market entry, rather than analyzing user welfare or the equilibrium decisions of companies.

Other related work includes the study of competition between algorithms (Immorlica et al., 2011; Kleinberg and Raghavan, 2021), retraining dynamics under user participation decisions (Hashimoto et al., 2018; Ginart et al., 2021; Dean et al., 2024a; Shekhtman and Dean, 2024; Su and Dean, 2024), the bargaining game between a foundation model company and a specialist (Laufer et al., 2024), and the market power of an algorithmic platform to shape user populations (Perdomo et al., 2020; Hardt et al., 2022; Mendler-Dünner et al., 2024).

Our work also relates to platform competition (Jullien and Sand-Zantman, 2021; Calvano and Polo, 2021a), the emerging area of competition policy and regulation of digital marketplaces (Stigler Committee, 2019; Vipra and Korinek, 2023; Hopkins et al., 2025; Competition and Markets Authority, 2024), the study of how antitrust policy impacts innovation in classical markets (Baker; Segal and Whinston, 2007), and industrial organization more broadly (Tirole, 1988). For example, recent work examines how increased public scrutiny from inclusion in the S&P 500 can harm firm performance (Bennett et al., 2023), how privacy regulation impacts firm competition (Gal and Aviv, 2020; Fallah et al., 2024), how regulatory inspections affect incentives to comply with safety constraints (Harrington, 1988; Fallah and Jordan, 2024), and how data-driven network effects can reduce innovation (Prüfer and Schottmüller, 2021).

Scaling laws and high-dimensional linear regression. Our work also contributes to an emerging line of work on scaling laws which study how model performance changes with training resources. Empirical studies have demonstrated that increases to scale often reliably improve model performance (e.g., Kaplan et al., 2020; Hernandez et al., 2021; Hoffmann et al., 2022), but have also identified settings where scaling behavior is more nuanced (e.g., Muennighoff et al., 2023; Gao et al., 2023). We build on a recent mathematical characterization of scaling laws based on high-dimensional linear regression (e.g., Hastie et al., 2022; Bordelon et al., 2020; Bahri et al., 2024; Cui et al., 2021; Wei et al., 2022; Bach, 2024; Wei, 2024; Patil et al., 2024; Bordelon et al., 2024; Mallinar et al., 2024; Lin et al., 2024; Atanasov et al., 2024). However, while these works focus on *single-objective* environments where all of the training data is labelled with outputs from a single predictor, we consider *multi-objective* environments where some fraction of the training data is labelled according to an alternate predictor.

We note that a handful of recent works similarly move beyond single-objective environments and study scaling laws where the training data comes a mixture of different data sources. Jain et al. (2024); Song et al. (2024) study high-dimensional ridge regression in a similar multi-objective environment to our setup. However, these results assume an *identity covariance* and focus on fixed regularization or no regularization. In contrast, we allow for richer covariance matrices that satisfy natural power scaling (Chapter 4.2.3), and we analyze optimally tuned regularization. Our analysis of these problem settings yields new insights about scaling behavior: for example, the scaling rate becomes slower with dataset size (Theorems 6-7). Other related works study scaling laws under mixtures of covariate distributions (Hashimoto, 2021), under data-quality heterogeneity (Goyal et al., 2024), under data addition (Shen et al., 2024), under mixtures of AI-generated data and real data (Dohmatob et al., 2024; Gerstgrasser et al., 2024), and with respect to the contribution of individual data points (Covert et al., 2024).

More broadly, our work relates to collaborative learning (Blum et al., 2017; Mohri et al., 2019; Sagawa et al., 2019; Haghtalab et al., 2022a), federated learning (see Yang et al., 2019, for a survey), optimizing data mixtures (e.g., Rolf et al., 2021; Xie et al., 2023), and adversarial robustness (e.g., Raghunathan et al., 2020). Finally, our work relates to non-monotone scaling laws in strategic environments (Jagadeesan et al., 2023b; Handina and Mazumdar, 2024), where increases to scale can worsen equilibrium social welfare.

4.2 Model

We define our linear-regression-based marketplace (Chapter 4.2.1), justify the design choices of our model (Chapter 4.2.2), and then delineate our statistical assumptions (Chapter 4.2.3).

4.2.1 Linear regression-based marketplace

We consider a marketplace where two companies fit linear regression models in a multi-objective environment.

Linear regression model. To formalize each company's machine learning pipeline, we consider the multi-objective, high-dimensional linear regression model described below. This multi-objective environment aims to capture how ML models are often trained to balance multiple objectives which are in tension with each other, and we consider linear regression since it has often accurately predicted scaling trends of large-scale machine learning models (see Chapter 4.2.2 for additional discussion).

More concretely, given an input $x \in \mathbb{R}^P$, let $\langle \beta_1, x \rangle$ be the output that targets performance maximization, and let $\langle \beta_2, x \rangle$ be the output that targets safety maximization. Given a linear predictor β, the performance loss is evaluated via a population loss, $L_1(\beta) = \mathbb{E}_{x \sim \mathcal{D}_F}[(\langle \beta_1, x \rangle - \langle \beta, x \rangle)^2]$, and the safety violation is captured by a loss $L_2(\beta) = \mathbb{E}_{x \sim \mathcal{D}_F}[(\langle \beta_2, x \rangle - \langle \beta, x \rangle)^2]$, where $\mathcal{D}_F$ is the input distribution.

The company implicitly determines how to balance β_1 and β_2 when determining how to label their training dataset. In particular, each company is given an unlabelled training dataset $X \in \mathbb{R}^{N \times P}$ with N inputs drawn from $\mathcal{D}_F$. To generate labels, they select the fraction $\alpha \in [0, 1]$ of training data to label according to each objective. They then sample a fraction α of the training data uniformly from X and label it as $Y_i = \langle \beta_1, X_i \rangle$; the remaining $1 - \alpha$ fraction is labelled as $Y_i = \langle \beta_2, X_i \rangle$. The company fits a ridge regression on the labelled training dataset with least-squares loss $\ell(y, y') = (y - y')^2$, and thus solves: $\hat{\beta}(\alpha, \lambda, X) = \arg\min_\beta \left(\frac{1}{N} \sum_{i=1}^N (Y_i - \langle \beta, X_i \rangle)^2 + \lambda ||\beta||_2^2 \right)$.

Marketplace. The marketplace contains two companies, an *incumbent company* I already in the market and a *new (entrant) company* E trying to enter the market. At a high level, each company $C \in \{I, E\}$ faces reputational damage if their safety violation exceeds their safety constraint τ_C. Each company C is given N_C unlabelled data points sampled from $\mathcal{D}_F$, and selects a mixture parameter α_C and regularizer λ_C to maximize their performance given their safety constraint τ_C. We assume that the incumbent company I faces a stricter safety constraint, $\tau_I < \tau_E$, due to increased public or regulatory scrutiny (see Chapter 4.2.2 for additional discussion).

When formalizing how the companies choose hyperparameters, we make the following simplications. First, rather than work directly with the performance and safety losses of the ridge regression estimator, we assume for analytic tractability that they approximate these losses by $L_1^* := L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ and $L_2^* := L_2^*(\beta_1, \beta_2, \mathcal{D}_F, \alpha)$ defined as follows.

- *Performance:* We define L_1^* to be a *deterministic equivalent* $L_1^{\texttt{det}}(\beta_1, \beta_2, \Sigma, \lambda, N, \alpha)$ which we derive in Lemma 10. The deterministic equivalent (cf. Hachem et al., 2007) is a tool from random matrix theory that is closely linked to the Marčenko-Pastur law (Marchenko and Pastur, 1967). Under standard random matrix assumptions (Assumption 7), the deterministic equivalent asymptotically approximates the loss $L_1(\hat{\beta}(\alpha, \lambda, X))$ when X is constructed from N i.i.d. samples from $\mathcal{D}_F$ (see Appendix B.4 for additional discussion).
- *Safety:* For analytic simplicity, in the main body of the paper, we define L_2^* to be the safety violation of the infinite-data ridgeless regression estimator with mixture parameter α.[2] In Appendix B.5, we instead define L_2^* analogously to L_1^*—i.e., as a deterministic equivalent $L_2^{\text{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$—and extend our model and results to this more complex setting.[3]

Second, we assume that $(\beta_1, \beta_2) \sim \mathcal{D}_W$ for some joint distribution $\mathcal{D}_W$ and that the companies take expectations when choosing hyperparameters, since it will be easier to specify assumptions

[2]For this specification, the dataset size N and the regularization parameter λ only affect L_1^* and not L_2^*, which simplifies our analysis in Chapters 4.3-4.4 and enables us to obtain tight characterizations.

[3]We directly extend our results in Chapter 4.3, and we also show relaxed versions of our results in Chapter 4.4.

in Chapter 4.4.3 over distributions of predictors.

Within this setup, a company C faces reputational damage if the safety violation exceeds a certain threshold:

$$\mathbb{E}_{(\beta_1,\beta_2)\sim\mathcal{D}_W}[L_2^*(\beta_1,\beta_2,\mathcal{D}_F,\alpha_C)] > \tau_C.$$

We assume that the safety thresholds for the two companies satisfy the following inequalities:

$$\tau_E >_{(A)} \tau_I \geq_{(B)} \mathbb{E}_{(\beta_1,\beta_2)\sim\mathcal{D}_W}[L_2^*(\beta_1,\beta_2,\mathcal{D}_F,0.5)]. \tag{4.1}$$

Here, inequality (A) captures the notion that the incumbent needs to achieve higher safety to avoid reputational damage. Inequality (B) guarantees that both companies, $C \in \{I, E\}$, can set the mixture parameter $\alpha_C \geq 0.5$ without facing reputational damage, and thus ensures that the safety constraint does not dominate the company's optimization task.[4]

The company selects $\alpha_C \in [0.5, 1]$ and $\lambda_C \in (0, 1)$ to maximize their performance subject to their safety constraint, as formalized by the following optimization program:[5]

$$(\alpha_C,\lambda_C) = \underset{\alpha\in[0.5,1],\lambda\in(0,1)}{\arg\min}\ \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1,\beta_2,\mathcal{D}_F,\lambda,N_C,\alpha)] \text{ s.t. } \mathbb{E}_{\mathcal{D}_W}[L_2^*(\beta_1,\beta_2,\mathcal{D}_F,\alpha)] \leq \tau_C.$$

Market-entry threshold. We define the market-entry threshold to capture the minimum number of data points N_E that the new company needs to collect to achieve better performance than the incumbent company while avoiding reputational damage.

Definition 1. *The market-entry threshold* $N_E^*(N_I,\tau_I,\tau_E,\mathcal{D}_W,\mathcal{D}_F)$ *is the minimum value of* $N_E \in \mathbb{Z}_{\geq 1}$ *such that* $\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1,\beta_2,\mathcal{D}_F,\lambda_E,N_E,\alpha_E)] \leq \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1,\beta_2,\mathcal{D}_F,\lambda_I,N_I,\alpha_I)]$.

The goal of our work is to analyze the function $N_E^*(N_I,\tau_I,\tau_E,\mathcal{D}_W,\mathcal{D}_F)$.

4.2.2 Model discussion

Now that we have formalized our statistical model, we discuss and justify our design choices in greater detail. We defer a discussion of limitations to Chapter 4.6.

Presence of competing objectives. Our multi-objective formulation is motivated by how ML models are often trained to balance multiple objectives which are in tension with each other. In some cases, the pretraining objective is in tension with the finetuning objective (Wei et al., 2023). For example, the fine-tuning of a language model to be more aligned with user intent can degrade performance—e.g., because the model hedges too much—which creates an "alignment tax" (Ouyang et al., 2022). In other cases, fine-tuning approaches themselves balance multiple objectives such as helpfulness (which can be mapped to performance in

[4]More specifically, inequality (B) ensures that the safety constraint still allows both companies to label 50% of their training data according to the performance-optimal outputs.

[5]Technically, the optimum might be achieved at $\lambda = 0$ or $\lambda = 1$, and the min should be replaced by an inf .

our model) and harmlessness (which can be mapped to safety in our model) (Bai et al., 2022). These objectives can be in tension with one another, for example if the user asks for dangerous information.

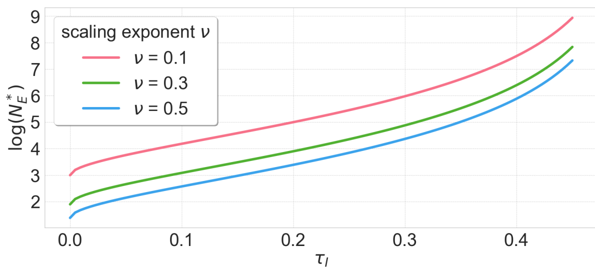
High-dimensional linear regression as a statistical model. We focus on high-dimensional linear regression due to its ability to capture scaling trends observed in large-scale machine learning models such as language models, while still retaining analytic tractability. In particular, in single-objective environments, scaling trends for high-dimensional linear regression recover the empirically observed power-law scaling of the loss with respect to the dataset size (Kaplan et al., 2020; Cui et al., 2021; Wei et al., 2022). Moreover, from an analytic perspective, the structural properties of high-dimensional linear regression make it possible to characterize the loss using random matrix machinery (see Appendix B.4).

Impact of market position on company constraint τ. Our assumption that $\tau_E > \tau_I$ (inequality (A) in (4.1)) is motivated by how large companies face greater reputational damage from safety violations than smaller companies. One driver of this unevenness in reputational damage is *regulation*: for example, recent regulation and policy (The White House, 2023; California Legislature, 2024) places stricter requirements on companies that use significant amounts of compute during training. In particular, these companies face more stringent compliance requirements in terms of safety assessments and post-deployment monitoring. Another driver of uneven reputational damage is *public perception*: we expect that the public is more likely to uncover safety violations for large companies, due to the large volume of user queries to the model. In contrast, for small companies, safety violations may be undetected or subject to less public scrutiny.

4.2.3 Assumptions on linear regression problem

To simplify our characterization of scaling trends, we follow prior work on high-dimensional linear regression (see, e.g., Cui et al., 2021; Wei et al., 2022) and make the following empirically motivated power-law assumptions. Let $\Sigma = \mathbb{E}_{x \sim \mathcal{D}_F}[xx^T]$ be the covariance matrix, and let λ_i and v_i be the eigenvalues and eigenvectors, respectively. We require the eigenvalues to decay with scaling exponent $\gamma > 0$ according to $\lambda_i = i^{-1-\gamma}$ for $1 \leq i \leq P$. For the alignment coefficients $\langle \beta_j, v_i \rangle$, it is cleaner to enforce power scaling assumptions in expectation, so that we can more easily define a correlation parameter. We require that for some $\delta > 0$, the alignment coefficients satisfy $\mathbb{E}_{\mathcal{D}_W}[\langle \beta_j, v_i \rangle^2] = i^{-\delta}$, where v_i is the ith eigenvector of Σ, for $j \in \{1, 2\}$ and $1 \leq i \leq P$. We also introduce a similar condition on the joint alignment coefficients, requiring that for some $\rho \in [0, 1)$, it holds that $\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle \langle \beta_2, v_i \rangle] = \rho \cdot i^{-\delta}$. Finally, we assume an overparameterized limit where the number of parameters $P \to \infty$ approaches infinity. Below, we provide an example which satisfies these assumptions.

Example 2. *Suppose that the covariance Σ is a diagonal matrix with diagonal given by*



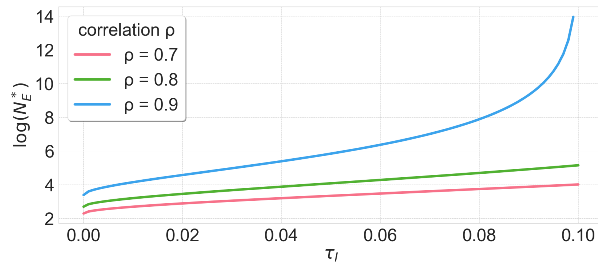


Figure 4.1: Market-entry threshold N_E^* as a function of the incumbent's safety constraint τ_I, when the incumbent has infinite data and entrant has no safety constraint (Theorem 5). The plots show varying values of the scaling exponent ν where the correlation parameter $\rho = 0.5$ is held fixed (left) and varying values of ρ where $\nu = 0.34$ is held fixed (right). The market-entry threshold N_E^* is finite. It is also higher when the constraint τ_I is weaker, when the correlation ρ is stronger, and when the scaling exponent ν is lower.

$\lambda_i = i^{-1-\gamma}$. *Let the joint distribution over* β_1 *and* β_2 *be a multivariate Gaussian such that:*

$$\mathbb{E}_{\mathcal{D}_W}[(\beta_{j_1})_{i_1}(\beta_{j_2})_{i_2}] = \begin{cases} 0 & \text{if } 1 \leq j_1, j_2 \leq 2, 1 \leq i_1 \neq i_2 \leq P \\ i_1^{-\delta} & \text{if } 1 \leq j_1 = j_2 \leq 2, 1 \leq i_1 = i_2 \leq P \\ \rho \cdot i_1^{-\delta} & \text{if } 1 \leq j_1 \neq j_2 \leq 2, 1 \leq i_1 = i_2 \leq P. \end{cases}$$

This implies that $\mathbb{E}_{\mathcal{D}_W}[\langle \beta_j, v_i \rangle^2] = i^{-\delta}$ *and* $\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle \langle \beta_2, v_i \rangle] = \rho \cdot i^{-\delta}$.

We adopt the random matrix theory assumptions on the covariance matrix and linear predictors from Bach (2024) (see Assumption 7 in Appendix B.4), which guarantee that the Marčenko-Pastur law holds (Marchenko and Pastur, 1967). That is, the covariance $(\hat{\Sigma} + \lambda I)^{-1}$ of the samples can be approximated by a deterministic quantity (see Appendix B.4.1 for a more detailed discussion). We leverage this Marčenko-Pastur law to derive a deterministic equivalent $L_1^{\mathtt{det}}$ for the performance loss $L_1(\hat{\beta}(\alpha, \lambda, X))$ of the ridge regression estimator (Lemma 10).

4.3 Warm Up: Infinite-Data Incumbent and Unconstrained Entrant

As a warmup, we analyze the market entry N_E^* threshold in a simplified environment where the incumbent has infinite data and the new company faces no safety constraint. In this result, we place standard power-law scaling assumptions on the covariance and alignment coefficients (Chapter 4.2.3) and we characterize the threshold N_E^* up to constants (Theorem 5; Figure 4.1).

Theorem 5. *Suppose that power-law scaling holds for the eigenvalues and alignment coefficients, with scaling exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, and suppose that $P = \infty$. Suppose that the incumbent company has infinite data (i.e., $N_I = \infty$), and that the entrant faces no constraint on their safety (i.e., $\tau_E = \infty$). Suppose that the safety constraint τ_I satisfies (4.1). Then, it holds that:*[6]

$$N_E^*(\infty, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F) = \Theta\left(\left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^{-2/\nu}\right),$$

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)] = \Theta(1 - \rho)$, and where $\nu := \min(2(1+\gamma), \delta + \gamma)$.

The intuition is as follows. The safety constraint τ_I forces the incumbent company to partially align their predictor with the safety objective β_2. Since β_1 and β_2 point in different directions, this reduces the performance of the incumbent along β_1 as a side effect, resulting in strictly positive loss with respect to performance. On the other hand, since the new company faces no safety constraint, the new company can optimize entirely for performance along β_1. This means that the new company can enter the market as long as their finite data error is bounded by the incumbent's performance loss. We formalize this intuition in the following proof sketch.

Proof sketch of Theorem 5. The incumbent chooses the *infinite-data* ridgeless estimator $\beta(\alpha, 0)$ with mixture parameter $\alpha \in [0, 1]$ tuned so the safety violation is τ_I (Lemma 135). The resulting performance loss is $\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}$. Since the new company has no safety constraint, they choose the *single-objective* ridge regression estimator where $\alpha = 1$ and where λ is chosen optimally.[7] Theorem 6 (or alternatively, existing analyses of high-dimensional linear regression (e.g., Cui et al., 2021; Wei et al., 2022)) demonstrate the loss follows a scaling law of the form $\inf_{\lambda > 0} L_1(\hat{\beta}(1, \lambda, X)) = \Theta\left(N^{-\nu}\right)$ where $\nu := \min(2(1+\gamma), \delta + \gamma)$. The full proof is in Appendix B.1. $\square$

Theorem 5 reveals that the market-entry threshold is *finite* as long as (1) the safety constraint τ_I places nontrivial restrictions on the incumbent company and (2) the safety and performance objectives are not perfectly correlated. This result captures the notion that the new company can enter the market even after the incumbent company has accumulated an infinite amount of data.

Theorem 5 further illustrates how the market-entry threshold changes with other parameters (Figure 4.1). When safety and performance objectives are more correlated (i.e., when ρ is higher), the market-entry threshold increases, which increases barriers to entry. When the safety constraint for the incumbent is weaker (i.e., when τ_I is higher), the market-entry threshold also increases. Finally, when the power scaling parameters of the covariance and alignment coefficients increase, which increases the scaling law exponent ν, the market-entry threshold decreases.

[6]Throughout the paper, we allow $\Theta()$ and $O()$ to hide implicit constant which depends on the scaling exponents γ, δ.

[7]We formally rule out the possibility that $\alpha \neq 1$ using our multi-objective scaling law in Theorem 6.

4.4 Generalized Analysis of the Market-entry Threshold

To obtain a more fine-grained characterization of the market-entry threshold, we now consider more general environments. Our key technical tool is *multi-objective scaling laws*, which capture the performance of ridge regression in high-dimensional, multi-objective environments with finite data (Chapter 4.4.1). Using these scaling laws, we characterize the market-entry threshold when the incumbent has finite data (Chapter 4.4.2) and when the new company has a safety constraint (Chapter 4.4.3).

Our results in this section uncover the following conceptual insights about market entry. First, our main finding from Chapter 4.3—that the new company can enter the market with significantly less data than the incumbent—applies in many cases to these generalized environments. Moreover, our characterizations of N_E^* exhibit a *power-law-like dependence* with respect to the incumbent's dataset size (Theorem 8) and the difference in safety requirement for the two companies (Theorem 9). Interestingly, the scaling exponent c is not a constant across the full regime and instead takes on up to three different values. As a consequence, the new company can afford to scale up their dataset at a slower rate as the incumbent's dataset size increases, but needs to scale up their dataset at a faster rate as the two safety constraints become closer together. Proofs are deferred to Appendix B.2.

4.4.1 Technical tool: Scaling laws in multi-objective environments

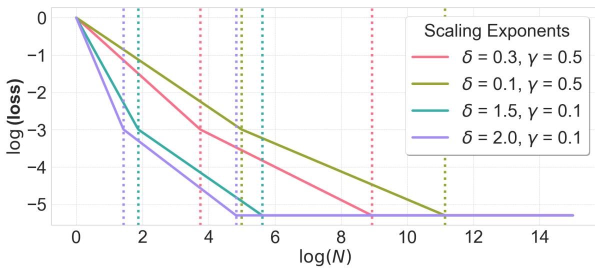
In this section, we give an overview of multi-objective scaling laws (see Chapter 4.5 for a more formal treatment and derivations). Our scaling laws capture how the ridge regression loss $L_1(\hat{\beta}(\alpha, \lambda, X))$ along the primary objective β_1 scales with the dataset size N, when the regularizer λ is optimally tuned to both N and problem-specified parameters. We show scaling laws for both the loss $\inf_{\lambda \in (0,1)} \mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X))]$ and the excess loss $\inf_{\lambda \in (0,1)}(\mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X)) - L_1(\beta(\alpha, 0))])$ where $\beta(\alpha, 0)$ is the infinite-data ridgeless regression estimator.

Scaling law for the loss. We first describe the scaling law for $\inf_{\lambda \in (0,1)} \mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X))]$ (Theorem 6; Figure 4.2a).

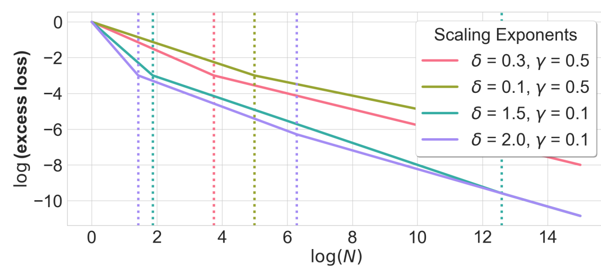
Theorem 6 (Informal Version of Corollary 12)**.** *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$. Suppose also that $P = \infty$ and $\alpha \geq 0.5$. Then, a deterministic equivalent for the expected loss under optimal regularization $\inf_{\lambda \in (0,1)} \mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X))]$ scales according to $N^{-\nu^*}$, where the scaling exponent ν^* is defined to be:*

$$\nu^* = \begin{cases} \nu & \text{if } N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \frac{\nu}{\nu+1} & \text{if } (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ 0 & \text{if } N \geq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}, \end{cases}$$

for $\nu := \min(2(1+\gamma), \delta+\gamma)$.



(a) Scaling law for loss (up to constants)



(b) Scaling law for excess loss (up to constants)

Figure 4.2: Data scaling laws for multi-objective environments where a fraction $\alpha = 0.9$ of the data is labelled according to the primary objective and a fraction $1 - \alpha = 0.1$ is labelled according to the secondary objective. The plots show, up to constants, the loss $\Theta(\inf_{\lambda \in (0,1)} \mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X))])$ (left, Theorem 6) and excess loss $\Theta(\inf_{\lambda \in (0,1)}(\mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X)) - L_1(\beta(\alpha, 0))]))$ (right, Theorem 7) as a function of the total number of training data points N. The loss and excess loss both take the form $N^{-\nu^*}$, but where the scaling exponent ν^* takes on *multiple (two or three) different values* depending on the size of N relative to other parameters. The scaling exponent is smaller when N is larger, thus demonstrating that the scaling rate becomes slower as the dataset size N increases.

Theorem 6 (Figure 4.2a) illustrates that the scaling rate becomes slower as the dataset size N increases. In particular, while the scaling exponent in single-objective environments is captured by a *single* value, Theorem 6 illustrates that the scaling exponent ν^* in multi-objective environments takes on *three different values*, depending on the size of N relative to other parameters. When N is small (the first regime), the scaling exponent $\nu^* = \nu$ is identical to that of the single-objective environment given by β_1. When N is a bit larger (the second regime), the scaling exponent *reduces* to $\nu^* = \nu/(\nu + 1) < \nu$. To make this concrete, if we take $\nu = 0.34$ to be an empirically estimated scaling law exponent for language models (Hoffmann et al., 2022), this would mean that $\nu^* \approx 0.34$ in the first regime and $\nu^* \approx 0.25$ in the second regime. Finally, when N is sufficiently large (the third regime), the scaling exponent reduces all the way to $\nu^* = 0$ and the only benefit of additional data is to improve constants on the loss.

Scaling law for the excess loss. We next turn to the excess loss, $\inf_{\lambda \in (0,1)}(\mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X)) - L_1(\beta(\alpha, 0))])$, which is normalized by the loss of the infinite-data ridgeless predictor $\beta(\alpha, 0)$. We show that the excess loss exhibits the same scaling behavior as the loss when N is sufficiently small, but exhibits different behavior when N is sufficiently large (Theorem 7; Figure 4.2b).

Theorem 7 (Informal Version of Corollary 14)**.** *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$. Suppose also that $P = \infty$ and $\alpha \geq 0.75$. Then, a deterministic equivalent for the expected loss under optimal regularization* $\inf_{\lambda \in (0,1)}(\mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X)) - L_1(\beta(\alpha, 0))])$ *scales according* *to $N^{-\nu^*}$, where the scaling exponent ν^* is defined to be:*

$$\nu^* = \begin{cases} \nu & \text{if } N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \frac{\nu}{\nu+1} & \text{if } (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}} \\ \frac{\nu'}{\nu'+1} & \text{if } N \geq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}, \end{cases}$$

for $\nu := \min(2(1+\gamma), \delta+\gamma)$ and $\nu' := \min(1+\gamma, \delta+\gamma)$.

Theorem 7 (Figure 4.2b) again shows that the scaling rate can become slower as the dataset size N increases, and again reveals three regimes of scaling behavior. While the first two regimes of Theorem 7 resemble the first two regimes of Theorem 6, the third regime of Theorem 7 (where $N \geq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}$) behaves differently. In this regime, the scaling exponent for the excess loss is $\frac{\nu'}{\nu'+1}$, rather than zero—this captures the fact that additional data can nontrivially improve the excess loss even in this regime, even though it only improves the loss up to constants. In terms of the magnitude of the scaling exponent $\frac{\nu'}{\nu'+1}$, it is *strictly smaller* than the scaling exponent $\frac{\nu}{\nu+1}$ when $\delta > 1$ and *equal* to the scaling exponent $\frac{\nu}{\nu+1}$ when $\delta \leq 1$.

4.4.2 Finite data for the incumbent

We compute N_E^* when the incumbent has finite data and the new company has no safety constraint (Theorem 8; Figure 4.3). The market-entry threshold N_E^* depends on the incumbent's dataset size N_I, the incumbent's performance loss G_I if they were to have infinite data but face the same safety constraint, the scaling exponents γ, δ, and the correlation coefficient ρ.

Theorem 8. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, and suppose that $P = \infty$. Assume that $\tau_E = \infty$. Suppose that the safety constraint τ_I satisfies (4.1). Then we have that $N_E^* = N_E^*(N_I, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F)$ satisfies:*

$$N_E^* := \begin{cases} \Theta\big(N_I\big) & \text{if } N_I \leq G_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ \Theta\left(N_I^{\frac{1}{\nu+1}} \cdot G_I^{-\frac{1}{2(\nu+1)}}(1-\rho)^{-\frac{1}{2(\nu+1)}}\right) & \text{if } G_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq G_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(G_I^{-\frac{1}{\nu}}\right) & \text{if } N_I \geq G_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}}, \end{cases}$$

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)] = \Theta(1-\rho)$, $G_I := (\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))})^2$, and $\nu = \min(2(1+\gamma), \delta+\gamma)$.

The market-entry threshold in Theorem 8 exhibits three regimes of behavior depending on N_I. In particular, the market-entry threshold takes the form $N_E^* = \Theta(N_I^c)$ where c decreases from 1 (in the first regime) to $\frac{1}{\nu+1}$ (in the second regime) to 0 (in the third regime) as N_I

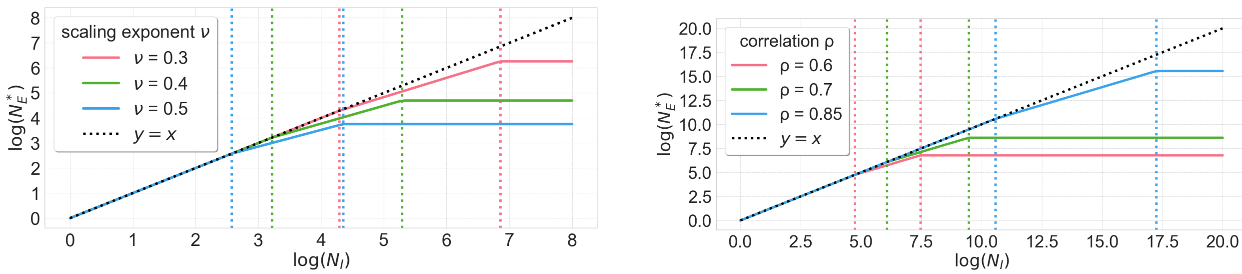


Figure 4.3: The market-entry threshold N_E^* as a function of the incumbent dataset size N_I, when the new company has no safety constraint (Theorem 8). The plots show varying values of the scaling exponent ν where the correlation parameter $\rho = 0.5$ is held fixed (left) and varying values of ρ where $\nu = 0.34$ is held fixed (right). When N_I is sufficiently large, the market-entry threshold N_E^* is asymptotically less than N_I (i.e., below the dotted black line). Each curve is the union of three line segments with slope decreasing in N_I. This demonstrates that the new company can afford to scale up their dataset at a slower rate than the incumbent, when the incumbent's dataset size is sufficiently large.

increases. To connect this to large-language-model marketplaces, we directly set $\nu = 0.34$ to be the empirically estimated scaling law exponent for language models (Hoffmann et al., 2022); in this case, the scaling exponent c ranges from 1 to 0.75 to 0. The fact that there are three regimes come from the scaling law derived in Theorem 6, as the following proof sketch illustrates.

Proof sketch. The key technical tool is the scaling law for the loss $\inf_{\lambda \in (0,1)} \mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X))]$ (Theorem 6), which has three regimes of scaling behavior for different values of N. We apply the scaling law to analyze the performance of the incumbent, who faces a safety constraint and has finite data. Analyzing the performance of the new company—who faces no safety constraint—is more straightforward, given that the new company can set $\alpha_E = 1$. We compute N_E^* as the number of data points needed to match the incumbent's performance level. The full proof is deferred to Appendix B.2.1. □

Theorem 8 reveals that the new company can enter the market with $N_E^* = o(N_I)$ data, as long as the incumbent's dataset is sufficiently large (i.e., $N_I \geq G_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}}$). The intuition is when there is sufficient data, the multi-objective scaling exponent is worse than the single-objective scaling exponent (Theorem 6). The incumbent thus faces a worse scaling exponent than the new company, so the new company can enter the market with asymptotically less data.

The three regimes in Theorem 8 further reveal that the market-entry threshold N_E^* scales at a slower rate as the incumbent's dataset size N_I increases (Figure 4.3). The intuition is that the multi-objective scaling exponent ν^* faced by the incumbent decreases as dataset size increases, while the single-objective scaling exponent faced by the new company is constant in dataset size (Theorem 6). The incumbent thus becomes less efficient at using additional data to improve performance, while the new company's efficiency in using additional data remains unchanged.

Theorem 8 also offers finer-grained insight into the market-entry threshold in each regime. In the first regime, where the incumbent's dataset is small, the threshold N_E^* matches the incumbent dataset size. This means the new company does need as much data as the incumbent to enter the market, even though the new company faces a less stringent safety constraint. In the second (intermediate) regime, the new company can enter with a dataset size proportional to $N_I^{1/(\nu+1)}$. This *polynomial speedup* illustrates that the new company can more efficiently use additional data to improve performance than the incumbent company. A caveat is that this regime is somewhat restricted in that the ratio of the upper and lower boundaries is bounded. In the third regime, where the incumbent's dataset size is large, the market-entry threshold N_E^* matches the market-entry threshold from Theorem 5 where the incumbent has *infinite* data.

4.4.3 Safety constraint for the new company

We compute N_E^* when the new company has a nontrivial safety constraint and the incumbent has infinite data. For this result, we strengthen the conditions on τ_E and τ_I from (4.1), instead requiring:

$$\tau_E >_{(A)} \tau_I \geq_{(B)} \mathbb{E}_{(\beta_1,\beta_2)\sim\mathcal{D}_W}[L_2^*(\beta_1,\beta_2,\mathcal{D}_F,0.75)], \tag{4.2}$$

where (4.2) replaces the 0.5 with a 0.75 in the right-most quantity.[8]

We state the result below (Theorem 9; Figure 4.4). The market-entry threshold N_E^* depends on the incumbent's safety constraint τ_I, the performance loss G_I (resp. G_E) if the incumbent (resp. new company) had infinite data and faced the same safety constraint, the difference $D = G_I - G_E$ in infinite-data performance loss achievable by the incumbent and new company, the scaling exponents γ, δ, and the correlation coefficient ρ.

Theorem 9. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, and suppose that $P = \infty$. Suppose that the safety constraints τ_I and τ_E satisfy (4.2). Then it holds that $N_E^* = N_E^*(\infty, \tau_I, \tau_E, \mathcal{D}_W, \mathcal{D}_F)$ satisfies:*

$$N_E^* := \begin{cases} \Theta(D^{-\frac{1}{\nu}}) & \text{if } D \geq G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(D^{-\frac{\nu+1}{\nu}} G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}}\right) & \text{if } G_E^{\frac{\nu}{2(\nu-\nu')}}(1-\rho)^{\frac{\nu}{2(\nu-\nu')}} \leq D \leq G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(\left(D \cdot G_E^{-\frac{1}{2}}(1-\rho)^{-\frac{1}{2}}\right)^{-\frac{\nu'+1}{\nu'}}\right) & \text{if } D \leq G_E^{\frac{\nu}{2(\nu-\nu')}}(1-\rho)^{\frac{\nu}{2(\nu-\nu')}}, \end{cases}$$

[8] Inequality (B) in (4.2) requires that the safety constraint still allows both company to label 75% of their training data according to performance-optimal outputs. We make this modification, since our analysis of multi-objective scaling laws for the *excess* loss assumes $\alpha \geq 0.75$ (see Chapter 4.5.3).

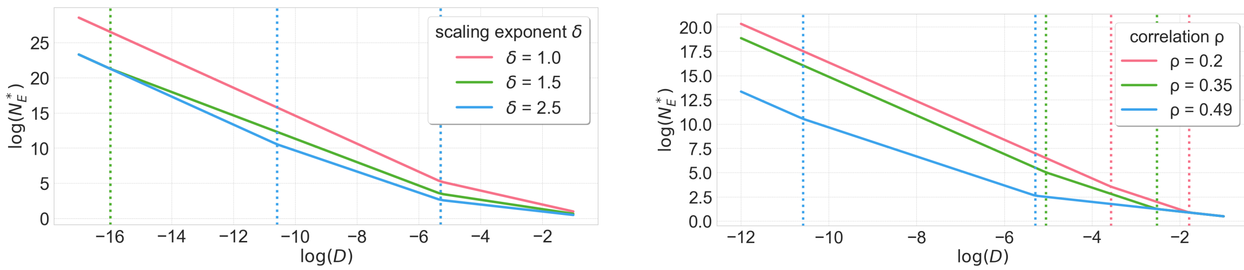


Figure 4.4: The market-entry threshold N_E^* as a function of the difference D between the infinite-data performance loss of the incumbent and new company, when the incumbent has infinite data (Theorem 9). The plots show varying values of the scaling exponent δ where the correlation parameter $\rho = 0.49$ is held fixed (left) and varying values of ρ where $\delta = 2.5$ is held fixed (right). The plots are shown in log space. The market-entry threshold is finite in all cases. Each curve is the union of multiple line segments with slope increasing in magnitude as $\log D$ decreases, demonstrating that the new company needs to scale up their dataset at a faster rate as the safety thresholds become closer together.

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta)] = \Theta(1-\rho)$, *where* $\nu = \min(2(1+\gamma), \delta + \gamma)$ *and* $\nu' = \min(1+\gamma, \delta+\gamma)$, *where* $G_I := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2$ *and* $G_E := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_E, L^*(\rho))}\right)^2$, *and where* $D := G_I - G_E$.

The market-entry threshold in Theorem 9 also exhibits three regimes of behavior depending on the difference D in the infinite-data performance loss achievable by the incumbent and new company. In particular, the market-entry threshold takes the form $N_E^* = \Theta(D^{-c})$ where c increases from $\frac{1}{\nu}$ to $\frac{\nu+1}{\nu}$ to $\frac{\nu'+1}{\nu'}$ as D decreases. (The third regime only exists when $\delta > 1$.) To connect this to large-language-model marketplaces, if we take $\nu = 0.34$ to be the empirically estimated scaling law exponent for language models (Hoffmann et al., 2022), then c would range from 2.94 to 3.94 to potentially even larger. The fact that there are three regimes come from the scaling law derived in Theorem 7, as the following proof sketch illustrates.

Proof sketch. The key technical tool is the scaling law for the *excess loss* $\inf_{\lambda \in (0,1)}(\mathbb{E}[L_1(\hat{\beta}(\alpha, \lambda, X)) - L_1(\beta(\alpha, 0))])$ (Theorem 7), which has three regimes of scaling behavior for different values of N. We apply the scaling law to analyze the performance of the new company, who faces a safety constraint and has finite data. Analyzing the performance of the incumbent—who has infinite data—is more straightforward, and the incumbent's performance loss is $G_I = D + G_E$. We compute the number of data points N_E^* needed for the new company to achieve an excess loss of D. The full proof is deferred to Appendix B.2.2. $\square$

Theorem 9 illustrates that the new company can enter the market with finite data N_E^*, as long as the safety constraint τ_E placed on the new company is strictly weaker than the constraint τ_I placed on the incumbent company (inequality (A) in (4.2)). This translates to the difference D being strictly positive. The intuition is that when the new company faces a weaker safety constraint, it can train on a greater number of data points labelled with the performance objective β_1, which improves performance.

The three regimes in Theorem 9 further reveal that the market-entry threshold N_E^* scales at a faster rate as the difference D between the two safety constraints decreases (Figure 4.3). The intuition is since the new company needs to achieve an excess loss of at most D, the new company faces a smaller multi-objective scaling exponent ν^* as D decreases (Theorem 7). The new company thus becomes less efficient at using additional data to improve performance.

4.5 Deriving Scaling Laws for Multi-Objective Environments

We formalize and derive our multi-objective scaling laws for the loss (Theorem 6) and excess loss (Theorem 7). Recall that the problem setting is high-dimensional ridge regression when a fraction α of the training data is labelled according to β_1 and the rest is labelled according to an alternate objective β_2. First, following the style of analysis of single-objective ridge regression (e.g., Cui et al., 2021; Wei et al., 2022), we first compute a *deterministic equivalent* of the loss (Chapter 4.5.1). Then we derive the scaling law under the power scaling assumptions on the eigenvalues and alignment coefficients in Chapter 4.2.3, both for the loss (Chapter 4.5.2) and for the excess loss (Chapter 4.5.3). Proofs are deferred to Appendix B.3.

4.5.1 Deterministic equivalent

We show that the loss of the ridge regression estimator can be approximated as a deterministic quantity. This analysis builds on the random matrix tools in Bach (2024) (see Appendix B.4). Note that our derivation of the deterministic equivalent does *not* place the power scaling assumptions on the eigenvalues or alignment coefficients; in fact, it holds for any linear regression setup which satisfies a standard random matrix theory assumption (Assumption 7).

We compute the following deterministic equivalent (proof deferred to Appendix B.3.5).[9]

Lemma 10. *Suppose that $N \geq 1$, $P \geq 1$, $\mathcal{D}_F$, β_1, and β_2 satisfy Assumption 7. Let Σ be the covariance matrix of $\mathcal{D}_F$, and let $\alpha \in [0,1]$ and $\lambda \in (0,1)$ be general parameters. Let $\Sigma_c = (\Sigma + cI)$ for $c \geq 0$, let $B^{\boldsymbol{sn}} = \beta_1 \beta_1^T$, let $B^{\boldsymbol{df}} = (\beta_1 - \beta_2)(\beta_1 - \beta_2)^T$, and let*

[9]Following Bach (2024), the asymptotic equivalence notation $u \sim v$ means that u/v tends to 1 as N and P go to ∞.

$B^{\mathtt{mx}} = (\beta_1 - \beta_2)\beta_1^T$. *Let* $\kappa = \kappa(\lambda, N, \Sigma)$ *from Definition 18. Then, it holds that*

$$L_1(\hat{\beta}(\alpha, \lambda, X)) \sim L_1^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) =: \frac{T_1 + T_2 + T_3 + T_4 + T_5}{Q},$$

where:

$$T_1 := \kappa^2 \cdot \mathrm{Tr}(\Sigma\Sigma_\kappa^{-2} B^{\mathtt{sn}}), \quad T_2 := (1-\alpha)^2 \left(\mathrm{Tr}\left(\Sigma_\kappa^{-2}\Sigma^3 B^{\mathtt{df}}\right)\right)$$

$$T_3 := 2(1-\alpha)\kappa \cdot \mathrm{Tr}\left(\Sigma_\kappa^{-2}\Sigma^2 B^{\mathtt{mx}}\right), \quad T_4 := -2(1-\alpha)\kappa\frac{1}{N}\mathrm{Tr}(\Sigma^2\Sigma_\kappa^{-2}) \cdot \mathrm{Tr}\left(\Sigma_\kappa^{-1}\Sigma B^{\mathtt{mx}}\right),$$

$$T_5 := (1-\alpha)\frac{1}{N}\mathrm{Tr}(\Sigma^2\Sigma_\kappa^{-2}) \cdot \left(\mathrm{Tr}\left(\Sigma B^{\mathtt{df}}\right) - 2(1-\alpha)\,\mathrm{Tr}\left(\Sigma_\kappa^{-1}\Sigma^2 B^{\mathtt{df}}\right)\right), \quad Q := 1 - \frac{1}{N}\mathrm{Tr}(\Sigma^2\Sigma_\kappa^{-2}).$$

Lemma 10 shows that the loss can be approximated by a deterministic quantity $L_1^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ which is sum of five terms, normalized by the standard degrees of freedom correction Q^{-1} (Bach, 2024). The sum $T_1 + T_2 + T_3$ is the loss of infinite-data ridge regression with regularizer κ. Terms T_4 and T_5 capture additional error terms.

In more detail, term T_1/Q captures the standard single-objective environment error for N data points (Bach, 2024): i.e., the population error of the single-objective linear regression problem with regularizer λ where all of the N training data points are labelled with β_i. Term T_2 is similar to the infinite-data ridgeless regression error but is slightly smaller due to regularization. Term T_3 is a cross term which is upper bounded by the geometric mean of term T_1 and term T_2. Term T_4 is another cross term which is subsumed by the other terms. Term T_5 captures an overfitting error which increases with the regularizer κ and decreases with the amount of data N.

From deterministic equivalents to scaling laws. In the following two subsections, using the deterministic equivalent from Lemma 10, we derive *scaling laws*. We make use of the the power scaling assumptions on the covariance and alignment coefficients described in Chapter 4.5.2, under which the deterministic equivalent takes a cleaner form (Lemma 150 in Appendix B.3). We note that strictly speaking, deriving scaling laws requires controlling the error of the deterministic equivalent relative to the actual loss; for simplicity, we do not control errors and instead directly analyze the deterministic equivalent.

4.5.2 Scaling law for the loss

We derive scaling laws for the loss $L_1^{\mathtt{det}} := L_1^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$. We first prove the following scaling law for a general regularizer λ (proof deferred to Appendix B.3.7).

Theorem 11. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents* $\gamma, \delta > 0$ *and correlation coefficient* $\rho \in [0, 1)$*, suppose that* $P = \infty$*. Assume that* $\alpha \geq 0.5$ *and* $\lambda \in (0, 1)$*. Let* $L_1^{\mathtt{det}} := L_1^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ *be the deterministic equivalent* *from Lemma 10. Let* $\nu := \min(2(1+\gamma), \delta+\gamma)$. *Then, the expected loss* $\mathbb{E}_{\mathcal{D}_W}[L_1^{\mathtt{det}}]$ *equals:*

$$\Theta\left(\underbrace{\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu})}_{\text{finite data error}} + \underbrace{(1-\alpha)^2 \cdot (1-\rho)}_{\text{mixture error}} + \underbrace{(1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho)}_{\text{overfitting error}}\right).$$

Theorem 11 illustrates that the loss is the sum of a *finite data error*, an *overfitting error*, and a *mixture error*. The finite data error for $L_1^{\mathtt{det}}$ matches the loss in the single-objective environment for N data points labelled with objective β_1. The mixture error equals the loss of the infinite-data ridgeless regression predictor $\beta(\alpha, 0)$. The overfitting error for $L_1^{\mathtt{det}}$ equals the error incurred when the regularizer λ is too small. This term is always at most $(1-\alpha)^{-1}$ times larger than the mixture error, and it is smaller than the mixture error when λ is sufficiently large relative to N.

Due to the overfitting error, the optimal loss is *not* necessarily achieved by taking $\lambda \to 0$ for multi-objective linear regression. In fact, if the regularizer decays too quickly as a function of N (i.e., if $\lambda = O(N^{-1-\gamma})$), then the error would converge to $(1-\alpha)(1-\rho)$, which is a factor of $(1-\alpha)^{-1}$ higher than the error of the infinite-data ridgeless predictor $\beta(\alpha, 0)$. The fact that $\lambda \to 0$ is suboptimal reveals a sharp disconnect between the multi-objective setting and the single-objective setting where no explicit regularization is necessary to achieve the optimal loss (see, e.g., Cui et al., 2021; Wei et al., 2022).[10]

In the next result, we compute the optimal regularizer and derive a scaling law under optimal regularization as a corollary of Theorem 11.

Corollary 12 (Formal version of Theorem 6)**.** *Consider the setup of Theorem 11. Then, the loss* $\inf_{\lambda \in (0,1)} \mathbb{E}_{\mathcal{D}_W}[L_1^{\mathtt{det}}]$ *under optimal regularization can be expressed as:*

$$\begin{cases} \Theta\left(N^{-\nu}\right) & \text{if } N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \Theta\left(\left(\frac{N}{(1-\alpha)(1-\rho)}\right)^{-\frac{\nu}{\nu+1}}\right) & \text{if } (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \Theta((1-\alpha)^2(1-\rho)) & \text{if } N \geq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}, \end{cases}$$

where $\nu := \min(2(1+\gamma), \delta+\gamma)$.

The scaling law exponent ν^* ranges from ν, to $\nu/(\nu+1)$, to 0 (Figure 4.2a). To better understand each regime, we provide intuition for when error term from Theorem 11 dominates, the form of the optimal regularizer, and the behavior of the loss.

- *Regime 1:* $N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$. Since N is small, the finite data error dominates regardless of λ. As a result, like in a single-objective environment, taking $\lambda = O(N^{-1-\gamma})$

[10] Tempered overfitting (Mallinar et al., 2022) can similarly occur in single-objective settings with *noisy observations*. In this sense, labelling some of the data with the alternate objective β_2 behaves qualitatively similarly to noisy observations.

recovers the optimal loss up to constants. Note that the loss thus behaves as if all N data points were labelled according to β_i: the learner benefits from *all* of the data, not just the data is labelled according to β_i.

- *Regime 2:* $(1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$. In this regime, the finite error term and overfitting error dominate. Taking $\lambda = \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{1+\gamma}{\nu+1}}\right)$, which equalizes the two error terms, recovers the optimal loss up to constants. The loss in this regime improves with N, but at a slower rate than in a single-objective environment.

- *Regime 3:* $N \geq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$. Since N is large, the mixture and the overfitting error terms dominate. Taking $\lambda = \Theta((N(1-\alpha))^{-1-\gamma})$, which equalizes the two error terms, recovers the optimal loss up to constants. The loss behaves (up to the constants) as if there were *infinitely many data points* from the mixture distribution with weight α. This is the minimal possible loss and there is thus no additional benefit for data beyond improving constants.

The full proof of Corollary 12 is deferred to Appendix B.3.8.

4.5.3 Scaling law for the excess loss

Now, we turn to scaling laws for the excess loss $\mathbb{E}_{\mathcal{D}_W}[L_1^{\texttt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) - L_1(\beta(\alpha, 0))]$, , which is normalized by the loss of the infinite-data ridgeless predictor $\beta(\alpha, 0)$. We first prove the following scaling law for a general regularizer λ, assuming that $\alpha \geq 0.75$ (proof deferred to Appendix B.3.9).[11]

Theorem 13. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, suppose that $P = \infty$. Assume that $\alpha \geq 0.75$ and $\lambda \in (0,1)$. Let $L_1^{\texttt{det}} := L_1^{\texttt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ be the deterministic equivalent from Lemma 10. Let $\nu := \min(2(1+\gamma), \delta+\gamma)$ and let $\nu' = \min(1+\gamma, \delta+\gamma)$ Then, the expected loss $\mathbb{E}_{\mathcal{D}_W}[L_1^{\texttt{det}} - L_1(\beta(\alpha, 0))]$ equals:*

$$\Theta\left(\underbrace{\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu})}_{\textit{finite data error}} + \underbrace{(1-\rho)(1-\alpha)\max(\lambda^{\frac{\nu'}{1+\gamma}}, N^{-\nu'})}_{\textit{mixture finite data error}} + \underbrace{(1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho)}_{\textit{overfitting error}}\right).$$

Theorem 13 illustrates that the loss is the sum of a *finite data error*, an *overfitting error*, and a *mixture finite data error*. In comparison with Theorem 11, the difference is that the mixture error is replaced by the mixture finite data error. Interestingly, the mixture finite

[11]The assumption that $\alpha \geq 0.75$ simplifies the closed-form expression for the deterministic equivalent of the excess loss in Lemma 150. We defer a broader characterization of scaling laws for the excess loss to future work.

data error exhibits a different asymptotic dependence with respect to λ and N than the finite data error: the asymptotic rate of decay scales with ν' rather than ν. In fact, the rate is *slower* for the mixture finite data error than the finite data error as long as $\delta > 1$ (since this means that $\nu' < \nu$).

Since the optimal excess loss is also not necessarily achieved by taking $\lambda \to 0$, we compute the optimal regularizer for the excess loss and derive a scaling law under optimal regularization as a corollary of Theorem 13.

Corollary 14 (Formal version of Theorem 7)**.** *Consider the setup of Theorem 13. The excess loss under optimal regularization can be expressed as:*

$$\inf_{\lambda \in (0,1)} (\mathbb{E}_{\mathcal{D}_W}[L_1^{\boldsymbol{det}} - L_1(\beta(\alpha, 0))])$$
$$= \begin{cases} \Theta\left(N^{-\nu}\right) & \text{if } N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \Theta\left(\left(\frac{N}{(1-\alpha)(1-\rho)}\right)^{-\frac{\nu}{\nu+1}}\right) & \text{if } (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}} \\ \Theta\left((1-\alpha)(1-\rho)N^{-\frac{\nu'}{\nu'+1}}\right) & \text{if } N \geq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}, \end{cases}$$

where $\nu := \min(2(1+\gamma), \delta+\gamma)$ *and* $\nu' = \min(1+\gamma, \delta+\gamma)$.

The scaling law exponent ν^* ranges from ν, to $\nu/(\nu+1)$, to $\nu'/(\nu'+1)$ (Figure 4.2b). The first two regimes behave similarly to Corollary 12, and the key difference arises in the third regime (when N is large). In the third regime ($N \geq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}$), the mixture finite data error and the overfitting error terms dominate. Taking $\lambda = \Theta\left(N^{-\frac{1+\gamma}{\nu'+1}}\right)$—which equalizes these two error terms—recovers the optimal loss up to constants. The resulting scaling behavior captures that in this regime, additional data meaningfully improves the *excess loss*, even though additional data only improves the loss in terms of constants. The full proof of Corollary 14 is deferred to Appendix B.3.10.

4.6 Discussion

We studied market entry in marketplaces for machine learning models, showing that pressure to satisfy safety constraints can reduce barriers to entry for new companies. We modelled the marketplace using a high-dimensional multi-objective linear regression model. Our key finding was that a new company can consistently enter the marketplace with significantly less data than the incumbent. En route to proving these results, we derive scaling laws for multi-objective regression, showing that the scaling rate becomes slower when the dataset size is large.

Potential implications for regulation. Our results have nuanced design consequences for regulators, who implicitly influence the level of safety that each company needs to achieve to avoid reputational damage. On one hand, our results suggest that placing greater scrutiny on dominant companies can encourage market entry and create a more competitive marketplace of companies. On the other hand, market entry does come at a cost to the safety objective: the smaller companies exploit that they can incur more safety violations while maintaining their reputation, which leads to a race to the bottom for safety. Examining the tradeoffs between market competitiveness and safety compliance is an important direction for future work.

Barriers to market entry for online platforms. While we focused on language models, we expect that our conceptual findings about market entry also extend to recommendation and social media platforms.

In particular, our motivation and modeling assumptions capture key aspects of these online platforms. Policymakers have raised concerns have been raised about barriers to entry for social media platforms (Stigler Committee, 2019), motivated by the fact that social media platforms such as X and Facebook each have over a half billion users (Statista, 2024; Ingram, 2024). Incumbent companies risk reputational damage if their model violates safety-oriented objectives—many recommendation platforms have faced scrutiny for promoting hate speech (European Union, 2022a), divisive content (Rathje et al., 2021), and excessive use by users (Hasan et al., 2018), even when recommendations perform well in terms of generating user engagement. This means that incumbent platforms must balance optimizing engagement with controlling negative societal impacts (Bengani et al., 2022). Moreover, new companies face less regulatory scrutiny, given that some regulations explicitly place more stringent requirements on companies with large user bases: for example the Digital Services Act (European Union, 2022a) places a greater responsibility on Very Large Online Platforms (with over 45 million users per month) to identify and remove illegal or harmful content.

Given that incumbent platforms similarly face more pressure to satisfy safety-oriented objectives, our results suggest that multi-objective learning can also reduce barriers to entry for new online platforms.

Limitations. Our model for interactions between companies and users makes several simplifying assumptions. For example, we focused entirely whether the new company can enter the market, which leaves open the question of whether the new company can survive in the long run. Moreover, we assumed that all users choose the model with the highest overall performance. However, different users often care about performance on different queries; this could create an incentive for specialization, which could also reduce barriers to entry and market concentration. Finally, we focused on direct interactions between companies and users, but in reality, downstream providers sometimes build services on top of a foundation model. Understanding how these market complexities affect market entry as well as long-term concentration is an interesting direction for future work.

Furthermore, our model also made the simplifying assumption that performance and safety trade off according to a multi-objective regression problem. However, not all safety objectives fit the mold of linear coefficients within linear regression. For some safety objectives such as privacy, we still expect that placing greater scrutiny on dominant companies could similarly reduce barriers to entry. Nonetheless, for other safety or societal considerations, we do expect that the implications for market entry might be fundamentally different. For example, if the safety objective is a multi-group performance criteria, and there is a single predictor that achieves zero accuracy on all distributions, then a dominant company with infinite data would be able to retain all users even if the company faces greater scrutiny. Extending our model to capture a broader scope of safety objectives is a natural direction for future work.

Chapter 5

Recommendation Platforms

This chapter is based on *Competition, Alignment, and Equilibria in Digital Marketplaces* (Jagadeesan et al., 2023c), which is on joint work with Michael I. Jordan and Nika Haghtalab.

5.1 Introduction

Recommendation systems are the backbone of numerous digital platforms—from web search engines to video sharing websites to music streaming services. To produce high-quality recommendations, these platforms rely on data which is obtained through interactions with users. This fundamentally links the quality of a platform's services to how well the platform can attract users.

What a platform must do to attract users depends on the amount of competition in the marketplace. If the marketplace has a single platform—such as Google prior to Bing or Pandora prior to Spotify—then the platform can accumulate users by providing any reasonably acceptable quality of service given the lack of alternatives. This gives the platform great flexibility in its choice of recommendation algorithm. In contrast, the presence of competing platforms makes user participation harder to achieve and intuitively places greater constraints on the recommendation algorithms. This raises the questions: *how does competition impact the recommendation algorithms chosen by digital platforms? How does competition affect the quality of service for users?*

Conventional wisdom tells us that competition benefits users. In particular, users vote with their feet by choosing the platform on which they participate. The fact that users have this power forces the platforms to fully cater to user choices and thus improves user utility. This phenomenon has been formalized in classical markets where firms produce homogenous products (Baye and Kovenock, 2008), where competition has been established to perfectly align market outcomes with user utility. Since user wellbeing is considered central to the healthiness of a market, perfect competition is traditionally regarded as the "gold standard" for a healthy marketplace: this conceptual principle underlies measures of market power (Lerner, 1934) and antitrust policy (Gellhorn, 1975).

In contrast, competition has an ambiguous relationship with user wellbeing in digital marketplaces, where digital platforms are data-driven and compete via recommendation algorithms that rely on data from user interactions. Informally speaking, these marketplaces exhibit an interdependency between user utility, the platforms' choices of recommendation algorithms, and the collective choices of other users. In particular, the size of a platform's user base impacts how much data the platform has and thus the quality of its service; as a result, an individual user's utility level depends on the number of users that the platform has attracted thus far. Having a large user base enables a platform to have an edge over competitors without fully catering to users, which casts doubt on whether classical alignment insights apply to digital marketplaces.

The ambiguous role of competition in digital marketplaces—which falls outside the scope of our classical understanding of competition power—has gained center stage in recent policymaking discourse. Indeed, several interdisciplinary policy reports (Stigler Committee, 2019; Crémer et al., 2019) have been dedicated to highlighting ways in which the structure of digital marketplaces fundamentally differs from that of classical markets. For example, these reports suggest that data accumulation can encourage market tipping, which leaves users particularly vulnerable to harm (as we discuss in more detail at the end of Chapter 5.1.1). Yet, no theoretical foundation has emerged to formally examine the market structure of digital marketplaces and assess potential interventions.

5.1.1 Our contributions

Our work takes a step towards building a theoretical foundation for studying competition in digital marketplaces. We present a framework for studying platforms that compete on the basis of learning algorithms, focusing on alignment with user utility at equilibrium. We consider a stylized duopoly model based on a multi-armed bandit problem where user utility depends on the incurred rewards. We show that *competition may no longer perfectly align market outcomes with user utility*. Nonetheless, we find that market outcomes exhibit *a weaker form of alignment*: the user utility is at least as large as the optimal utility in a population with only one user. Interestingly, there can be multiple equilibria, and the gap between the best equilibria and the worst equilibria can be substantial.

Model. We consider a market with two platforms and a population of users. Each platform selects a bandit algorithm from a class $\mathcal{A}$. After the platforms commit to algorithms, each user decides which platform they wish to participate on. Each user's utility is the (potentially discounted) cumulative reward that they receive from the bandit algorithm of the platform that they chose. Users arrive at a Nash equilibrium.[1] Each platform's utility is the number of users who participate on that platform, and the platforms arrive at a Nash equilibrium. The platforms either maintain *separate data repositories* about the rewards of their own users, or the platforms maintain a *shared data repository* about the rewards of all users.

[1] In Chapter 5.2, we will discuss subtleties that arise from having multiple Nash equilibria.

Alignment results. To formally consider alignment, we introduce a metric—that we call the *user quality level*—that captures the utility that a user would receive when a given pair of competing bandit algorithms are implemented and user choices form an equilibrium. Table 5.1 summarizes the alignment results in the case of a single user and multiple users. A key quantity that appears in the alignment results is $R_{A'}(n)$, which denotes the expected utility that a user receives from the algorithm A' when n users all participate in the same algorithm.

For the case of a single user, an *idealized* form of alignment holds: the user quality level at any equilibrium is the optimal utility $\max_{A'} R_{A'}(1)$ that a user can achieve within the class of algorithms $\mathcal{A}$. Idealized alignment holds regardless of the informational assumptions on the platform.

The nature of alignment fundamentally changes when there are multiple users. At a high level, we show that idealized alignment breaks down since the user quality level is no longer guaranteed to be the global optimum, $\max_{A'} R_{A'}(N)$, that cooperative users can achieve. Nonetheless, a weaker form of alignment holds: the user quality level nonetheless never falls below the single-user optimum $\max_{A'} R_{A'}(1)$. Thus, the presence of other users cannot make a user worse off than if they were the only participant, but users may not be able to fully benefit from the data provided by others.

More formally, consider the setting where the platforms have separate data repositories. We show that there can be many qualitatively different Nash equilibria for the platforms. The user quality level across all equilibria actually spans the full set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$; i.e., any user quality level is realizable in some Nash equilibrium of the platforms and its associated Nash equilibrium of the users (Theorem 16). Moreover, the user quality level at any equilibrium is contained in the set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ (Theorem 17). When the number of users N is large, the gap between $\max_{A'} R_{A'}(1)$ and $\max_{A'} R_{A'}(N)$ can be significant since the latter is given access to N times as much data at each time step than the former. The fact that the single-user optimum $\max_{A'} R_{A'}(1)$ is realizable means that the market outcome might only exhibit a weak form of alignment. The intuition behind this result is that the performance of an algorithm is controlled not only by its efficiency in transforming information to action, but also by the level of data it has gained through its user base. Since platforms have separate data repositories, a platform can thus make up for a suboptimal algorithm by gaining a significant user base. On the other hand, the global optimal user quality level $R_{A'}(N)$ is nonetheless realizable—this suggests that equilibrium selection could be used to determine when bad equilibria arise and to nudge the marketplace towards a good equilibrium.

What if the platforms were to share data? At first glance, it might appear that with data sharing, a platform can no longer make up for a suboptimal algorithm with data, and the idealized form of alignment would be recovered. However, we construct two-armed bandit problem instances where every symmetric equilibrium for the platforms has user quality level strictly below the global optimal $\max_{A'} R_{A'}(N)$ (Theorems 18-19). The mechanism for this suboptimality is that the global optimal solution requires “too much” exploration. If other users engage in their “fair share” of exploration, an individual user would prefer to explore less and free-ride off of the data obtained by other users. The platform is thus forced to explore

	Single user	**Multiple users**
Separate data repositories	$\max_{A'} R_{A'}(1)$	$[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$
Shared data repository	$\max_{A'} R_{A'}(1)$	subset of $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ (*strict* subset in safe-risky arm problem)

Table 5.1: User quality level of the Nash equilibrium for the platforms. A marketplace with a single user exhibits *idealized alignment*, where the user quality level is maximized. A marketplace with multiple users can have equilibria with a vast range of user quality levels—although *weak alignment* always holds—-and there are subtle differences between the separate and shared data settings.

less, which drives down the user quality level. To formalize this, we establish a connection to strategic experimentation (Bolton and Harris, 1999). We further show that although all of the user quality levels in $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ may not be realizable, the user quality level at any symmetric equilibria is still guaranteed to be within this set (Theorem 21).

Connection to policy reports. Our work provides a mathematical explanation of phenomena documented in recent policy reports (Stigler Committee, 2019; Crémer et al., 2019). The first phenomena that we consider is *market dominance from data accumulation*. The accumulation of data has been suggested to result in winner-takes-all-markets where a single player dominates and where market entry is challenging (Stigler Committee, 2019). The data advantage of the dominant platform can lead to lower quality services and lower user utility. Theorems 16-17 formalize this mechanism. We show that once a platform has gained the full user base, market entry is impossible and the platform only needs to achieve weak alignment with user utility to retain its user base (see discussion in Chapter 5.4.2). The second phenomena that we consider is the *impact of shared data access*. While the separate data setting captures much of the status quo of proprietary data repositories in digital marketplaces, sharing data access has been proposed as a solution to market dominance (Crémer et al., 2019). Will shared data access deliver on its promises? Theorems 18-19 highlight that sharing data does not solve the alignment issues, and uncovers free-riding as a mechanism for misalignment.

5.1.2 Related work

We discuss the relation between our work and research on *competing platforms*, *incentivizing exploration*, and *strategic experimentation*.

Competing platforms. Aridor et al. (2025) examine the interplay between competition and exploration in bandit problems in a duopoly economy with fully myopic users. They focus on platform regret, showing that platforms must both choose a greedy algorithm at equilibrium and thus illustrating that competition is at odds with regret minimization. In contrast, we take a user-centric perspective and investigate alignment with user utility. Interestingly, the findings in Aridor et al. (2025) and our findings are not at odds: the result in Aridor et al. (2025) can be viewed as alignment (since the optimal choice for a fully myopic user results in regret in the long run), and our results similarly recover idealized alignment in the special case when users are fully myopic. Going beyond the setup of Aridor et al. (2025), we investigate non-myopic users and allow multiple users to arrive at every round, and we show that alignment breaks down in this general setting.

Outside of the bandits framework, another line of work has also studied the behavior of competing learners when users can choose between platforms. Ben-Porat and Tennenholtz (2017; 2019) study equilibrium predictors chosen by competing *offline* learners in a PAC learning setup. Other work has focused on the dynamics when multiple learners apply out-of-box algorithms, showing that specialization can emerge (Ginart et al., 2021; Dean et al., 2024a) and examining the role of data purchase (Kwon et al., 2022); however, these works do not consider which algorithms the learners are incentivized to choose to gain users. In contrast, we investigate *equilibrium* bandit algorithms chosen by *online* learners, each of whom aims to maximize the size of its user base. The interdependency between the platforms' choices of algorithms, the data available to the platforms, and the users' decisions in our model drives our alignment insights.

Other aspects of competing platforms that have been studied include competition under exogeneous network effects (Rysman, 2009; Weyl and White, 2014), experimentation in price competition (Bergemann and Välimäki, 2000), dueling algorithms which compete for a single user (Immorlica et al., 2011), competition when firms select scalar innovation levels whose cost decreases with access to more data (Prüfer and Schottmüller, 2021), and measures of a digital platform's power in a marketplace (Hardt et al., 2022).

Incentivizing exploration. This line of work has examined how the availability of outside options impacts bandit algorithms. Kremer et al. (2014) show that Bayesian Incentive Compatibility (BIC) suffices to guarantee that users will stay on the platform. Follow-up work (e.g., Mansour et al. (2015); Sellke and Slivkins (2021)) examines what bandit algorithms are BIC. Frazier et al. (2014) explore the use of monetary transfers.

Strategic experimentation. This line of work has investigated equilibria when a population of users each choose a bandit algorithm. Bolton and Harris (1999; 2000a;b) analyze the equilibria in a risky-safe arm bandit problem: we leverage their results in our analysis of equilibria in the shared data setting. Strategic experimentation (see Hörner and Skrzypacz (2017) for a survey) has investigated exponential bandit problems (Keller et al., 2005), the impact of observing actions instead of payoffs (Rosenberg et al., 2007), and the impact of cooperation (Brânzei and Peres, 2021).

5.2 Model

We consider a duopoly market with two platforms performing a multi-armed bandit learning problem and a population of N users, $u_1, \ldots, u_N$, who choose between platforms. Platforms commit to bandit algorithms, and then each user chooses a single platform to participate on for the learning task.

5.2.1 Multi-armed bandit setting

Consider a Bayesian bandit setting where there are k arms with priors $\mathcal{D}_1^{\text{Prior}}, \ldots, \mathcal{D}_k^{\text{Prior}}$. At the beginning of the game, the mean rewards of arms are drawn from the priors $r_1 \sim \mathcal{D}_1^{\text{Prior}}, \ldots, r_k \sim \mathcal{D}_k^{\text{Prior}}$. These mean rewards are unknown to both the users and the platforms but are shared across the two platforms. If the user's chosen platform recommends arm i, the user receives reward drawn from a noisy distribution $\mathcal{D}^{\text{Noise}}(r_i)$ with mean r_i.

Let $\mathcal{A}$ be a class of bandit algorithms that map the information state given by the posterior distributions to an arm to be pulled. The *information state* $\mathcal{I} = [\mathcal{D}_1^{\text{Post}}, \ldots, \mathcal{D}_k^{\text{Post}}]$ is taken to be the set of posterior distributions for the mean rewards of each arm. We assume that each algorithm $A \in \mathcal{A}$ can be expressed as a function mapping the information state $\mathcal{I}$ to a distribution over arms in $[k]$.[2] We let $A(\mathcal{I})$ denote this distribution over arms $[k]$.

Running example: risky-safe arm bandit problem. To concretize our results, we consider the *risky-safe arm bandit problem* as a running example. The noise distribution $\mathcal{D}^{\text{Noise}}(r_i)$ is a Gaussian $N(r_i, \sigma^2)$. The first arm is a *risky arm* whose prior distribution $\mathcal{D}_1^{\text{Prior}}$ is over the set $\{l, h\}$, where l corresponds to a "low reward" and h corresponds to a "high reward." The second arm is a *safe arm* with known reward $s \in (l, h)$ (the prior $\mathcal{D}_2^{\text{Prior}}$ is a point mass at s). In this case, the information state $\mathcal{I}$ permits a one-dimensional representation given by the posterior probability $p(\mathcal{I}) := \mathbb{P}_{X \sim \mathcal{D}_1^{\text{Post}}}[X = h]$ that the risky arm is high reward.

We construct a natural algorithm class as follows. For a measurable function $f : [0, 1] \to [0, 1]$, let A_f be the associated algorithm defined so $A_f(\mathcal{I})$ is a distribution that is 1 with probability $f(p(\mathcal{I}))$ and 2 with probability $1 - f(p(\mathcal{I}))$. We define

$$\mathcal{A}_{\text{all}} := \{A_f \mid f : [0, 1] \to [0, 1] \text{ is measurable}\}$$

to be the class of all randomized algorithms. This class contains Thompson sampling ($A_{f_{\text{TS}}}$ is given by $f_{\text{TS}}(p) = p$), the Greedy algorithm ($A_{f_{\text{Greedy}}}$ is given by $f_{\text{Greedy}}(p) = 1$ if $ph + (1-p)l \geq s$ and $f_{\text{Greedy}}(p) = 0$ otherwise), and mixtures of these algorithms with uniform exploration. We consider restrictions of the class $\mathcal{A}_{\text{all}}$ in some results.

[2]This assumption means that an algorithm's choice is independent of the time step conditioned on $\mathcal{I}$. Classical bandit algorithms such as Thompson sampling (Thompson, 1933), finite-horizon UCB (Lai and Robbins, 1985), and the infinite-time Gittins index (Gittins, 1979) fit into this framework. This assumption is *not* satisfied by the infinite time horizon UCB.

5.2.2 Interactions between platforms, users, and data

The interactions between the platform and users impact the data that the platform receives for its learning task. The platform action space $\mathcal{A}$ is a class of bandit algorithms that map an information state $\mathcal{I}$ to an arm to be pulled. The user action space is $\{1, 2\}$. For $1 \leq i \leq N$, we denote by $p^i \in \{1, 2\}$ the action chosen by user u_i.

Order of play. The platforms commit to algorithms A_1 and A_2 respectively, and then users simultaneously choose their actions $p^1, \ldots, p^N$ prior to the beginning of the learning task. We emphasize that user i participates on platform p^i for the *full duration* of the learning task. (In Appendix C.2.2, we discuss the assumption that users cannot switch platforms between time steps.)

Data sharing assumptions. In the *separate data repositories* setting, each platform has its own (proprietary) data repository for keeping track of the rewards incurred by its own users. Platforms 1 and 2 thus have separate information states given by $\mathcal{I}_1 = [\mathcal{D}^{\text{Post}}_{1,1}, \ldots, \mathcal{D}^{\text{Post}}_{1,k}]$ and $\mathcal{I}_2 = [\mathcal{D}^{\text{Post}}_{2,1}, \ldots, \mathcal{D}^{\text{Post}}_{2,k}]$, respectively. In the *shared data repository* setting, the platforms share an information state $\mathcal{I}_{\text{shared}} = [\mathcal{D}^{\text{Post}}_{1}, \ldots, \mathcal{D}^{\text{Post}}_{k}]$, which is updated based on the rewards incurred by users of both platforms.[3]

Learning task. The learning task is determined by the choice of platform actions A_1 and A_2, user actions $p^1, \ldots, p^n$, and specifics of data sharing between platforms. At each time step:

1. Each user u_i arrives at platform p^i. The platform p^i recommends arm $a_i \sim A_i(\mathcal{I})$ to that user, where $\mathcal{I}$ denotes the information state of the platform. (The randomness of arm selection is fully independent across users and time steps.) The user u_i receives noisy reward $\mathcal{D}^{\text{Noise}}(r_{a_i})$.

2. After providing recommendations to all of its users, platform 1 observes the rewards incurred by users in $S_1 := \{i \in [N] \mid p^i = 1\}$. Platform 2 similarly observes the rewards incurred by users in $S_2 := \{i \in [N] \mid p^i = 2\}$. Each platform then updates their information state $\mathcal{I}$ with the corresponding posterior updates.

3. A platform may have access to external data that does not come from users. To capture this, we introduce background information into the model. Both platforms observe the same *background information* of quality $\sigma_b \in (0, \infty]$. In particular, for each arm i, the platforms observe the same realization of a noisy reward $\mathcal{D}^{\text{Noise}}(r_i)$. When $\sigma_b = \infty$, we say that there is *no background information* since the background information is uninformative. The corresponding posterior updates are then used to update the information state ($\mathcal{I}$ in the case of shared data; $\mathcal{I}_1$ and $\mathcal{I}_2$ in the case of separate data).

[3] In web search, recommender systems can query each other, effectively building a shared information state.

In other words, platforms receive information from users (and background information), and users receive rewards based on the recommendations of the platform that they have chosen.

5.2.3 Utility functions and equilibrium concept

User utility is generated by rewards, while the platform utility is generated by *user participation.*

User utility function. We follow the standard discounted formulation for bandit problems (e.g. (Gittins and Jones, 1979; Bolton and Harris, 1999)), where the utility incurred by a user is defined by the expected (discounted) cumulative reward received across time steps. The discount factor β parameterizes the extent to which agents are myopic. Let $\mathcal{U}(p^i; \mathbf{p}^{-i}, A_1, A_2)$ denote the utility of a user u_i if they take action $p^i \in \{1, 2\}$ when other users take actions $\mathbf{p}^{-i} \in \{1, 2\}^{N-1}$ and the platforms choose A_1 and A_2. For clarity, we make this explicit in the case of discrete time setup with horizon length $T \in [1, \infty]$. Let $a_i^t = a_i^t(A_1, A_2, \mathbf{p})$ denote the arm recommended to user u_i at time step t. The utility is defined to be

$$\mathcal{U}(p^i; \mathbf{p}^{-i}, A_1, A_2) := \mathbb{E}\left[\sum_{t=1}^{T} \beta^t r_{a_i^t}\right]$$

where the expectation is over randomness of the incurred rewards and the algorithms. In the case of continuous time, the utility is

$$\mathcal{U}(p^i; \mathbf{p}^{-i}, A_1, A_2) := \mathbb{E}\left[\int e^{-\beta t} d\pi(t)\right]$$

where the $\beta \in [0, \infty)$ denotes the discount factor and $d\pi(t)$ denotes the payoff received by the user.[4] In both cases, observe that the utility function is *symmetric* in user actions.

The utility function implicitly differs in the separate and shared data settings, since the information state evolves differently in these two settings. When we wish to make this distinction explicit, we denote the corresponding utility functions by $\mathcal{U}^{\text{separate}}$ and $\mathcal{U}^{\text{shared}}$.

User equilibrium concept. We assume that after the platforms commit to algorithms A_1 and A_2, the users end up at a pure strategy Nash equilibrium of the resulting game. More formally, let $\mathbf{p} \in \{1, 2\}^N$ be a *pure strategy Nash equilibrium for the users* if $p_i \in \arg\max_{p \in \{0,1\}} \mathcal{U}(p; \mathbf{p}^{-i}, A_1, A_2)$ for all $1 \le i \le N$. The existence of a pure strategy Nash equilibrium follows from the assumption that the game is symmetric and the action space has 2 elements (Cheng et al., 2004).

One subtlety is that there can be multiple equilibria in this general-sum game. For example, there are always at least 2 (pure strategy) equilibria when platforms play any

[4] For discounted utility, it is often standard to introduce a multiplier of β for normalization (see e.g. (Bolton and Harris, 1999)). The utility $\mathcal{U}(p^i; \mathbf{p}^{-i}, A_1, A_2)$ could have equivalently be defined as $\mathbb{E}\left[\int \beta e^{-\beta t} d\pi(t)\right]$ without changing any of our results.

(A, A), i.e., commit to the same algorithm — one equilibrium where all users choose the first platform, and another where all users choose the second platform). Interestingly, there can be multiple equilibria even when one platform chooses a "worse" algorithm than the other platform. We denote by $\mathcal{E}_{A_1,A_2}$ the set of pure strategy Nash equilibria when the platforms choose algorithms A_1 and A_2. We simplify the notation and use $\mathcal{E}$ when A_1 and A_2 are clear from the context. In Chapter C.2.1, we discuss our choice of solution concept, focusing on what the implications would have been of including mixed Nash equilibria in $\mathcal{E}$.

Platform utility and equilibrium concept. The utility of the platform roughly corresponds to the number of users who participate on that platform. This captures that in markets for digital goods, where platform revenue is often derived from advertisement or subscription fees, the number of users serviced is a proxy for platform revenue.

When formalizing the utility that a platform receives, the fact that there can be several user equilibria for a given choice of platform algorithms creates ambiguity. To resolve this, we consider the worst-case user equilibrium for the platform, and we define platform utility to be the *minimum number of users that a platform would receive at any pure strategy equilibrium for the users.* More formally, when platform 1 chooses algorithm A_1 and platform 2 chooses algorithm A_2, the utilities of platform 1 and platform 2 are given by:

$$v_1(A_1; A_2) := \min_{\mathbf{p} \in \mathcal{E}} \sum_{i=1}^{N} \mathbb{1}[p^i = 1] \quad \text{and} \quad v_2(A_2; A_1) = \min_{\mathbf{p} \in \mathcal{E}} \sum_{i=1}^{N} \mathbb{1}[p^i = 2]. \tag{5.1}$$

The minimum over equilibria $\mathbf{p} \in \mathcal{E}$ in (5.1) captures a worst-case perspective where a platform makes decisions based on the worst possible utility that they might receive by choosing a given algorithm.[5]

The equilibrium concept for the platforms is a *pure strategy Nash equilibrium*, and we often focus on *symmetric* equilibria. We discuss the existence of such an equilibrium in Chapters 5.4-5.5. We note that at equilibrium, the utility for the platforms is typically 0, aligning with classical economic intuition. In particular, platforms earning zero equilibrium utility in our model mirrors firms earning zero equilibrium profit in price competition (Baye and Kovenock, 2008). However, there is an important distinction: platform utility *ex-post* (after users choose between platforms) may no longer be 0 and in fact may be as large as N, while firm profit in price competition remains 0 ex-post.

5.3 Formalizing the Alignment of a Market Outcome

The *alignment* of an equilibrium outcome for the platforms is measured by the amount of user utility that it generates. In Chapter 5.3.1 we introduce the *user quality level* to formalize

[5] Interestingly, if we were to define the platform utility in (5.1) to be a maximum over equilibria $\mathbf{p} \in \mathcal{E}$, this would induce degenerate behavior: any symmetric solution (A, A) would maximize platform utility and thus be an equilibrium. In contrast, formalizing the platform utility as the minimum over equilibria avoids this degeneracy.

alignment. In Chapter 5.3.2, we show an idealized form of alignment for $N = 1$ (Theorem 15). In Chapter 5.3.3, we turn to the case of multiple users and discuss benchmarks for the user quality level. In Chapter 5.3.4, we describe mild assumptions on $\mathcal{A}$ that we use in our alignment results for multiple users.

5.3.1 User quality level

Given a pair of platform algorithms $A_1 \in \mathcal{A}$ and $A_2 \in \mathcal{A}$, we introduce the *user quality level* $Q(A_1, A_2)$ to measure the alignment between platform algorithms and user utility. Informally speaking, the user quality level $Q(A_1, A_2)$ captures the utility that a user would receive when the platforms choose algorithms A_1 and A_2 and when user choices form an equilibrium.

When formalizing the user quality level, the potential multiplicity of user equilibria creates ambiguity (like in the definition of platform utility in (5.1)), and different users potentially receiving different utilities creates further ambiguity. We again take a worst-case perspective and formalize the user quality level as the minimum over equilibria $\mathbf{p} \in \mathcal{E}$ and over users $1 \leq i \leq N$.

Definition 2 (User quality level)**.** *Given algorithms A_1 and A_2 chosen by the platforms, the user quality level is defined to be $Q(A_1, A_2) := \min_{\mathbf{p} \in \mathcal{E}, 1 \leq i \leq N} \mathcal{U}(p_i; \mathbf{p}^{-i}, A_1, A_2)$.*

Interestingly, our insights about alignment would remain unchanged if we were to define the user quality level based on an arbitrary user equilibrium and user, rather than taking a minimum. More specifically, our alignment results (Theorems 16, 17, 18, 19, 21) would still hold if $Q(A_1, A_2)$ were defined to be $\mathcal{U}(p_i; \mathbf{p}^{-i}, A_1, A_2)$ for any arbitrarily chosen $\mathbf{p} \in \mathcal{E}$ and $1 \leq i \leq N$.[6] This demonstrates that our alignment results are independent of the particularities of how the user quality level is formalized.

To simplify notation in our alignment results, we introduce the *reward function* which captures how the utility that a given algorithm generates changes with the number of users who contribute to its data repository. For an algorithm $A \in \mathcal{A}$, let the reward function $R_A : [N] \to \mathbb{R}$ be defined by:

$$R_A(n) := \mathcal{U}^{\text{separate}}(1; \mathbf{p}_{n-1}, A, A),$$

where $\mathbf{p}_{n-1}$ corresponds to a vector with $n - 1$ coordinates equal to one.

[6]For most of results (Theorems 16, 18, 19, 21), the reason that the results remain unchanged is that at a symmetric solution (A, A), the user utility turns out to be the same for all $\mathbf{p} \in \mathcal{E}$ and $1 \leq i \leq N$ (this follows by definition for the shared data setting and follows from Lemma 175 for the separate data setting). The reason that Theorem 17, which considers asymmetric solutions, remains unchanged is that the lower bound holds for the worst-case $\mathbf{p} \in \mathcal{E}$ and $1 \leq i \leq N$ (and thus for any $\mathbf{p} \in \mathcal{E}$ and $1 \leq i \leq N$) and the proof of the upper bound applies more generally to any selection of $\mathbf{p} \in \{1, 2\}^N$ and $1 \leq i \leq N$.

5.3.2 Idealized alignment result: The case of a single user

When there is a single user, the platform algorithms turn out to be perfectly aligned with user utilities at equilibrium. To formalize this, we consider the optimal utility that could be obtained by a user across any choice of actions by the platforms and users (not necessarily at equilibrium): that is, $\max_{p\in\{1,2\},A_1\in\mathcal{A},A_2\in\mathcal{A}}\mathcal{U}(p;\emptyset,A_1,A_2)$. Using the setup of the single-user game, we can see that this is equal to $\max_{A\in\mathcal{A}}\mathcal{U}(1;\emptyset,A,A)=\max_{A\in\mathcal{A}}R_A(1)$. We show that the user quality level always meets this benchmark (we defer the proof to Appendix C.3).

Theorem 15. *Suppose that $N=1$, and consider either the separate data setting or the shared data setting. If (A_1,A_2) is a pure strategy Nash equilibrium for the platforms, then the user quality level $Q(A_1,A_2)$ is equal to* $\max_{A\in\mathcal{A}}R_A(1)$.

Theorem 15 shows that in a single-user market, two firms is sufficient to perfectly align firm actions with user utility—this stands in parallel to classical Bertrand competition in the pricing setting (Baye and Kovenock, 2008).

Proof sketch of Theorem 15. There are only 2 possible pure strategy equilibria: either the user chooses platform 1 and receives utility $R_{A_1}(1)$ or the user chooses platform 2 and receives utility $R_{A_2}(1)$. If one platform chooses a suboptimal algorithm for the user (i.e. an algorithm A' where $R_{A'}(1)<\max_{A\in\mathcal{A}}R_A(1)$), then the other platform will receive the user (and thus achieve utility 1) if they choose a optimal algorithm $\arg\max_{p\in\{1,2\}}R_{A_p}(1)$. This means that (A_1,A_2) is a pure strategy Nash equilibrium if and only if $A_1\in\arg\max_{A'\in\mathcal{A}}R_{A'}(1)$ or $A_2\in\arg\max_{A'\in\mathcal{A}}R_{A'}(1)$. The user thus receives utility $\max_{A\in\mathcal{A}}R_{A'}(1)$. We defer the full proof to Appendix C.3.

5.3.3 Benchmarks for user quality level

In the case of multiple users, this idealized form of alignment turns out to break down, and formalizing alignment requires a more nuanced consideration of benchmarks. We define the *single-user optimal utility* of $\mathcal{A}$ to be $\max_{A\in\mathcal{A}}R_A(1)$. This corresponds to maximal possible user utility that can be generated by a platform who only serves a single user and thus relies on this user for all of its data. On the other hand, we define the *global optimal utility* of $\mathcal{A}$ to be $\max_{A\in\mathcal{A}}R_A(N)$. This corresponds to the maximal possible user utility that can be generated by a platform when all of the users in the population are forced to participate on the same platform. The platform can thus maximally enrich its data repository in each time step.

5.3.4 Assumptions on $\mathcal{A}$

While our alignment results for a single user applied to arbitrary algorithm classes, we require mild assumptions on $\mathcal{A}$ in the case of multiple users to endow the equilibria with basic structure.

Information monotonicity requires that an algorithm A's performance in terms of user utility does not worsen with additional posterior updates to the information state. Our first two instantations of information monotonicity—strict information monotonicity and information constantness—require that the user utility of A grow monotonically in the number of other users participating in the algorithm. Our third instantation of information monotonicity—side information monotonicity—requires that the user utility of A not decrease if other users also update the information state, regardless of what algorithm is used by the other users. We formalize these assumptions as follows:

Assumption 1 (Information monotonicity). *For any given discount factor β and number of users N, an algorithm $A \in \mathcal{A}$ is* ***strictly information monotonic*** *if $R_A(n)$ is strictly increasing in n for $1 \leq n \leq N$. An algorithm A is* ***information constant*** *if $R_A(n)$ is constant in n for $1 \leq n \leq N$. An algorithm A is* ***side information monotonic*** *if for every measurable function f mapping information states to distributions over $[k]$ and for every $1 \leq n \leq N-1$, it holds that $\mathcal{U}^{shared}(1; \mathbf{2}_n, A, f) \geq R_A(1)$ where $\mathbf{2}_n \in \{1,2\}^n$ has all coordinates equal to 2.*

While information monotonicity places assumptions on each algorithm in $\mathcal{A}$, our next assumption places a mild restriction on how the utilities generated by algorithms in $\mathcal{A}$ relate to each other. *Utility richness* requires that the set of user utilities spanned by $\mathcal{A}$ is a sufficiently rich interval.

Assumption 2 (Utility richness). *A class of algorithms $\mathcal{A}$ is* ***utility rich*** *if the set of utilities $\{R_A(N) \mid A \in \mathcal{A}\}$ is a contiguous set, the supremum of $\{R_A(N) \mid A \in \mathcal{A}\}$ is achieved, and there exists $A' \in \mathcal{A}$ such that $R_{A'}(N) \leq \max_{A \in \mathcal{A}} R_A(1)$.*

Discussion of assumptions. Utility richness holds (almost) without loss of generality, by taking the closure of an algorithmic class under the operation of mixing with the pure exploration algorithm (see Lemma 25). On the other hand, not all algorithms are information monotone. Nevertheless, we show that information monotonicity is satisfied for several algorithms for the risky-safe arm setup, including any nondegenerate algorithm under undiscounted rewards (see Lemma 22) and Thompson sampling under discounted rewards (see Lemma 23). These results are of independent interest, and more broadly, understanding information monotonicity is crucial for investigating the incentive properties of bandit algorithms: indeed prior work (e.g. Aridor et al. (2025)) has explored variants of this assumption. We defer a detailed discussion of these assumptions to Chapter 5.6.

5.4 Separate data repositories

We investigate alignment when the platforms have separate data repositories. In Chapter 5.4.1, we show that there can be many qualitatively different equilibria for the platforms and characterize the alignment of these equilibria. In Chapter 5.4.2, we discuss factors that drive the level of misalignment in a marketplace.

5.4.1 Multitude of equilibria and the extent of alignment

In contrast with the single user setting, the marketplace can exhibit multiple equilibria for the platforms. As a result, to investigate alignment, we investigate the *range* of achievable user quality levels. Our main finding is that the equilibria in a given marketplace can exhibit a vast range of alignment properties. In particular, *every* user quality level in between the single-user optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(1)$ and the global optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(N)$ can be realized by some equilibrium for the platforms.

Theorem 16. *Suppose that each algorithm in $\mathcal{A}$ is either strictly information monotonic or information constant (Assumption 1), and suppose that $\mathcal{A}$ is utility rich (Assumption 2). For every $\alpha \in [\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N)]$, there exists a symmetric pure strategy Nash equilibrium (A, A) in the separate data setting such that $Q(A, A) = \alpha$.*

Nonetheless, there is a baseline (although somewhat weak) form of alignment achieved by all equilibria. In particular, every equilibrium for the platforms has user quality level at least the single-user optimum $\max_{A' \in \mathcal{A}} R_{A'}(1)$.

Theorem 17. *Suppose that each algorithm in $\mathcal{A}$ is either strictly information monotonic or information constant (see Assumption 1). In the separate data setting, at any pure strategy Nash equilibrium (A_1, A_2) for the platforms, the user quality level lies in the following interval:*

$$Q(A_1, A_2) \in \left[\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N)\right].$$

An intuition for these results is that the performance of an algorithm depends not only on how it transforms information to actions, but also on the amount of information to which it has access. A platform can make up for a suboptimal algorithm by attracting a significant user base: if a platform starts with the full user base, it is possible that *no single user* will switch to the competing platform, even if the competing platform chooses a stricter better algorithm. However, if a platform's algorithm is highly suboptimal, then the competing platform will indeed be able to win the full user base.

Proof sketches of Theorem 16 and Theorem 17. The key idea is that pure strategy equilibria for users take a simple form. Under strict information monotonicity, we show that every pure strategy equilibrium $p^* \in \mathcal{E}^{A_1, A_2}$ is in the set $\{[1, \ldots, 1], [2, \ldots, 2]\}$ (Lemma 175). The intuition is that the user utility strictly grows with the amount of data that the platform has, which in turn grows with the number of other users participating on the same platform. It is often better for a user to switch to the platform with more users, which drives all users to a single platform in equilibrium.

The reward functions $R_{A_1}(\cdot)$ and $R_{A_2}(\cdot)$ determine which of these two solutions are in $\mathcal{E}^{A_1, A_2}$. It follows from definition that $[1, \ldots, 1]$ is in $\mathcal{E}_{A_1, A_2}$ if and only if $R_{A_1}(N) \geq R_{A_2}(1)$. This inequality can hold even if A_2 is a better algorithm in the sense that $R_{A_2}(n) > R_{A_1}(n)$ for all n. The intuition is that the performance of an algorithm is controlled not only by its efficiency in choosing the possible action from the information state, but also by the size of its user base. The platform with the worse algorithm can be better for users if it has accrued enough users.

This characterization of the set $\mathcal{E}_{A_1,A_2}$ enables us to reason about the platform equilibria. To prove Theorem 16, we show that (A, A) is an equilibrium for the platforms as long as $R_A(N) \geq \max_{A'} R_{A'}(1)$. This, coupled with utility richness, enables us to show that every utility level in $[\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N)]$ can be realized. To prove Theorem 17, we first show platforms can't both choose highly suboptimal algorithms: in particular, if $R_{A_1}(N)$ and $R_{A_2}(N)$ are both below the single-user optimal $\max_{A' \in \mathcal{A}} R_{A'}(1)$, then (A_1, A_2) is not in equilibrium. Moreover, if one of the platforms chooses an algorithm A where $R_A(N) < \max_{A' \in \mathcal{A}} R_{A'}(1)$, then all of the users will choose the other platform in equilibrium. The full proofs are deferred to Appendix C.4.

5.4.2 What drives the level of misalignment in a marketplace?

The existence of multiple equilibria makes it more subtle to reason about the alignment exhibited by a marketplace. The level of misalignment depends on two factors: first, the size of the range of realizable user quality levels, and second, the selection of equilibrium within this range. We explore each of these factors in greater detail.

How large is the range of possible user quality levels? Both the algorithm class and the structure of the user utility function determine the size of the range of possible user quality levels. We informally examine the role of the user's discount factor on the size of this range.

First, consider the case where users are fully non-myopic (so their rewards are undiscounted across time steps). The gap between the single-user optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(1)$ and global optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(N)$ can be substantial. To gain intuition for this, observe that the utility level $R_{A'}(N)$ corresponds to the algorithm A' receiving N times as much as data at every time step than the utility level $R_{A'}(1)$. For example, consider an algorithm A' whose regret grows according to $\sqrt{T}$ where T is the number of samples collected, and let $\text{OPT} := \mathbb{E}_{r_1 \sim \mathcal{D}_1^{\text{Prior}}, \ldots r_k \sim \mathcal{D}_k^{\text{Prior}}}[\max_{1 \leq i \leq k} r_i]$ be the expected maximum reward of any arm. Since utility and regret are related up to additive factors for fully non-myopic users, then we have that $R_{A'}(1) \approx \text{OPT} - \sqrt{T}$ while $R_{A'}(N) \approx \text{OPT} - \sqrt{NT}$.

At the other extreme, consider the case where users are fully myopic. In this case, the range collapses to a *single point*. The intuition is that the algorithm generates the same utility for a user regardless of the number of other users who participate: in particular, $R_{A'}(1)$ is equal to $R_{A'}(N)$ for any algorithm $A' \in \mathcal{A}$. To see this, we observe that the algorithm's behavior beyond the first time step does not factor into user utility, and the algorithm's selection at the first time is determined before it receives any information from users. Put differently, although $R_{A'}(N)$ can receives N times more information, there is a delay before the algorithm sees this information. Thus, in the case of fully myopic users, the user quality level is always equal to the global optimal user utility $\max_A R_A(N)$ so idealized alignment is actually recovered. When users are partially non-myopic, the range is no longer a single point, but the range is intuitively smaller than in the undiscounted case.

Which equilibrium arises in a marketplace?. When the gap between the single-user optimal and global optimal utility levels is substantial, it becomes ambiguous what user quality level will be realized in a given marketplace. Which equilibria arises in a marketplace depends on several factors.

One factor is the secondary aspects of the platform objective that aren't fully captured by the number of users. For example, suppose that the platform cares about the its reputation and thus is incentivized to optimize for the quality of the service. This could drive the marketplace towards higher user quality levels. On the other hand, suppose that the platform derives other sources of revenue from recommending certain types of content (e.g. from recommending advertisements). If these additional sources of revenue are not aligned with user utility, then this could drive the marketplace towards lower user quality levels.

Another factor is the mechanism under which platforms arrive at equilibrium solutions, such as *market entry*. We informally show that market entry can result in the the *worst possible user utility* within the range of realizable levels. To see this, notice that when one platform enters the marketplace shortly before another platform, all of the users will initially choose the first platform. The second platform will win over users only if $R_{A_2}(1) > R_{A_1}(N)$, where A_2 denotes the algorithm of the second platform and A_1 denotes the algorithm of the first platform. In particular, the platform is susceptible to losing users only if $R_{A_1}(N) < \max_{A' \in \mathcal{A}} R_{A'}(1)$. Thus, the worst possible equilibrium can arise in the marketplace, and this problem only worsens if the first platform enters early enough to accumulate data beforehand. This finding provides a mathematical backing for the barriers to entry in digital marketplaces that are documented in policy reports (Stigler Committee, 2019).

This finding points to an interesting direction for future work: *what equilibria arise from other natural mechanisms?*

5.5 Shared data repository

What happens when data is shared between the platforms? We show that both the nature of alignment and the forces that drive misalignment fundamentally change. In Chapter 5.5.1, we show a construction where the user quality levels do not span the full set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$. Despite this, in Chapter 5.5.2, we establish that the user quality level at any symmetric equilibrium continues to be at least $\max_{A'} R_{A'}(1)$.

5.5.1 Construction where global optimal is not realizable

In contrast with the separate data setting, the set of user quality levels at symmetric equilibria for the platforms does not necessarily span the full set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$. To demonstrate this, we show that in the risky-safe arm problem, every symmetric equilibrium (A, A) has user quality level $Q(A, A)$ strictly below $\max_{A'} R_{A'}(N)$.

Theorem 18. *Let the algorithm class $\mathcal{A}_{all}^{cont} \subseteq \mathcal{A}_{all}$ consist of the algorithms A_f where $f(0) = 0$, $f(1) = 1$, and f is continuous at 0 and 1. In the shared data setting, for any choice of prior $p \in (0, 1)$ and any background information quality $\sigma_b \in (0, \infty)$, there exists an undiscounted risky-safe arm bandit setup (see Setup 1) such that the set of realizable user quality levels for algorithm class $\mathcal{A}_{all}^{cont}$ is equal to a singleton set:*

$$\{Q(A, A) \mid (A, A) \textit{ is a symmetric equilibrium for the platforms } \} = \{\alpha^*\}$$

where

$$\max_{A' \in \mathcal{A}} R_{A'}(1) < \alpha^* < \max_{A' \in \mathcal{A}} R_{A'}(N).$$

Theorem 19. *In the shared data setting, for any discount factor $\beta \in (0, \infty)$ and any choice of prior $p \in (0, 1)$, there exists a discounted risky-safe arm bandit setup with no background information (see Setup 2) such that the set of realizable user quality levels for algorithm class $\mathcal{A}_{all}$ is equal to a singleton set:*

$$\{Q(A, A) \mid (A, A) \textit{ is a symmetric equilibrium for the platforms } \} = \{\alpha^*\}$$

where

$$\max_{A' \in \mathcal{A}} R_{A'}(1) \leq \alpha^* < \max_{A' \in \mathcal{A}} R_{A'}(N).$$

Theorems 18 and 19 illustrate examples where there is no symmetric equilibrium for the platforms that realizes the global optimal utility $\max_{A'} R_{A'}(N)$—regardless of whether users are fully non-myopic or have discounted utility. These results have interesting implications for shared data access as an intervention in digital marketplace regulation (e.g. see Crémer et al. (2019)). At first glance, it would appear that data sharing would resolve the alignment issues, since it prevents platforms from gaining market dominance through data accumulation. However, our results illustrate that the platforms may still not align their actions with user utility at equilibrium.

Comparison of separate and shared data settings. To further investigate the efficacy of shared data access as a policy intervention, we compare alignment when the platforms share a data repository to alignment when the platforms have separate data repositories, highlighting two fundamental differences. We focus on the undiscounted setup (Setup 1) analyzed in Theorem 18; in this case, the algorithm class $\mathcal{A}_{\text{all}}^{\text{cont}}$ satisfies information monotonicity and utility richness (see Lemma 22) so the results in Chapter 5.4.1 are also applicable.[7] The first difference in the nature of alignment is that there is a unique symmetric equilibrium for the shared data setting, which stands in contrast to the range of equilibria that arose in the separate data setting. Thus, while the particularities of equilibrium selection significantly impact alignment in the separate data setting (see Chapter 5.4.2), these particularities are irrelevant from the perspective of alignment in the shared data setting.

[7]In the discounted setting, not all of the algorithms in $\mathcal{A}_{\text{all}}$ necessarily satisfy the information monotonicity requirements used in the alignment results for the separate data setting. Thus, Theorem 19 cannot be used to directly compare the two settings.

The second difference is that the user quality level of the symmetric equilibrium in the shared data setting is in the *interior* of the range $[\max_{A \in \mathcal{A}} R_A(1), \max_{A \in \mathcal{A}} R_A(N)]$ of user quality levels exhibited in the separate data setting. The alignment in the shared data setting is thus *strictly better* than the alignment of the worst possible equilibrium in the separate data setting. Thus, if we take a pessimistic view of the separate data setting, assuming that the marketplace exhibits the worst-possible equilibrium, then data sharing does help users. On the other hand, the alignment in the shared data setting is also *strictly worse* than the alignment of the best possible equilibrium in the separate data setting. This means if that we instead take an optimistic view of the separate data setting, and assume that the marketplace exhibits this best-case equilibrium, then data sharing is actually harmful for alignment. In other words, when comparing data sharing and equilibrium selection as regulatory interventions, data sharing is worse for users than maintaining separate data and applying an equilibrium selection mechanism that shifts the market towards the best equilibria.

Mechanism for misalignment. Perhaps counterintuitively, the mechanism for misalignment in the shared data setting is that a platform must perfectly align its choice of algorithm with the preferences of a user (given the choices of other users). In particular, the algorithm that is optimal for one user given the actions of other users is different from the algorithm that would be optimal if the users were to cooperate. This is because exploration is costly to users, so users don't want to perform their fair share of exploration, and would rather *free-ride* off of the exploration of other users. As a result, a platform who chooses an algorithm with the global optimal strategy cannot maintain its user base. We formalize this phenomena by establishing a connection with *strategic experimentation*, drawing upon the results of Bolton and Harris (1999; 2000a;b) (see Appendix C.5.2 for a recap of the relevant results).

Proof sketches of Theorem 18 and Theorem 19. The key insight is that the symmetric equilibria of our game are closely related to the equilibria of the following game G. Let G be an N player game where each player chooses an algorithm in $\mathcal{A}$ within the same bandit problem setup as in our game. The players share an information state $\mathcal{I}$ corresponding to the posterior distributions of the arms. At each time step, all of the N players arrive at the platform, player i pulls the arm drawn from $A_i(\mathcal{I})$, and the players all update $\mathcal{I}$. The utility received by a player is given by their discounted cumulative reward.

We characterize the symmetric equilibria of the original game for the platforms.

Lemma 20. *The solution (A, A) is in equilibrium if and only if A is a symmetric pure strategy equilibrium of the game G described above.*

Moreover, the user quality level $Q(A, A)$ is equal to $R_A(N)$, which is also equal to the utility achieved by players in G when they all choose action A.

In the game G, the global optimal algorithm $A^* = \arg\max_{A' \in \mathcal{A}} R_{A'}(N)$ corresponds to the solution when all N players cooperate rather than arriving at an equilibrium. Intuitively, all of the players choosing A^* is not an equilibrium because exploration comes at a cost to utility, and thus players wish to "free-ride" off of the exploration of other players. The value

$\max_{A' \in \mathcal{A}} R_{A'}(N)$ corresponds to the cooperative maximal utility that can be obtained the N players.

To show Theorem 19, it suffices to analyze structure of the equilibria of G. Interestingly, Bolton and Harris (1999; 2000a;b)—in the context of strategic experimentation—studied a game very similar to G instantiated in the risky-safe arm bandit problem with algorithm class $\mathcal{A}_{\text{all}}$. We provide a recap of the relevant aspects of their results and analysis in Appendix C.5.2. At a high level, they showed that there is a unique symmetric pure strategy equilibrium and showed that the utility of this equilibrium is strictly below the global optimal. We can adopt this analysis to conclude that the equilibrium player utility in G is strictly below $R_A(N)$. The full proof is deferred to Appendix C.5.

5.5.2 Weak alignment

Although not all values in $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ can be realized, we show that the user quality level at any symmetric equilibrium is always at least $\max_{A'} R_{A'}(1)$.

Theorem 21. *Suppose that every algorithm in $\mathcal{A}$ is side information monotonic (Assumption 1). In the shared data setting, at any symmetric equilibrium (A, A), the user quality level $Q(A, A)$ is in the interval $[\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N)]$.*

Theorem 21 demonstrates that the free-riding effect described in Chapter 5.5.1 cannot drive the user quality level below the single-user optimal. Recall that the single-user optimal is also a lower bound on the user quality level for the separate data setting (see Theorem 17). This means that regardless of the assumptions on data sharing, the market outcome exhibits a weak form of alignment where the user quality level is at least the single-user optimal.

Proof sketch of Theorem 21. We again leverage the connection to the game G described in the proof sketch of Theorem 19. The main technical step is to show that at any symmetric pure strategy equilibrium A, the player utility $R_A(N)$ is at least $\max_{A' \in \mathcal{A}} R_{A'}(1)$ (Lemma 179). Intuitively, since A is a best response for each player, they must receive no more utility by choosing $A^* \in \arg\max_{A' \in \mathcal{A}} R_{A'}(1)$. The utility that they would receive from playing A^* if there were no other players in the game is $R_{A^*}(1) = \max_{A' \in \mathcal{A}} R_{A'}(1)$. The presence of other players can be viewed as background updates to the information state, and the information monotonicity assumption on A guarantees that these updates can only improve the player's utility in expectation. The full proof is deferred to Appendix C.5.

5.6 Algorithm classes $\mathcal{A}$ that satisfy our assumptions

We describe several different bandit setups under which the assumptions on $\mathcal{A}$ described in Chapter 5.3.4 are satisfied. **Discussion of information monotonicity (Assumption 1).**

We first show that in the undiscounted, continuous-time, risky-safe arm bandit setup, the information monotonicity assumptions are satisfied for essentially any algorithm (proof is deferred to Appendix C.6).

Lemma 22. *Consider the undiscounted, continuous-time risky-safe arm bandit setup (see Setup 1). Any algorithm $A \in \mathcal{A}_{all}^{cont}$ satisfies strict information monotonicity and side information monotonicity.*

While the above result focuses on undiscounted utility, we also show that information monotonicity can also be achieved with discounting. In particular, information monotonicity is satisfied by `ThompsonSampling` (proof is deferred to Appendix C.6).

Lemma 23. *For the discrete-time risky-safe arm bandit problem with finite time horizon, prior $p \in (0,1)$, $N = 2$ users, and no background information (see Setup 3), `ThompsonSampling` is strictly information monotonic and side information monotonic for any discount factor $\beta \in (0,1]$.*

In fact, we actually show in the proof of Lemma 23 that the ε-`ThompsonSampling` algorithm that explores uniformly with probability ε and applies `ThompsonSampling` with probability $1-\varepsilon$ also satisfies strict information monotonicity and side information monotonicity.

These information monotonicity assumptions become completely unrestrictive for fully myopic users, where user utility is fully determined by the algorithm's performance at the first time step, before any information updates are made. In particular, *any* algorithm is information constant and side-information monotonic.

More broadly, understanding information monotonicity and its variants is crucial for investigating the incentive properties of bandit algorithms: indeed prior work (e.g. Aridor et al. (2025); Mansour et al. (2022); Sellke and Slivkins (2021)) has explored variants of this assumption. Since these works focus on fully myopic users that may arrive at any time step, they require a different information monotonicity assumption, that they call *Bayes monotonicity* (Aridor et al., 2025). (An algorithm satisfies Bayes monotonicity if its expected reward is non-decreasing *in time*.) Bayes monotonicity is strictly speaking incomparable to our information monotonicity assumptions; in particular, Bayes monotonicity does not imply either strict information monotonicity or side information monotonicity.

Discussion of utility richness (Assumption 2). At an intuitive level, as long as the algorithm class reflects a range of exploration levels, it will satisfy utility richness.

We first show that in the undiscounted setup in Theorem 18, the algorithm class satisfies utility richness (proof in Appendix C.6).

Lemma 24. *Consider the undiscounted, continuous-time risky-safe arm bandit setup (see Setup 1). The algorithm class $\mathcal{A}_{all}^{cont}$ satisfies utility richness.*

Since the above result focuses on a particular bandit setup, we also describe a general operation to transform an algorithm class into one that satisfies utility richness. In particular, the closure of an algorithm class under mixtures with uniformly random exploration satisfies utility richness (proof in Appendix C.6).

Lemma 25. *Consider any discrete-time setup with finite time horizon and bounded mean rewards. For $A \in \mathcal{A}$, let A_ε be the algorithm that chooses an arm at random w/ probability ε. Suppose that the reward $R_A(N)$ of every algorithm $A \in \mathcal{A}$ is at least $R_{A_1}(N)$ (the reward of uniform exploration), and suppose that the supremum of $\{R_A(N) \mid A \in \mathcal{A}\}$ is achieved. Then, the algorithm class $\mathcal{A}_{closure} := \{A_\varepsilon \mid A \in \mathcal{A}, \varepsilon \in [0,1]\}$ satisfies utility richness.*

Example classes that achieve information monotonicity and utility richness. Together, the results above provide two natural bandit setups that satisfy strict information monotonicity, side information monotonicity, and utility richness.

1. The algorithm class $\mathcal{A}_{\text{all}}^{\text{cont}}$ in the undiscounted, continuous-time risky-safe arm bandit setup with any $N \geq 1$ users (see Setup 1).
2. The class of ε-Thompson sampling algorithms in the discrete time risky-safe arm bandit setup with discount factor $\beta \in (0, 1]$, $N = 2$ users, and no background information (see Setup 3).

These setups, which span the full range of discount factors, provide concrete examples where our alignment results are guaranteed to apply.

5.7 Discussion

Towards investigating competition in digital marketplaces, we present a framework for analyzing competition between two platforms performing multi-armed bandit learning through interactions with a population of users. We propose and analyze the *user quality level* as a measure of the alignment of market equilibria. We show that unlike in typical markets of products, competition in this setting does not perfectly align market outcomes with user utilities, both when the platforms maintain separate data repositories and when the platforms maintain a shared data repository.

Our framework further allows to compare the separate and shared data settings, and we show that the nature of misalignment fundamentally depends on the data sharing assumptions. First, different mechanisms drive misalignment: when platforms have separate data repositories, the suboptimality of an algorithm can be compensated for with a larger user base; when the platforms share data, a platform can't retain its user base if it chooses the global optimal algorithm since users wish to free-ride off of the exploration of other users. Another aspect that depends on the data sharing assumptions is the specific form of misalignment exhibited by market outcomes. The set of realizable user quality levels ranges from the single-user optimal to the global optimal in the separate data setting; on the other hand, in the shared data setting, neither of these endpoints may be realizable. These differences suggest that data sharing performs worse as a regulatory intervention than a well-designed equilibrium selection mechanism.

More broadly, our work provides a mathematical explanation of phenomena documented in recent policy reports and reveals that competition has subtle consequences for users in digital marketplaces that merit further inquiry. We hope that our work provides a starting point for building a theoretical foundation for investigating competition and designing regulatory interventions in digital marketplaces.

Chapter 6

The Power of a Digital Platform

This chapter is based on *"Performative Power"* (Hardt et al., 2022), which is joint work with Moritz Hardt and Celestine Mendler-Dünner.

6.1 Introduction

Digital platforms pose a well-recognized challenge for antitrust enforcement. Traditional market definitions, along with associated notions of competition and market power, map poorly onto digital platforms. A core challenge is the difficulty of precisely modeling the interactions between the market participants, products, and prices. An authoritative report, published by the Stigler Committee (2019), details the many challenges associated with digital platforms, among them: *"Pinpointing the locus of competition can also be challenging because the markets are multisided and often ones with which economists and lawyers have little experience. This complexity can make market definition another hurdle to effective enforcement."* Published the same year, a comprehensive report from the European Commission (Crémer et al., 2019) calls for *"less emphasis on analysis of market definition, and more emphasis on theories of harm and identification of anti-competitive strategies."*

Our work responds to this call by developing a normative and technical proposal for reasoning about power in digital economies, while relaxing the reliance on market definition. Our running example is a digital content recommendation platform. The platform connects content creators with viewers, while monetizing views through digital advertisement. Key to the business strategy of a firm operating a digital content recommendation platform is its ability to predict revenue for content that it recommends or ranks highly. Often framed as a supervised learning task, the firm trains a statistical model on observed data to predict some proxy of revenue, such as clicks, views, or engagement. Better predictions enable the firm to more accurately identify content of interest and thus increase profit.

A second way of increasing profit is more subtle. The platform can use its predictions to *steer* participants towards modes of consumption and production that are easier to predict and monetize. For example, the platform could reward consistency in the videos created by content creators, so that the audience and the popularity of their videos becomes more predictable. Similarly, the platform could recommend addictive content to viewers, appealing to behavioral weaknesses in order to drive up viewer engagement. How potent such a strategy is depends on the extent to which the firm is able to steer participants, which we argue reveals a salient power relationship between the platform and its participants.

6.1.1 Our contribution

We introduce the notion of *performative power* that quantifies a firm's ability to steer a population of participants. We argue that the sensitivity of participant behavior to algorithmic changes in the platform provides an important indicator of the firm's power. Performative power is a causal statistical notion that directly quantifies how much participants change in response to actions by the platform, such as updating a predictive model. In doing so it avoids market specifics, such as the number of firms involved, products, and monetary prices. Neither does it require a competitive equilibrium notion as a reference point. Instead, it focuses on where rubber meets the road: the algorithmic actions of the platform and their causal powers.

We first investigate the role of performative power in optimization. In particular, we build on recent developments in performative prediction (Perdomo et al., 2020) to articulate the fundamental difference between learning and steering in prediction. We show that under low performative power, a firm cannot do better than standard supervised learning on observed data. Intuitively, this means the firm optimizes its loss function *ex-ante* on data it observes without the ability to steer towards data it would prefer. We interpret this optimization strategy as analogous to the firm being a price-taker, an economic condition that arises under perfect competition in classical market models. We contrast this optimization strategy with a firm that performs *ex-post* optimization and benefits from steering towards data it prefers. Formally, we provide an upper-bound on the distance between the two solution concepts in terms of performative power.

Then, to study the qualitative properties of performative power we consider the concrete algorithmic market model of strategic classification. Strategic classification models participants as best-responding agents that change their features rationally in response to a predictor with the goal of achieving a better prediction outcome. In this simple setting, we show that the willingness of participants to invest in changing their features governs the performative power of the firm. We investigate the role of different economic factors by extending the standard model to incorporate competing firms and outside options. We highlight two key observations:

- A *monopoly* firm can have significant performative power. In this case, performative power is derived because participants are willing to incur a cost up to the utility of using the service in order to adjust to the firm's predictor. Moreover, performative power is maximized if a monopoly firm has the ability to personalize decisions to individual users.

- Performative power decreases in the presence of *competition* and *outside options*. In particular, when firms compete for participants, offering services that are perfect substitutes for each other, then even two firms can lead to zero performative power. This result stands in analogy with the classical Bertrand competition.

On the empirical side, we propose a causal design to identify performative power in the context of a recommender system arranging content into display slots. This design, we call *discrete display design (DDD)*, establishes a connection between performative power and the causal effect of display position on consumption. To derive a lower bound on performative power, DDD constructs a hypothetical algorithmic action that aggregates the causal effects of display position across the population. This allows us to repurpose reported causal effects of display position as lower bounds on performative power. It also charts out a concrete empirical strategy for understanding power in digital economies, both experimentally and observationally.

Finally, we examine the potential role of performative power in competition policy. We contrast performative power with traditional measures of market power, describe how performative power can capture complex behavioral patterns, and discuss the role that performative power might play in ongoing antitrust debates.

6.1.2 Related work

Our notion of performative power builds on the development of performativity in prediction by Perdomo et al. (2020). Performativity captures that the predictor can influence the data-generating process, a dependency ruled out by the traditional theory of supervised learning. A growing line of work on performative prediction, e.g., (Mendler-Dünner et al., 2020; Drusvyatskiy and Xiao, 2023; Izzo et al., 2021; Dong et al., 2023; Miller et al., 2021; Brown et al., 2022b; Li and Wai, 2022; Ray et al., 2022; Jagadeesan et al., 2022; Wood et al., 2021), has studied different optimization challenges and solution concepts in performative prediction. Rather than viewing performative effects as an additional challenge for the learning algorithm, we argue that performativity reveals a salient power relationship between the decision maker and the population. From an optimization perspective, our work demonstrates that sufficiently high performative power is necessary for performative optimization approaches to achieve lower risk compared with standard supervised learning.

The *strategic classification* setup we use for our case study was proposed in (Brückner et al., 2012; Hardt et al., 2016) and is closely related to a line of work in the economics community (Frankel and Kartik, 2022; Ball, 2025; Hennessy and Goodhart, 2023; Frankel and Kartik, 2019). A long line of work on strategic classification makes the assumption that performative effects are the result of individuals manipulating their features so as to best respond to the deployment of a predictive model. The focus has been on describing participant behavior in response to a single firm acting in isolation. Our extensions incorporate additional market factors into the model, such as outside options or the choice between competing firms, which we believe are helpful for gaining a better understanding of strategic interactions in digital economies. Beyond the case of a single classifier, recently, Narang et al. (2023) and Piliouras and Yu (2023) analyzed settings with multiple firms that simultaneously apply retraining algorithms in performative environments. Similar to our analysis in Chapter 6.3, these works study the solution concept of a Nash equilibrium, however, with a focus on proving convergence to equilibrium solutions, whereas we are interested in how these equilibria interact with performative power. Ginart et al. (2021) study another model of feedback loops arising from competition between machine learning models.

There is extensive literature on the topic of competition on digital platforms that we do not attempt to survey here. For starting points, see, for example, recent work by Bergemann and Bonatti (2024), a survey by Calvano and Polo (2021b), a discussion by Parker et al. (2019), the reports already mentioned (Stigler Committee, 2019; Crémer et al., 2019), as well as a macroeconomic perspective on the topic (Syverson, 2019).

6.2 Performative power

Fix a set $\mathcal{U}$ of participants interacting with a designated firm, where each $u \in \mathcal{U}$ is associated with a data point $z(u)$. Fix a metric $\mathrm{dist}(z, z')$ over the space of data points. Let $\mathcal{F}$ denote the set of actions a firm can take. We think of an action $f \in \mathcal{F}$ as a predictor that the firm can deploy at a fixed point in time. For each participant $u \in \mathcal{U}$ and action $f \in \mathcal{F}$, we denote by $z_f(u)$ the potential outcome random variable representing the counterfactual data of participant u if the firm were to take action f.

Definition 3 (Performative Power)**.** *Given a population $\mathcal{U}$, an action set $\mathcal{F}$, potential outcome pairs $(z(u), z_f(u))$ for each unit $u \in \mathcal{U}$ and action $f \in \mathcal{F}$, and a metric* dist *over the space of data points, we define the* performative power *of the firm as*

$$\mathrm{P} := \sup_{f \in \mathcal{F}} \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}\left[\mathrm{dist}\left(z(u), z_f(u)\right)\right] ,$$

where the expectation is over the randomness in the potential outcomes.

The expression inside the supremum generalizes an average treatment effect, corresponding to scalar valued potential outcomes and the absolute value as metric. We could generalize other causal quantities such as heterogeneous treatment effects, but this avenue is not subject of our paper. The definition takes a supremum over possible actions a firm can take at a specific point in time. We can therefore lower bound performative power by estimating the causal effect of any given action $f \in \mathcal{F}$.

Having specified the sets $\mathcal{F}$ and $\mathcal{U}$, estimating performative power amounts to causal inference involving the potential outcome variables $z_f(u)$ for unit $u \in \mathcal{U}$ and action $f \in \mathcal{F}$. In an observational design, an investigator is able to identify performative power without an experimental intervention on the platform. We propose and apply one such observational design in Chapter 6.5. In an experimental design, the investigator deploys a suitably chosen action to estimate the effect. Neither route requires understanding the specifics of the market in which the firm operates. It is not even necessary to know the firm's objective function, how it optimizes its objective, and whether it successfully achieves its objective. In practice, the dynamic process that generates the potential outcome $z_f(u)$ may be highly complex, but this complexity does not enter the definition. Consequently, the definition applies to complex multisided digital economies that defy mathematical specification. To make this abstract concept of performative power more concrete, we instantiate it in a concrete example.

6.2.1 Running example: Digital content recommendation

Consider a digital content recommendation platform, such as the video sharing services YouTube or Twitch. The platform aims to recommend channels that generate high revenue, personalized to each viewer. Towards this goal, the platform collects data to build a predictor f for the value of a channel c to a viewer with preferences p. Let $x = (x_c, x_p)$ be the features used for the prediction task that capture attributes x_c of the channel and the attributes x_p of the viewer preferences. Let y be the target variable, such as *watch time*, that acts as a proxy for the monetary value of showing a channel to a specific viewer. For concreteness, take the supervised learning loss $\ell(f(x), y)$ incurred by a predictor f to be the squared loss $(f(x) - y)^2$.

When defining performative power, participants could either be viewers or content creators. The definition is flexible and applies to both. By selecting the units $\mathcal{U}$, which features to include in the data point z, and how to specify the distance metric dist, we can pinpoint the power relationship we would like to investigate.

Content creators. The predictor f can affect the type of videos that content creators stream on their channels. For example, content creators might strategically adjust various features of their content relevant for the predicted outcome, such as the length, type or description of their videos, to improve their ranking. Thus, by changing how it predicts the monetary value of a channel, the platform can induce changes in the content on the channel. To measure this source of power, we let the participants $\mathcal{U}$ be content creators and suppose that each content creator $u \in \mathcal{U}$ maintains a channel of videos. Let the data point $z(u)$ correspond to features x_c characterizing the channel c created by content creator u. Let dist be a metric over features of content. The resulting instantiation of performative power measures the changes in content induced by potential implementations $\mathcal{F}$ of the prediction function and thus captures a power relationship between the platform and the content creators. In Chapter 6.4, we investigate this form of performative power from a theoretical perspective by building on the setup of *strategic classification*.

Viewers. The predictor f can shape the consumption patterns of viewers. In particular, viewers tend to follow recommendations when deciding what content to consume (e.g. (Ursu, 2018)). Thus, by changing which content it recommends to a user, the platform can induce changes in the target variable: how much time the user spends watching content on a given channel. Let's suppose that we wish to investigate the effect of the predictor on viewer consumption of a certain genre of content (e.g. radical content). To formalize this source of power, we let the units $\mathcal{U}$ be viewers. Let the data point $z(u)$ correspond to how long the viewer u spends watching content in the genre of interest. Let $\mathrm{dist}(z, z') = |z - z'|$ capture the difference in watch time. The resulting instantiation of performative power measures the changes in consumption of a given genre of content induced by a set of prediction functions $\mathcal{F}$ the firm could implement. In Chapter 6.5, we propose an observational design to identify this quantity by establishing a formal connection to the causal effect of display position.

6.3 Learning versus steering

Performative power enters the firm's optimization problem and has direct consequences for how a firm can achieve low risk. Instead of identifying the best action f while treating data as fixed, high performative power enables the firm to *steer* the population towards data that it prefers. In the following, we elucidate the role of performative power in the optimization strategy of a firm and the equilibria attained in an economy of predictors.

6.3.1 Optimization strategies

We focus on predictive accuracy as the optimization objective of the firm. Hence, the goal of the firm is to choose a predictive model f that suffers small loss $\ell(f(x), y)$ measured over instances (x, y). To elucidate the role of steering we distinguish between the *ex-ante* loss $\ell(f(x(u)), y(u))$ and the *ex-post* loss $\ell(f(x_f(u)), y_f(u))$. The former describes the loss that the firm can optimize when building the predictor. The latter describes the loss that the firm observes after deploying f. More formally, the *ex-post risk* that the firm suffers after deploying f on a population $\mathcal{U}$ is given by

$$\frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \ell(f(x_f(u)), y_f(u)) \,. \tag{6.1}$$

Expression (6.1) is an instance of what Perdomo et al. (2020) call *performative risk* of a predictor. That is the loss a predictor incurs on the distribution over instances it induces. To simplify notation we adopt their conceptual device of a distribution map: let $\mathcal{D}(\theta)$ map a predictive model, characterized by model parameters θ, to a distribution over data instances.

To express our setting within the framework of performative prediction, we assume the predictive model f is parameterized by a parameter vector $\theta \in \Theta$. We let a data instance correspond to $z(u) = (x(u), y(u))$ for $u \in \mathcal{U}$ so we can capture performativity in the features as well as in the labels. Then, the *aggregate distribution* over data $\mathcal{D}(\theta)$ corresponds to the distribution over the potential outcome variable $z_\theta(u)$ after the firm takes action θ, where the randomness comes from u being uniformly drawn from $\mathcal{U}$ as well as randomness in the potential outcomes. The firm's ex-post risk (6.1) from deploying predictor f_θ corresponds to the performative risk:

$$\mathrm{PR}(\theta) := \mathop{\mathbb{E}}_{z \sim \mathcal{D}(\theta)} [\ell(\theta;\, z)]$$

where the loss typically corresponds to the mismatch between the predicted label and the true label: $\ell(\theta; z) = \ell(f_\theta(x), y)$ for $z = (x, y)$.

In performative risk minimization, observe that θ arises in two places in the objective: in the distribution $\mathcal{D}(\theta)$ and in the loss $\ell(\theta; z)$. Thus, for any choice of model ϕ, we can decompose the performative risk $\mathrm{PR}(\theta)$ as:

$$\mathrm{PR}(\theta) = \mathrm{R}(\phi, \theta) + (\mathrm{R}(\theta, \theta) - \mathrm{R}(\phi, \theta)) \tag{6.2}$$

where $\mathrm{R}(\phi, \theta) := \mathbb{E}_{z\sim\mathcal{D}(\phi)}\, \ell(\theta; z)$ denotes the loss of a model θ on the distribution $\mathcal{D}(\phi)$. This tautology highlights the difference between learning and steering and we differentiate between the following two optimization approaches:

Ex-ante optimization. Ex-ante optimization focuses on optimizing the first term in the decomposition (6.2). For any ϕ, the resulting minimizer can be computed statistically:

$$\theta_{\mathrm{SL}} := \arg\min_{\theta\in\Theta} \mathrm{R}(\phi, \theta).$$

Let f_ϕ be any previously chosen model, then employing supervised learning on historical data sampled from $\mathcal{D}(\phi)$ corresponds to what we call ex-ante optimization.

Ex-post optimization. In contrast to ex-ante optimization, *ex-post optimization* accounts for the impact of the model on the distribution. It trades-off the two terms in (6.2), and directly optimizes the performative risk

$$\theta_{\mathrm{PO}} := \arg\min_{\theta\in\Theta} \mathrm{PR}(\theta).$$

Solving this problem exactly, and finding the performative optimum θ_{PO} requires optimization over the distribution map $\mathcal{D}(\theta)$.

In the context of digital content recommendation, ex-ante optimization corresponds to training the model θ on historical data collected by the platform, whereas ex-post optimization selects θ based on randomized experiments, A/B testing or explicit modeling of $\mathcal{D}(\theta)$. It holds that $\mathrm{PR}(\theta_{\mathrm{PO}}) \leq \mathrm{PR}(\theta_{\mathrm{SL}})$, because in ex-post optimization the firm can choose to steer the population towards more predictable behavior. High ex-post predictability may be an objective worth pursuing for firms relying on predictive optimization (Shmueli and Tafti, 2020), as speculated on in popular science writing (Ward, 2022).

Remark 1 (Generalizing to other objectives)**.** *Note that we focus on predictive accuracy as an objective function. Nonetheless, the conceptual distinction between learning and steering applies to general optimization objectives. Ex-ante optimization corresponds to optimizing on historical data, whereas ex-post optimization corresponds to implicitly or explicitly optimizing over the counterfactuals.*

6.3.2 Gain of ex-post optimization is bounded by a firm's performative power

We show that the gain of ex-post optimization over ex-ante optimization can be bounded by the firm's performative power with respect to the set of actions Θ and the data vector $z = (x, y)$. Intuitively, if the firm's performative power is low, then the distributions $\mathcal{D}(\theta)$ and $\mathcal{D}(\phi)$ for any $\theta, \phi \in \Theta$ are close to one another. This distributional closeness, coupled with a regularity assumption on the loss, means that the second term in (6.2) should be small. Thus, using the ex-ante approach of minimizing the first term produces a near-optimal ex-post solution, as we demonstrate in the following result:

Proposition 26. *Let* P *be the performative power of a firm with respect to the action set* Θ*. Let* L_z *be the Lipschitzness of the loss in* z *with respect to the metric* dist*. Let* θ_{PO} *be the ex-post solution and* θ_{SL} *be the ex-ante solution computed from* $\mathcal{D}(\phi)$ *for any past deployment* $\phi \in \Theta$*. Then, we have that:*

$$\mathrm{PR}(\theta_{\mathrm{SL}}) \leq \mathrm{PR}(\theta_{\mathrm{PO}}) + 4L_z\mathrm{P}.$$

If ℓ *is* γ*-strongly convex, we can further bound the distance between* θ_{SL} *and* θ_{PO} *in parameter space as:*

$$\|\theta_{\mathrm{SL}} - \theta_{\mathrm{PO}}\|_2 \leq \sqrt{\frac{8L_z\mathrm{P}}{\gamma}}.$$

Proposition 26 illustrates that the gain achievable through ex-post optimization is bounded by performative power. Thus, a firm with small performative power cannot do much better than ex-ante optimization and might be better off sticking to classical supervised learning practices instead of engaging with ex-post optimization.

6.3.3 Ex-post optimization in an economy of predictors

The result in Proposition 26 studies the optimization strategy of a single firm in isolation. In this section, we investigate the interaction between the strategies of multiple firms that optimize simultaneously over the same population. We consider an idealized marketplace where C firms all engage in ex-post optimization and we assume all exogenous factors remain constant. Let $\mathcal{D}(\theta^1, \ldots, \theta^{i-1}, \theta^i, \theta^{i+1}, \ldots, \theta^C)$ be the distribution over $z(u)$ induced by each firm $i \in [C]$ deploying model f_{θ^i}. Let ℓ_i denote the loss function chosen by firm i. We say a set of predictors $[f_{\theta^1}, \ldots, f_{\theta^C}]$ is a *Nash equilibrium* if and only if no firm has an incentive to unilaterally deviate from their predictor using ex-post optimization:

$$\theta^i \in \arg\min_{\theta \in \Theta} \mathop{\mathbb{E}}_{z \sim \mathcal{D}(\theta^1, \ldots, \theta^{i-1}, \theta, \theta^{i+1}, \ldots, \theta^C)} [\ell_i(\theta;\, z)].$$

First, we show that at the Nash equilibrium, the suboptimality of each predictor f_{θ^i} on the induced distribution depends on the performative power of the respective firm.

Proposition 27. *Suppose that the economy is in a Nash equilibrium $(\theta^1, \ldots, \theta^C)$, and firm i has performative power P_i with respect to the action set Θ. Let L_z be the Lipschitzness of the loss ℓ_i in z with respect to the metric* dist*. Then, it holds that:*

$$\mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell_i(\theta^i;\, z)] \leq \min_{\theta} \mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell_i(\theta;\, z)] + L_z \mathrm{P}_i\,,$$

where $\mathcal{D} = \mathcal{D}(\theta^1, \ldots, \theta^C)$ is the distribution induced at the equilibrium. If ℓ_i is γ-strongly convex, then we can also bound the distance between θ^i and $\arg\min_{\theta\in\Theta} \mathbb{E}_{z\sim\mathcal{D}}[\ell_i(\theta;\, z)]$ in parameter space.

Proposition 27 implies that if the performative power of all firms is small ($\mathrm{P}_i \to 0\ \forall i$), then the equilibrium becomes indistinguishable from that of a static, non-performative economy with distribution $\mathcal{D}$ over content. However, there is an interesting distinction between such a Nash equilibrium and the static setting: if the firms were to pursue a different strategy and decided to collude—for example, because of common ownership (Azar et al., 2018) —then they would be able to significantly shift the distribution.

Mixture economy. Next, we analyze the behavior of multiple firms optimizing simultaneously. We consider a *mixture economy*, where all of the firms share a common loss function ℓ and performative power is uniformly distributed across firms. Let $z(u), z_\theta^{C=1}(u)$ denote the pair of counterfactual outcomes before and after the deployment of θ in a hypothetical monopoly economy where a single firm holds all the performative power. Let $\mathcal{D}^{C=1}(\theta)$ be the distribution map associated with the variables $z_\theta^{C=1}(u)$ for $u \in \mathcal{U}$. In a uniform mixture economy, we assume that each participant $u \in \mathcal{U}$ uniformly chooses one of the C firms. Consequently, the counterfactual $z_\theta(u)$ associated with one firm changing its predictor to θ is equal to $z(u)$ with probability $1 - 1/C$ and $z_\theta^{C=1}(u)$ otherwise. We can apply Proposition 27 to analyze the equilibria in the limit as $C \to \infty$.

Corollary 28. *Suppose that all firms $i \in [C]$ share the same loss function $\ell_i = \ell$. Let θ^* be a symmetric Nash equilibrium in the mixture economy with C platforms. As $C \to \infty$, it holds that:*

$$\mathop{\mathbb{E}}_{z\sim\mathcal{D}(\theta^*,\ldots,\theta^*)}[\ell(\theta^*;\, z)] \;\to\; \min_{\theta} \mathop{\mathbb{E}}_{z\sim\mathcal{D}^{C=1}(\theta^*)}[\ell(\theta;\, z)].$$

Corollary 28 demonstrates that a symmetric equilibrium approaches a *performatively stable point* of $\mathcal{D}^{C=1}$ as the number of firms in the economy grows and the performative power of each individual firm diminishes. In contrast, if all C firms collude, their performative power adds up and they would obtain the performative power of a monopoly platform. As a consequence, the firms would take advantage of their collective power and all choose a *performatively optimal point* of $\mathcal{D}^{C=1}$—recovering the equilibrium in a monopoly economy with a single firm. Since performatively optimal and performatively stable points can be arbitrarily far apart in general (Miller et al., 2021), a competitive economy of optimizing firms can exhibit a significantly different equilibrium from that of the monopoly or collusive economy.

6.4 Performative power in strategic classification

We now turn to a stylized market model and investigate how performative power depends on the economy in which the firm operates. Specifically, we use *strategic classification* (Hardt et al., 2016) as a test case for our definition. In strategic classification, participants strategically adapt their features with the goal of achieving a favorable classification outcome. Hence, performative power is determined by the degree to which a firm's classifier can impact participant features. We use this concrete market setting to examine the qualitative behavior of performative power in the presence of competition and outside options.

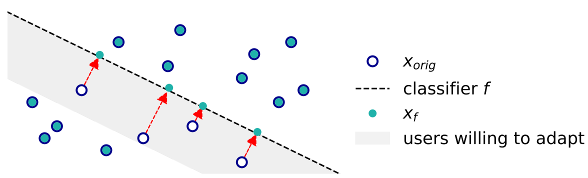
6.4.1 Strategic classification setup

Let $x(u)$ be the *features* and $y(u)$ the *binary label* describing a participant $u \in \mathcal{U}$. A firm chooses a binary predictor $f : \mathbb{R}^m \to \{0, 1\}$ and incurs loss $\ell(f(x), y) = |f(x) - y|$. Let $\mathcal{D}_{\text{orig}}$ denote the base distribution over features and labels $(x_{\text{orig}}(u), y_{\text{orig}}(u))$ absent any strategic adaptation, which we assume is continuous and supported everywhere. Let $\mathcal{D}(f)$ be the distribution over potential outcomes $(x_f(u), y_f(u))$ that arises from the response of participant u to the deployment of a model f. We assume that participant u incurs a cost $c(x_{\text{orig}}(u), x')$ for changing their features to x'. In line with the standard strategic classification setup, the cost for feature changes is measured relative to the *original* features. We further assume that c is a metric, in particular, any feature change that deviates from the original features results in nonnegative cost for participants. Further, we assume the label does not change, i.e., $y_f(u) = y_{\text{orig}}(u)$.

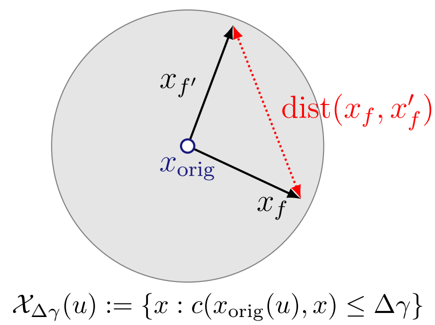
Instantiation of performative power. We measure performative power over the data vector $z(u) = x(u)$, reflecting that strategic behavior impacts the feature vector that enters the prediction function. Then, the choice of distance metric enables us to define how to weight specific feature changes. For instance, in our running example of digital content recommendations where participants correspond to content creators, performative power measures how much the content of each channel changes with changes in the recommendation algorithm. If we are interested in the burden on *content creators*, we choose the distance metric to be aligned with the cost function c of producing a piece of content. However, if we are interested in measuring the impact of changes in content on viewers, a distance metric that reflects harm to viewers might be more appropriate. We keep this distance metric abstract in our analysis.

6.4.2 Performative power in the monopoly setting

Consider a single firm that offers utility $\gamma > 0$ to its participants for a positive classification. We assume that participants want to use the service regardless of what classifier the firm chooses. In addition, assume there is an *outside options* at utility level $\beta > 0$. This decreases the budget participants are willing to invest to their *surplus utility* $\Delta\gamma = \max(0, \gamma - \beta)$. We adopt the following standard rationality assumption on participant behavior.



(a)



(b)

Figure 6.1: Illustrations for 2-dimensional strategic classification example. (left) Participants behave differently depending on their relative position to the decision boundary. (right) Visualization of participant expenditure constraint $\mathcal{X}_{\Delta\gamma}(u)$.

Assumption 3 (Participant Behavior Specification)**.** *Let $\Delta\gamma \geq 0$ be the surplus utility that a participant can expect from a positive classification outcome from classifier f over any outside option. Then, a participant $u \in \mathcal{U}$ with original features $x_{orig}(u)$ will change their features according to*

$$x_f(u) = \arg\max_{x'} \left(\Delta\gamma f(x') - c(x_{orig}(u), x')\right) .$$

Assumption 3 guarantees that a participant will change their features if and only if the cost of a feature change is no larger than $\Delta\gamma$. Furthermore, if participants change their features, then they will expend the minimal cost required to achieve a positive outcome. For $\beta = 0$ this recovers the typical strategic classification setup proposed by Hardt et al. (2016). This specification of participant behavior allows us to bound performative power in terms of the cost function c and the distance function dist. Namely, the potential values that $x_f(u)$ can take on is restricted to

$$\mathcal{X}_{\Delta\gamma}(u) := \{x : c(x_{\mathrm{orig}}(u), x) \leq \Delta\gamma\} . \tag{6.3}$$

Thus, the effect of a change to the decision rule on an individual participant u can be upper bounded by the distance between $x(u)$ and the most distant point in $\mathcal{X}_{\Delta\gamma}$. Aggregating these unilateral effects yields a bound on performative power:

Lemma 29. *The performative power* P *of the firm with respect to any set of predictors $\mathcal{F}$ can be upper bounded as:*

$$\mathrm{P} \leq \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sup_{x' \in \mathcal{X}_{\Delta\gamma}(u)} \mathrm{dist}(x(u), x') \tag{6.4}$$

If the firm action space $\mathcal{F}$ is restricted to a parameterized family, the upper bound in Lemma 29 need not be tight. In particular, a typical decision rule, such as a linear threshold classifier, does not impact all participants $u \in \mathcal{U}$ equally (the amount of change that the firm can induce with a decision rule f on an individual u depends the relative position of their features $x_{\text{orig}}(u)$ to the decision boundary, as we visualize in Figure 6.1). Thus, the firm can't necessarily extract the full utility from all participants simultaneously. We quantify this gap for a 1-dimensional example in Appendix D.1.3.

Personalization. Interestingly, the ability to fully *personalize* decisions to each user maximizes a firm's performative power. To capture this, let the first coordinate of the features $x(u)$ be the index of the user in the population and suppose that this coordinate is immutable. In this case, we can precisely pin down the performative power as follows:

Proposition 30. *Consider a population $\mathcal{U}$ of users. Suppose that the first coordinate is immutable: that is, $c(x, x') = \infty$ if $x_1 \neq x'_1$ and $(x_{\text{orig}}(u))_1 = (x(u))_1 = i$ where i is the index of user u. Then, the performative power with respect to the set $\mathcal{F}$ of all functions from $\mathbb{R}^m$ to $\{0, 1\}$ is given by:*

$$\mathrm{P} = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sup_{x' \in \mathcal{X}_{\Delta\gamma}(u)} \mathrm{dist}(x(u), x').$$

Proposition 30 demonstrates that when firms can fully personalize their decisions to each user, the upper bound in Lemma 29 is in fact tight. In particular, the firm is able to extract maximum utility from each user, despite the heterogeneity in the population.

Value of the service. We investigate the role of $\Delta\gamma$ in the upper bound of Lemma 29. Recall that user behavior is determined by the cost c of changing features relative to $x_{\text{orig}}(u)$, performative power is measured as the distance from the current state $x(u)$ with respect to dist (see Figure 6.1). The Lipschitz constant

$$L := \sup_{x,x'} \frac{\mathrm{dist}(x, x')}{c(x, x')}$$

allows us to relate the two metrics and derive a simpler bound:

Corollary 31. *The performative power P of a firm in the monopoly setup with respect to any set of predictors $\mathcal{F}$ can be bounded as:*

$$\mathrm{P} \leq 2L\,\Delta\gamma. \tag{6.5}$$

where $\Delta\gamma$ measures the surplus utility offered by the service of the firm over outside options.

Corollary 31 makes explicit that $\Delta\gamma > 0$ is a prerequisite for a firm to have any performative power, even in a monopoly economy. This qualitative behavior of performative power is in line with common intuition in economics that monopoly power relies on the firm offering a service that is superior to existing options.

6.4.3 Firms competing for participants

We next consider a model of *competition* between two firms where participants always choose the firm that offers higher utility. In this model of perfectly elastic demand, we demonstrate how the presence of competition reduces the performative power of a firm. In particular, we will show that for a natural constraint on the firm's action set, each firm's performative power can drop to zero at equilibrium, regardless of how much utility participants derive from the firm's service.

To model competition in strategic classification, we specify participant behavior as follows: Given that the first firm deploys f_1 and the second firm deploys f_2, then participant u will choose the first firm if $\max_{x'} (f_1(x') - c(x_{\text{orig}}(u), x')) > \max_{x'} (f_2(x') - c(x_{\text{orig}}(u), x'))$, and choose f_2 otherwise. A participant tie-breaks in favor of the lower threshold, randomizing if they are equal. After choosing firm $i \in \{1, 2\}$, they change their features according to Assumption 3 as $x_f(u) = \arg\max_{x'} (\gamma f_i(x') - c(x_{\text{orig}}(u), x'))$, where γ is the utility of a positive outcome.

We assume that the firm chooses their classifier based on the following utility function. For a rejected participant, the firm receives utility 0 and for an accepted participant, the firm receives utility $\alpha > 0$ if they have a positive label and utility $-\alpha$ if they have a negative label. We assume that the firm's action set is constrained to models for which it derives non-negative utility. More specifically, if f_θ denotes the model deployed by the competing firm, let the action set $\mathcal{F}^+(\theta)$ of this firm denote the set of models that yield non-negative utility for the firm.

For simplicity, focus on a 1-dimensional setup where $\mathcal{F}$ is the set of threshold functions. We assume that the cost function $c(x, x')$ is continuous in both of its arguments, strictly increasing in x' for $x' > x$, strictly decreasing for $x' < x$, and satisfies $\lim_{x' \to \infty} c(x, x') = \infty$. Furthermore, we assume that the posterior $p(x) = \mathbb{P}_{\mathcal{D}_{\text{orig}}}[Y = 1 \mid X = x]$ is strictly increasing in x with $\lim_{x \to -\infty} p(x) = 0$, and $\lim_{x \to \infty} p(x) = 1$.

We show that the presence of competition drives the performative power of each firm to zero.

Proposition 32. *Consider the 1-dimensional setup with two competing firms specified above. Suppose that the economy is at a symmetric Nash equilibrium (θ^*, θ^*). If $L < \infty$, then the performative power of either firm with respect to the action set $\mathcal{F}^+(\theta^*)$ is*

$$\mathrm{P} = 0.$$

The intuition behind Proposition 32 is that performative power of a firm purely arises from how much larger the current threshold θ is than the minimum threshold a firm can deploy within their action set $\mathcal{F}^+(\theta)$. At the Nash equilibrium (where both firms best-respond with respect to their utility functions taking their own performative effects into account), the firms deploy exactly the minimum threshold within their action set. The formal proof of the result can be found in Appendix D.2.8.

Proposition 32 bears an intriguing resemblance to well-known results on market power under Bertrand competition in economics (see e.g., (Baye and Kovenock, 2008)) that show

how a state of zero power is reached in classical pricing economies with only two competing firms.

6.5 Discrete display design

Now that we have examined the theoretical properties of performative power, we turn to the question of *measuring* performative power from observational data. We focus on our running example of digital content recommendation and propose an observational design to measure the recommender system's ability to shape consumption patterns through the arrangement of content.

6.5.1 The causal effect of position

We assume that there are C pieces of content $\mathcal{C} = \{0, 1, 2, \ldots, C-1\}$ that the platform can present in m display slots. We make the convention that item 0 corresponds to leaving the display slot empty. We focus on the case of two display slots ($m = 2$) since it already encapsulates the main idea. The first display slot is more desirable as it is more likely to catch the viewer's attention. Researchers have investigated the causal effect of position on consumption, often via quasi-experimental methods such as regression discontinuity designs, but also through experimentation in the form of A/B tests.

Definition 4 (Causal effect of position)**.** *Let the treatment $T \in \{0, 1\}$ be the action of flipping the content in the first and second display slots for a viewer u, and let the potential outcome variable $Y_t(u)$ indicate whether, under the treatment $T = t$, viewer u consumes the content that is initially in the first display slot. We call the corresponding average treatment effect*

$$\beta = \Big| \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}\left[Y_1(u) - Y_0(u)\right] \Big|$$

the causal effect of position in a population of viewers $\mathcal{U}$, where the expectation is taken over the randomness in the potential outcomes.

For example, Narayanan and Kalyanam (2015) estimate the causal effect of position in search advertising, where advertisements are displayed across a number of ordered slots whenever a keyword is searched. They measured the causal effect of position on click-through rate of participants.

6.5.2 From causal effect of position to performative power

The identification strategy we propose, called *discrete display design (DDD)*, derives a lower bound on performative power by repurposing existing measures of the causal effect of position. Note that we focus on content recommendation in this section, the design however can be generalized to other settings where the firm's action corresponds to a discrete decision of how to display content. Setting up the DDD involves two steps: First, we need to instantiate the definition of performative power with a suitable action set which we choose such that one of the firm's actions result in swapping the position of content items, and second, we plug in the causal effect of position to lower bound performative power.

While the first step is mostly a technical exercise, the second step relies on a crucial assumption. In particular, it involves relating the unilateral causal effect of position to performative power that quantifies the effect of an action on the entire population of viewers. Thus, for being able to extrapolate the effect from a single viewer to the population DDD relies on a non-interference assumption. In the advertising example, this means that the ads shown to one viewer do not influence the consumption behavior of another viewer. We investigate the two steps in detail:

Step 1: Instantiating performative power. Let the units $\mathcal{U}$ be the set of viewers. For each viewer $u \in \mathcal{U}$ let $z(u) \in \mathbb{R}^C$ be the distribution over content items $\mathcal{C}$ consumed by viewer u, represented as a histogram. More formally, let $z(u)$ be a vector in the C-dimensional probability simplex where the ith coordinate is the probability that viewer u consumes content item i. The metric $\mathrm{dist}(z, z')$ is the ℓ_1-distance $\mathrm{dist}(z, z') = \sum_{i=0}^{C-1} |z[i] - z'[i]|$.

The decision space $\mathcal{F}$ of the firm corresponds to its decisions of how to arrange content in the $m = 2$ display slots. It is natural to decompose this decision into a continuous score function s followed by a discrete conversion function κ that maps scores into an allocation. The score function $s\colon \mathcal{U} \to \mathbb{R}^C$ maps the viewer to a vector of scores, where each coordinate is an estimate of the quality of the match between the viewer and the corresponding piece of content. The conversion function $\kappa\colon \mathbb{R}^C \to \mathcal{C}^2$ takes as input the vector of scores and outputs an ordered list of items with the top 2 scores. We assume the platform displays these 2 items in order and the conversion function κ is fixed. Hence, we identify the firm's action space with the set of feasible score functions $\mathcal{S} \subseteq \mathcal{U} \to \mathbb{R}^C$.

To define the reference state $z(u)$, we think of s_{curr} as being the score function currently deployed by the platform. Let δ be the maximum difference in the highest score and second highest score for any user under s_{curr}. Consider the set $\mathcal{S}$ of local perturbations to the scoring function s_{curr} defined as

$$\mathcal{S} := \left\{ s\colon \mathcal{U} \to \mathbb{R}^C \mid \forall u \in \mathcal{U}\colon \|s(u) - s_{\mathrm{curr}}(u)\|_\infty \leq \delta \right\}.$$

Notably, there exists an $s_{\mathrm{swap}} \in \mathcal{S}$ that is capable of swapping the order of the first and second highest scoring item under s_{curr} for any user $u \in \mathcal{U}$ simultaneously. We denote the counterfactual variable corresponding to a score function $s \in \mathcal{S}$ as $z_s(u)$. Given this specification, performative power with respect to the action set $\mathcal{S}$ can be bounded by the causal effect of s_{swap} as follow

$$\mathrm{P} = \sup_{s \in \mathcal{S}} \frac{1}{\mathcal{U}} \sum_{u \in \mathcal{U}} \|z_{s_{\mathrm{curr}}}(u) - z_s(u)\|_1 \geq \frac{1}{\mathcal{U}} \sum_{u \in \mathcal{U}} \|z_{s_{\mathrm{curr}}}(u) - z_{s_{\mathrm{swap}}}(u)\|_1. \tag{6.6}$$

Step 2: Lower bounding performative power. To relate the lower bound on performative power from (6.6) to the causal effect of position, let $i_{\text{top}}(u) = \kappa \circ s_{\text{curr}}(u)[1]$ denote the coordinate of the item displayed to user u in the first display slot under s_{curr}. Then, we can lower bound each term in the sum (6.6) as

$$\|z_{s_{\text{curr}}}(u) - z_{s_{\text{swap}}}(u)\|_1 \geq |z_{s_{\text{curr}}}(u)[i_{\text{top}}(u)] - z_{s_{\text{swap}}}(u)[i_{\text{top}}(u)]|.$$

Now, to enable us to study the effect of changing s_{curr} to s_{swap} independently for each user we place the following non-interference assumption on the counterfactual variables which closely relates to the stable unit treatment value assumption (SUTVA) (Imbens and Rubin, 2015) prevalent in causal inference.

Assumption 4 (No interference across units)**.** *For any $u \in \mathcal{U}$ and any pair of scoring functions $s_1, s_2 \in \mathcal{S}$, if $\kappa(s_1(u)) = \kappa(s_2(u))$, it also holds that $z_{s_1}(u) = z_{s_2}(u)$.*

The assumption requires that there are no spill-over or peer effects and the content a viewer consumes only depends on the content recommended to them and not the content recommended to other viewers. The last step is to see that the effect of a unilateral change to the consumption of item $i_{\text{top}}(u)$ under s_{swap} exactly corresponds to what we defined as the causal effect of position. Aggregating these unilateral causal effects across all viewers in the population we obtain a lower bound on performative power. The proof of Theorem 33 can be found in Appendix D.2.11.

Theorem 33. *Let* P *be performative power as instantiated above. If Assumption 4 holds, then performative power is at least as large as the causal effect of position*

$$\mathrm{P} \geq \beta.$$

As a case study, consider the search advertisement marketplace of Narayanan and Kalyanam (2015). We can leverage Theorem 33 to relate the findings of their observational causal design to performative power. In particular, Narayanan and Kalyanam (2015) examine position effects in search advertising, where ads are displayed across a number of ordered slots whenever a keyword is searched. They found that the effect of showing an ad in display slot 1 versus display slot 2 corresponds to 0.0048 clicks per impression (see Table 2 in their paper). By treating each incoming keyword query as a distinct "viewer" u, this number exactly corresponds to what we defined as the causal effect of position. Thus, we can apply Theorem 33 to get $P \geq 0.0048$. Putting this into context; the mean click-through rate in display slot 2 is 0.023260. Hence, the lower bound 0.0048 is a 21% percent increase relative to the baseline. The firm thus has a substantial ability to shape what advertisements users click on.

6.6 Discussion

We discuss the potential role of performative power in competition policy and antitrust enforcement. The complexity of digital marketplaces has made it necessary to develop new approaches for evaluating and regulating these economies. One challenge is that traditional measures of market power—such as the Lerner Index (Lerner, 1934), or the Herfindahl–Hirschman Index (HHI)—are based on classical pricing markets for homogeneous goods, but these markets map poorly to digital economies. In particular, these measures struggle to appropriately capture the multi-sided nature of digital economies, to describe the multi-dimensionality of interactions, and to account for the role of behavioral weaknesses of consumers—such as tendencies for single-homing, vulnerability to addiction, and the impact of framing and nudging on participant behavior (e.g. Thaler and Sunstein, 2008; Fogg, 2002). We further expand on this in Appendix D.1.

By focusing on directly observable statistics, performative power could be particularly helpful in markets that resist a clean mathematical specification. Performative power is sensitive to the market nuances without explicitly modeling them. For example, suppose that as a result of uncertainty about market boundaries, a regulator failed to account for a competitor in a marketplace. Performative power would still implicitly capture the impact of the competitor and indicate the market is more competitive than suspected.

We leave open the question of how to best instantiate performative power in a given marketplace. Conceptually, we view performative power as a tool to flag market situations that merit further investigation, since it corresponds to "potential for harm to users". However, if a regulator wishes to draw fine-grained conclusions about consumer harm, it is crucial to appropriately instantiate the choice of action set $\mathcal{F}$, the definition of a population $\mathcal{U}$, and the specification of the features z. As an example, we show in Appendix D.1.2 how to closely relate performative power into consumer harm for strategic classification. In general, however, harm and power are two distinct normative concepts, and going from performative power to consumer harm thus requires additional substantive arguments.

Part III

Incentives for Digital Content Creation

Chapter 7

Overview

In content recommendation ecosystems, ML models shape incentives for digital content creation. For example, consider recommendation models, which determine what content to show to consumers; since content creators often want to maximize the exposure of their content, recommendation models influence how creators are incentivized to design their content, and thus shape the supply of content available on the platform. As another example, consider generative models, which continue to make it cheaper to produce digital content; the resulting technology improvements influence whether digital content is produced by content creators or directly by consumers themselves.

However, these supply-side effects are largely neglected when designing and evaluating the ML models deployed in content recommendation ecosystems. Specifically, the classical view of a recommendation model is that it selects which content from a fixed content supply to show consumers. This perspective—which underpins information retrieval as well as standard recommendation approaches such as matrix factorization (Koren et al., 2009) and two-tower embedding models (Yi et al., 2019)—treats the content supply as static. Moreover, generative models tend to be evaluated based on whether they output well-received content and how they impact creator productivity (e.g., (Zhou and Lee, 2024)). This perspective does not capture how these models influence the ecosystem-level structure of the supply-side market—including which types of users are incentivized to create content in the ecosystem.

7.1 Our contributions

This part investigates how the interactions between content creators and ML models shape digital content creation. We characterize how a recommendation model shapes the supply of content available on the platform, and how these supply-side effects amplify the impact of details of the recommendation model. Moreover, we show how the costs associated with using generative models influence whether consumers are incentivized to directly create content on their own. We describe this in more detail below:

- In Chapter 8, we characterize how a recommendation model shapes creator incentives to specialize content to niche consumers vs. create mainstream content. We focus on the role of *learned embeddings*, motivated by how standard recommendation models learn D-dimensional embeddings for consumers and content, and select recommendations based on linear scores. We show that specialization occurs if and only if learned consumers embeddings are sufficiently heterogeneous relative to costs of synthesizing many consumer preferences, and we empirically connect this finding to matrix factorization.

- In Chapter 9, we characterize how a recommender system shapes creator incentives to invest in clickbait vs. quality. We theoretically and empirically show that regardless of the engagement metric optimized by the recommender system, the content supply exhibits a positive correlation between clickbait and quality. We then theoretically show that these supply-side effects can lead engagement-based optimization to perform worse than simple baselines in terms of both realized engagement and user welfare.

- In Chapter 10, we investigate how the costs of using generative models impact whether creators or consumers produce content. We show that consumers are incentivized to bypass creators as long as costs are sufficiently high or sufficiently low. We also characterize the downstream impact of creator disintermediation: for example, we show that the presence of creators can counterintuitively lead to lower content quality, even though creators benefit from economies of scale.

7.2 Methodological theme

In this part, a common methodological theme is again to leverage economic models, but with an eye towards the details of ML models in content recommendation ecosystems.

In Chapter 8, we generalize *product selection models* (Hotelling, 1929) to high dimensions, by drawing upon models for *product characteristics* (Berry, 1994). This bears similarity to Part II, where product selection models also served as a foundation for some of those works. In fact, the connection runs even deeper: Chapter 8 and Chapter 3 both investigate how learned embeddings shape specialization. However, the takeaways go in *opposite directions*: increasing the informativeness of the consumer embeddings learned by the recommendation model in Chapter 8 leads to greater specialization, whereas increasing the quality of the embeddings learned by the pretrained model in Chapter 3 leads to lower specialization. This reversal traces back to how learned embeddings interact with the product space in different ways in the two ecosystems: this is because products capture digital content in this part, while products capture ML models in Part II. This comparison underscores the need to perform detail-sensitive analyses of ecosystem-level outcomes, which helps justify the methodology of this thesis more broadly.

In Chapter 9, we develop a model for product selection that draws upon ideas from strategic classification (Hardt et al., 2016) which has roots in *contract theory*. In Chapter 10, we build on standard economic models of *supply chains* but focus on how technology improvements lead to disintermediation.

7.3 Other co-authored work

In other co-authored work which is not included in this thesis, we further investigate incentives for digital content creation.

Specifically, we focus on ecosystems where creators frequently upload new content, and the platform deploys a learning algorithm to change what content is recommended over time. In Hu et al. (2023) (led by Xinyan Hu), we investigate how the learning algorithm shapes creator effort over time (i.e., whether creators free-ride off of their reputation or maintain consistent effort), and we design incentive-aware learning algorithms to incentivize the creation of a high-quality content supply. In Dai et al. (2024) (led by Jessica Dai), we investigate how the learning algorithm's treatment of probabilistic feedback influences whether creators are incentivized to produce reactive vs. unreactive content, and we design black-box algorithmic transformations which steer the content supply towards each extreme.

In a broader position paper (Dean et al., 2024b), we use content recommender systems as a case study to argue for the development of formal interaction models that capture how AI systems and users shape one another.

Chapter 8

Specialized vs. Homogenized Content

This chapter is based on *"Supply Side Equilibria in Recommender Systems"* (Jagadeesan et al., 2023a), which is joint work with Nikhil Garg and Jacob Steinhardt.

8.1 Introduction

Algorithmic recommender systems have disrupted the production of digital goods such as movies, music, and news. In the music industry, artists have changed the length and structure of songs in response to Spotify's algorithm and payment structure (Hodgson, 2021). In the movie industry, personalization has led to low-budget films catering to specific audiences (McDonald, 2019), in some cases constructing data-driven "taste communities" (Adalian, 2018). Across industries, recommender systems shape how producers decide what content to create, influencing the supply side of the digital goods market. This raises the questions: *What factors drive and influence the supply-side marketplace? What content will be produced at equilibrium?*

Intuitively, supply-side effects are induced by the multi-sided interaction between producers, the recommendation algorithm, and users. Users tend to follow recommendations when deciding what content to consume (Ursu, 2018)—thus, recommendations influence how many users consume each digital good and impact the profit (or utility) generated by each content producer. As a result, content producers shape their content to maximize appearance in recommendations; this creates competition between the producers, which can be modeled as a game. However, understanding such producer-side effects has been difficult, both empirically and theoretically. This is a pressing problem, as these gaps in understanding have hindered the regulation of digital marketplaces (Stigler Committee, 2019).

At a high level, there are two primary challenges that complicate theoretical analyses of these supply-side effects. (1) Digital goods such as movies have many attributes and thus must be embedded in a *multi-dimensional* continuous space, leading to a large producer action space. This multi-dimensionality is a departure from traditional economic models of price and spatial competition. (2) A core aspect of such marketplaces is the potential

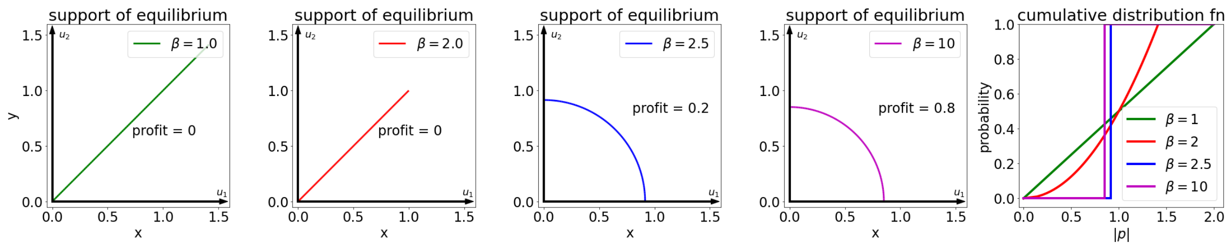


Figure 8.1: A symmetric equilibrium for different settings of β, for 2 users located at the standard basis vectors e_1 and e_2, $P = 2$ producers, and producer cost function $c(p) = \|p\|_2^\beta$. The first 4 plots show the support of the equilibrium μ. As β increases, there is a phase transition from a single-genre equilibrium to an equilibrium with infinitely many genres (Theorem 46). This illustrates how the cost function influences whether or not specialization occurs. The profit also transitions from zero to positive, demonstrating how specialization reduces the competitiveness of the marketplace (Propositions 52-53). The last plot shows the cumulative distribution function of $\|p\|$ where $p \sim \mu$, which is a step function for the multi-genre equilibria: all equilibria thus exhibit either pure vertical differentiation or pure horizontal differentiation.

for specialization: that is, different producers may produce different items at equilibrium. Incentives to specialize depend on the level of *heterogeneity of user preferences* and the cost structure for producing goods (whether it is more expensive to produce items that are good in multiple dimensions). As a result, supply-side equilibria have the potential to exhibit rich economic phenomena, but pose a challenge to both modeling and analysis.

We introduce a simple game-theoretic model for supply-side competition in personalized recommender systems. Our model captures the multi-dimensional space of producer decisions, rich structures of production costs, and general configurations of users. Users and digital goods are represented as D-dimensional vectors in $\mathbb{R}^D_{\geq 0}$, and the inferred user value of a digital good p for a user with vector u is equal to the inner product $\langle u, p \rangle$. The platform has $N \geq 1$ users and $P \geq 2$ producers: each user $i \in [N]$ is associated with a fixed vector $u_i \in \mathbb{R}^D_{\geq 0}$, and each producer $j \in [P]$ *chooses* a single digital good $p_j \in \mathbb{R}^D_{\geq 0}$ to create. The recommendation algorithm is personalized and shows each user the good with maximum inferred user value for them, so user i is recommended the digital good created by producer $j^*(i) = \arg\max_{1 \leq j \leq P} \langle u_i, p_j \rangle$. The goal of a producer to maximize their profit, which is equal to the number of users who are recommended their content minus the (one-time) cost of producing the content. We consider producer cost functions of the form $c(p) := \|p\|^\beta$, where $\|\cdot\|$ is an arbitrary norm and the exponent β is at least 1. Our model can be viewed as high-dimensional variant of a competitive facility location game (ReVelle and Eiselt, 2005) as we describe Chapter 8.1.1.

In this model, producers face a complex choice of what content p to create. To understand this choice better, let's first focus on a single user u. A producer can increase their chance of winning u with two levers: (1) improving the content's *quality* (vector norm $\|p\|$) or (2) aligning the content's *genre* (direction $p/\|p\|$) with the user vector u. As to how these levers impact the chance of winning other users, improving quality simultaneously improves the producer's chance of winning every user; however, aligning the genre with one user can worsen the alignment of the genre with other users. This creates tradeoffs between alignment with different users, which producers must balance when selecting the genre of their content: producers may choose a niche genre that perfectly caters to a specific user or subgroups of users, or choose a generic genre that somewhat caters to all of the users.

To ground our investigation of these complex producer choices, we focus on one particular economic phenomena—*the potential for specialization*—in this chapter. Specialization, which occurs when different producers create different genres of content at equilibrium, has several economic consequences. For example, whether specialization occurs, as well as the form that specialization takes, determines the diversity of content available on the platform. Moreover, specialization influences the competitiveness of the marketplace by reducing the amount of competition in each genre. This raises the questions:

> *Under what conditions does specialization occur at equilibrium? What form does specialization take? What is its impact on market competitiveness?*

Before mathematically studying these questions, we need to specify the equilibrium concept and formalize specialization. We focus on symmetric mixed Nash equilibria, which we show are guaranteed to exist in Chapter 8.2.1. These symmetric equilibria can be represented as a distribution μ over $\mathbb{R}^D_{\geq 0}$ and are thus more tractable than general asymmetric equilibria. Although we focus on symmetric equilibria, we can nonetheless capture specialization—which is an asymmetric concept—in terms of the support of the equilibrium distribution μ. We say that *specialization* occurs at an equilibrium μ if and only if the support of μ has more than one *genre* (direction).[1] The particular form of specialization exhibited by μ is further captured by the number and set of genres in the support of μ. See Figure 8.1 for a depiction of markets with a single-genre equilibrium and markets with a multi-genre equilibrium.

With this formalization, we investigate specialization and its consequences on the supply-side market. We analyze how the specialization exhibited at equilibrium varies with user vector geometry $(u_1, \ldots, u_N)$ and producer cost function parameters ($\|\cdot\|$ and β). Our main results provide insight into each of the questions from above: we characterize when specialization occurs, analyze the form that specialization takes, and investigate the impact of specialization on market competitiveness.

Characterization of when specialization occurs. We first provide a tight geometric characterization of when a marketplace has a single-genre equilibria versus has all multi-genre equilibria (Theorem 38). Interestingly, the occurrence of specialization depends on the geometry of the users as well as the cost function parameters, but does *not* depend on the number of producers P. For example, in the concrete instance depicted in Figure 8.1,

[1] See Chapter 8.2.3 for a mathematical definition of specialization.

single-genre equilibria exist exactly when $\beta \leq 2$. Conceptually, larger β make producer costs more superlinear, which eventually discentivizes producers from attempting to perform well on all dimensions at once.

In Chapter 8.3, we show several corollaries of Theorem 38 that elucidate the role of β in concrete instances and characterize the direction of single-genre equilibria. We also provide an empirical analysis using the MovieLens-100K dataset (Harper and Konstan, 2015) that offers additional qualitative intuition for our theoretical results (Figure 8.5). The empirical analysis also explicitly connects our model to recommender systems performing nonnegative matrix factorization: the embedding dimension D corresponds the number of factors used in matrix factorization, and the user vectors and content vectors correspond to the embeddings learned by matrix factorization.

Form of specialization. For further economic insight, we focus on the concrete setting of two equally sized populations of users with cost function $c(p) = \|p\|_2^\beta$. We first show that all equilibria must have either one or infinitely many genres (Theorem 46). Producers thus do not simply randomize between genres aligned with the two user vectors; instead, producers randomize across infinitely many genres of content that balance the preferences of the two populations in different ways. In several examples, the equilibrium spans all possible genres (e.g. see Figure 8.1 and Figure 8.6).

We also recover equilibria in an infinite-producer limit for any 2 user vectors (Theorem 49; see Figure 8.2). Interestingly, these equilibria have two genres: thus, even though two-genre equilibria do not exist for finite P by Theorem 46, they turn out to re-emerge in the limit. The resulting equilibrium distribution also has complex structure, e.g., the support consists of countably infinite disjoint line segments.

Impact of specialization on market competitiveness. Finally, we study how specialization affects the equilibrium profit level of producers, which provides insight into market competitiveness. We show that producers can achieve positive profit at a multi-genre equilibrium (Chapter 52). The marketplace can therefore exhibit monopolistic behavior; the intuition is that specialization reduces competition along each genre of content. We confirm this intuition by showing that without specialization (i.e. at single-genre equilibria), the producer profit is always zero (Proposition 53). This analysis of equilibrium profit establishes a distinction between single- and multi-genre equilibria, which parallels classical distinctions between markets with homogeneous goods and markets with differentiated goods.[2] Our results thus formalize how the supply-side market of a recommender system can resemble a market with homogeneous goods or with differentiated goods, depending on whether or not specialization occurs.

Technical tools. En route to our results, we develop technical tools to analyze the complex, multi-dimensional behavior of producers. We highlight two tools here which may be of broader interest.

[2]See Anderson et al. (1992) for a textbook treatment.

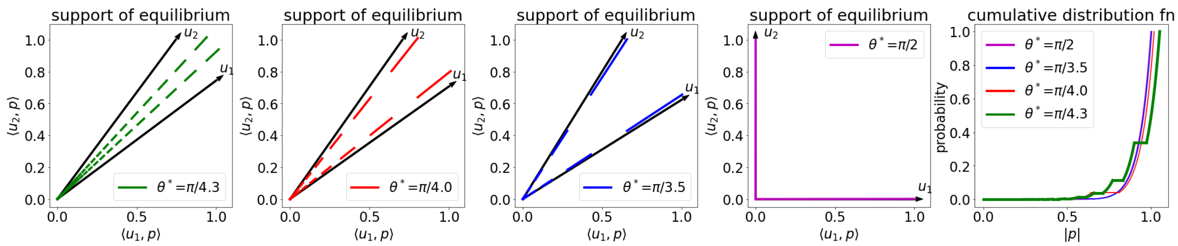


Figure 8.2: A symmetric equilibrium for different settings of θ^*, for 2 users located at u_1 and u_2 such that $\theta^* = \cos^{-1}\left(\frac{\langle u_1, u_2\rangle}{\|u_1\|\|u_2\|}\right)$, for producer cost function $c(p) = \|p\|_2^\beta$ with $\beta = 7$, and for $P = \infty$ producers (see Theorem 49). The first 4 plots show the support of the equilibrium in a reparameterized space: note that the the x-axis is $\langle u_1, p\rangle$ and the y-axis is $\langle u_2, p\rangle$, i.e., the inferred user values for good p. These equilibria have 2 genres: thus, although two-genre equilibria do not exist for any finite P (Theorem 46), they do exist in the infinite-producer limit. The last plot shows the cumulative distribution function of the conditional quality distribution (i.e. the distribution of the maximum quality along a genre). The support consists of countably infinite disjoint intervals, with the property that at most one of the genres achieves a given utility for a given user.

1. To analyze when specialization occurs, we draw a connection to minimax theory in optimization. In particular, we show that the existence of a single-genre equilibrium is equivalent to strong duality holding for a certain optimization program that we define (Lemma 39). This allows us to leverage techniques from optimization theory to provide a necessary and sufficient condition for genre formation (Theorem 38).

2. To analyze the properties of equilibria in concrete instances, we provide a decoupling lemma in terms of the equilibrium's support and its one-dimensional marginals (Lemma 50). This produces one-dimensional functional equations that make solving for the underlying equilibrium more tractable. We apply this decoupling lemma to analyze the form of specialization in the concrete setting of two equally sized populations of users with cost function $c(p) = \|p\|_2^\beta$.

Other technical ideas underlying our results include formalizing the formation of genres—which intuitively captures heterogeneity across producers—in terms of the support of a symmetric equilibrium distribution and applying the technology of discontinuous games Reny (1999) to establish the existence of symmetric mixed equilibria.

Summary of our results. Our simple model yields a nuanced picture of supply-side equilibria in recommender systems. Our results provide insight into specialization and its implications, and en route to proving these results, we develop a technical toolkit to analyzing the multi-dimensional behavior of producers. More broadly, our model and results open the door to investigating how recommender systems shape the diversity and quality of content created by producers, and we outline several directions for future work in Chapter 8.6.

8.1.1 Related Work

Our work is related to research on *societal effects in recommender systems*, *models of competition in economics and operations research*, and *strategic effects induced by algorithmic decisions*.

Supply-side effects of recommender systems. A line of work in the machine learning literature has studied supply-side effects from a theoretical perspective, but existing models do not capture the infinite, multi-dimensional decision space of producers. Ben-Porat and Tennenholtz (2018) study supply-side effects with a focus on mitigating strategic effects by content producers; Ben-Porat et al. (2020), building on Basat et al. (2017), also studied supply-side equilibria with a focus on convergence of learning dynamics for producers. The main difference from our work is that producers in these models choose a topic from a *finite* set of options; in contrast, our model captures the infinite, multi-dimensional producer decision space that drives the emergence of genres. Moreover, we focus on the structure of equilibria rather than the convergence of learning.

In concurrent and independent work, Hron et al. (2022) study a related model for supply-side competition in recommender systems where producers choose content embeddings in $\mathbb{R}^D$. One main difference is that, rather than having a cost on producer content, they constrain producer vectors to the ℓ_2 unit ball (this corresponds to our model when $\beta \to \infty$ and the norm is the ℓ_2-norm, although the limit behaves differently than finite β). Additionally, Hron et al. incorporate a softmax decision rule to capture exploration and user non-determinism, whereas we focus entirely on hardmax recommendations. Thus, our model focuses on the role of producer costs while Hron et al.'s focuses on the role of the recommender environment. At a technical level, Hron et al. study the existence of different types of equilibria and the use of behaviour models for auditing, whereas we analyze the economic phenomena exhibited by symmetric mixed strategy Nash equilibria, with a focus on specialization.

Other work has studied the emergence of filter bubbles (Flaxman et al., 2016), the ability of users to reach different content (Dean et al., 2020), the shaping of user preferences (Adomavicius et al., 2013), and stereotyping (Guo et al., 2021).

Models of competition in microeconomics and operations research. Our model and research questions relate to classical models of competition in economic theory; however, particular aspects of recommender systems—high-dimensionality of digital goods, rich structure of producer costs, and user geometry—are not captured by these classical models. For example, in *price competition*, producers set a *price*, but do not decide what good to produce (e.g. Bertrand competition, see (Baye and Kovenock, 2008) for a textbook treatment). Price is a one-dimensional quantity, but producer decisions in our model are multi-dimensional.

Another line of work on *product selection* has investigated how producers choose goods (i.e., content) at equilibrium (see Anderson et al. (1992) for a textbook treatment). For example, in *competitive facility (spatial) location models* (see ReVelle and Eiselt (2005) for a survey), producers choose a direction in a low-dimensional space (e.g., $\mathbb{R}^1$ in (Hotelling, 1929; d'Aspremont et al., 1979) and $\mathbb{S}^1$ in (Salop, 1979)), and users typically receive utility based on the negative of the Euclidean distance. In contrast, producers in our model jointly select the direction and magnitude of their content, and users receive utility based on inner product. Since some variants of spatial location models additionally allow producers to set prices, it may be tempting to draw an analogy between the quality $\|p\|$ in our model and the price in these models. However, this analogy breaks down because production costs in our model can be highly nonlinear in the quality $\|p\|$ (i.e. when the cost function exponent β is greater than 1). In fact, this nonlinear structure creates tradeoffs between excelling in different dimension; these tradeoffs underpin our specialization results (Theorem 38).

Other related work has investigated *supply function equilibria* (e.g. (Grossman, 1981)), where the producer chooses a function from quantity to prices, rather than what content to produce, and the *pure characteristics model* (e.g. (Berry, 1994)), where attributes of users and producers are also embedded in $\mathbb{R}^D$ like in our model, but which focuses on demand estimation for a fixed set of content, rather than analyzing the content that arises at equilibrium in the marketplace. Recent work in economics has allowed for endogenous product choice (e.g. (Wollmann, 2018)) and also studied specialization (e.g. (Vogel, 2008; Perego and Yuksel, 2022)), though with different modeling choices than our work.

Strategic classification. A line of work of *strategic classification* (Brückner et al., 2012; Hardt et al., 2016) has studied how algorithmic decisions induce participants to strategically change their features to improve their outcomes, but with different assumptions on participant behavior. In particular, the models for participant behavior in this line of work (e.g. Kleinberg and Raghavan (2020); Jagadeesan et al. (2021); Ghalme et al. (2021)) generally do not capture competition between participants. One exception is Liu et al. (2022), where participants compete to appear higher in a single ranked list; in contrast, the participants in our model simultaneously compete for users with *heterogeneous* preferences.

8.2 Model and Preliminaries

We introduce a game-theoretic model for supply-side competition in recommender systems. Consider a platform with $N \geq 1$ heterogeneous users who are offered personalized recommendations and $P \geq 2$ producers who strategically decide what digital good to create.

Embeddings of users and digital goods. Each user i is associated with a D-dimensional embedding u_i that captures their preferences. We assume that $u_i \in \mathbb{R}^D_{\geq 0} \setminus \left\{\vec{0}\right\}$—i.e., the coordinates of each embedding are nonnegative and each embedding is nonzero. While user vectors are fixed, producers *choose* what content to create. Each producer j creates a single digital good, which is associated with a content vector $p_j \in \mathbb{R}^D_{\geq 0}$. The inferred user value of good p for user u is $\langle u, p \rangle$.

Personalized recommendations. After the producers decide what content to create, the platform offers personalized recommendations to each user. We consider a stylized model where the platform has complete knowledge of the user and content vectors. The platform recommends to each user the content with the maximal inferred user value for them, assigning them to the producer who created this content. Mathematically, the platform assigns a user u to the producer j^*, where $j^*(u;\ p_{1:P}) = \arg\max_{1 \leq j \leq P} \langle u, p_j \rangle$. If there are ties, the platform sets $j^*(u;\ p_{1:P})$ to be a producer chosen uniformly at random from the argmax.

Producer cost function. Each producer faces a *fixed (one-time) cost* for producing content p, which depends on the magnitude of p. Since the good is digital and thus cheap to replicate, the production cost does not scale with the number of users. We assume that the cost function $c(p)$ takes the form $\|p\|^\beta$, where $\|\cdot\|$ is any norm and the exponent β is at least 1. The magnitude $\|p\|$ captures the *quality* of the content: in particular, if a producer chooses content λp, they win at least as many users as if they choose $\lambda' p$ for $\lambda' < \lambda$. (This relies on the fact that all vectors are in the positive orthant.) The norm and β together encode the cost of producing a content vector v, and reflect cost tradeoffs for excelling in different dimensions (for example, producing a movie that is both a drama and a comedy). Large β, for instance, means that this cost grows superlinearly. In Chapter 8.3, we will see that these tradeoffs capture the extent to which producers are incentivized to specialize.

Producer profit. A producer receives profit equal to the number of users who are recommended their content minus the cost of producing the content. The profit of producer j is equal to:

$$\mathcal{P}(p_j; p_{-j}) = \mathbb{E}\Big[\Big(\sum_{i=1}^{n} \mathbb{1}[j^*(u_i;\ p_{1:P}) = j]\Big) - \|p_j\|^\beta\Big], \tag{8.1}$$

where $p_{-j} = [p_1, \ldots, p_{j-1}, p_{j+1}, \ldots, p_P]$ denotes the content produced by all of the other producers and where the expectation comes from the randomness over platform recommendations in the case of ties.

8.2.1 Equilibrium concept and existence of equilibrium

We study the Nash equilibria of the game between producers. In particular, each producer j chooses a (random) strategy over content, given by a probability measure μ_j over the content embedding space $\mathbb{R}^D_{\geq 0}$. The strategies $(\mu_1, \ldots, \mu_P)$ form a *Nash equilibrium* if no producer—given the strategies of other producers—can chose a different strategy where they achieve higher expected profit: that is, for every $j \in [P]$ and every $p_j \in \text{supp}(\mu_j)$, it holds that $p_j \in \arg\max_{p \in \mathbb{R}^D_{\geq 0}} \mathbb{E}_{p_{-j} \sim \mu_{-j}}[\mathcal{P}(p; p_{-j})]$. A salient feature of our model is that there are discontinuities in the producer utility function in equation (8.1), since the function $\arg\max_{1 \leq i \leq P} \langle u_i, p_j \rangle$ changes discontinuously with the producer vectors p_j. Due to these discontinuities, pure strategy equilibria do not exist.[3]

Proposition 34. *For any set of users and any $\beta \geq 1$, a pure strategy equilibrium does not exist.*

[3]A Nash equilibrium $(\mu_1, \mu_2, \ldots, \mu_P)$ is a *pure strategy equilibrium* if each μ_j contains only one vector in its support; otherwise, it is a *mixed strategy equilibrium*.

The intuition is that if two producers are tied, then a producer can increase their utility by infinitesimally increasing the magnitude of their content.

Since pure strategy equilibria do not exist, we must turn to mixed strategy equilibria. Using the technology of equilibria in discontinuous games (Reny, 1999), we show that a mixed strategy equilibrium exists. In fact, because of the symmetries in the producer utility functions, we can actually show that a *symmetric* mixed strategy equilibrium (i.e. an equilibrium where $\mu_1 = \ldots = \mu_P$) exists.

Proposition 35. *For any set of users and any $\beta \geq 1$, a symmetric mixed equilibrium exists.*

Interestingly, symmetric mixed equilibria must exhibit significant randomness across different content embeddings. (Note that every symmetric equilibrium must exhibits some randomization, since pure strategy equilibria do not exist.) In particular, we show that a symmetric mixed equilibrium cannot contain point masses.

Proposition 36. *For any set of users and any $\beta \geq 1$, every symmetric mixed equilibrium is atomless.*

Proposition 36 implies that a symmetric mixed equilibrium has *infinite support*. The randomness can come from randomness over *quality* $\|p\|$ as well as randomness over *genres* $p/\|p\|$.

We take the symmetric mixed equilibria of this game as the main object of our study, since they are both tractable to analyze and rich enough to capture asymmetric solution concepts such as specialization. In terms of tractability, a symmetric mixed equilibrium (unlike an asymmetric equilibrium) can be represented as a *single* distribution μ. Despite this simplicity, we can still study specialization—which is an asymmetric concept—within the family of symmetric equilibria as we formalize in Chapter 8.2.3.

8.2.2 Warmup: Homogeneous Users

To gain intuition for the structure of μ, let's focus on a simple one-dimensional setting with one user. We show that the equilibria take the following form (see Figure 8.3):

Example 3 (1-dimensional setup)**.** *Let $D = 1$, and suppose that there is a single user $u_1 = 1$. Suppose the cost function is $c(p) = |p|^\beta$. The unique symmetric mixed equilibrium μ is supported on the full interval $[0, 1]$ and has cumulative distribution function $F(p) = (p/N)^{\beta/(P-1)}$. We defer the derivation to Appendix E.2.4.*

Since $D = 1$ in Example 3, content is specified by a single value $p \in \mathbb{R}^{\geq 0}$. Since the user will be assigned to the content with the highest value of p, we can interpret p as the *quality* of the content. For a producer, setting p to be larger increases the likelihood of being assigned to users, at the expense of a greater cost of production.

The equilibrium changes substantially with the parameters β and P. First, for any fixed P, the equilibrium distribution for higher values of β stochastically dominates the equilibrium

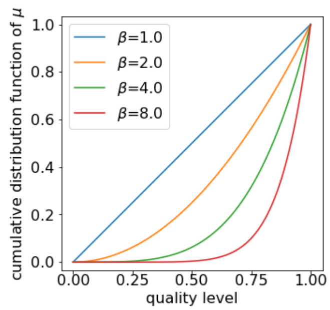


Figure 8.3: Cumulative distribution function (cdf) of the symmetric equilibrium μ for 1-dimensional setup (Example 3) with $P = 2$ producers. The equilibrium μ interpolates from a uniform distribution to a point mass as the exponent β increases.

distribution for lower values of β (see Figure 8.3). The intuition is that increasing β lowers production costs for content with a given quality, so producers must produce higher quality content at equilibrium. Similarly, for any fixed value of β, the equilibrium distribution for lower values of P stochastically dominates the equilibrium distribution for higher values of P. This is because when more producers enter the market, any given producer is less likely to win users (i.e. a producer only wins a user with probability $1/P$ if all producers choose the same vector), so they cannot expend as high of a production cost.

We next translate these insights about the equilibria for one-dimensional marketplaces to higher-dimensional marketplaces with a population of *homogeneous* users. If all users are embedded at the same vector $u \in \mathbb{R}^D_{\geq 0}$, then the producer's decision about what direction of content to choose is trivial: they would choose a direction in $\arg\max_{\|p\|=1}\langle p, u\rangle$. As a result, the producer's decision again boils down to a one-dimensional decision: choosing the quality $\|p\|$ of the content.

Corollary 37. *Suppose that there is a single population of N users, all of whose embeddings are at the same vector u. Then, there is a symmetric mixed Nash equilibrium μ supported on $\left\{qp^* \mid q \in [0, N^{\frac{1}{\beta}}]\right\}$ where $p^* \in \arg\max_{\|p\|=1}\langle p, u\rangle$. The cumulative distribution function of $q = \|p\| \sim \mu$ is $F(q) = (q/N)^{\beta/(P-1)}$.*

Corollary 37 relies on the fact that when users are homogeneous, there is no tension between catering to one user and catering to other users.

8.2.3 Specialization and the formation of genres

In contrast, when users are heterogeneous, there are inherent tensions between catering to one user and catering to other users. As a result, the producer make nontrivial choices not only about the quality of the content (Chapter 8.2.2), but also the *genre* of content as reflected by its direction in $\mathbb{R}^d$. This can lead to *specialization*, which is when different producers create goods tailored to different users; alternatively, all producers might still produce the same genre of content at equilibrium and thus only exhibit differentiation on the axis of quality.

To formalize specialization, we need to disentangle two forms of differentiation: (1) differentiation along direction (genre), and (2) differentiation along magnitude (quality). We define specialization as differentiation along genres, and not as differentiation along quality. To focus on the former, we define *genres* as the set of *directions* that arise at a symmetric mixed Nash equilibrium μ:

$$\text{Genre}(\mu) := \left\{ \frac{p}{\|p\|} \mid p \in \text{supp}(\mu) \right\}, \tag{8.2}$$

where we normalize by $\|p\|$ to separate out the quality (norm) from the genre (direction). The set of genres $\text{Genre}(\mu)$ captures the set of content that may arise on the platform in some realization of randomness of the producers' strategies. When an equilibrium has a single genre, all producers cater to an average user, and only a single type of content appears on the platform. On the other hand, when an equilibrium has multiple genres, many types of digital content are likely to appear on the platform.

We thus say that *specialization* occurs at an equilibrium μ if and only if μ has more than one *genre* (i.e., if and only if $|\text{Genre}(\mu)| > 1$). When specialization does occur at μ, the form of specialization is further captured by the number of genres $|\text{Genre}(\mu)|$ and other properties of the set of genres $\text{Genre}(\mu)$. Note that we define specialization in terms of the *support* of a symmetric mixed equilibrium distribution. In this definition, we implicitly interpret the randomness in the producer strategies as differentiation between producers; this formalization of specialization obviates the need to reason about asymmetric equilibria, thus making the model much more tractable to analyze.

8.2.4 Model discussion

Our model is one of the simplest possible that studies specialization in the supply-side marketplace. In particular, although many classical models[4] (e.g. spatial location models with specific user distributions and costs based on the Euclidean distance) permit closed-form equilibria, they elide important aspects of supply-side markets—such as the multi-dimensionality of producer decisions, the joint selection of genre and quality, and the structure of producer costs—which significantly influence the form that specialization takes. Our model incorporates these aspects at the cost of not having general closed-form equilibria; we nonetheless develop technical tools to study specialization without relying on closed-form solutions (while also obtaining closed forms in special cases). On the other side of the spectrum, we do not aim to provide a fully general model of product selection, production, and pricing. Instead, our model adds assumptions specific to recommender systems that provide sufficient structure to derive precise properties of specialization.

Our formalization of user preferences and the producer decision space is motivated by distinguishing aspects of content recommender systems. First, the infinite, high-dimensional

[4]See Anderson et al. (1992) for a textbook treatment.

content embedding space captures that digital goods can't be cleanly clustered into categories, but rather, are often mixtures of different dimensions (e.g. a movie can be both a drama and a comedy). Furthermore, the bilinear (dot product) form of inferred user values is motivated by standard recommendation algorithms: for example, matrix factorization assumes that the inferred user values are inner products between user vectors and content attributes vectors (Koren et al., 2009). We explicitly connect our model to matrix factorization in our empirical analysis in Chapter 8.3.4.

Our assumptions on the structure of producer costs allow us to study specialization, while retaining mathematical tractability. The family of producer cost functions is stylized, but flexible, in that it accommodates arbitrary powers of arbitrary norms and it can capture both specialization and homogenization (Theorem 38). The assumption that all producers share the same cost function is also simplifying, but, potentially surprisingly, still allows us to study specialization. In particular, specialization occurs in a rich class of marketplaces (Corollary 43), *despite* the fact that producers have symmetric utility functions; we anticipate that the tendency towards specialization would only be amplified if producers could have different cost functions.

We hope that the simplicity of our model, and its ability to capture specialization, make it a useful starting point to further study the impact of recommender systems on production; we highlight some potential directions in Chapter 8.6.

8.3 When does specialization occur?

In order to investigate whether specialization occurs in a given marketplace, we investigate when the set of genres $\text{Genre}(\mu)$ of an equilibrium μ contains more than one direction. We distinguish between two regimes of marketplaces based on whether or not a single-genre equilibrium exists:

1. A marketplace is in the *single-genre regime* if there exists an equilibrium μ such that $|\text{Genre}(\mu)| = 1$. All producers thus create content of the same genre.

2. A marketplace is in the *multi-genre regime* if all equilibria μ satisfy $|\text{Genre}(\mu)| > 1$. Producers thus necessarily differentiate in the genre of content that they produce.

To understand these regimes, we ask: *what conditions on the user vectors $u_1, \ldots, u_N$ and the cost function parameters $\|\cdot\|$ and β determine which regime the marketplace is in?*

In Chapter 8.3.1, we give necessary and sufficient conditions for all equilibria to have multiple genres (Theorem 38). In Chapter 8.3.2, we show several corollaries of Theorem 38. In Chapter 8.3.3, we show that the location of the single-genre equilibrium (in cases where it exists) maximizes the Nash social welfare. In Chapter 8.3.4, we provide an empirical analysis using the MovieLens-100K dataset (Harper and Konstan, 2015) that provides additional intuition for our theoretical results.

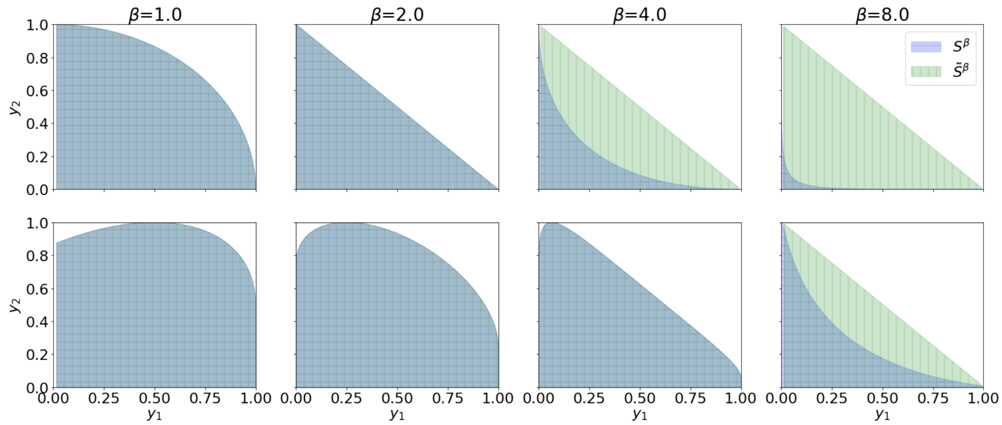


Figure 8.4: The sets $\mathcal{S}^\beta$ and $\bar{\mathcal{S}}^\beta$ for two configurations of user vectors (rows) and settings of β (columns). The user vectors are $[1, 0], [0, 1]$ (top, same as Figure 8.1) and $[1, 0], [0.5, 0.87]$ (bottom). $\mathcal{S}^\beta$ transitions from convex to non-convex as β increases, though the transition point depends on the user vectors. When $\mathcal{S}^\beta$ is convex, a single vector p can more easily satisfy both users at low cost.

8.3.1 Characterization of single-genre and multi-genre regimes

We first provide a tight geometric characterization of when a marketplace is in the single-genre regime versus in the multi-genre regime. More formally, let $\mathbf{U} = [u_1; \cdots ; u_N]$ be the $N \times D$ matrix of user vectors, and let $\mathcal{S}$ denote the image of the unit ball under $\mathbf{U}$:

$$\mathcal{S} := \left\{ \mathbf{U}p \mid \|p\| \leq 1, p \in \mathbb{R}^D_{\geq 0} \right\} \tag{8.3}$$

Each element of $\mathcal{S}$ is an N-dimensional vector, which represents the inferred user values for some unit-norm producer p. Additionally, let $\mathcal{S}^\beta$ be the image of $\mathcal{S}$ under coordinate-wise powers, i.e. if $(z_1, \ldots, z_N) \in \mathcal{S}$ then $(z_1^\beta, \ldots, z_N^\beta) \in \mathcal{S}^\beta$. We show that genres emerge when $\mathcal{S}^\beta$ is sufficiently different from its convex hull $\bar{\mathcal{S}}^\beta$:

Theorem 38. *Let* $\mathbf{U} := [u_1; \cdots ; u_N]$, *let* $\mathcal{S}$ *be* $\left\{ \mathbf{U}p \mid \|p\| \leq 1, p \in \mathbb{R}^D_{\geq 0} \right\}$, *and let* $\mathcal{S}^\beta$ *be the image of* $\mathcal{S}$ *under coordinate-wise powers. Then, there is a symmetric equilibrium* μ *with* $|\mathrm{Genre}(\mu)| = 1$ *if and only if*

$$\max_{y \in \mathcal{S}^\beta} \prod_{i=1}^{N} y_i = \max_{y \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^{N} y_i. \tag{8.4}$$

Otherwise, all symmetric equilibria have multiple genres. Moreover, if (8.4) *holds for some* β*, it also holds for every* $\beta' \leq \beta$.

Theorem 38 relates the existence of a single-genre equilibrium to the convexity of the set $\mathcal{S}^\beta$. As a special case, the condition in Theorem 38 always holds if $\mathcal{S}^\beta$ is convex, but is strictly speaking weaker than convexity. Interestingly, the condition depends on the geometry of the user embeddings and the cost function but *not* on the number of producers. Intuitively, convexity of $\mathcal{S}^\beta$ relates to the ease with which a vector p can satisfy all users simultaneously, at low cost—each dimension of $\mathcal{S}$ corresponds to a user's utility. In Figure 8.4, we display the sets $\mathcal{S}^\beta$ and $\bar{\mathcal{S}}^\beta$ for different configurations of user vectors and different settings of β.

Theorem 38 further shows that the boundary between the single-genre and multi-genre regimes can be represented by a *threshold* defined as follows

$$\beta^* := \sup \left\{ \beta \geq 1 \mid \max_{y \in \mathcal{S}^\beta} \prod_{i=1}^N y_i = \max_{y \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^N y_i \right\}.$$

where single-genre equilibria exist exactly when $\beta \leq \beta^*$. Conceptually, larger β make producer costs more superlinear, which eventually discentivizes producers from attempting to perform well on all dimensions at once.

Proof techniques for Theorem 38. Since the single-genre equilibrium does not admit a straightforward closed-form solution, we must implicitly reason about its existence when proving Theorem 38. To do so, we draw a connection to minimax theory in optimization. Our main lemma shows that the existence of a single-genre equilibrium is equivalent to strong duality holding for the following minmax problem:

Lemma 39 (Informal)**.** *There exists a symmetric equilibrium μ with $|\mathrm{Genre}(\mu)| = 1$ if and only if:*

$$\inf_{y \in \mathcal{S}^\beta} \left(\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{y_i} \right) = \sup_{y' \in \mathcal{S}^\beta} \left(\inf_{y \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{y_i} \right). \tag{8.5}$$

To prove Theorem 38 from Lemma 39, we analyze when strong duality holds. Note that while the objective in (8.5) is convex in y and linear (concave) in y', the constraints on y and y' through the set $\mathcal{S}^\beta$ can be non-convex. It turns out that we can eliminate the non-convexity in the constraint on y for free, by reparameterizing to the space of content vectors $p \in \mathbb{R}^D_{\geq 0}$ with unit norm. On the other hand, to handle the non-convexity in the constraint on y', we need to convexify the optimization program by replacing $\mathcal{S}^\beta$ with its convex hull $\bar{\mathcal{S}}^\beta$. By Sion's min-max theorem, we can flip sup and inf in this convexified version of the left-hand side of (8.5). The remaining technical step is to relate the resulting expression to the right-hand side of (8.5), which we defer to Appendix E.3.1.

To prove Lemma 39, we first characterize the cumulative distribution function of quality at a single-genre equilibria as $F(q) \propto q^\beta$ (Lemma 185). Then we show that y corresponds to an equilibrium direction if and only if $\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{y_i} \leq N$, which means that there exists an equilibrium direction if and only if the left-hand side of (8.5) is at most N. We also show that the dual the right-hand side of (8.5) is always equal to N, which allows us to prove Lemma 39.

8.3.2 Corollaries of Theorem 38

To further understand the condition in equation (8.4), we consider a series of special cases that provide intuition for when single-genre equilibria exist (proofs deferred to Chapter E.3.2). First, let us consider $\beta = 1$, in which case the cost function is a norm. Then $\mathcal{S}^1 = \mathcal{S}$ is convex, so a single-genre equilibrium always exists.

Corollary 40. *The threshold β^* is always at least 1. That is, if $\beta = 1$, there exists a single-genre equilibrium.*

The economic intuition behind Corollary 40 is that norms incentivize averaging rather than specialization.

We next take a closer look at how the choice of norm affects the emergence of genres. For cost functions $c(p) = \|p\|_q^\beta$, we show that $\beta^* \geq q$ for any set of user vectors, with equality achieved at the standard basis vectors.

Corollary 41. *Let the cost function be $c(p) = \|p\|_q^\beta$. For any set of user vectors, it holds that $\beta^* \geq q$. If the user vectors are equal to the standard basis vectors $\{e_1, \ldots, e_D\}$, then β^* is equal to q.*

Corollary 41 illustrates that the threshold β^* relates closely to the convexity of the cost function and whether the cost function is superlinear. In particular, the cost function must be sufficiently nonconvex for all equilibria to be multi-genre. For example, for the ℓ_∞-norm, where producers only pay for the highest magnitude coordinate, it is never possible to incentivize specialization: there exists a single-genre equilibrium regardless of β. On the other hand, for norms where costs aggregate nontrivially across dimensions, specialization is possible.

In addition to the choice of norm, the geometry of the user vectors also influences whether multiple genres emerge. To illustrate this, we first show that in a concrete market instance with 2 equally sized populations of users, the threshold depends on the cosine similarity between the two user vectors:

Corollary 42. *Suppose that there are N users split equally between two linearly independently vectors $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$, and let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2 \rangle}{\|u_1\|_2 \|u_2\|}\right)$. Let the cost function be $c(p) = \|p\|_2^\beta$. Then it holds that:*

$$\beta^* = \frac{2}{1 - \cos(\theta^*)}.$$

Corollary 42 demonstrates the threshold β^* increases as the angle θ^* between the users decreases (i.e. as the users become closer together), because it is easier to simultaneously cater to all users. In particular, β^* interpolates from 2 when the users are orthogonal to ∞ when the users point in the same direction.

Finally, we consider general configurations of users and cost functions, and we upper bound β^*:

Corollary 43. *Let $\|\cdot\|_*$ denote the dual norm of $\|\cdot\|$, defined to be $\|p\|_* = \max_{\|p\|=1, p\in\mathbb{R}^D_{\geq 0}} \langle q, p\rangle$. Let $Z := \| \sum_{n=1}^N \frac{u_n}{\|u_n\|_*} \|_*$. Then,*

$$\beta^* \leq \frac{\log(N)}{\log(N) - \log(Z)}. \tag{8.6}$$

In equation (8.6), the upper bound on the threshold β^* increases as Z increases. As an example, consider the cost function $c(p) = \|p\|_2^\beta$. We see that if the user vectors point in the same direction, then $Z = N$ and the right-hand side of (8.6) is ∞. On the other hand, if $u_1, \ldots, u_n$ are orthogonal, then $Z = \sqrt{N}$ and the right-side of (8.6) is 2, which exactly matches the bound in Corollary 41. In fact, for *random* vectors $u_1, \ldots, u_N$ drawn from a truncated gaussian distribution, we see that $Z = \tilde{O}(\sqrt{N})$ in expectation, in which case the right-hand side of (8.6) is close to 2 as long as N is large. Thus, for many (but not all) choices of user vectors, even small values of β are enough to induce multiple genres. In Chapter 8.3.4, we compute the right-hand side of (8.6) on user embeddings generated from the MovieLens dataset for different cost functions.

8.3.3 Location of single-genre equilibrium

We next study where the single-genre equilibrium is located, in cases where it exists. As a consequence of the proof of Theorem 38, we can show that the location of the single-genre equilibrium maximizes the *Nash social welfare* Nash et al. (1950) of the users.

Corollary 44. *If there exists μ with $|\mathrm{Genre}(\mu)| = 1$, then the corresponding producer direction maximizes Nash social welfare of the users:*

$$\mathrm{Genre}(\mu) = \operatorname*{arg\,max}_{\|p\|=1 | p\in\mathbb{R}^D_{\geq 0}} \sum_{i=1}^N \log(\langle p, u_i\rangle). \tag{8.7}$$

Corollary 44 demonstrates that the single-genre equilibrium directions maximizes the Nash social welfare Nash et al. (1950) for users. Interestingly, this measure of welfare for *users* is implicitly maximized by *producers* competing with each other in the marketplace. Properties of the Nash social welfare are thus inherited by single-genre equilibria. In particular, since the Nash social welfare corresponds to the logarithm of the geometric mean of the inferred user values, the Nash social welfare strikes a compromise between fairness (balancing inferred user values of different users) and efficiency (the sum of the inferred user values achieved across all users)—this means that the single-genre equilibria exhibit the same tradeoff between fairness and efficiency.

We note that this welfare result relies on the assumption that all producers choose the same direction of content. In particular, at multi-genre equilibria, the Nash social welfare could be even higher due to specialization leading to personalization. On the other hand, the reduced amount of competition at multi-genre equilibria may end up lowering the quality of goods. We defer an in-depth analysis of the welfare implications of supply-side competition to future work.

8.3.4 Empirical analysis on the MovieLens dataset

We provide an empirical analysis of supply-side equilibria using the MovieLens-100K dataset and recommendations based on nonnegative matrix factorization (NMF). In particular, we compute the single-genre equilibrium direction for different cost functions as well as estimates of β^* (i.e., the threshold where specialization starts to occur) for different values of the dimension D. These experiments provide qualitative insights that offer additional intuition for our theoretical results.

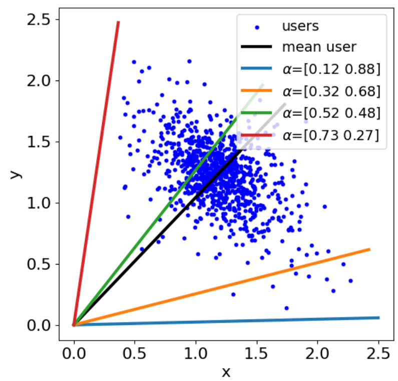
We focus on the rich family of cost functions $c_{q,\alpha,\beta}(p) = \| [p_1 \cdot \alpha_1, \ldots, p_D \cdot \alpha_D] \|_q^\beta$ parameterized by weights $\alpha \in \mathbb{R}_{\geq 0}^D$, parameter $q \geq 1$, and cost function exponent $\beta \geq 1$. The weights $\alpha \in \mathbb{R}_{\geq 0}^D$ capture asymmetries in the costs of different dimensions (a higher value of α_i means that dimension i is more costly). The parameters q and β together capture the tradeoffs between improving along a single dimension versus simultaneously improving among many dimensions. To isolate the impact of each parameter, we either fix $q = 2$ and vary α (and β) or we fix $\alpha = [1, 1, \ldots, 1]$ and vary q (and β).

Setup. The MovieLens 100K dataset consists of 943 users, 1682 movies, and 100,000 ratings (Harper and Konstan, 2015). For $D \in \{2, 3, 4, 5, 6, 8, 10, 20, 40\}$, we obtain D-dimensional user embeddings by running NMF (with D factors) using the scikit-surprise library. We calculate the single-genre equilibrium genre $p^* = \arg\max_{\|p\|=1 | p \in \mathbb{R}_{\geq 0}^D} \sum_{i=1}^N \log(\langle p, u_i \rangle)$ (Corollary 44) by solving the optimization program, using the cvxpy library for $q = 2$ and projected gradient descent for $q \neq 2$. We calculate the upper bound $\beta_u := \frac{\log(N)}{\log(N) - \log\left(\| \sum_{n=1}^N \frac{u_n}{\|u_n\|_*} \|_*\right)} \geq \beta^*$ from Corollary 43. We calculate another estimate β_e by binary searching and estimating whether (8.4) holds at each candidate value β as follows: we estimate $\max_{y \in \mathcal{S}^\beta} \prod_{i=1}^N y_i$ using the cvxpy library and we estimate $\max_{y \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^N y_i$ by taking $\bar{S}^\beta$ to be the convex hull of randomly drawn points. For computational reasons, when computing the estimate β_e, we consider a restricted dataset consisting of $N \in \{20, 30, 40\}$ randomly chosen users and focus on $q = 2$. See Appendix E.1 for details of the empirical setup.[5]

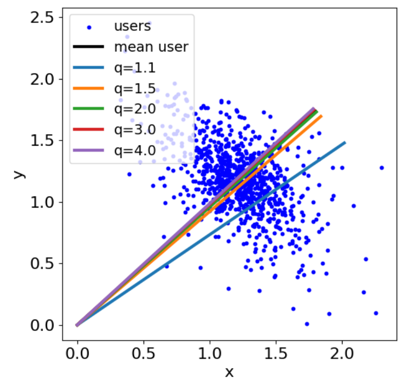
Single-genre equilibrium direction p^*. Figures 8.5a, 8.5b, 8.5d, and 8.5e show the direction of the single-genre equilibrium p^* across different cost functions. These plots uncover several properties of the genre p^*. First, the genre generally does not coincide with the arithmetic mean of the users. Moreover, the genre varies significantly with the weights α. In particular, the magnitude of the dimension p_i is higher if α_i is *lower*, which aligns with the intuition that producers invest more in cheaper dimensions. In contrast, the genre turns out to not change significantly with the norm parameter. Altogether, these insights illustrate that how the genre can be influenced by specific aspects of producer costs.

Threshold β^* where specialization starts to occur. Figures 8.5c and 8.5f show the value of β_e and β_u across different values of D, q, and N. As the dimension D increases, the estimate β_e and the upper bound β_u both generally decrease, indicating that specialization is more likely to occur. The intuition is that D amplifies the heterogeneity of user embeddings, subsequently

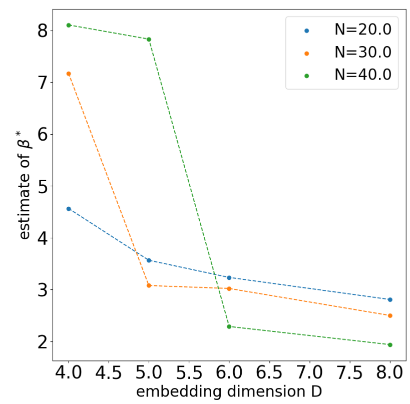
[5] The code is available at `https://github.com/mjagadeesan/supply-side-equilibria` .



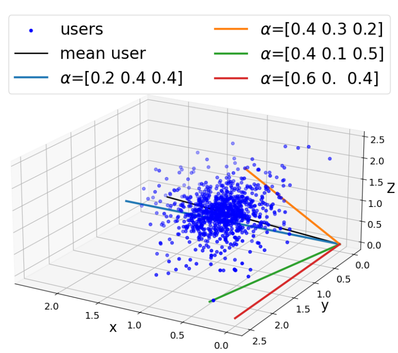
(a) p^* for $D = 2$ and $q = 2$



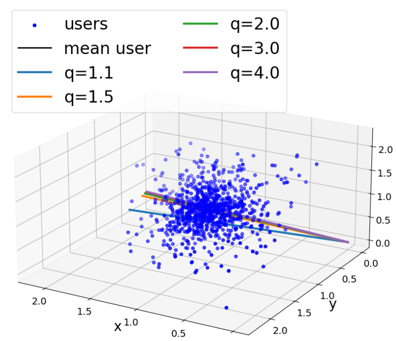
(b) p^* for $D = 2$ and $\alpha = [1, 1]$



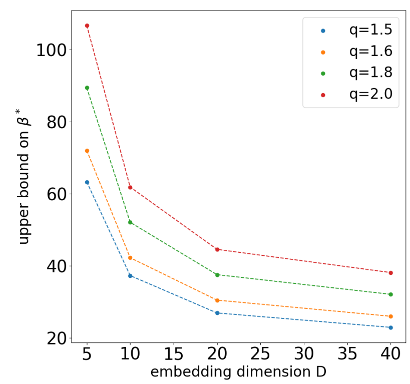
(c) β_e for $\alpha = [1, \ldots, 1]$ and $q = 2$



(d) p^* for $D = 3$ and $q = 2$



(e) p^* for $D = 3$ and $\alpha = [1, 1, 1]$



(f) β_u for $\alpha = [1, \ldots, 1]$

Figure 8.5: Empirical analysis of supply-side equilibria on the MovieLens-100K dataset. *Plots (a), (b), (d), (e):* Single-genre equilibrium direction p^* (computed using Corollary 44) for different cost function weights $\alpha \in \mathbb{R}^D_{\geq 0}$ and parameters $q \geq 1$ as well as for different dimensions $D \geq 1$. Interestingly, the single-genre equilibrium direction is generally not aligned with the arithmetic mean and places a higher weight on cheaper dimensions. *Plots (c) and (f):* Estimates β_e and β_u of the threshold β^* where specialization starts to occur for different values of norm parameter q and the number of users N. Observe that higher values of D make specialization more likely to occur.

increasing the likelihood of specialization. This insight has an interesting consequence for platform design: *the platform can influence the level of specialization by tuning the number of factors D used in matrix factorization.* Producer costs also impact whether specialization occurs: as the norm q increases, the value of β_u increases and specialization is less likely to occur.

8.4 Equilibrium structure for two equally sized populations of users

We next investigate the form of specialization exhibited by multi-genre equilibria, focusing on the case of two equally sized populations and producer cost functions given by powers of the ℓ_2 norm. More formally, there are N users split equally between two linearly independently vectors $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$, and the the cost function is $c(p) = \|p\|_2^\beta$. We establish *structural properties* of the equilibria (see Chapter 8.4.1). We next concretely compute the equilibria μ in several special instances that permit *closed-form solutions* (see Chapter 8.4.2-8.4.3). We then provide an overview of proof techniques, which involves developing machinery to characterize these equilibria (see Chapter 8.4.4).

8.4.1 Structural properties of equilibria

We first establish properties about the *support* of the equilibrium distributions μ. First, we show that the support of cannot contain an ε-ball for any ε and is thus 1-dimensional.

Proposition 45. *Suppose that there are N users split equally between two linearly independently vectors $u_1, u_2 \in \mathbb{R}^2_{\geq 0}$, and let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2\rangle}{\|u_1\|_2\|u_2\|}\right)$ be the angle between the user vectors. Let the cost function be $c(p) = \|p\|_2^\beta$, and let $P \geq 2$. Let μ be a symmetric Nash equilibrium such that the distributions $\langle u_1, p\rangle$ and $\langle u_2, p\rangle$ over $\mathbb{R}_{\geq 0}$ are absolutely continuous. As long as $\beta \neq 2$ or $\theta^* \neq \pi/2$, the support of μ does not contain an ℓ_2-ball of radius ε for any $\varepsilon > 0$.*[6]

Proposition 45 demonstrates that the support of μ must be a union of 1-dimensional curves. In the single-genre regime, the support is always a line segment through the origin. In the multi-genre regime, however, the support can be curves with different shapes (see Figure 8.6 for specific examples). We will later characterize where these curves are increasing or decreasing in terms of the location of the curve, the angle $\theta^* = \cos^{-1}\left(\frac{\langle u_1, u_2\rangle}{\|u_1\|\|u_2\|}\right)$, and the cost function parameter β (Lemma 195).

We next show that all equilibria must have either one or infinitely many genres, dictated by whether β is above or below the critical value β^* (see Figure 8.1):

Theorem 46. *Suppose that there are N users split equally between two linearly independently vectors $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$, and let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2\rangle}{\|u_1\|_2\|u_2\|}\right)$ be the angle between the user vectors. Let the cost function be $c(p) = \|p\|_2^\beta$. Let μ be a a distribution on $\mathbb{R}^d$ such that the distributions $\langle u_1, p\rangle$ and $\langle u_2, p\rangle$ over $\mathbb{R}_{\geq 0}$ over $\mathbb{R}_{\geq 0}$ for $p \sim \mu$ are absolutely continuous and twice continuously differentiable within their supports. There are two regimes based on β and θ^*:*

[6] The case of $\beta = 2$ and $\theta^* = \pi/2$ is degenerate and permits a range of possible equilibria.

1. *If* $\beta < \beta^* = \frac{2}{1-\cos(\theta^*)}$ *and if* μ *is a symmetric mixed equilibrium, then* μ *satisfies* $|\text{Genre}(\mu)| = 1$.

2. *If* $\beta > \beta^* = \frac{2}{1-\cos(\theta^*)}$, *if* $|\text{Genre}(\mu)| < \infty$, *and if the conditional distribution of* $\|p\|$ *along each genre is continuously differentiable, then* μ *is not an equilibrium.*

Theorem 46 provides a tight characterization of when specialization occurs in a marketplace: specialization occurs *if and only if* β is above β^* (subject to some mild continuity conditions). The threshold β^* can thus be interpreted as a *phase transition* at which the equilibrium transitions from single-genre to infinitely many genres (see Figure 8.1). More specifically, the first part of Theorem 46 strengthens Theorem 38 to show that *all* equilibria are single-genre when $\beta < \beta^*$, which means that producers are *never* incentivized to specialize in this regime. The equality condition $\beta = \beta^*$ captures the transition point where both single-genre and multi-genre equilibria can exist.

In the multi-genre regime where $\beta \geq \beta^*$, Theorem 46 shows that producers do not fully personalize content to either of the two users u_1 and u_2, or even choose between finitely many types of content. Rather, producers choose infinitely many types of content that balance the preferences of the two populations in different ways. The lack of coordination between producers—as captured by a symmetric mixed Nash equilibrium—is what drives this result. Producers do not know exactly what content other producers will create in a given realization of the randomness, which results in a diversity of content on the platform.

8.4.2 Closed-form equilibria for the standard basis vectors

We next compute the equilibria in the special case of user vectors located at the standard basis vectors, and we analyze the form of specialization that the equilibria exhibit. For ease of notation, for the remainder of the section, we assume these populations each consist of a *single* user (these results can be easily adapted to the case of $N/2$ users in each population).

Interestingly, all of these multi-genre equilibria exhibit the following *relaxation of pure horizontal differentiation*: producers can differentiate along genre, but the genre of content fully specifies the content's quality. More specifically, for any genre $p^* \in \text{Genre}(\mu)$, the set $\text{Genre}(\mu) \cap \left\{q \cdot p^* \mid q \in \mathbb{R}^{\geq 0}\right\}$ contains exactly one single element.[7] This stands in contrast to single-genre equilibria, which by definition exhibit pure vertical differentiation.[8]

We first explicitly compute the equilibria in the case of $P = 2$ producers (see Figure 8.1).

Proposition 47. *Suppose that there are 2 users located at the standard basis vectors* $e_1, e_2 \in \mathbb{R}^2$, *and the cost function is* $c(p) = \|p\|_2^\beta$. *For* $P = 2$ *and* $\beta \geq \beta^* = 2$, *there is an equilibrium* μ *supported on the quarter-circle of radius* $(2\beta^{-1})^{1/\beta}$, *where the angle* $\theta \in [0, \pi/2]$ *has density* $f(\theta) = 2\cos(\theta)\sin(\theta)$.

[7] Pure horizontal differentiation is not satisfied, since content in different genres may not have the same quality (see Figure 8.6).

[8] Pure vertical differentiation is when producers only differentiate along quality, not along direction.

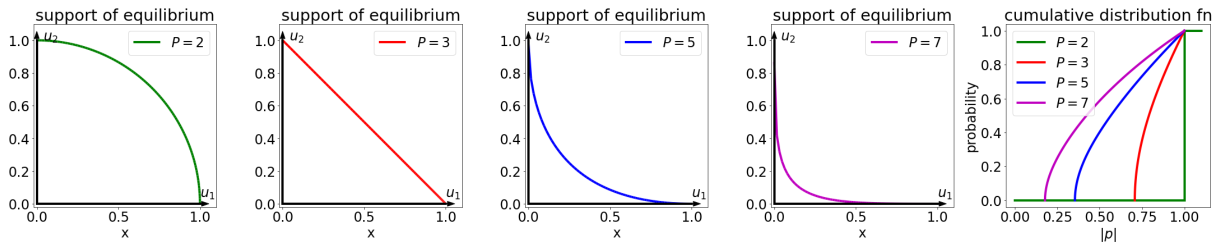


Figure 8.6: A symmetric equilibrium for different number of producers P, for 2 users located at the standard basis vectors e_1 and e_2, for producer cost function $c(p) = \|p\|_2^\beta$ with $\beta = 2$ (see Proposition 48). The first 4 plots show the support of an equilibrium μ. As P increases, the support goes from concave, to a line segment, to convex. The last plot shows the cumulative distribution function of $\|p\|$ for $p \sim \mu$. The distribution for lower P stochastically dominates the distribution for higher values of P. All of these equilibria either exhibit pure vertical differentiation or a relaxed form of horizontal differentiation where the genre fully specifies the content's quality (but not pure horizontal differentiation, which would require that quality is constant across genres).

Proposition 47 demonstrates the support of the equilibrium distribution is a quarter circle with radius $(2\beta^{-1})^{1/\beta}$. This equilibrium exhibits pure horizontal differentation (as well as the relaxation of pure horizontal differentiation that we described above). Since all (x, y) in the support have the same radius, producers always expend the same cost regardless of the realization of randomness in their strategy. Since $c(p) = \|p\|_2^\beta$, producers pay a cost of $2\beta^{-1}$. The cost of production therefore goes to 0 as $\beta \to \infty$. This enables producers achieving *positive profit* at equilibrium (see Corollary 51) as we describe in more detail in Chapter 8.5.

We next vary the number of producers P while fixing $\beta = 2$ (see Figure 8.6).

Proposition 48. *Suppose that there are 2 users located at the standard basis vectors $e_1, e_2 \in \mathbb{R}^2$, with cost function $c(p) = \|p\|_2^\beta$. For $\beta = 2$, there is a multi-genre equilibrium μ with support equal to*

$$\left\{\left(x, (1 - x^{\frac{2}{P-1}})^{\frac{P-1}{2}}\right) \mid x \in [0, 1]\right\}, \tag{8.8}$$

and where the distribution of x has cdf equal to $\min(1, x^{2/(P-1)})$.

Proposition 48 demonstrates that for different values of P, the support of the equilibrium μ follows different curves connecting $[1, 0]$ and $[0, 1]$. Note that these equilibria exhibit the relaxation of pure horizontal differentiation that we described earlier. Moreover, the curve is concave for $P = 2$, a line segment for $P = 3$, and convex for all $P \geq 4$. Indeed, as P increases, the support converges to the union of the two coordinate axes.

8.4.3 Closed-form equilibria in an infinite-producer limit

Motivated by the support collapsing onto the standard basis vectors for $P \to \infty$ in Proposition 48, we investigate equilibria in a "limiting marketplace" where $P \to \infty$. In the infinite-producer limit, we show that a *two-genre* equilibrium exists, regardless of the geometry of the 2 user vectors, and we characterize the equilibrium distribution μ (see Figure 8.2). Interestingly, these equilibria do not exhibit pure vertical differentiation or (the relaxation of) pure horizontal differentiation.

Formalizing the infinite-producer limit is subtle: the distribution of any single producer approaches a point mass at 0, but the distribution of the *winning* producer turns out to be non-degenerate. To get intuition for this, let's revisit the one-dimensional setup of Example 3. The cumulative distribution function $F(p) = (p/N)^{\beta/(P-1)}$ of a single producer as $P \to \infty$ approaches $F(p) = 1$ for any $p > 0$—this corresponds to a point mass at 0.[9] On the other hand, the cumulative distribution function of the *winning* producer $F^{\max}(p) = (p/N)^{\beta P/(P-1)}$ approaches $(p/N)^{\beta}$, which is a well-defined function.

When we formalize the infinite-producer limit for $N \geq 1$ users, we leverage the intuition that the distribution function of the winning producer is non-degenerate. In particular, we specify infinite-producer equilibria in terms of three properties—the *genres*, the *conditional quality distributions for each genre* (i.e. the distribution of the maximum quality $\|p\|$ along a genre, conditional on all of the producers choosing that genre), and the *weights* (i.e. the probability that a producer chooses each genre). We defer a formal treatment to Definition 20 in Chapter E.4.5.

In the infinite-producer limit, we show the following 2-genre distribution is an equilibrium. For ease of notation, we again assume these populations each consist of a *single* user (these results can be easily adapted to the case of $N/2$ users in each population).

Theorem 49. *[Informal version of Theorem 198] Suppose that there are 2 users located at two linearly independently vectors $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$, let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2\rangle}{\|u_1\|_2 \|u_2\|}\right)$ be the angle between them. Suppose we have cost function $c(p) = \|p\|_2^{\beta}$, $\beta > \beta^* = \frac{2}{1-\cos(\theta^*)}$, and $P = \infty$ producers. Then, there exists an equilibrium with two genres:*

$$\left\{[\cos(\theta^G + \theta_{min}), \sin(\theta^G + \theta_{min})], [\cos(\theta^* - \theta^G + \theta_{min}), \sin(\theta^* - \theta^G + \theta_{min})]\right\}$$

where $\theta^G := \arg\max_{\theta \leq \theta^*/2} \left(\cos^{\beta}(\theta) + \cos^{\beta}(\theta^* - \theta)\right)$ and where
$\theta_{min} := \min\left(\cos^{-1}\left(\frac{\langle u_1, e_1\rangle}{\|u_1\|}\right), \cos^{-1}\left(\frac{\langle u_2, e_1\rangle}{\|u_2\|}\right)\right)$.

For each genre, the conditional quality distribution (i.e. the distribution of the maximum quality $\|p\|$ along a genre, conditional on all of the producers choosing that genre) has cdf given by a countably-infinite piecewise function, where each piece is either constant or grows proportionally to $\|p\|^{2\beta}$.

[9] The intuition is that the expected number of users that the producer wins at a symmetric equilibrium is N/P, which approaches 0 in the limit; thus, the production cost that a producer can afford to expend must approach 0 in the limit.

Theorem 49 reveals that finite-genre equilibria (that have more than one genre) re-emerge in the limit as $P \to \infty$, although they do not exist for any finite P (see Figure 8.2). For users located at the standard basis vectors, Theorem 49 formalizes the intuition from Proposition 48 that the equilibrium converges to a distribution supported on the standard basis vectors. This means that at $P = \infty$, producers either entirely personalize their content to the first user or entirely personalize their content to the second user, but do not try to appeal to both users at the same time.

Interestingly, the set of genres is *not* equal to the set of two users unless users are orthogonal. As shown in Figure 8.2, the two-genres are located within the interior of the convex cone formed by the two users. This means that producers always attempt to cater their content to both users at the same time, although they either place a greater weight on one user or the other user, depending on which genre they choose. The location of these two genres changes for different values of β. When β approaches the single-genre threshold, the genres both collapse onto the single-genre direction $\theta^*/2$. On the other hand, when β approaches ∞, the genres converge to the two users.

Finally, the support of the equilibrium distribution consists of countably infinite disjoint line segments with interesting economic interpretations. First, observe that the cdf of the conditional quality distributions of each genre (see the last panel of Figure 8.2) has gaps in its support: it is a countably-infinite piecewise function, where each piece is either constant or grows proportionally to $q^{2\beta}$. The level of "bumpiness" of the cdf decreases as θ^* increases: for the limiting case of $\theta = \pi/2$, it converges to the smooth function $F^{\max}(q) = q^{2\beta}$. Moreover, the regions of zero density of each of the two genres are actually staggered, so that at most one of the genres can achieves a given utility for a given user. In particular, for each user u_i, it never holds that $\langle u_i, p\rangle = \langle u_i, p'\rangle$ for $p \neq p'$: that is, the utility level fully specifies the genre of the content. The closed-form expression of the density (see Theorem 198) formally establishes these properties.

8.4.4 Overview of proof techniques

To prove our results in this section, our first step is establish a useful characterization of equilibria that enables us to separately account for the geometry of the users and the number of producers. This takes the form of necessary and sufficient conditions that decouple in terms of two quantities: a set of *marginal distributions* H_i, and the *support* $S \subseteq \mathbb{R}^N_{\geq 0}$.

Lemma 50. *Let $\mathbf{U} = [u_1; u_2; \ldots; u_N]$ be the $N \times D$ matrix of users vectors. Given a set $S \subseteq \mathbb{R}^N_{\geq 0}$ and distributions $H_1, \ldots, H_N$ over $\mathbb{R}_{\geq 0}$, suppose that the following conditions hold:*

(C1) Every $z^* \in S$ is a maximizer of the equation:

$$\max_{z \in \mathbb{R}^D_{\geq 0}} \sum_{i=1}^{N} H_i(z_i) - c_{\mathbf{U}}(z), \tag{8.9}$$

where $c_{\mathbf{U}}(z) := \min\left\{c(p) \mid p \in \mathbb{R}^D_{\geq 0}, \mathbf{U}p = z\right\}$.

(C2) There exists a random variable Z with support S, such that the marginal distribution Z_i has cdf equal to $H_i(z)^{1/(P-1)}$.

(C3) Z is distributed as $\mathbf{U}Y$ with $Y \sim \mu$, for some distribution μ over $\mathbb{R}^D_{\geq 0}$.

Then, the distribution μ from (C3) is a symmetric mixed Nash equilibrium. Moreover, every symmetric mixed Nash equilibrium μ is associated with some $(H_1, \ldots, H_N, S)$ that satisfy (C1)-(C3).

The set S captures the support of the realized inferred user values $[\langle u_1, p\rangle, \ldots, \langle u_N, p\rangle]$ for $p \sim \mu$. The distribution H_i captures the distribution of the maximum inferred user value $\max_{1 \leq j \leq P-1}\langle u_i, p_j\rangle$ for user u_i.

The conditions in Lemma 50 help us identify and analyze the equilibria in concrete instantations, including in the 2 user vector setting that we focus on in this section.

- (C1) places conditions on H_1, H_2, and S in terms of the induced cost function $c_{\mathbf{U}}$. We use the first-order and second-order conditions of equation (8.9) at $z = [z_1, z_2]$ to determine the necessary densities $h_1(z_1)$ and $h_2(z_2)$ of H_1 and H_2 for z to be in the support S.
- (C2) restricts the relationship between H_1, H_2, and S for a given value of P, which we instantiate in two different ways, depending on whether the support is a single curve or whether the distribution μ has finitely many genres.
- (C3) holds essentially without loss of generality when u_1 and u_2 are linearly independent.

The proofs of our results in this section boil down to leveraging these conditions.

8.5 Impact of specialization on market competitiveness

Having studied the phenomenon of specialization, we next study its economic consequences on the resulting marketplace of digital goods. We show that producers can achieve *positive profit at equilibrium*, even though producers are competing with each other.

More formally, we can quantify the producer profit at equilibrium as follows. At a symmetric equilibria μ, all producers receive the same expected profit given by:

$$\mathcal{P}^{\text{eq}}(\mu) := \mathbb{E}_{p_1,\ldots,p_P\sim\mu}[\mathcal{P}(p_1; p_{-1})] = \mathbb{E}\Big[\Big(\sum_{i=1}^{N} \mathbb{1}[j^*(u_i;\ p_{1:P}) = 1]\Big) - \|p_1\|^\beta\Big], \tag{8.10}$$

where expectation in the last term is taken over $p_1, \ldots, p_P \sim \mu$ as well as randomness in recommendations. Intuitively, the equilibrium profit of a marketplace provides insight about market competitiveness. Zero profit suggests that competition has driven producers to expend their full cost budget on improving product quality. Positive profit, on the other hand, suggests that the market is not yet fully saturated and new producers have incentive to enter the marketplace.

We show a sufficient condition for positive profit in terms of the user geometry and the cost function, and we show this result relies on the equilibrium exhibiting specialization (see Chapter 8.5.1). Using this analysis of profit, we draw a connection between recommender systems and markets with homogeneous/differentiated goods and discuss implications for market saturation (see Chapter 8.5.2).

8.5.1 Positive equilibrium profit

To gain intuition, let us revisit two users located at the standard basis vectors and cost function $c(p) = \|p\|_2^\beta$. We can obtain the following characterization of profit.

Corollary 51. *Suppose that there are 2 users located at the standard basis vectors $e_1, e_2 \in \mathbb{R}^2$, and the cost function is $c(p) = \|p\|_2^\beta$. For $P = 2$ and $\beta \geq \beta^* = 2$, there is an equilibrium μ where $\mathcal{P}^{eq}(\mu) = 1 - \frac{2}{\beta}$.*

Corollary 51 shows that there exist equilibria that exhibit strictly positive profit for any $\beta \geq 2$. The intuition is that (after sampling the randomness in μ), different producers often produce different genres of content. This reduces the amount of competition along any single genre. Producers are thus no longer forced to maximize quality, enabling them to generate a strictly positive profit.

We generalize this finding to sets of many users and producers and to arbitrary norms. In particular, we provide the following sufficient condition under which the profit at equilibrium is strictly positive.

Proposition 52. *Suppose that*

$$\max_{\|p\| \leq 1} \min_{1 \leq i \leq N} \left\langle p, \frac{u_i}{||u_i||} \right\rangle < N^{-P/\beta}. \tag{8.11}$$

Then for any symmetric equilibrium μ, the profit $\mathcal{P}^{eq}(\mu)$ is strictly positive.

Proposition 52 provides insight into how the geometry of the users and structure of producer costs impact whether producers can achieve positive profit. To interpret Proposition 52, let us examine the quantity $Q := \max_{\|p\| \leq 1} \min_{i=1}^N \langle p, \frac{u_i}{||u_i||} \rangle$ that appears on the left-hand side of (8.11). Intuitively, Q captures how easy it is to produce content that appeals simultaneously to all users. It is larger when the users are close together and smaller when they are spread out. For any set of vectors we see that $Q \leq 1$, with strict inequality if the set of vectors is non-degenerate. The right-hand side of (8.11), on the other hand, goes to 1 as $\beta \to \infty$. Thus, for any non-degenerate set of users, if β is sufficiently large, the condition in Proposition 52 is met and producer profit is strictly positive. The value of β at which positive profit is guaranteed by Proposition 52 decreases as the user vectors become more spread out.

Although Proposition 52 does not explicitly consider specialization, we show that specialization is nonetheless central to achieving positive profit at equilibrium. To illustrate this, we show that at a single-genre equilibrium, the profit is zero whenever there are at least $P \geq 2$ producers.

Proposition 53. *If μ is a single-genre equilibrium, then the profit $\mathcal{P}^{eq}(\mu)$ is equal to* 0.

This draws a distinction between profit in the single-genre regime (where there is no specialization) and the multi-genre regime (where there is specialization).

8.5.2 Economic consequences

We describe two interesting economic consequences of our analysis of equilibrium profit.
Connection to markets with homogeneous and heterogeneous goods

The distinction between equilibrium profit in the single- and multi-genre equilibria parallels the classical distinctions in economics between markets with homogeneous goods and markets with differentiated goods (see (Baye and Kovenock, 2008) for a textbook treatment).

Single-genre equilibria resemble markets with homogeneous goods where firms compete on price. If a firm sets their price above the zero profit level, they can be undercut by other firms and lose their users. The possibility of undercutting drives the profit to zero at equilibrium. Similarly, in the market that we study, when there is no specialization, producers all compete along the same direction, which drives profit to zero. The analogy is not exact: in our model, producers play a distribution of quality and thus might be out-competed in a given realization.

Multi-genre equilibria resemble markets with differentiated goods. In these markets, product differentiation reduces competition between firms, since firms compete for different users. This leads to local monopolies where firms can set prices above the zero profit level. Similarly, in the market that we study, specialization by producers leads to product differentiation and thus induces monopolistic behavior where the profit is positive. More specifically, specialization limits competition within each genre and can enable producers to set the quality of their goods below the zero profit level.

Our results formalize how the supply-side market of a recommender system can resemble a market with homogeneous goods or a market with differentiated goods, depending on whether specialization occurs. An empirical analysis could quantify where on this spectrum a given recommender system is located, and regulatory policy could seek to shift a recommender system towards one of the regimes.

When is a marketplace saturated? Our results provide insight about the number of producers needed for a market to be saturated and fully competitive. Theorem 52 reveals that the marketplace of digital goods may need far more than 2 producers in order to be saturated. Nonetheless, the equilibrium profit does approach 0 as the number of producers in the marketplace goes to ∞: this is because the cumulative profit of all producers is at most

N and producers achieve the same profit, so $\mathcal{P}^{\text{eq}}(\mu) \leq N/P$. Perfect competition is therefore recovered in the infinite-producer limit.

8.6 Discussion and Future Directions

We presented a model for supply-side competition in recommender systems. The rich structure of production costs and the heterogeneity of users enable us to capture marketplaces that exhibit a wide range of forms of specialization. Our main results characterize when specialization occurs, analyze the form of specialization, and show that specialization can reduce market competitiveness. More broadly, we hope that our work serves as a starting point to investigate how recommendations shape the supply-side market of digital goods, and we propose several directions for future work.

One direction for future work is to further examine the economic consequences of specialization. Several of our results take a step towards this goal: Corollary 44 illustrates that single-genre equilibria occur at the direction that maximizes the Nash user welfare, and Proposition 52 shows that specialization can lead to positive producer profit. These results leave open the question of how the welfare of users and producers relate to one another. Characterizing the welfare at equilibrium would elucidate whether specialization helps producers at the expense of users or helps all market participants.

Another direction for future work is to further characterize the equilibrium structure. Our analysis in Chapter 8.4.1 provides insight into the equilibrium structure in the case of two homogeneous users: we showed that finite genre equilibria do not exist outside of the single-genre regime (Theorem 46), and we provided closed-form expression for the equilibria in special cases (Propositions 47-48 and Theorem 49). It would be interesting to extend these insights to general configurations of users.

Finally, we hope that future work extends our model to incorporate additional aspects of content recommender systems. For example, although we focus on perfect recommendations that match each user to their favorite content, we envision that this assumption could be relaxed in several ways: e.g., the platform may have imperfect information about users, users may not always follow platform recommendations, and producers may learn their best-responses over repeated interactions with the platform. Moreover, although we assume that producers earn fixed per-user revenue, this assumption could be relaxed to let producers set prices.

Addressing these questions would further elucidate the market effects induced by supply-side competition, and inform our understanding of the societal effects of recommender systems.

Chapter 9

Clickbait vs. Quality Content

This chapter is based on *"Clickbait vs. Quality: How Engagement-Based Optimization Shapes the Content Landscape in Online Platforms"* (Immorlica et al., 2024), which is joint work with Nicole Immorlica and Brendan Lucier.

9.1 Introduction

Content recommendation platforms typically optimize *engagement metrics* such as watch time, clicks, retweets, and comments (e.g., Smith (2021); Twitter (2023)). Since engagement metrics increase with content quality, one might hope that engagement-based optimization would lead to desirable recommendations. However, engagement-based optimization has led to a proliferation of clickbait (YouTube, 2019), incendiary content (Munn, 2020), divisive content (Rathje et al., 2021) and addictive content (Bengani et al., 2022). A driver of these negative outcomes is that engagement metrics not only reward *quality*, but also reward *gaming tricks* such as clickbait that worsen the user experience.

In this chapter, we examine how engagement-based optimization shapes the landscape of content available on the platform. We focus on the role of strategic behavior by content creators: competition to appear in a platform's recommendations influences what content they are incentivized to create (Ben-Porat and Tennenholtz, 2018; Jagadeesan et al., 2023a; Hron et al., 2022). In the case of engagement-based optimization, we expect that creators strategically decide how much effort to invest in quality versus how much effort to spend on gaming tricks, both of which increase engagement. For example, since the engagement metric for Twitter includes the number of retweets (Twitter, 2023)—which includes both quote retweets (where the retweeter adds a comment) and non-quote retweets (without any comment)—creators can either increase quote retweets by using offensive or sensationalized language (Milli et al., 2023) or increase non-quote retweets by putting more effort into the quality of their content (Example 4). When the engagement metric for video content includes total watch time (Smith, 2021), creators may either increase the "span" of their videos—by investing in quality—or instead increase the "moreishness" by leveraging behavioral weaknesses of users such as temptation (Kleinberg et al., 2024) (Example 5). When the engagement metric includes clicks, creators can rely on clickbait headlines (YouTube, 2019) or actually improve content quality (Example 6).

Intuitively, creators must balance two opposing forces when incorporating quality and gaming tricks in the content that they create. On one hand, it is expensive for creators to invest in quality, but it may be much cheaper to utilize gaming tricks that also increase engagement. On the other hand, gaming tricks generate disutility for users, which might discourage them from engaging with the content even if it is recommended by the platform. This raises the questions:

> *Under engagement-based optimization, how do creators balance between quality and gaming tricks at equilibrium? What is the resulting impact on the content landscape and on the downstream performance of engagement-based optimization?*

To investigate these questions, we propose and analyze a game between content creators competing for user consumption through a platform that performs engagement-based optimization. We model the content creator as jointly choosing investment in quality and utilization of gaming tricks. Both quality and gaming tricks increase engagement from consumption, and utilizing gaming tricks is relatively cheaper for the creators than investing in quality. However, gaming decreases user utility, while quality increases user utility, and a user will not consume the content if their utility from consumption is negative. We study the Nash equilibrium in the game between the content creators.

We first examine the balance between gaming tricks and quality amongst content created at equilibrium (Chapter 9.3). Interestingly, we find that there is a *positive correlation* between gaming and investment at equilibrium: higher-quality content typically exhibits *higher* levels of gaming tricks. We prove that equilibria exhibit this positive correlation (Figure 9.1; Theorem 56), and we also empirically validate this finding on a Twitter dataset (Milli et al., 2025) (Figure 9.2 and Table 9.1). These results suggest that gaming tricks and quality should be viewed as *complements*, rather than substitutes.

Accounting for how the platform's metric shapes the content landscape at equilibrium, we then analyze the downstream performance of engagement-based optimization (Chapter 9.4). We uncover striking properties of engagement-based optimization along several performances axes and discuss implications for platform design (Figure 9.3).

- **Content Quality.** First, we examine the average quality of content consumed by users and show that it can *decrease* as gaming tricks become more costly for creators (Figure 9.3a; Theorem 57). In other words, as it becomes more difficult for content creators to game the engagement metric, the average content quality at equilibrium becomes worse. From a platform design perspective, this suggests that increasing the transparency of the platform's metric (which intuitively reduces gaming costs for creators) may improve the average quality of content consumed by users.

- **User Engagement.** Next, we examine the realization of user engagement metrics at the equilibrium of content generation and user consumption. Even though engagement-based optimization perfectly optimizes engagement on a fixed content landscape, engagement-based optimization can perform worse than other baselines (e.g., optimizing directly for quality) at equilibrium (Figure 9.3b; Theorem 60). From a platform design perspective, this suggests that even if the platform's true objective is realized engagement, the platform might still prefer approaches other than engagement-based optimization when accounting for the way content creators will respond.

- **User Welfare.** Finally, we examine the user welfare at equilibrium. We show that engagement-based optimization can lead to lower user welfare at equilibrium than even the conservative baseline of randomly recommending content (Figure 9.3c; Theorem 62). From a platform design perspective, this suggests that engagement-based optimization may not retain users in a competitive marketplace in the long-run.

Altogether, these results illustrate the importance of factoring in the endogeneity of the content landscape when assessing the downstream impacts of engagement-based optimization.

9.1.1 Related Work

Our work connects to research threads on *content creator competition in recommender systems* and *strategic behavior in machine learning.*

Content-creator competition in recommender systems. An emerging line of work has proposed game-theoretic models of content creator competition in recommender systems, where content creators strategically choosing what content to create (Basat et al., 2017; Ben-Porat and Tennenholtz, 2018; Ben-Porat et al., 2020) or the quality of their content (Ghosh and McAfee, 2011; Qian and Jain, 2024). Some models embed content in a continuous, multi-dimensional action space, characterizing when specialization occurs (Jagadeesan et al., 2023a) and the impact of noisy recommendations (Hron et al., 2022). Other models capture that content creators compete for engagement (Yao et al., 2023a) and general functions of platform "scores" across the content landscape (Yao et al., 2023b). These models have also been extended to dynamic settings, including where the platform learns over time (Ghosh and Hummel, 2013; Liu and Ho, 2018; Hu et al., 2023) and where content creators learn over time (Ben-Porat et al., 2020; Prasad et al., 2023). Notably, Buening et al. (2024) study a dynamic setting where the platform learns over time and content creators strategically choose the probability of feedback (clickrate) of their content. However, while these works all assume that creator utility depends only on winning recommendations (or only on content scores according to the platform metric (Yao et al., 2023a;b)), our model incorporates *misalignment* between the platform's (engagement) metric and user utility.[1] In particular, our model and

[1] A rich line of work (e.g., (Ekstrand and Willemsen, 2016; Milli et al., 2021; Kleinberg et al., 2024; Stray et al., 2021)) has identified sources of misalignment between engagement metrics and user utility and broader issues with inferring user preferences from observed behaviors; these sources of misalignment motivated us to incorporate gaming tricks which increase engagement but reduce user utility into our model.

insights rely on the fact that creators only derive utility if their content is recommended *and* the content generates nonnegative user utility.

Several other works study content creator competition under different modelling assumptions: e.g., where content quality is fixed and all creator actions are gaming (Milli et al., 2023), where content creators have fixed content but may dynamically leave the platform over time (Mladenov et al., 2020; Ben-Porat and Torkan, 2023; Huttenlocher et al., 2023), where the recommendation algorithm biases affect market concentration but content creators have fixed content (Calvano et al., 2023; Castellini et al., 2023), where the platform designs a contract determining payments and recommendations (Zhu et al., 2023), where the platform creates its own content (Aridor and Gonçalves, 2022), and where the platform designs badges to incentivize user-generated content (Immorlica et al., 2015). This line of work also builds on Hotelling models of product selection from economics (e.g. (Hotelling, 1929; Salop, 1979), see Anderson et al. (1992) for a textbook treatment).

Strategic behavior in machine learning. A rich line of work on *strategic classification* (e.g. (Brückner et al., 2012; Hardt et al., 2016)) focuses primarily on agents strategically adapting their features in classification problems, whereas our work focuses on agents competing to win users in recommender systems. Some works also consider improvement (e.g. (Kleinberg and Raghavan, 2020; Haghtalab et al., 2021; Ahmadi et al., 2022)), though also with a focus on classification problems. One exception is (Liu et al., 2022), which studies ranking problems; however, the model in (Liu et al., 2022) considers all effort as improvement, whereas our model distinguishes between clickbait and quality. Other topics studied in this research thread include shifts to the population in response to a machine learning predictor (e.g. (Perdomo et al., 2020)), strategic behavior from users (e.g. (Haupt et al., 2023)), and incentivizing exploration (e.g., (Kremer et al., 2014; Frazier et al., 2014; Sellke and Slivkins, 2021)).

9.2 Model

We study a stylized model for content recommendation in which an online platform recommends to each user a single piece of digital content within the content landscape available on the platform.[2] There are $P \geq 2$ content creators who each create a single piece of content and compete to appear in recommendations. Building on the models of Ben-Porat and Tennenholtz (2018); Jagadeesan et al. (2023a); Hron et al. (2022); Yao et al. (2023b), the content landscape is *endogeneously* determined by the multi-dimensional actions of the content creators.

9.2.1 Creator Costs, User Utility, and Platform Engagement

Since our focus is on investment versus gaming, we project pieces of digital content into 2 dimensions $w = [w_{\text{costly}}, w_{\text{cheap}}] \in \mathbb{R}^2_{\geq 0}$. The more costly dimension w_{costly} denotes a measure

[2] Our model can also capture a stream of content (e.g., see Example 5), even though we abstract away from this by focusing on one recommendation at a time.

of the content's *quality*, whereas the cheap dimension w_{cheap} reflects the extent of *gaming tricks* present in the content. These measures are normalized so that $w = [0, 0]$ represents content generated by a creator who exerted no effort on quality or gaming.

We specify below how the costly and cheap dimensions impact *creator costs*, *user utility*, and *platform engagement*. Using these specifications, we then provide additional intuition for the *qualitative interpretation of quality and gaming tricks* in our model.

Creator Costs. Each content creator pays a (one-time) cost of $c(w) \geq 0$ to create content $w \in \mathbb{R}^2_{\geq 0}$. We assume that c is continuously differentiable in w and satisfies the following additional assumptions. First, investing in quality content is costly: $(\nabla(c(w)))_1 > 0$ for all $w \in \mathbb{R}^2_{\geq 0}$. Moreover, engaging in gaming tricks is either always free or always incurs a cost: either $(\nabla(c(w)))_2 > 0$ for all $w \in \mathbb{R}^2_{\geq 0}$ or $(\nabla(c(w)))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$. Furthermore, creators have the option to opt out by not investing costly effort in either gaming tricks or quality: $c([0, 0]) = 0$. Finally, costs go to ∞ in the limit: $\sup_{w_{\text{costly}}} c([w_{\text{costly}}, 0]) = \infty$.

User Utility. Each user has a type $t \in \mathcal{T} \subseteq \mathbb{R}_{\geq 0}$ that reflects the user's relative tolerance for gaming tricks. We assume that the type space $\mathcal{T}$ is finite. A user with type $t \in \mathcal{T}$ receives utility $u(w, t) \in \mathbb{R}$ from consuming content $w \in \mathbb{R}^2_{\geq 0}$, where the utility function is normalized so that the user's outside option offers 0 utility. We assume that u is continuously differentiable in w for each $t \in \mathcal{T}$ and satisfies the following additional assumptions. Users derive positive utility from w_{costly} and negative utility from w_{cheap}:

- For each $t \in \mathcal{T}$ and $w_{\text{costly}} \in \mathbb{R}_{\geq 0}$: the utility $u([w_{\text{costly}}, w_{\text{cheap}}], t)$ is strictly decreasing in w_{cheap} and approaches $-\infty$ as $w_{\text{cheap}} \to \infty$.

- For each $t \in \mathcal{T}$ and $w_{\text{cheap}} \in \mathbb{R}_{\geq 0}$: the utility $u([w_{\text{costly}}, w_{\text{cheap}}], t)$ is strictly increasing in w_{costly} and approaches ∞ as $w_{\text{costly}} \to \infty$.

Furthermore, higher types are more likely to have a nonnegative user utility than lower types, which captures that higher types are less sensitive to gaming tricks than lower types:

- For any $w \in \mathbb{R}^2_{\geq 0}$ and $t, t' \in \mathcal{T}$ such that $t' > t$: if $u(w, t) \geq 0$, then it holds that $u(w, t') \geq 0$.

Engagement. If a user chooses to consume content w, this interaction generates platform engagement $M^{\text{E}}(w) \in \mathbb{R}$. The engagement metric $M^{\text{E}}(w)$ depends on the content w but is independent of the user's type t (conditional on the user choosing to consume the content). We assume that M^{E} is continuously differentiable in w and satisfies the following additional assumptions. First, both cheap gaming tricks and investment in quality increase the engagement metric: $(\nabla M^{\text{E}}(w))_1, (\nabla M^{\text{E}}(w))_2 > 0$ for all $w \in \mathbb{R}^2_{\geq 0}$. Moreover, the engagement metric is nonnegative: $M^{\text{E}}(w) \geq 0$ for all $w \in \mathbb{R}^2_{\geq 0}$. Finally, the relative cost of gaming tricks versus costly investment is less than the relative benefit: $\frac{(\nabla c(w))_2}{(\nabla c(w))_1} < \frac{(\nabla M^{\text{E}}(w))_2}{(\nabla M^{\text{E}}(w))_1}$ for all $w \in \mathbb{R}^2_{\geq 0}$. In other words, it is more cost-effective for a creator to increase the engagement metric via gaming than via quality, for a user who would choose to consume the content either way.

Qualitative Interpretation of Quality and Gaming Tricks. With this formalization of creator costs, user utility, and platform engagement in place, we turn to the qualitative interpretation of quality as measured by w_{costly} and gaming tricks as measured by w_{cheap}. Both quality and gaming tricks reflect effort by creators that increases engagement; however, quality captures effort that is beneficial to users (increases user utility), whereas gaming tricks captures effort that is harmful to users (reduces user utility). Moreover, since a creator can simultaneously invest effort into both quality and gaming tricks, a single piece of digital content can exhibit both gaming tricks and quality at the same time. In fact, high-quality content which also exhibits a sufficient level of gaming tricks can generate arbitrarily low user utility, which illustrates that quality does *not* capture a user's level of appreciation of the content. We defer further discussion of quality and gaming tricks to Chapter 9.2.3, where we instantiate our model within several real-world examples.

9.2.2 Timing and Interaction between the Platform, Users, and Content Creators

The interaction between the platform, users, and content creators defines a game that proceeds in stages. The timing of the game is as follows:

Stage 1: Each content creator $i \in [P]$ simultaneously chooses what content $w_i \in \mathbb{R}^2_{\geq 0}$ to create. These choices give rise to a content landscape $\mathbf{w} = (w_1, \ldots, w_P)$.

Stage 2: A user with type $t \sim \mathcal{T}$ is uniformly drawn and comes to the platform.

Stage 3: The platform observes the user's type and evaluates content w according to a metric $M : \mathbb{R}^2_{\geq 0} \to \mathbb{R}$ that maps each piece of content w_i to a score $M(w_i)$. The platform optimizes M over content available in the content landscape that generates nonnegative utility for the user. More formally, the platform selects content creator

$$i^*(M; \mathbf{w}) \in \arg\max_{i \in [P]} (M(w_i) \cdot \mathbb{1}[u(w_i, t) \geq 0]),$$

breaking ties uniformly at random, and recommends the content $w_{i^*(M;\mathbf{w})}$ to the user.

Stage 4: The user consumes the recommended content $w_{i^*(M;\mathbf{w})}$ iff $u(w_{i^*(M;\mathbf{w})}, t) \geq 0$ (i.e., if and only if the content is at least as appealing as their outside option).

We assume that content creators know the user utility function u and the distribution of $\mathcal{T}$ but do not know the specific realization of $t \sim \mathcal{T}$ in **Stage 2**. On the other hand, the platform can observe the realization $t \sim \mathcal{T}$. The platform can also observe the full content landscape $\mathbf{w}$ and knows the user utility function u. This provides the platform with sufficient information to solve the optimization problem $\arg\max_{i \in [P]}(M(w_i) \cdot \mathbb{1}[u(w_i, t) \geq 0])$ in **Stage**

3.[3] The user knows their own type t and the utility function u, and can also observe the content w recommended to them, so they can evaluate whether $u(w_{i^*(M;\mathbf{w})}, t) \geq 0$.[4]

Equilibrium decisions of content creators. The recommendation process defines a game played between the content creators, who strategically choose their content $w_i \in \mathbb{R}^2_{\geq 0}$ in **Stage 1**. We assume that values are normalized so that a content creator receives a value of 1 for being shown to a user. Since the goods are digital, production costs are one-time and incurred regardless of whether the user consumes the content. Creator i's expected utility is therefore

$$U_i(w_i; \mathbf{w}_{-i}) := \mathbb{E}[\mathbb{1}[i^*(M; \mathbf{w}) = i]] - c(w), \tag{9.1}$$

where the expectation is over any randomness in user types $\mathcal{T}$. We allow content creators to randomize over their choice of content, and write $\mu_i \in \Delta(\mathbb{R}^2_{\geq 0})$ for such a mixed strategy. A (mixed) Nash equilibrium $(\mu_1, \ldots, \mu_P)$, for $\mu_i \in \Delta(\mathbb{R}^2_{\geq 0})$, is a profile of mixed strategies that are mutual best-responses. Since the content creators are symmetric in our model, we will focus primarily on symmetric mixed Nash equilibria in which each creator employs the same mixed strategy, which must exist (see Theorem 54 below). Note that the Nash equilibrium specifies the distribution over content landscapes $\mathbf{w}$.

The platform's choice of metric M in Stage 3. We primarily focus on *engagement-based optimization* where $M = M^{\mathrm{E}}$, meaning that the platform optimizes for engagement. As a benchmark, we also consider *investment-based optimization* where $M(w) = M^{\mathrm{I}}(w) := w_{\mathrm{costly}}$ does not reward gaming tricks; however, note that this baseline is idealized, since w_{costly} is not always identifiable from observable data in practice. As another baseline, we consider *random recommendations* where $M(w) = M^{\mathrm{R}}(w) := 1$ which captures choosing uniformly at random from all content that generates nonnegative user utility.

9.2.3 Running examples

We provide instantation of our models that serves as running examples throughout the paper.

Example 4. *Consider an online platform such as Twitter which uses retweets as one of the terms its objective (Twitter, 2023). However, Twitter does not differentiate between quote retweets (where the retweeter adds a comment) and non-quote retweets (where there is no added comment). Creators can cheaply increase quote retweets by increasing the offensiveness*

[3]The platform may be able to evaluate $\arg\max_{i \in [P]}(M(w_i) \cdot \mathbb{1}[u(w_i, t) \geq 0])$ with less information. For example, if $M = M^{\mathrm{E}}$, then $M^{\mathrm{E}}(w)$ can typically be estimated from observable data such as user behavior patterns without knowledge of w_{costly} and w_{cheap}. Moreover, since $\mathbb{1}[u(w_i, t) \geq 0]$ captures the event that users click on the content w_i, if the platform has a predictor for clicks, this would provide them an estimate of $\mathbb{1}[u(w_i, t) \geq 0]$.

[4]In reality, users may not always be able to perfectly observe w_{costly} and w_{cheap} (or gauge their own utility) without consuming the content. Our model makes the simplifying assumption that user choice is noiseless.

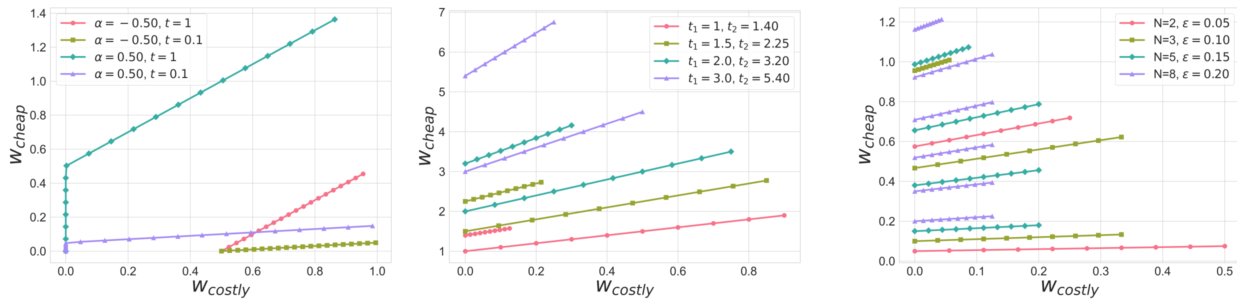
or sensationalism of the content (Milli et al., 2023), or increase non-quote retweets by actually improving content quality. As a stylized model for this, let w_{cheap} be the offensiveness of the content and let w_{costly} capture costly investment into content quality. Let the utility function of a user with type $t \in \mathcal{T} \subseteq \mathbb{R}_{>0}$ be the linear function $u(w,t) = w_{costly} - (w_{cheap}/t) + \alpha$, where $\alpha \in \mathbb{R}$ is the baseline utility from no effort and t captures the user's tolerance to offensive content. Let the platform metric $M^E(w) = w_{costly} + w_{cheap}$ and cost function $c(w) = w_{costly} + \gamma \cdot w_{cheap}$ for $\gamma \in [0,1)$ also be linear functions. The platform metric captures the idea that the platform does not distinguish between different types of retweets; the cost function captures the idea that it is relatively easier for creators to insert sensationalism into tweets, which requires just a few word changes, compared to improving content quality, which might require, for example, time-intensive fact-checking.

Example 5. *Consider an online platform such as TikTok (Smith, 2021) that incorporates watch time into its objective. Creators can increase watch time by: a) creating "moreish" content that keeps users watching a video stream even after they are deriving disutility from it, or b) increasing "span" by increasing the amount of substantive content, as modelled in Kleinberg et al. (2024). More formally, let $w_{costly} := \frac{q}{1-q}$ be a reparameterized version of the span $q \in [0,1]$, let $w_{cheap} := \frac{p}{1-p}$ be a reparameterized version of the moreishness $p \in [0,1]$.*[5] *For a given user, let v be the value derived from each time step from watching substantive content, let W be the outside option for each time step, and let $t := v/W - 1 > 0$ capture the shifted ratio. In this notation, the engagement metric M^E and user utility u from Kleinberg et al. (2024) take the following form: $M^E([w_{cheap}, w_{costly}]) := w_{costly} + w_{cheap} + 1$ and $u(w,t) := W \cdot t \cdot (w_{costly} - w_{cheap}/t + 1)$. We further specify the cost function based on a linear combination of the expected amount of "span" time and the expected amount of "moreish" time that the user consumes: $c(w) := w_{costly} + \gamma \cdot w_{cheap}$ where $\gamma \in [0,1)$ specifies the cost of increasing moreishness relative to increasing span.*[6]

Example 6. *Consider an online platform such as YouTube that historically used clicks as one of the terms in their objective (YouTube, 2019). Creators can cheaply increase clicks by leveraging clickbait titles or thumbnails or by increasing the quality of their content. As a stylized model for this, let w_{cheap} capture how flashy or sensationalized the title or thumbnail is, and let w_{costly} capture the quality of the content. The number of clicks $M^E(w)$ increases with both clickbait w_{cheap} and quality w_{costly}, and user utility $u(w,t)$ increases with quality w_{costly} and decreases with clickbait w_{cheap}. A user quits the platform if their utility falls below zero. (This means the event $\mathbb{1}[u(w,t) \geq 0]$ captures that the user does not quit the platform, rather than the event that the user clicks the content, for this particular example.)*

[5] In the model in Kleinberg et al. (2024), users have two modes: System 1 (the "addicted" mode) and System 2 (the "rational" mode). Roughly speaking, the moreishness p is the probability that the user continues to watch the video stream while in System 1, and the span q is the analogous probability for System 2.

[6] While Example 4 and Example 5 differ in terms of real-world interpretations, the functional forms in the two examples are very similar. In particular, the cost functions $c(w)$ are identical, and the engagement functions M^{E} are identical up to a scalar shift of 1. The user utility u in Example 5 is equal to the user utility u in Example 4 with $\alpha = 1$ and with a multiplicative shift of $W \cdot t$.



(a) $|\mathcal{T}| = 1$ (b) $|\mathcal{T}| = 2$ (c) $|\mathcal{T}| = N$

Figure 9.1: Support of a symmetric mixed equilibrium for engagement-based optimization in Chapter 4. The parameter settings are $\gamma = 0.1$ (left), $\alpha = 1$, $\gamma = 0$, $\mathcal{T} = \{t_1, t_2\}$ (middle), and $\alpha = 1$, $\gamma = 0$, $\mathcal{T} = \mathcal{T}_{N,\varepsilon}$ (right). The support exhibits positive correlation between gaming tricks w_{cheap} and investment in quality w_{costly} (Proposition 55 and Theorem 56). For homogeneous users (left), the slope varies with the type t and the intercept varies with the baseline utility α (Theorem 67). For heterogeneous users with N well-separated types (right), the support consists of N' disjoint line segments with varying slopes and intercepts, where $N' < N$ in several cases (Theorem 70).

9.2.4 Equilibrium existence and overview of equilibrium characterization results

We show that a symmetric mixed equilibrium exists for engagement-based optimization for arbitrary setups.

Theorem 54. *Let $\mathcal{T} \subseteq \mathbb{R}_{\geq 0}$ be any finite type space. Then a symmetric mixed equilibrium exists in the game between content creators with $M = M^E$.*

Since the game has an infinite action space and has discontinuous utility functions, the proof of Theorem 54 relies on equilibrium existence technology for *discontinuous* games (Reny, 1999). We defer the full proof to Chapter F.2.

Although the symmetric mixed equilibrium does not appear to permit a clean closed-form characterization in general, we compute closed-form expressions for a symmetric mixed equilibrium $\mu^{\text{e}}(P, c, u, \mathcal{T})$ under further structural assumptions (Figure 9.1; Chapter 9.6). When users are *homogeneous* (i.e. $\mathcal{T} = \{t\}$), we compute a symmetric mixed equilibrium for all possible settings of P, c, and u (Figure 9.1a; Theorem 67). We also consider *heterogeneous* users (i.e. where $|\mathcal{T}| > 1$) under further restrictions: we assume the gaming tricks are costless, place a linearity assumption on the costs c and engagement metric M^{E} that is satisfied by Examples 4-5, and focus on the case of $P = 2$ creators. We compute a symmetric mixed equilibrium for arbitrary type spaces $\mathcal{T} = \{t_1, t_2\}$ with two types (Figure 9.1b; Theorem 71) and for arbitrarily large type spaces with sufficiently "well-separated" types such as $\mathcal{T}_{N,\varepsilon} := \{(1+\varepsilon)(1+1/N)^{i-1} - 1 \mid 1 \leq i \leq N\}$ (Figure 9.1c; Theorem 70).

We also provide closed-form expressions for a symmetric mixed equilibrium for investment-based optimization $\mu^{\mathrm{i}}(P, c, u, \mathcal{T})$ and random recommendations $\mu^{\mathrm{r}}(P, c, u, \mathcal{T})$ under certain structural assumptions (Chapter 9.5).

9.3 Positive correlation between quality and gaming tricks

When the platform optimizes engagement metrics M^{E}, each content creator *jointly* determines how much to utilize gaming tricks and invest in quality. The creators' equilibrium decisions of how to balance gaming and quality in turn determine the properties of content in the content landscape. In this section, we show that there is a positive correlation between gaming and quality: that is, content that exhibits higher levels of gaming typically exhibits higher investment in quality. We prove that the equilibria satisfy this property (Chapter 9.3.1), and we empirically validate this property on a dataset (Milli et al., 2025) of Twitter recommendations (Chapter 9.3.2).

9.3.1 Theoretical analysis of balance between gaming and quality

We theoretically analyze the balance of gaming and quality at equilibrium as follows. Since the content landscape $\mathbf{w} = [w_1, \ldots, w_P]$ at equilibrium consists of content $w_i \sim \mu_i$ for $i \in [P]$, the set of content that shows up in the content landscape with nonzero probability is equal to $\cup_{i \in [P]} \mathrm{supp}(\mu_i)$. We examine the relationship between the quality w_{costly} and the level of gaming w_{cheap} for $w \in \cup_{i \in [P]} \mathrm{supp}(\mu_i)$.

For general type spaces, we show that the set $\cup_{i \in [P]} \mathrm{supp}(\mu_i)$ of content is contained in a union of curves, each exhibiting "positive correlation" between w_{cheap} and w_{costly} (Figure 9.1).

Proposition 55. *Let $\mathcal{T} \subseteq \mathbb{R}_{\geq 0}$ be any finite type space, and suppose that gaming is not costless (i.e. $(\nabla(c(w)))_2 > 0$ for all $w \in \mathbb{R}^2_{\geq 0}$). There exist weakly increasing functions $f_t : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ for each $t \in \mathcal{T}$ such that at any (mixed) Nash equilibrium $(\mu_1, \mu_2, \ldots, \mu_P)$ in the game with $M = M^E$, the set of content $\cup_{i \in [P]} \mathit{supp}(\mu_i)$ is contained in the following set:*

$$\cup_{i \in [P]} \mathit{supp}(\mu_i) \subseteq \left(\cup_{t \in \mathcal{T}} \underbrace{\{(f_t(w_{cheap}), w_{cheap}) \mid w_{cheap} \geq 0\}}_{(A)} \right) \cup \underbrace{\{(0,0)\}}_{(B)}.$$

Proposition 55 guarantees positive correlation within each of (at most) $|\mathcal{T}|$ curves in the support. While this does not guarantee positive correlation across the full support in general, it does imply this global form of positive correlation for the *homogeneous users*. We make this explicit in the following corollary of Proposition 55.

Theorem 56. *Suppose that users are homogenous (i.e. $\mathcal{T} = \{t\}$) and gaming is not costless (i.e. $\nabla(c(w))_2 > 0$ for all $w \in \mathbb{R}^2_{\geq 0}$). Let $(\mu_1, \mu_2, \ldots, \mu_P)$ be any (mixed) Nash equilibrium in the game with $M = M^E$, and let $w^1, w^2 \in \cup_{i \in [P]} supp(\mu_i)$ be any two pieces of content in the support. If $w^2_{cheap} \geq w^1_{cheap}$, then $w^2_{costly} \geq w^1_{costly}$.*

Theorem 56 shows that a creator's investment in quality content weakly increases with the creator's utilization of gaming tricks. This illustrates a positive correlation between gaming tricks and investment in quality in the content landscape. Perhaps surprisingly, this positive correlation indicates that even high-quality content on the content landscape will have clickbait headlines or exhibit other gaming tricks. Thus, gaming tricks and investment should be viewed as *complements* rather than substitutes.

We provide a proof sketch of Proposition 55 (Theorem 56 follows immediately as a corollary).

Proof sketch of Proposition 55. Let us first interpret the two types of sets in Proposition 55. For each t, the set (A) is a one-dimensional curve specified by f_t where the costly component is weakly increasing in the cheap component. We construct f_t to be the *minimum-investment function*

$$f_t(w_{\text{cheap}}) = \inf \{w_{\text{costly}} \mid w_{\text{costly}} \geq 0, u([w_{\text{costly}}, w_{\text{cheap}}], t) \geq 0\}.$$

The value $f_t(w_{\text{cheap}})$ captures the minimum investment level in quality needed to achieve nonnegative utility for type t users, given w_{cheap} utilization of gaming tricks. For example, the function f_t takes the following form in Chapter 4:

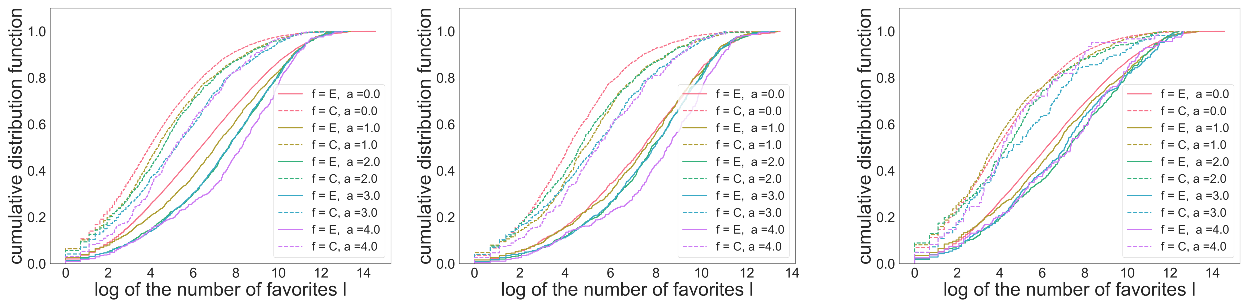
Example 4 (Continued)**.** *The function f_t can be taken to be $f_t(w_{cheap}) = \max(0, (w_{cheap}/t) - \alpha)$ (this follows from Lemma 200 in Chapter F.1 and Lemma 207 in Chapter F.3). As t increases (and users becomes more tolerant to gaming tricks), the slope of f_t decreases. As a result, an increase in utilization of gaming tricks results in less of an increase in investment in quality.*

The set (B) of costless actions captures creators "opting out" of the game by not expending any costly effort in producing their content.

We show that the set (A) captures all of the content that a creator might reasonably select if they are optimizing for being recommended to a user with type t. In particular, if a creator is optimizing for type-t users, they will invest the minimum amount in quality to maintain nonnegative utility on those users. We further show that when best-responding to the other content creators, a creator will either optimize for winning one of the user types $t \in \mathcal{T}$ or opt out by expending no costly effort. We defer the full proof to Chapter F.3. $\square$

9.3.2 Empirical analysis on Twitter dataset

We next provide empirical validation for the positive correlation between gaming and investment on a Twitter dataset (Milli et al., 2025). The dataset consists of survey responses from 1730 participants, each of whom was asked several questions about each of the top ten tweets



(a) $\mathcal{G} = \{P, \neg P\}$ (b) $\mathcal{G} = \{P\}$ (c) $\mathcal{G} = \{\neg P\}$

Figure 9.2: Cumulative distribution function $H_{a,f,\mathcal{G}}$ of the number of favorites ($w_{\text{costly}} = l$) conditioned on different angriness levels ($w_{\text{cheap}} = a$) on a dataset (Milli et al., 2025) of tweets from the engagement-based feeds ($f = E$) and chronological feeds ($f = C$). The tweet genre is unrestricted (left), restricted to political tweets (middle), and restricted to not political tweets (right). The cdf for higher values of a appears to stochastically dominate the cdf for lower values of a, suggesting a positive correlation between w_{cheap} and w_{costly}. The stochastic dominance is more pronounced for political tweets than for non-political tweets, and it occurs for engagement-based and chronological feeds.

in their personalized and chronological feeds. Using the user survey responses, we associate each tweet with a tuple:

$$(f, g, a, l) \in \{E, C\} \times \{P, \neg P\} \times \{0, 1, 2, 3, 4\} \times \mathbb{Z}_{\geq 0}.$$

The *feed* $f \in \{E, C\}$ captures whether the tweet was in the user's engagement-based feed ($f = E$) or chronological feed ($f = C$). The *genre* $g \in \{P, \neg P\}$ captures whether the user labelled the content as in the political genre ($g = P$) or not ($g = \neg P$). The *angriness level* $a \in \{0, 1, 2, 3, 4\}$ captures the reader's evaluation of how angry the author appears in their tweet, rated numerically between 0 and 4.[7] The *number of favorites* $l \in \mathbb{Z}_{\geq 0}$ captures the number of favorites (i.e. "heart reactions") of the tweet. Let D be the multiset D of tuples from the tweets in the dataset, and let $\mathcal{D}$ be the distribution where (f, g, a, l) is drawn uniformly from the multiset D.

We map this empirical setup to Chapter 4 as follows. Since w_{cheap} is intended to capture the offensiveness of content in Chapter 4, we estimate w_{cheap} by the angriness level a. Since w_{costly} is intended to capture the costly investment into content quality in Chapter 4, we estimate w_{costly} by the number of favorites l. We expect that increasing author angriness w_{cheap} decreases user utility, drawing upon intuition from Munn (2020) that incendiary or divisive content drives engagement by provoking outrage in users. Furthermore, we expect

[7]The survey question asked: "How is [author-handle] feeling in their tweet?" (Milli et al., 2025)

	$\mathcal{G} = \{P, \neg P\}$	$\mathcal{G} = \{P\}$	$\mathcal{G} = \{\neg P\}$
$f = E$	0.131	0.092	0.048
	$(2 \cdot 10^{-76})$	$(2 \cdot 10^{-10})$	$(3 \cdot 10^{-9})$
$f = C$	0.086	0.138	0.004
	$(2 \cdot 10^{-33})$	$(4.49 \cdot 10^{-19})$	$(3 \cdot 10^{-1})$

Table 9.1: Correlation coefficient $\rho_{f,\mathcal{G}}$ (with p-value $p_{f,\mathcal{G}}$ in parentheses) between the number of favorites ($w_{\text{costly}} = l$) and the angriness level ($w_{\text{cheap}} = a$) on a dataset (Milli et al., 2025) of tweets from the engagement-based feeds ($f = E$) and chronological feeds ($f = C$) and across political (P) and non-political ($\neg P$) tweets. The correlation coefficient is positive (though weak) and statistically significant in all cases except for non-political tweets in the chronological feed. Moreover, correlations are stronger for political than for non-political tweets.

that higher quality content would generally receive more favorites w_{costly} and lead to higher user utility (if the author angriness level is held constant).[8]

We analyze the relationship between the number of favorites (w_{costly}) and the angriness (w_{cheap}) with two different approaches:

- *Stochastic dominance of conditional distributions:* Given an angriness level $a \in \{0, 1, 2, 3, 4\}$, feed $f \in \{E, C\}$ and subset of genres $\mathcal{G} \subseteq \{P, \neg P\}$, consider the random variable $\ln(L)$ where (F, G, A, L) is drawn from the conditional distribution $\mathcal{D} \mid (A = a, F = f, G \in \mathcal{G})$. We let $H_{a,f,\mathcal{G}}$ denote the cumulative density function of this random variable. We visually examine the extent to which $H_{a,f,\mathcal{G}}$ stochastically dominates $H_{a',f,\mathcal{G}}$ when $a > a'$.

- *Correlation coefficient:* Given a feed $f \in \{E, C\}$ and subset of genres $\mathcal{G} \subseteq \{P, \neg P\}$, we compute the multiset

$$S_{f,\mathcal{G}} := \{(a, l) \mid (f, g, a, l) \in D \mid f = f', g' \in \mathcal{G}\} \tag{9.2}$$

  We compute the Spearman's rank correlation coefficient $\rho_{f,\mathcal{G}} \in [-1, 1]$ of the multiset $S_{f,\mathcal{G}}$ and a corresponding p-value $p_{f,\mathcal{G}}$.[9]

Stochastic dominance of conditional distributions. Figure 9.2 shows the cumulative distribution function $H_{a,f,\mathcal{G}}$ for different values of a, f, and $\mathcal{G}$. The primary finding is that in

[8]The dataset (Milli et al., 2025) also includes other author and reader emotions besides author angriness (such as author happiness or reader sadness). The reason that we focus on author angriness is that we believe it to most closely match the interpretation of "gaming tricks" in our model: while we expect increasing author angriness to decrease user utility (as described above), we might not expect increasing other emotions, such as author happiness, to decrease user utility.

[9]The p-value is for a one-sided hypothesis test with null hypothesis that a and l have no ordinal correlation, calcuated using the scipy.stats.spearmanr Python library.

all of the plots, the cdf for higher values of a visually appears to stochastically dominate the cdf for lower values of a. This stochastic dominance reflects a higher author's angriness level $w_{\text{cheap}} = a$ leads to higher numbers of favorites $w_{\text{costly}} = l$, thus suggesting that content with higher levels of gaming w_{cheap} also exhibit higher quality w_{costly}.

Interestingly, the stochastic dominance is most pronounced when $\mathcal{G} = \{P, \neg P\}$ and $\mathcal{G} = \{P\}$, but less pronounced when $\mathcal{G} = \{\neg P\}$. This aligns with the intuition that increasing author angriness more effectively increases engagement for political tweets than for non-political tweets.[10] Moreover, within $\mathcal{G} = \{P, \neg P\}$ and $\mathcal{G} = \{P\}$, the stochastic dominance occurs for both $f = E$ and $f = C$. We view each of $f = E$ and $f = C$ as capturing a different slice of the content landscape: the fact that stochastic dominance occurs in two different slices suggests it broadly occurs in the content landscape.

Correlation coefficient. Table 9.1 shows $\rho_{f,\mathcal{G}}$ for different genres of tweets and feeds. Interestingly, the correlation coefficient is positive in all cases, which suggests that content with higher levels of gaming tend to exhibit higher levels of investment in quality. However, the correlation is somewhat weak: this may be due to angriness ratings being incomparable across different survey participants. Nonetheless, the correlation is stronger for political content, which again aligns with the intuition that increasing author angriness is more effective in increasing engagement for political tweets.[11]

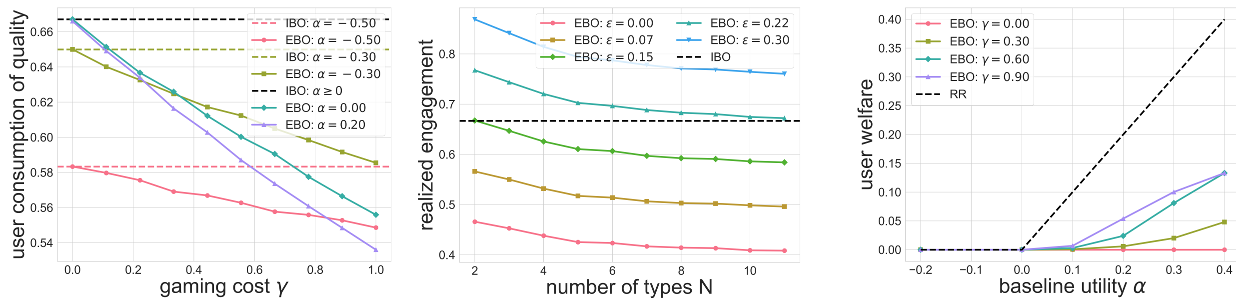
9.4 Performance of engagement-based optimization at equilibrium

In this section, taking into account the structure of the the content landscape at equilibrium, we investigate the downstream performance of engagement-based optimization. As baselines, we consider investment-based optimization (an idealized baseline that optimizes directly for quality $M^{\mathrm{I}}(w) = w_{\text{costly}}$) and random recommendations (a trivial baseline that results in randomly choosing from content that achieves nonnegative user utility). We highlight striking aspects of these comparisons (Figure 9.3), considering three qualitatively different performance axes: user consumption of quality (Chapter 9.4.1), realized engagement (Chapter 9.4.2), and user utility (Chapter 9.4.3).

Our comparisons take into account the *endogeneity of the content landscape*: i.e., that the content landscape at equilibrium depends on the choice of metric. The possibility of multiple equilibria casts ambiguity on which equilibrium to consider. To resolve this ambiguity, we will focus on the (symmetric mixed) equilibria from our characterization results in Chapter 9.5

[10] For non-political tweets, we expect other types of gaming tricks are employed.

[11] For many other emotions (both positive and negative) measured in (Milli et al., 2025), the analogous correlation coefficients are also positive. For negative emotions, these coefficients could also be interpreted as correlations between gaming tricks and quality within our model. On the other hand, for positive emotions, where increasing the level of the positive emotion might *increase* (rather than decrease) user utility, the resulting correlation coefficient does not have a clear interpretation within our model.



(a) User consumption of quality (b) Realized engagement (c) User welfare

Figure 9.3: Equilibrium performance of engagement-based optimization (EBO) in Chapter 4 with $P = 2$ creators along several performance axes (left to right). The performance is numerically estimated from 100,000 samples from the equilibrium distributions (Chapter 9.6). The parameter settings are $\mathcal{T} = \{1\}$ (left), $\mathcal{T} = \mathcal{T}_{N,\varepsilon}$, $\alpha = 1$, and $\gamma = 0$ (middle), and $\mathcal{T} = \{5\}$ (right). The equilibrium performance of investment-based optimization (IBO) and random recommendations (RR) are analytically computed from the equilibrium distributions (Theorem 65 and Theorem 66) and shown as baselines. User consumption of quality can decrease with gaming costs (left; Theorems 57-58), realized engagement can be lower for EBO than for IBO (middle; Theorem 60), and user welfare can be lower for EBO than for RR (right; Theorem 62).

and Chapter 9.6. That is, throughout this section we will focus on equilibria $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ (for engagement-based optimization), $\mu^{\mathrm{i}}(P, c, u, \mathcal{T})$ (for investment-based optimization), and $\mu^{\mathrm{r}}(P, c, u, \mathcal{T})$ (for random recommendations).

For ease of exposition, in this section, we focus on Example 4 for different parameter settings of the baseline utility α, the gaming cost level γ, the number of creators P, and the type space $\mathcal{T}$. The results in this section directly translate to other instantiations of our model including Chapter 5.

9.4.1 User consumption of quality

We first consider the average quality of content consumed by users (formalized below), focusing on the case of homogeneous users in Chapter 4. We show that as gaming costs increase, the performance of engagement-based optimization *worsens*; in fact, unless gaming is *costless*, engagement-based optimization performs strictly worse than investment-based optimization.

We formalize user consumption of quality by:

$$\mathrm{UCQ}(M; \mathbf{w}) := \mathbb{E}\left[M^{\mathrm{I}}(w_{i^*(M;\mathbf{w})}) \cdot \mathbb{1}[u(w_{i^*(M;\mathbf{w})}, t) \geq 0]\right],$$

which only counts content quality if the content is actually consumed by the user. Taking into account the endogeneity of the content landscape, the user consumption of quality at a symmetric mixed Nash equilibrium μ^M is:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^M)^P}\left[\mathrm{UCQ}(M;\mathbf{w})\right].$$

The following result shows that in Chapter 4 the average user consumption of quality strictly *decreases* as gaming costs (parameterized by γ) become more expensive (Figure 9.3a).

Theorem 57. *Suppose that users are homogeneous (i.e. $\mathcal{T} = \{t\}$). For any sufficiently large baseline utility $\alpha > -1$, bounded gaming costs $\gamma \in [0, 1)$, and any number of creators $P \geq 2$, the user consumption of quality $\mathbb{E}_{\mathbf{w}\sim(\mu^e(P,c,u,\mathcal{T}))^P}[UCQ(M^E;\mathbf{w})]$ for engagement-based optimization is strictly decreasing in γ.*

Proof sketch of Theorem 57. For sufficiently large values of w_{costly}, creators compete their utility down to 0, so the only remaining strategic choice is how they choose to trade off effort spent on gaming versus investment. If gaming is costly, then creators need to expend more of their effort on gaming to achieve a desired increase in engagement, so they will necessarily devote less effort to investment in quality. In contrast, if gaming is costless, creators devote all of their effort to investment. To formalize this intuition, we explicitly compute user consumption of quality using the equilibrium characterization. We defer the proof to Chapter F.6.1. □

Theorem 57 thus has a striking consequence for platform design: to improve user consumption of quality, it can help to reduce the costs of gaming tricks as much as possible. One concrete approach for reducing gaming costs is to increase the *transparency* of the platform's metric, for example by publishing the metric in an interpretable manner. In particular, if a content creator does not have access to the platform's metric, they would have to expend effort to learn the metric to game it; on the other hand, transparency would reduce these costs. Perhaps countuitively, our results suggest that increasing transparency can *improve* user consumption of quality in the presence of strategic content creators.[12] In particular, our results suggest the recent trend of recommender systems publishing their algorithms (e.g., Twitter (2023)) may improve user consumption of quality content, and encourage the continued release of recommendation algorithms more broadly.

To further understand the impact of gaming costs γ, we compare the performance of engagement-based optimization with the performance of investment-based optimization (which does not depend on γ). We treat the performance of investment-based optimization as an "idealized baseline" for UCQ: the reason is that for any *fixed* content landscape $\mathbf{w}$, investment-based optimization maximizes the $\mathrm{UCQ}(M;\mathbf{w})$ across all possible metrics M,

[12] This finding bears some resemblance to results in the strategic classification literature (Ghalme et al., 2021; Bechavod et al., 2022). For example, Ghalme et al. (2021) shows that transparency is the optimal policy in terms of optimizing the decision-maker's accuracy. However, a lack of transparency is suboptimal in Ghalme et al. (2021) because it prevents the decision-maker from being able to fully anticipating strategic behavior; in contrast, a lack of transparency is suboptimal in our setting because it leads effort to be spent on figuring how to game the classifier rather than investing in quality.

because the objectives exactly align. The following result shows that engagement-based optimization performs strictly worse than investment-based optimization unless gaming tricks are *costless* (Figure 9.3a).

Theorem 58. *Suppose that users are homogeneous (i.e. $\mathcal{T} = \{t\}$). For any sufficiently large baseline utility $\alpha > -1$, bounded gaming costs $\gamma \in [0, 1)$, and any number of creators $P \geq 2$, it holds that:*

$$\mathbb{E}_{\mathbf{w} \sim (\mu^e(P,c,u,\mathcal{T}))^P}[UCQ(M^E; \mathbf{w})] \leq \mathbb{E}_{\mathbf{w} \sim (\mu^i(P,c,u,\mathcal{T}))^P}[UCQ(M^I; \mathbf{w})],$$

with equality if and only if $\gamma = 0$.

Theorem 58 illustrates that reducing the gaming costs to 0 is *necessary* for engagement-based optimization to perform as well as the idealized baseline. This serves as a further motivation for a social planner to try to reduce gaming costs as much as possible, for example through increased transparency as discussed above.

We caution that reducing gaming costs to 0 is not *sufficient* to guarantee that engagement-based optimization performs as well as investment-based optimization, if users are heterogeneous. The following result shows that for heterogeneous users, the average user consumption of quality of engagement-based optimization can be significantly lower than the average user consumption of quality of investment-based optimization.

Proposition 59. *For any $N \geq 2$, there exists an instance with $\gamma = 0$ and a type space $\mathcal{T}$ of N well-separated types such that the average user consumption of quality of engagement-based optimization is less than the average user consumption of quality of investment-based optimization:*

$$\mathbb{E}_{\mathbf{w} \sim (\mu^e(P,c,u,\mathcal{T}))^P}[UCQ(M^E; \mathbf{w})] \leq \frac{1}{N} < \frac{2}{3} = \mathbb{E}_{\mathbf{w} \sim (\mu^i(P,c,u,\mathcal{T}))^P}[UCQ(M^I; \mathbf{w})].$$

Proposition 59 illustrates that it is possible for the average user consumption of quality of engagement-based optimization to approach 0 as $N \to \infty$ while the performance of investment-based optimization stays constant and nonzero, when gaming costs are reduced to 0. This result suggests that other interventions, beyond reducing gaming costs, may be necessary to ensure that engagement-based optimization does not substantially degrade the overall quality of content being consumed by users.

9.4.2 Realized engagement

We next consider how much engagement is realized by user consumption patterns, when accounting for the fact that users only consume recommendations that generate nonnegative utility for them. We show that even though engagement-based optimization maximizes realized engagement on any *fixed* content landscape, engagement-based optimization can be suboptimal at equilibrium, when taking into account the endogeneity of the content landscape.

We formalize realized engagement by

$$\mathrm{RE}(M; \mathbf{w}) := \mathbb{E}[M^{\mathrm{E}}(w_{i^*(M;\mathbf{w})}) \cdot \mathbb{1}[u(w_{i^*(M;\mathbf{w})}, t) \geq 0]].$$

When taking into account the endogeneity of the content landscape, the realized engagement at a symmetric mixed Nash equilibrium μ^M is $\mathbb{E}_{\mathbf{w}\sim(\mu^M)^P}[\mathrm{RE}(M; \mathbf{w})]$.

To show the suboptimality of engagement-based optimization at equilibrium, we construct an instance where engagement-based optimization performs strictly worse than investment-based optimization (Figure 9.3b).

Theorem 60. *For sufficiently large N, there exists an instance with a type space $\mathcal{T}$ of N well-separated types such that the realized engagement of engagement-based optimization is less than the realized engagement of investment-based optimization:*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(P,c,u,\mathcal{T}))^2}[RE(M^E; \mathbf{w})] < \mathbb{E}_{\mathbf{w}\sim(\mu^i(P,c,u,\mathcal{T}))^2}[RE(M^I; \mathbf{w})].$$

Proof sketch of Theorem 60. The main ingredient of the proof of Theorem 60 is constructing and analyzing an instance where engagement-based optimization achieves a low realized engagement. We construct the instance to be Example 4 with costless gaming ($\gamma = 0$), baseline utility $\alpha = 1$, and $P = 2$ creators, and type space $\mathcal{T}_{N,\varepsilon} := \{(1+\varepsilon)(1+1/N)^{i-1} - 1 \mid 1 \leq i \leq N\}$ for sufficiently small $\varepsilon > 0$ and sufficiently large N.[13]

One key aspect of $\mathcal{T}_{N,\varepsilon}$ is that the heterogeneity in user types segments the market and significantly reduces investment in quality. In particular, since user types are well-separated in the type space $\mathcal{T}_{N,\varepsilon}$, creators can't realistically compete for multiple types at the same time and must choose a single type to focus on. At a high-level, this segments the market, and a single creator can only hope to win $O(1/N)$ users and thus only invests $O(1/N)$ in costly effort. (Note that there are some subtleties, because a creator who targets a lower type might win a higher type if none of the other creators target the higher type in that particular realization of randomness.) In the limit as $N \to \infty$, we show that the investment in quality approaches 0.

However, for engagement-based optimization, a low investment in quality does not directly imply a low realized engagement. This is because if users are high type (and thus highly tolerant of gaming tricks), creators can utilize a high level of gaming tricks without investing at all in quality, while still maintaining nonnegative utility for these users. Thus, to show that the realized engagement is low, we must consider the distribution over user types. The construction of $\mathcal{T}_{N,\varepsilon}$ appropriately balances two forces: (1) making the user types sufficiently well-separated to reduce the investment in quality, and (2) making the user types as low as possible to reduce engagement from gaming tricks.

To analyze the realized engagement of this construction, we first upper bound $\mathrm{RE}(M^{\mathrm{E}}; \mathbf{w})$ by the maximum engagement achieved by any content in the content landscape $\max_{w\in\mathbf{w}} M^{\mathrm{E}}(w)$. It then remains to analyze the engagement distribution of $M^{\mathrm{E}}(w)$ for w in the equilibrium

[13] While Theorem 60 focuses on the limit as $N \to \infty$, the numerical estimates shown in Figure 9.3b suggest that the result applies for any $N \geq 2$.

distribution. Although the cdf of the engagement distribution is messy for any given value of N, it approaches a continuous distribution in the limit. This enables us to show that:

$$\limsup_{\varepsilon \to 0} \limsup_{N \to \infty} \mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2}[\mathrm{RE}(M^{\mathrm{E}}; \mathbf{w})] < \mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2}[\mathrm{RE}(M^{\mathrm{I}}; \mathbf{w})].$$

We defer the proof to Chapter F.7.1.

□

Theorem 60 has an interesting platform design consequence: even if the platform wants to optimize realized engagement (e.g., because their revenue comes from advertising), engagement-based optimization is not necessarily the optimal approach. In particular, Theorem 60 illustrates potential benefits of using investment-based optimization, even though M^{I} does not directly reward engagement. A practical challenge is that investment-based optimization is often difficult for the platform to implement because w_{costly} may not be directly observable; nonetheless, the platform may be able to perform a noisy version of investment-based optimization by collecting (sparse) feedback about content quality from users. Theorem 60 raises the possibility that noisy versions of investment-based optimization may be worthwhile for the platform to pursue, even if the platform's goal is to maximize realized engagement.

As a caveat, Theorem 60 does rely on users types being heterogeneous. In fact, the following result shows that for homogeneous users, engagement-based optimization performs at least as well as investment-based optimization in terms of realized engagement.

Proposition 61. *Suppose that users are homogeneous ($\mathcal{T} = \{t\}$). For any sufficiently large baseline utility $\alpha > -1$, bounded gaming costs $\gamma \in [0, 1)$, and any number of creators $P \geq 2$, the realized engagement of engagement-based optimization is at least as large as the the realized engagement of investment-based optimization:*

$$\mathbb{E}_{\mathbf{w} \sim (\mu^e(P,c,u,\mathcal{T}))^P}[RE(M^E; \mathbf{w})] \geq \mathbb{E}_{\mathbf{w} \sim (\mu^i(P,c,u,\mathcal{T}))^P}[RE(M^I; \mathbf{w})].$$

While Proposition 61 does show that engagement-based optimization can generate nontrivial realized engagement when users are homogeneous, we expect that when the user base exhibits sufficient diversity in tolerance towards gaming tricks, investment-based optimization would be an appealing alternative to engagement-based optimization.

9.4.3 User welfare

Finally, we consider user utility realized by user consumption patterns, which can be interpreted as *user welfare*. We show that engagement-based optimization can alarmingly perform worse than random recommendations in terms of user welfare.[14]

We formalize user welfare by

$$\mathrm{UW}(M; \mathbf{w}) := \mathbb{E}[u(w_{i^*(M;\mathbf{w})}, t) \cdot \mathbb{1}[u(w_{i^*(M;\mathbf{w})}, t) \geq 0]].$$

[14]We view random recommendations as a conservative baseline, since M^{R} does not reward investment or gaming.

Taking into account the endogeneity of the content landscape, the user welfare at a symmetric mixed Nash equilibrium μ^M is $\mathbb{E}_{\mathbf{w}\sim(\mu^M)^P}[\mathrm{UW}(M;\mathbf{w})]$.

The following result shows that for homogeneous users, engagement-based optimization always performs at least as poorly as random recommendations, and can even perform strictly worse than random recommendations under certain conditions (Figure 9.3c).

Theorem 62. *Suppose that users are homogeneous (i.e. $\mathcal{T} = \{t\}$), and that gaming costs $\gamma \in [0, 1)$ are bounded. If baseline utility $\alpha > 0$ is positive, the user welfare of engagement-based optimization is strictly lower than the user welfare of random recommendations:*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(P,c,u,\mathcal{T}))^P}[UW(M^E;\mathbf{w})] < \mathbb{E}_{\mathbf{w}\sim(\mu^r(P,c,u,\mathcal{T}))^P}[UW(M^R;\mathbf{w})].$$

If baseline utility $\alpha \leq 0$ is nonpositive, engagement-based optimization and random recommendations both result in zero user welfare:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(P,c,u,\mathcal{T}))^P}[UW(M^E;\mathbf{w})] = \mathbb{E}_{\mathbf{w}\sim(\mu^r(P,c,u,\mathcal{T}))^P}[UW(M^R;\mathbf{w})] = 0.$$

Proof sketch of Theorem 62. We first focus on the simple case where gaming tricks are free ($\gamma = 0$) and the baseline utility is positive ($\alpha \geq 0$). For engagement-based optimization, creators will increase gaming tricks until the user utility drops down to 0, which means the user welfare at equilibrium is 0. In contrast, for random recommendations, creators do not expend effort on either gaming tricks or investment; thus, the user welfare at equilibrium is $u([0, 0], t) > 0$, which is strictly higher than the user welfare for engagement-based optimization. The other cases, though a bit more involved, follow from similar intuition: for engagement-based optimization, creators choose the balance between gaming tricks and investment in quality that drives user utility as close to zero as possible, whereas for random recommendations, creators choose the minimum amount of investment to achieve nonzero user utility. We defer the full proof to Chapter F.8.1. □

From a platform design perspective, Theorem 62 highlights the pitfalls of engagement-based optimization for users. In particular, the user welfare of engagement-based optimization can fall below the conservative baseline where users randomly select content on their own (and the content landscape shifts in response). This suggests that engagement-based optimization may not retain users in the long-run, especially in a competitive marketplace with multiple platforms.

It is important to note that for *heterogeneous users*, engagement-based optimization does not always perform as poorly as random recommendations. In the following result, we turn to Example 5 and construct instances with 2 user types where engagement-based optimization outperforms random recommendations.[15]

[15] While all of our other results in Chapter 9.4 apply to Examples 4 and 5, Propositions 63-64 only apply to Example 5. The extra factor of t in the utility function formalization in Example 5 turns out to be necessary for these results.

Proposition 63. *Consider Chapter 5. There exist instances with 2 types where the user welfare of engagement-based optimization is higher than the user welfare of random recommendations:*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(2,c,u,\mathcal{T}))^2}[UW(M^E;\mathbf{w})] > \mathbb{E}_{\mathbf{w}\sim\left(\mu^i_{2,c,u,\mathcal{T}}\right)^2}[UW(M^R;\mathbf{w})].$$

On the other hand, we also construct instances with two types where user welfare of random recommendations outperforms the user welfare of engagement-based optimization, thus behaving similarly to the case of homogeneous users.

Proposition 64. *Consider Chapter 5. There exist instances with 2 types where the user welfare of engagement-based optimization is lower than the user welfare of random recommendations:*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(2,c,u,\mathcal{T}))^2}[UW(M^E;\mathbf{w})] < \mathbb{E}_{\mathbf{w}\sim\left(\mu^i_{2,c,u,\mathcal{T}}\right)^2}[UW(M^R;\mathbf{w})].$$

Interestingly, the construction in Proposition 63 relies on the types *not* being too well-separated while the construction in Proposition 64 relies on the types being sufficiently well-separated. This raises the interesting question of characterizing the relative performance of random recommendation and engagement-based optimization in greater generality, which we defer to future work.

9.5 Equilibrium characterization results for baseline approaches

Within our analysis in Chapter 9.4, we leveraged closed-form equilibrium characterizations for investment-based optimization and random recommendations in several cases. In this section, we state these characterizations. To state our characterizations, we define a distribution $\mu^{\text{i}}(P, c, u, \mathcal{T})$ for investment-based optimization and a distribution $\mu^{\text{r}}(P, c, u, \mathcal{T})$ for random recommendations.

Since neither baseline approach directly incentivizes gaming tricks, the distributions $\mu^{\text{i}}(P, c, u, \mathcal{T})$ and $\mu^{\text{r}}(P, c, u, \mathcal{T})$ both satisfy $w_{\text{cheap}} = 0$ for all w in the support (i.e., the marginal distribution of W_{cheap} is a point mass at 0). We can thus convert the two-dimensional action space into a one-dimensional action space specified by w_{costly}, where the cost function is

$$C^I_b(w_{\text{costly}}) := c([w_{\text{costly}}, 0]) \tag{9.3}$$

and the utility function is:

$$U^I_b(w_{\text{costly}}, t) := u([w_{\text{costly}}, 0], t). \tag{9.4}$$

We place the following structural assumptions on the type space and utilities which simplify the equilibrium structure. For each type $t \in \mathcal{T}$, let β_t be the minimum level of investment needed to achieve nonnegative utility:

$$\beta_t := \min \left\{ w_{\text{costly}} \mid U_b^I(w_{\text{costly}}) \geq 0 \right\}.$$

We assume that either (1) users are homogeneous ($\mathcal{T} = 1$), or (2) users are heterogeneous and no user requires investment to achieve nonnegative utility (i.e., $\beta_t = 0$ for all $t \in \mathcal{T}$).

We now specify the marginal distribution of quality W_{costly} for investment-based optimization (Chapter 9.5.1) and random recommendations (Chapter 9.5.2). We defer the proofs to Chapter F.5.

9.5.1 Characterization for investment-based optimization

We first consider investment-based optimization where $M = M^{\text{I}}$.

When users are homogeneous ($\mathcal{T} = \{t\}$), we define the marginal distribution of W_{costly} for $\mu^{\text{i}}(P, c, u, \mathcal{T})$ by:

$$\mathbb{P}[W_{\text{costly}} \leq w_{\text{costly}}] = \begin{cases} \min\left(1, C_b^I(\beta_t)\right)^{1/(P-1)} & \text{if } 0 \leq w_{\text{costly}} \leq \beta_t \\ \min\left(1, C_b^I(w_{\text{costly}})\right)^{1/(P-1)} & \text{if } w_{\text{costly}} \geq \beta_t. \end{cases}$$

When users are heterogeneous and $\beta_t = 0$ for all $t \in \mathcal{T}$, we define the marginal distribution of W_{costly} for $\mu^{\text{i}}(P, c, u, \mathcal{T})$ by:

$$\mathbb{P}[W_{\text{costly}} \leq w_{\text{costly}}] = \min\left(1, C_b^I(w_{\text{costly}})\right)^{1/(P-1)}.$$

We show that $\mu^{\text{i}}(P, c, u, \mathcal{T})$ is a symmetric mixed equilibrium.

Theorem 65. *Suppose that either (a) $|\mathcal{T}| = 1$ or (b) $\beta_t = 0$ for all $t \in \mathcal{T}$. Then, the distribution $\mu^{\text{i}}(P, c, u, \mathcal{T})$ is a symmetric mixed Nash equilibrium in the game with $M = M^I$.*

9.5.2 Characterization for random recommendations

We next consider random recommendations where $M = M^{\text{R}}$.

First, we consider the case where users are homogeneous (i.e., $\mathcal{T} = \{t\}$). Let κ be minimum cost to achieve 0 user utility, truncated at 1: that is, $\kappa := \min\left(1, C_b^I(\beta_t)\right)$. Let the probability ν be defined as follows: $\nu = 0$ if $\kappa \leq 1/P$, and otherwise $\nu \in [0, 1]$ is the unique value such that that $\sum_{i=0}^{P-1} \nu^i = P \cdot \kappa$. We define the marginal distribution of W_{costly} for $\mu^{\text{r}}(P, c, u, \mathcal{T})$ by

$$\mathbb{P}[W_{\text{costly}} = w_{\text{costly}}] = \begin{cases} \nu & \text{if } w_{\text{costly}} = 0 \\ 1 - \nu & \text{if } w_{\text{costly}} = \beta_t \\ 0 & \text{otherwise.} \end{cases}$$

When users are heterogeneous and $\beta_t = 0$ for all $t \in \mathcal{T}$, we define the marginal distribution of W_{costly} for $\mu^{\text{r}}(P, c, u, \mathcal{T})$ to be a point mass at $w_{\text{costly}} = 0$.

We show that $\mu^{\text{r}}(P, c, u, \mathcal{T})$ is a symmetric mixed equilibrium.

Theorem 66. *Suppose that either (a) $|\mathcal{T}| = 1$ or (b) $\beta_t = 0$ for all $t \in \mathcal{T}$. Then, the distribution $\mu^{\mathrm{r}}(P, c, u, \mathcal{T})$ is a symmetric mixed Nash equilibrium in the game with $M = M^R$.*

9.6 Equilibrium characterization for engagement optimization

Within our analysis of the performance of engagement-based optimization in Chapter 9.4, we implicitly leveraged closed-form characterizations of the symmetric mixed equilibria for engagement-based optimization in several concrete instantiations. In this section, we state these closed-form characterizations. To state our characterizations, we define a distribution $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ over $\mathbb{R}^2_{\geq 0}$, when $\mathcal{T}$ has a single type (Definition 5), when $\mathcal{T}$ has two types under further assumptions (Definition 8), and when $\mathcal{T}$ has N well-separated types under further assumptions (Definition 7). We will show $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ is a symmetric mixed Nash equilibria for engagement-based optimization in each case.

To simplify the notation in our specification of $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$, we convert the two-dimensional action space into the following union of $|T|$ one-dimensional curves that specifies the support of the equilibria. We define the *minimum-investment functions* $f_t : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, as follows:

$$f_t(w_{\mathrm{cheap}}) := \inf \left\{ w_{\mathrm{costly}} \mid w_{\mathrm{costly}} \geq 0, u([w_{\mathrm{costly}}, w_{\mathrm{cheap}}], t) \geq 0 \right\}, \tag{9.5}$$

so f_t captures the amount of investment needed to offset the disutility from w_{cheap} level of gaming tricks for users of type t. Within each one-dimensional curve, the content w is entirely specified by the cheap component w_{cheap}, which motivates us to define a one-dimensional cost function for content along each curve:

$$C_t(w_{\mathrm{cheap}}) := c([f_t(w_{\mathrm{cheap}}), w_{\mathrm{cheap}}]). \tag{9.6}$$

For example, the functions f_t and C_t take the following form in Chapter 4:

Example 4 (Continued)**.** *The functions f_t and C_t are as follows:*

$$f_t(w_{cheap}) = \max(0, (w_{cheap}/t) - \alpha)$$
$$C_t(w_{cheap}) = \begin{cases} w_{cheap}(\gamma + 1/t) - \alpha & \text{if } w_{cheap} > \max(0, t \cdot \alpha) \\ w_{cheap} \cdot \gamma & \text{if } w_{cheap} \leq t \cdot \alpha. \end{cases}$$

As t increases (and users becomes more tolerant to gaming tricks), the slope of f_t and C_t both decrease. The minimum-investment f_t is independent of γ, but the cost function increases with γ.

In Chapter 9.6.1, we focus on homogeneous users. In Chapter 9.6.2, we state additional assumptions for the case of heterogeneous users. In Chapter 9.6.3, we focus on well-separated types, and in Chapter 9.6.4 we consider two arbitrary types. We defer proofs to Chapter F.4.

9.6.1 Equilibrium characterization for homogeneous users

We first focus on the case where $\mathcal{T} = \{t\}$ has a single type (Figure 9.1a).

Definition 5. *We define the distribution $(W_{costly}, W_{cheap}) \sim \mu^{e}(P, c, u, \mathcal{T})$ over $\mathbb{R}^2_{\geq 0}$ as follows. Let f_t be defined by (9.5), and let C_t be defined by (9.6). The marginal distribution W_{cheap} is defined by:*

$$\mathbb{P}[W_{cheap} \leq w_{cheap}] = \left(\min(1, C_t(w_{cheap}))\right)^{1/(P-1)}.$$

For each $w_{cheap} \in supp(W_{cheap})$, the conditional distribution $W_{costly} \mid W_{cheap} = w_{cheap}$ is defined as follows: if $w_{cheap} > 0$, then $W_{costly} \mid W_{cheap} = w_{cheap}$ is a point mass at $f_t(w_{cheap})$; if $w_{cheap} = 0$, then $W_{costly} \mid W_{cheap} = w_{cheap}$ is a point mass at 0.

For example, the distribution takes the following form within Example 4.

Example 4 (Continued)**.** *Let $P = 2$, $\gamma = 0.1$, $\alpha = 0.5$, and $|\mathcal{T}| = 1$. Then, W_{cheap} and W_{costly} are both distributed as uniform distributions and $\mu^{e}(P, c, u, \mathcal{T})$ is supported on a line segment (Figure 9.1a).*

We prove that $\mu^{e}(P, c, u, \mathcal{T})$ is a symmetric mixed equilibrium.

Theorem 67. *If $|\mathcal{T}| = 1$, the distribution $\mu^{e}(P, c, u, \mathcal{T})$ is a symmetric mixed equilibrium in the game with $M = M^E$.*

In fact, we further prove that $\mu^{e}(P, c, u, \mathcal{T})$ is the unique symmetric mixed equilibrium when gaming tricks are costly.

Theorem 68. *Suppose that $|\mathcal{T}| = 1$ and gaming is costly (i.e. $(\nabla c(w))_2 > 0$ for all $w \in \mathbb{R}^2_{\geq 0}$). Then, if μ is a symmetric mixed equilibrium in the game with $M = M^E$, it holds that $\mu = \mu^{e}(P, c, u, \mathcal{T})$.*

The fact that $\mu^{e}(P, c, u, \mathcal{T})$ is the unique symmetric mixed equilibrium under costly gaming tricks and homogeneous users provides additional justification for our focus on $\mu^{e}(P, c, u, \mathcal{T})$ in Chapter 9.4.

We do note that although $\mu^{e}(P, c, u, \mathcal{T})$ is unique within the class of symmetric equilibrium, there typically do exist asymmetric equilibria. For example, if $P = 3$, the mixed strategy profile where μ_1 is a point mass at $[0, 0]$ and $\mu_2 = \mu_3 = \mu^{e}(2, c, u, t)$ is an equilibrium. Extending our analysis and results to asymmetric equilibria is an interesting direction for future work.

9.6.2 Additional assumptions for characterization results for multiple types

In our characterization results for heterogeneous users, we require the following additional assumptions. One key assumption is the following linearity condition on the *induced cost* *function* given by the optimization program:

$$C_t^E(m) := \min_w c(w) \text{ s.t. } u(w,t) \geq 0, M^{\mathrm{E}}(w) \geq m. \tag{9.7}$$

which captures the minimum production cost to create content with engagement at least m and nonnegative user utility.

Assumption 5 (Linearity of cost functions). *We assume that there exists coefficients $a_t > 0$ for $t \in \mathcal{T}$, intercept $b > 0$, and shift parameter $s \geq -1 \cdot \min_{w \in \mathbb{R}^2_{\geq 0}} M^E(w)$ such that:*

1. *The coefficients a_t are strictly decreasing: $a_{t_1} > a_{t_2}$ for all $t_2 > t_1$.*
2. *The induced cost function is a nonnegative part of a linear function: that is, $C_t^E(m) = \max(0, a_t(m+s) - 1)$ for all $m \in \mathbb{R}$.*

Assumption 5 guarantees that there is a linear relationship between costs and engagement. Apart from Assumption 5, we further assume that gaming tricks are costless (that is, $(\nabla(c(w)))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$) and that $u([0,0],t) \geq 0$ for all $t \in \mathcal{T}$ (i.e. no costly effort is required to meet the user utility constraint for any user).

These assumptions are satisfied by the linear functional forms in Chapter 4 and Chapter 5 with specific parameter settings.

Example 4 (Continued). *For this setup with $\alpha = 1$ and $\gamma = 0$, the cost function assumptions are satisfied for $a_t = \frac{1}{1+t}$ and $s = 1$.*

Example 5 (Continued). *For this setup with $\gamma = 0$, the cost function assumptions are satisfied for $a_t = \frac{1}{1+t} = \frac{W}{v}$ and $s = 0$.*

9.6.3 Characterization for N well-separated types

Interestingly, even under the assumptions in Chapter 9.6.2, the symmetric mixed equilibrium structure is already complex for the case of 2 arbitrary types (as we will show in Chapter 9.6.4). Nonetheless, the equilibrium structure turns out to be significantly cleaner under a "well-separated" assumption on the types: $a_{t_1} \geq 1.5 a_{t_2}$. This motivates us to restrict to "well-separated" types in our analysis of type spaces $\mathcal{T}$ of arbitrary size. The appropriate generalization of the 2-type condition turns out to be:

$$a_{t_1} \geq \left(1 + \frac{1}{N}\right) a_{t_2} \geq \ldots \geq \left(1 + \frac{1}{N}\right)^{N'-1} a_{N'} > 0.$$

As a warmup, let's first consider the case of 2 well-separated types satisfying $a_{t_1} \geq 1.5 a_{t_2}$. The equilibrium is a mixture of 2 distributions, one for each type. The distribution for type t_i looks similar to the equilibrium distribution in Definition 5 for homogeneous users of type t_i, with the modification that there is a factor of 2 multiplier on the cumulative density function of W_{cheap}.

Definition 6. *Let $\mathcal{T} = \{t_1, t_2\}$ be a type space consisting of two types. Furthermore, suppose that gaming tricks are costless (that is, $(\nabla(c(w)))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$) and suppose that $u([0,0], t) \geq 0$ for all $t \in \mathcal{T}$. Suppose that Assumption 5 holds with coefficients satisfying $a_{t_1} \geq 1.5 a_{t_2} > 0$. We define the distribution $(W_{cheap}, W_{costly}) \sim \mu^e(P, c, u, \mathcal{T}, M^E)$ to be be a mixture of the following 2 distributions $(W^1_{cheap}, W^1_{costly})$ and $(W^2_{cheap}, W^2_{costly})$, where the mixture weights are 0.5 and 0.5. Let f_t be defined by (9.5), and let C_t be defined by (9.6). The random variables W^1_{cheap} and W^2_{cheap} are defined by:*

$$\mathbb{P}[W^1_{cheap} \leq w_{cheap}] = \min\left(2 \cdot C_{t_1}(w_{cheap}), 1\right)$$
$$\mathbb{P}[W^2_{cheap} \leq w_{cheap}] = \min\left(4 \cdot C_{t_2}(w_{cheap}), 1\right),$$

and where for $1 \leq i \leq 2$ and $w_{cheap} \in supp(W^i_{costly})$, the distribution $W^i_{costly} \mid W^i_{cheap} = w_{cheap}$ is a point mass at $f_{t_i}(w_{cheap})$.

Proposition 69. *Let $\mathcal{T} = \{t_1, t_2\}$ be a type space consisting of two types. Furthermore, suppose that gaming tricks are costless (that is, $(\nabla(c(w)))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$) and suppose that $u([0,0], t) \geq 0$ for all $t \in \mathcal{T}$. Suppose that Assumption 5 holds with coefficients satisfying $a_{t_1} \geq 1.5 a_{t_2} > 0$. Let $\mu^e(P, c, u, \mathcal{T}, M^E)$ be defined according to Definition 6. Then, $\mu^e(P, c, u, \mathcal{T}, M^E)$ is a symmetric mixed equilibrium in the game with $M = M^E$.*

We are now ready to generalize Definition 6 to $N \geq 2$ "well-separated" types. The distribution is again $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ a mixture of distributions: however, it is surprisingly not a mixture of N distributions, but rather a mixture of $N' \leq N$ distributions corresponding to the first N' types $t_1, \ldots, t_{N'}$. The distribution for t_i again looks similar to the equilibrium distribution in Definition 5 for homogeneous users with type t_i, but again with a multiplicative rescaling on the cdf of W_{cheap}. The multiplicative rescaling is N for $N'-1$ out of N' types.

Definition 7. *Let $\mathcal{T} = \{t_1, \ldots, t_N\}$ be a type space consisting of N types, let $P = 2$, suppose that gaming tricks are costless (that is, $(\nabla(c(w)))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$) and suppose that $u([0,0], t) \geq 0$ for all $t \in \mathcal{T}$. Suppose that Assumption 5 holds with coefficients satisfying*

$$a_{t_1} \geq \left(1 + \frac{1}{N}\right) a_{t_2} \geq \ldots \geq \left(1 + \frac{1}{N}\right)^{N'-1} a_{N'} > 0.$$

We define the distribution $(W_{cheap}, W_{costly}) \sim \mu^e(P, c, u, \mathcal{T})$ to be a mixture of the following N' distributions $\left\{(W^i_{cheap}, W^i_{costly})\right\}_{1 \leq i \leq N'}$, where $N' \in \mathbb{Z}_{\geq 1}$ is the minimum number such that $\sum_{i=1}^{N'} \frac{1}{N-i+1} \geq 1$. The mixture weight α^i on $(W^i_{cheap}, W^i_{costly})$ is

$$\alpha^i := \begin{cases} \frac{1}{N-i+1} & \text{if } 1 \leq i \leq N'-1 \\ 1 - \sum_{i'=1}^{N'-1} \frac{1}{N-i'+1} & \text{if } i = N'. \end{cases}$$

The random vectors $(W^i_{cheap}, W^i_{costly})$ are defined as follows. Let f_t be defined by (9.5), and let C_t be defined by (9.6). The marginal distribution of W^i_{cheap} is defined so that $\mathbb{P}[W^i_{cheap} \leq w_{cheap}]$

equals:

$$\begin{cases} \min\left(N \cdot C_{t_i}(w_{cheap}), 1\right) & \text{if } 1 \le i \le N' - 1 \\ \min\left(\frac{N}{N-N'+1} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)^{-1} \cdot C_{t_i}(w_{cheap}), 1\right) & \text{if } i = N'. \end{cases}$$

For each $1 \le i \le N'$ *and* $w_{cheap} \in supp(W^i_{cheap})$*, the conditional distribution* $W^i_{costly} \mid W^i_{cheap} = w_{cheap}$ *is a point mass at* $f_{t_i}(w_{cheap})$*.*

For the cases of N well-separated types, we show that $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ is a symmetric mixed Nash equilibrium.

Theorem 70. *Let* $\mathcal{T} = \{t_1, \ldots, t_N\}$ *be a type space consisting of* N *types, let* $P = 2$*, suppose that gaming tricks are costless (that is,* $(\nabla(c(w)))_2 = 0$ *for all* $w \in \mathbb{R}^2_{\ge 0}$*), and suppose that* $u([0,0], t) \ge 0$ *for all* $t \in \mathcal{T}$*. Suppose that Assumption 5 holds with coefficients satisfying*

$$a_{t_1} \ge \left(1 + \frac{1}{N}\right) a_{t_2} \ge \ldots \ge \left(1 + \frac{1}{N}\right)^{N'-1} a_{N'} > 0. \tag{9.8}$$

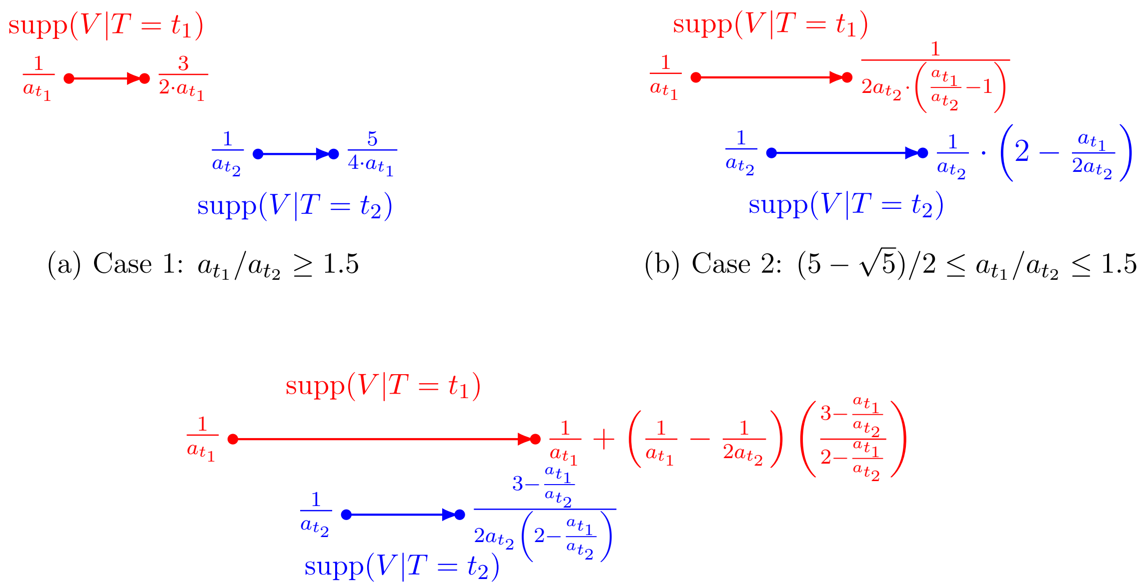
Let $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ *be defined according to Definition 7. Then,* $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ *is a symmetric mixed Nash equilibrium in the game with* $M = M^E$*.*

9.6.4 Characterization for 2 types

For the case of 2 arbitrary types, it is cleaner to work in the following reparametrized space $S := \left\{(M^{\mathrm{E}}([w_{\mathrm{costly}}, w_{\mathrm{cheap}}]) - s, t) \mid t \in \mathcal{T}, u([w_{\mathrm{costly}}, w_{\mathrm{cheap}}], t) = 0\right\} \subseteq \mathbb{R} \times \{t_1, t_2\}$ than directly over the content space $\mathbb{R}_{\ge 0}$. We map each $(v, t) \in S$ to the unique content $h(v, t) \in \mathbb{R}^2_{\ge 0}$ of the form $h(v, t) = [f_t(w_{\mathrm{cheap}}), w_{\mathrm{cheap}}]$ such that $M^{\mathrm{E}}([f_t(w_{\mathrm{cheap}}), w_{\mathrm{cheap}}]) = v - s$. Conceptually, $h(v, t)$ captures content with engagement $v - s$ optimized for winning type t. In our characterization, rather than define directly a distribution over content, we instead define a random vector (V, T) over S, which corresponds a distribution W over content defined so $W \mid (V, T) = (v, t)$ is a point mass at $h(v, t)$.

We split our characterization into three cases depending on the relationship between $\mathrm{supp}(V \mid T = t_1)$ and $\mathrm{supp}(V \mid T = t_2)$ (see Figure 9.4). When types are well-separated, it turns out that $\mathrm{supp}(V \mid T = t_1)$ and $\mathrm{supp}(V \mid T = t_2)$. When types are closer together, the supports are two overlapping line segments, and when types are very close together, the support $\mathrm{supp}(V \mid T = t_2)$ is contained in the support $\mathrm{supp}(V \mid T = t_1)$ and $\mathrm{supp}(V \mid T = t_2)$. We formally define the characterization as follows.

Definition 8. *Let* $\mathcal{T} = \{t_1, t_2\}$ *be a type space consisting of two types, let* $P = 2$*, suppose that gaming tricks are costless (that is,* $(\nabla(c(w)))_2 = 0$ *for all* $w \in \mathbb{R}^2_{\ge 0}$*), and suppose that* $u([0,0], t) \ge 0$ *for all* $t \in \mathcal{T}$*. Suppose that Assumption 5 holds with parameters* b*,* s*, and* $a_{t_1} > a_{t_2} > 0$*. We define the distribution* $W \sim \mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ *as follows. Let* f_t *be defined by*



(a) Case 1: $a_{t_1}/a_{t_2} \geq 1.5$

(b) Case 2: $(5-\sqrt{5})/2 \leq a_{t_1}/a_{t_2} \leq 1.5$

(c) Case 3: $1 < a_{t_1}/a_{t_2} \leq (5-\sqrt{5})/2$

Figure 9.4: The support of (V, T) in Definition 8 for different values of a_{t_1}/a_{t_2}. The red line shows the support of $V \mid T = t_1$, and the blue line shows the support of $V \mid T = t_2$. If a_{t_1} and a_{t_2} are sufficiently far apart (Case 1), then the supports are disjoint. When a_{t_1} and a_{t_2} become closer (Case 2), the supports start to overlap, and when a_{t_1} and a_{t_2} are sufficiently close (Case 3), the support of $V \mid T = t_2$ is contained in the support of $V \mid T = t_1$.

(9.5). *Below we define a random vector* (V, T) *over* $\mathbb{R} \times \{t_1, t_2\}$*; the distribution* $W \mid (V, T) = (v, t)$ *is a point mass at* $W = [f_t(w_{cheap}), w_{cheap}]$ *such that* $M^E([f_t(w_{cheap}), w_{cheap}]) = v - s$.

Case 1 *(*$a_{t_1}/a_{t_2} \geq 1.5$*)**:*** *We define the random vector* (V, T) *so where* V *has density* g *defined to be:*

$$g(v) := \begin{cases} 0 & \textit{if } v \leq \frac{1}{a_{t_1}} \\ a_{t_1} & \textit{if } \frac{1}{a_{t_1}} \leq v \leq \frac{3}{2a_{t_1}} \\ 2a_{t_2} & \textit{if } \frac{1}{a_{t_2}} \leq v \leq \frac{5}{4a_{t_2}} \\ 0 & \textit{if } v \geq \frac{5}{4a_{t_2}} \end{cases}$$

and where $T \mid V = v$ *is distributed according to*

$$\begin{cases} \mathbb{P}[T = t_1 \mid V = v] = 1 & \textit{if } v \in \left[\frac{1}{a_{t_1}}, \frac{3}{2a_{t_1}}\right] \\ \mathbb{P}[T = t_1 \mid V = v] = 0 & \textit{if } v \in \left[\frac{1}{a_{t_2}}, \frac{5}{4a_{t_2}}\right] \end{cases}$$

Case 2 *(*$(5-\sqrt{5})/2 \leq a_{t_1}/a_{t_2} \leq 1.5$*)* *We define the random vector* (V, T) *so where* V

has density g defined to be:

$$g(v) := \begin{cases} 0 & \text{if } v \leq \frac{1}{a_{t_1}} \\ a_{t_1} & \text{if } \frac{1}{a_{t_1}} \leq v \leq \frac{1}{a_{t_2}} \\ 2a_{t_2} & \text{if } \frac{1}{a_{t_2}} \leq v \leq \frac{1}{a_{t_2}}\left(2 - \frac{a_{t_1}}{2a_{t_2}}\right) \\ 0 & \text{if } v \geq \frac{1}{a_{t_2}}\left(2 - \frac{a_{t_1}}{2a_{t_2}}\right) \end{cases}$$

and where $T \mid V = v$ is distributed according to

$$\begin{cases} \mathbb{P}[T = t_1 \mid V = v] = 1 & \text{if } v \in \left[\frac{1}{a_{t_1}}, \frac{1}{a_{t_2}}\right] \\ \mathbb{P}[T = t_1 \mid V = v] = \frac{a_{t_1}}{a_{t_2}} - 1 & \text{if } v \in \left[\frac{1}{a_{t_2}}, \frac{1}{2a_{t_2}\cdot\left(\frac{a_{t_1}}{a_{t_2}}-1\right)}\right] \\ \mathbb{P}[T = t_1 \mid V = v] = 0 & \text{if } v \in \left[\frac{1}{2a_{t_2}\cdot\left(\frac{a_{t_1}}{a_{t_2}}-1\right)}, \frac{1}{a_{t_2}}\left(2 - \frac{a_{t_1}}{2a_{t_2}}\right)\right] \end{cases}$$

***Case 3** ($1 < a_{t_1}/a_{t_2} \leq (5-\sqrt{5})/2$) We define the random vector (V, T) so where V has density g defined to be:*

$$g(v) := \begin{cases} 0 & \text{if } v \leq \frac{1}{a_{t_1}} \\ a_{t_1} & \text{if } \frac{1}{a_{t_1}} \leq v \leq \frac{1}{a_{t_2}} \\ 2a_{t_2} & \text{if } \frac{1}{a_{t_2}} \leq v \leq \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)} \\ a_{t_1} & \text{if } \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)} \leq v \leq \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2-\frac{a_{t_1}}{a_{t_2}}}\right) \\ 0 & \text{if } v \geq \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2-\frac{a_{t_1}}{a_{t_2}}}\right). \end{cases}$$

and where $T \mid V = v$ is distributed according to

$$\begin{cases} \mathbb{P}[T = t_1 \mid V = v] = 1 & \text{if } v \in \left[\frac{1}{a_{t_1}}, \frac{1}{a_{t_2}}\right] \\ \mathbb{P}[T = t_1 \mid V = v] = \frac{a_{t_1}}{a_{t_2}} - 1 & \text{if } v \in \left[\frac{1}{a_{t_2}}, \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right] \\ \mathbb{P}[T = t_1 \mid V = v] = 1 & \text{if } v \in \left[\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}, \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2-\frac{a_{t_1}}{a_{t_2}}}\right)\right] \end{cases}$$

For the case of 2 types, we show that $\mu^{\mathrm{e}}(P, c, u, \mathcal{T}, M^{\mathrm{E}})$ is a symmetric mixed equilibrium.

Theorem 71. *Let $\mathcal{T} = \{t_1, t_2\}$ be a type space with 2 types, let $P = 2$, suppose that gaming tricks are costless (that is, $(\nabla(c(w)))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$) and suppose that $u([0,0],t) \geq 0$ for all $t \in \mathcal{T}$. Suppose that Assumption 5 holds with coefficients $a_{t_1} > a_{t_2} > 0$. Let $\mu^e(P,c,u,\mathcal{T})$ be defined according to Definition 8. Then $\mu^e(P,c,u,\mathcal{T})$ is a symmetric mixed equilibrium in the game with $M = M^E$.*

9.7 Discussion

In this chapter, we study content creator competition for engagement-based recommendations that reward both quality and gaming tricks (e.g. clickbait). Our model further captures that a user only tolerates gaming tricks in sufficiently high-quality content, which also shapes content creator incentives. Our first result (Theorem 56) suggests that gaming and quality are complements for the content creators, which we empirically validate on a Twitter dataset. We then analyze the downstream performance of engagement-based optimization at equilibrium. We show that higher gaming costs can lead to lower average consumption of quality (Theorem 57), engagement-based optimization can be suboptimal even in terms of (realized) engagement (Theorem 60), the user welfare of engagement-based optimization can fall below that of random recommendations (Theorem 62).

More broadly, our results illustrate how content creator incentives can influence the downstream impact of a content recommender system, which poses challenges when evaluating a platform's metric. In particular, there is a disconnect between how a platform's engagement metric behaves on a fixed content landscape and how the same metric behaves on an endogeneous content landscape shaped by the metric. Interestingly, this disconnect manifests in both performance measures relevant to the platform and performance measures relevant to society as a whole. We hope that our work encourages future evaluations of recommendation policies— both of platform metrics and societal impacts—to carefully account for content creator incentives.

Chapter 10

Impact of Generative Models

This chapter is based on *"Flattening Supply Chains: When do Technology Improvements lead to Disintermediation?"* (Ali et al., 2025) which is joint work with Nagee Ali, Nicole Immorlica, and Brendan Lucier.

10.1 Introduction

The digital economy thrives on the ability to distribute content and services at scale. Consider an artist or other creator that uses the latest technology to develop high-quality content, which can be shared through online distribution platforms for a portion of advertising or subscription revenue. A key feature of these platforms is that they enable creators to distribute content to a broad audience at negligible marginal cost.

As technological innovations lead to new tools for content creation, the fixed costs of content production also continue to fall. For example, generative AI tools are making it cheaper to produce art and other digital content.[1] At first glance, it might appear that these technology improvements would benefit creators by reducing the cost of creating high-quality content. However, technology improvements can also threaten *disintermediation*: if the cost of content development gets sufficiently low, consumers may consider bypassing creators in favor of using the underlying technology themselves. Since content creators typically have no control over pricing on these platforms, the only way for them to retain their audience is to provision quality content. Improvements in technology make this strategy cheaper for content creators, but also make self-production more enticing for consumers.

In this chapter, we investigate how shifts in production technology impact disintermediation and, through it, the overall welfare and the landscape of content quality enjoyed at the equilibrium of the content market. We study the strategic choices and pressures faced by intermediaries that sit between an underlying production technology and content consumers (Figure 10.1). The intermediaries can produce content at a chosen level of quality to

[1]We use the term *Generative AI tools* to refer to a wide range of generative models such as language models, text-to-image models, and text-to-video models.

be distributed at scale using a fixed distribution platform. Our focus is the relationship between the underlying production technology (which determines costs), the quality of content produced, and users' choice to use (or not) the intermediary.

Intermediaries provide a social good by unlocking economies of scale: multiple consumers can derive benefit from a one-time investment in content quality.[2] As long as there are multiple consumers, it is socially optimal for an intermediary to create content and distribute it broadly. If the underlying technology improves and production costs decrease then, at this socially optimal solution, the intermediary will produce at a higher level of quality. However, depending on the cost of production, equilibrium forces can push against this first-best outcome. On one hand, if producing high-quality content is too expensive, being a creator that earns some fixed subscription fee or advertising revenue stream may simply not be profitable. On the other hand, if production technology improves to the point of being extremely cheap and easy to use, then consumers are likely to prefer self-made solutions that bypass platform fees or the nuisance cost of advertising. Anticipation of these issues leads to a distortion in the level of quality produced by the intermediary, who must stave off these potential market failures.

Model and results overview. To focus on the competitive disintermediation pressure between the intermediary and the consumers that form their target audience, we study a model with a single intermediary that offers content to a population of consumers. We think of the intermediary as having built an audience for their content, perhaps by developing a niche or reputation, and our model abstracts away from horizontal differentiation and other inter-creator competitive forces that would shape that audience. Consumers pay a fee to access the intermediary's content, and we likewise view the fee structure as determined by the distribution channel(s) used by the intermediary and not the intermediary itself. This is motivated by a lack of market power on the part of any single creator of content: rather than directly influencing market prices or advertising deals, the intermediary controls the quality of the content they produce.

The intermediary contracts with one of multiple possible suppliers of a production technology, which maps quality levels to production costs. The intermediary then chooses to produce their content at a strategically-chosen level of quality. Crucially, the production technology is also available to the consumers, who can choose whether to consume the content created by the intermediary or to bypass it and create their own content. If the consumers consume the content of the intermediary, the intermediary receives a fixed fee per consumer.

We analyze the subgame perfect equilibrium of the resulting game, which we show to be effectively unique. We focus on how this equilibrium changes as the production technology improves, modeled as a multiplicative shift in the cost of production. We summarize our findings as follows:

[2]There are many other reasons intermediaries are a desirable feature of markets, particularly in the context of content production. For example, when intermediaries distribute the same content to multiple consumers, it creates a sense of community among the audience which has arguably additional positive externalities. We leave investigation of these forces open for future work.

- **Disintermediation at the extremes of production technology.** We find that the intermediary stays in the market at a bounded range of technology levels, whereas disintermediation (i.e., all consumers choosing to bypass the intermediary) occurs when the production costs are either too high or too low (Chapter 10.3; Figure 10.2).

- **Intermediary is welfare-improving.** When the intermediary is present in the market, social welfare at equilibrium is higher than in a counterfactual where the intermediary does not exist (Figure 10.3a). The equilibrium welfare is increasing as the production technology improves, and there is at most one "bliss point" where equilibrium welfare matches the first-best (Chapter 10.4.4; Figure 10.7).

- **Intermediary extracts all gains to social welfare.** Any increase in welfare over the counterfactual with no intermediary is captured by the intermediary itself. In other words, consumers and suppliers are indifferent to the intermediary's presence (Chapter 10.4.3; Figure 10.3b). We emphasize that this full extraction of the surplus improvement occurs even though the price of consuming the intermediary's content is fixed: the only strategic knob available to the intermediary is the quality of content they produce. The intermediary's utility is "inverse U-shaped" as a function of the technology level: it first rises as production technology improves, but eventually levels out and begins to drop as the impact of potential disintermediation grows stronger, until eventually the intermediary's utility falls to 0 and disintermediation occurs (Chapter 10.4.2; Figure 10.5).

- **Intermediary can raise or lower quality, and reduces the sensitivity to technology improvements.** We also consider the impact of the intermediary and technology costs on the quality of content produced (Chapter 10.4.1; Figure 10.3c). Cheaper production costs lead to higher equilibrium levels of quality. However, whenever the intermediary is being used at equilibrium, their presence dampens the sensitivity of quality to technology costs. When production costs are high, the intermediary produces higher-quality content than what consumers would produce themselves. As technology improves and production costs go down, quality improvements lag, ultimately reaching a point where the intermediary is producing lower-quality content than what consumers would produce. Once production costs drop low enough that disintermediation occurs, the quality of content increases sharply as the intermediary leaves the market (Figure 10.4).

Model extensions. Finally, as a robustness check, we consider extensions to our base model (Chapter 10.5). First, we consider what changes if the production technology is controlled by a monopolist who can strategically set the relative price of production to maximize revenue, given the underlying technology cost (Chapter 10.5.1; Figure 10.8a). We still find that the intermediary stays in the market at intermediate levels of technology. However, the range of technology levels supporting an intermediary is shifted to be lower: the intermediary can survive when the technology costs are lower, but is less likely to enter when the technology costs are high.

We also consider a variation of the model where marginal distribution is not free, but rather occurs at a reduced cost that is still proportional to content creation costs (Chapter 10.5.2; Figure 10.8b). This can capture settings where the intermediary produces content that might need to be specialized for each consumer at a small marginal cost. For example, the intermediary might be a graphic designer with a suite of standard wedding invitation designs that can be tweaked for any particular client. We again find that the intermediary stays in the market at intermediate levels of technology. However, the range of technology levels supporting an intermediary reduces in width.

Finally, we note the importance of our modeling assumption that the intermediary does not have full control over both the content quality and the fee structure. We consider a variation where the intermediary can charge a fee that increases linearly with content quality (Chapter 10.5.3). In this case, disintermediation can be avoided entirely, at all technology levels, as long as the marginal fee is not too high. We conclude that the nature of the fee structure in digital content distribution, and in particular that individual creators may not be able to differentiate on price, has a substantial impact on the market's sensitivity to underlying technology changes.

10.1.1 Related work

Our work connects with research threads on the *economic ramifications of generative AI tools*, *content creator incentives in recommender systems*, and *fragility of endogenous supply chain networks*.

Economic ramifications of generative AI tools. Our work contributes to an emerging literature on the impact of generative AI tools on labor. Many of these models consider how generative AI substitutes for or is complementary to workers at a macroeconomic level (Acemoglu, 2025; Ide and Talamas, 2024). Our paper views AI tools as complementing creators through cost reduction while also threatening to substitute for them via disintermediation.

In field experiments across specific domains as diverse as software development (Cui et al. (2024); Yeverechyahu et al. (2024)), customer service (Brynjolfsson et al. (2025)), research (Toner-Rodgers (2024)) and art (Zhou and Lee (2024)), generative AI also shows promising gains. For example, the last of these papers showed that for artists, adoption of generative AI tools resulted in 25% more artworks and 25% more "favorites" per view. Our paper tries to understand the implication of these demonstrated effects (that quality production is "cheaper" with the introduction of AI), especially in the artistic domain as consumers bypass experts and use AI to satisfy their own individual demand.

Content creator incentives in recommender systems. Our work relates to a rich line of work on content creator incentives in recommender systems. Most of these works have focused on the *impact of the recommendation algorithm* on the supply of digital content (e.g., (Ghosh and McAfee, 2011; Ben-Porat and Tennenholtz, 2018; Jagadeesan et al., 2023a;

Hron et al., 2022; Yao et al., 2023a)), pricing decisions by creators (e.g., (Calvano et al., 2023; Castellini et al., 2023)), and creator participation decisions (e.g., Ghosh and McAfee (2011); Mladenov et al. (2020); Ben-Porat and Torkan (2023); Huttenlocher et al. (2023)). In contrast, we study the impact of generative AI tools on whether creators can survive in the market.

Similar to our work, a handful of recent works study how generative AI tools affect the digital content economy. For example, Yao et al. (2024); Esmaeili et al. (2024) study competition between creators and generative models, capturing how the quality of the generative model depends on the quality of human-generated content. Taitler and Ben-Porat (2025a;b) study the interplay between participation on human-based platforms and the generative model-provider's strategic retraining decisions. Raghavan (2024) studies how competition between creators who use generative AI tools affects content diversity. Burtch et al. (2024) empirically study how LLMs affect participation in online knowledge communities. In contrast to these works, we focus on how generative AI tools lead to implicit competition between creators and consumers: specifically, we analyze when consumers are incentivized to bypass creators and create content on their own.[3]

Fragility of endogenous supply chain networks. Our work concerns the study of a specific supply chain network – one where there is a single intermediary positioned between firms providing improving production technologies and consumers. A long line of literature in economics, computer science and operations research considers more complex supply chain networks and how they react to shocks such as the severance of links or the bankruptcy of nodes (Acemoglu and Tahbaz-Salehi, 2024; Bimpikis et al., 2019; Blume et al., 2013; Elliott et al., 2022). Similar to our work, these papers allow relationships to form (or be severed) endogenously, but these decisions are driven by the desire to be robust to failures as opposed to changing production technology. Furthermore, in contrast to our work where quality selection is the only lever available to the intermediary, these papers sometimes assume price setting capability. The general message of these papers is that equilibrium supply chain networks can be inefficient and small shocks can cause disproportionate disruptions, suggesting an endogenous fragility. Our work also demonstrates a discrete jump in some comparative statics, namely provisioned quality and disintermediation (i.e., network structure), from small changes. However these outcomes are not disastrous for overall welfare.

Some papers also consider how changes in production technology impact networks (Acemoglu and Azar, 2020) (see also cited papers on economic ramifications of generative AI tools above). These papers tend to show that improvements in technology reduce prices as they diffuse through a fixed production network and lead to a denser endogenous production network.

[3]Our work focuses on a monopolist creator, and also does not account for interdependence between quality of the generative model and the quality of human-generated content.

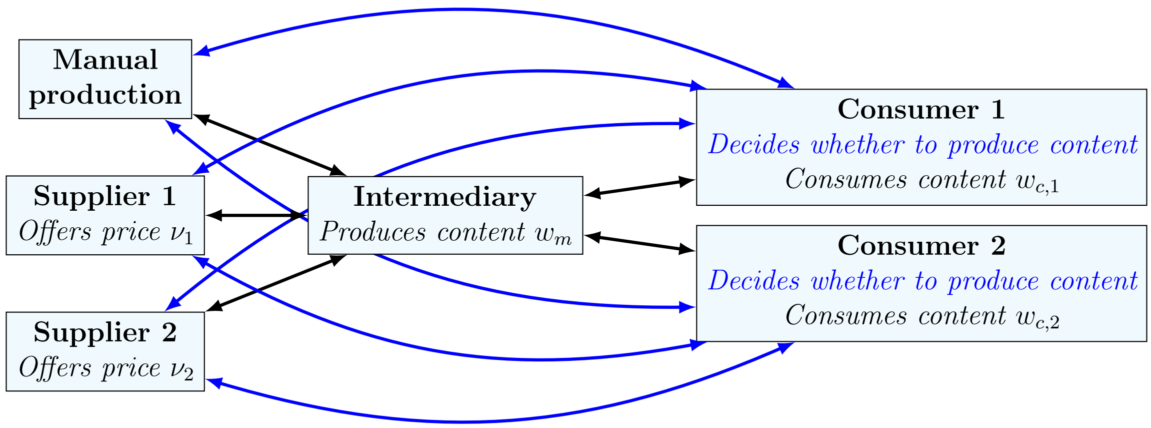


Figure 10.1: Our model for a digital content supply chain with suppliers, a intermediary, and consumers (Chapter 10.2). The supplier offers a technology to produce content, the intermediary produces content, and the consumers consume content. The suppliers also offer the technology to the consumers, so the consumers have the option to directly produce content and bypass the intermediary (the blue arrows).

10.2 Model

We develop the following model of a digital content supply chain (Figure 10.1). There are $P \geq 2$ homogeneous suppliers[4], a monopolist intermediary (a content creator), and $C \geq 2$ homogeneous consumers.[5] We embed content into a single dimension that captures content quality. We normalize our quality measure according to consumer utility: we say that content has quality $w \in \mathbb{R}_{\geq 0}$ if a consumer derives utility w from consuming it. Producing content incurs a cost that scales with its quality: we write $g(w)$ for the cost of (manually) producing content of quality w, where g is an increasing function. We will place technical conditions on cost function g in Chapter 10.3.2.

The suppliers offer a technology that assists with digital content creation. The technology automates some—though not necessarily all—aspects of content creation. Furthermore, the technology is available not only to the intermediary, but also to consumers. The supplier incurs supply-side costs for operating the technology that scale with content quality: it costs the supplier $\nu^* \cdot g(w)$ to operate the technology to assist in creating content with quality w, where $\nu^* > 0$ describes the supplier's relative marginal cost. When the technology is used for content creation, the additional human labor required to create content is captured by the

[4]See Chapter 10.5.1 for a generalization to a monopolist supplier.

[5]While the consumer population is finite, we emphasize that they are homogeneous and do not interact strategically with each other in our model. We model consumers as finite, rather than a large-market continuum, to allow an apples-to-apples comparison of production costs borne by consumers and by the intermediary.

human-driven production costs, which scale with content quality w according to $\nu^H \cdot g(w)$ where $\nu^H \geq 0$. Alternatively, content can be created manually (i.e., without the use of the technology), and these manual production costs scale with content quality w according to $\nu_0 \cdot g(w)$. It will be notationally convenient to define $\nu_0 = 1$, so that $g(w)$ is the cost of creating content manually.

Each supplier sets a price ν (which can be different from ν^*) and charges $\nu \cdot g(w)$ for operating the technology to assist in creating content with quality w. The intermediary selects which supplier to link to (or whether to create content manually), and also strategically chooses the quality of the content that they produce. Each consumer then strategically decides whether to consume the content that the intermediary has produced, paying a fee to the intermediary in order to do so, or to directly produce content themselves for their own personal consumption.

To further motivate our model, we illustrate how it provides a framework to study the impact of generative AI on content creation supply chains.

Example 7. *Consider the ongoing trend of generative AI being integrated into content creation supply chains. To capture this, let the suppliers correspond to companies which serve generative models (e.g., text-to-image models and text-to-video models) to users. Let the function $g(w)$ capture the number of tokens needed to produce content with quality w.*[6] *Let the supply-side costs $\nu^* \cdot g(w)$ capture company's inference costs from querying the model, which tend to scale with the number of generated tokens. Let the pricing structure of suppliers capture setting a price-per-token and charging users based on the number of tokens, which is a common approach used in practice.*[7] *In contrast, the marginal transaction fee α can represent either a fixed platform fee (perhaps in the form of advertising) or a market rate for the commissioned service of the creator's content.*

At a conceptual level, our model captures how generative AI models reduce the expertise needed for content creation, as reflected by how consumers and the intermediary can both leverage generative AI to assist with content creation. Nonetheless, our model also accounts the possibility of human-driven production costs which persist even in the presence of generative AI.

Stages of the Game. The game proceeds in the following stages:

1. Each supplier $i \in [P]$ chooses a price ν_i for their technology.

2. The intermediary choose a content quality w_m.[8] To produce w_m, they either choose a provider $i_m \in [P]$ whose technology to use, or they decide to create content manually

[6] We might expect that the number of tokens $g(w)$ to increase with content quality if the user needs to go back and forth over more rounds of dialogue to obtain higher-quality, or if higher-quality content requires more tokens to generate than lower-quality content.

[7] See `https://openai.com/api/pricing/`.

[8] We could have included an option for the intermediary to opt out. However, this is already implicitly captured by the intermediary producing content with quality 0.

$i_m = 0$. If $i_m = 0$, they incur a manual production cost of $\nu_0 \cdot g(w_m)$; if $i_m \in [P]$, they pay $\nu_{i_m} \cdot g(w_m)$ to supplier i and also incur a human-driven production cost of $\nu^H \cdot g(w_m)$.

3. Each consumer $j \in [C]$ chooses a mode of consumption $a_j \in \{M, D\}$.[9]
   - If they choose $a_j = M$ (the intermediary option), they pay the fee $\alpha > 0$ to the intermediary and consume $w_{c,j} := w_m$.
   - If they choose $a_j = D$ (the direct creation option), they instead choose quality of the content $w_{c,j}$ which they will produce and consume. To produce the content, they either choose a provider $i_j \in [P]$ whose technology to use, or they decide to create content manually $i_j = 0$. If $i_j = 0$, they incur a manual production cost of $\nu_0 \cdot g(w_{c,j})$. If $i_j \in [P]$, they pay $\nu_{i_j} \cdot g(w_{c,j})$ to supplier i and also incur a human-driven production cost of $\nu^H \cdot g(w_{c,j})$.

This game captures several features of digital content creation. First, observe that the intermediary incurs the same production costs regardless of how many consumers consume the content: this captures how digital content is typically free to distribute, regardless of how many consumers consume the content.[10] Moreover, observe the intermediary (the content creator) receives the same fee from consumers regardless of content quality. This captures how when content creators rely on online platforms to share their content with consumers, it is common for creators to be rewarded based on exposure. This exposure is proportional to the size of the creator's audience; our model abstracts away from inter-creator forces and content types that would determine the size of that audience. On the consumer side, this fee could capture either the subscription fees paid to the platform or disutility from being shown advertisements.

10.2.1 Utility functions

We specify the utility functions of the suppliers, intermediary, and consumers. Each supplier i derives profit equal to their revenue from usage of their technology minus supply-side costs:

$$\underbrace{1[i_m = i] \cdot (\nu_i \cdot g(w_m) - \nu^* \cdot g(w_m))}_{\text{profit from intermediary usage}} + \sum_{j=1}^{C} \underbrace{1[a_j = D] \cdot 1[i_j = i] \cdot (\nu_i \cdot g(w_{c,j}) - \nu^* \cdot g(w_{c,j}))}_{\text{profit from consumer } j\text{'s usage}}.$$

The intermediary derives utility equal to their revenue from consumer fees minus production costs:

$$\sum_{j=1}^{C} \underbrace{1[a_j = M] \cdot \alpha}_{\text{revenue from consumer } j} - \underbrace{1[i_m \in [P]] \cdot \left(\nu_{i_m} \cdot g(w_m) + \nu^H \cdot g(w_m)\right)}_{\text{costs if technology is used}} - \underbrace{1[i_m = 0] \cdot (\nu_0 \cdot g(w_m))}_{\text{manual costs}}.$$

[9] We could have also included an option for the consumer to opt out. However, the consumer would never be incentivized to opt out, since they can always choose the direct creation option D and create content $w = 0$ with quality level 0 for free.

[10] See Chapter 10.5.2 for an extension to the case of nonzero marginal costs.

Each consumer j derives utility equal to the quality of the content they consume minus fees paid to the intermediary or production costs, depending on their chosen mode of consumption:

$$\underbrace{w_{c,j}}_{\text{quality}} - \underbrace{1[a_j = M] \cdot \alpha}_{\text{intermediary fee}} - \underbrace{1[a_j = D] \cdot \left(1[i_j \in [P]] \cdot (\nu_{i_j} \cdot g(w_{c,j}) + \nu^H \cdot g(w_{c,j})) + 1[i_j = 0] \cdot (\nu_0 \cdot g(w_{c,j}))\right)}_{\text{production costs}}.$$

10.2.2 Equilibrium concept and equilibrium existence

We focus on the pure strategy *subgame perfect equilibria* in the game between suppliers, the intermediary, and users. The following result shows that a pure strategy equilibrium exists (proof deferred to Appendix G.2).

Theorem 72. *There exists a pure strategy equilibrium in the game between suppliers, the intermediary, and consumers.*

When we place additional structure on tiebreaking, we can also show a partial uniqueness result. In particular, we assume the following structure on tiebreaking: each consumer j tiebreaks in favor of the intermediary (i.e., $a_j = M$), the intermediary tiebreaks in favor of producing higher-quality content over lower-quality content, the intermediary and consumers tiebreak in favor of suppliers over manual production, the intermediary and consumers tiebreak in favor of suppliers with a lower index.

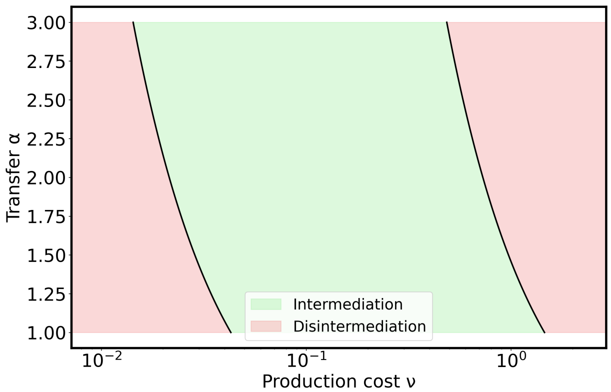
Theorem 73. *Under the tiebreaking assumptions described above, the actions of the intermediary and consumers are the same at every pure strategy equilibrium. Moreover, the production cost $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$ is the same at every pure strategy equilibrium.*

We will focus on this class of equilibria throughout our analysis in Chapters 10.3-10.4. The formal equilibrium construction is deferred to Appendix G.2, but let us briefly describe its structure. At equilibrium, all suppliers will select the same price ν_i, which (due to competitive pressure) will be equal to the supplier's marginal production cost ν^*. There are then two cases, corresponding to two types of equilibria.

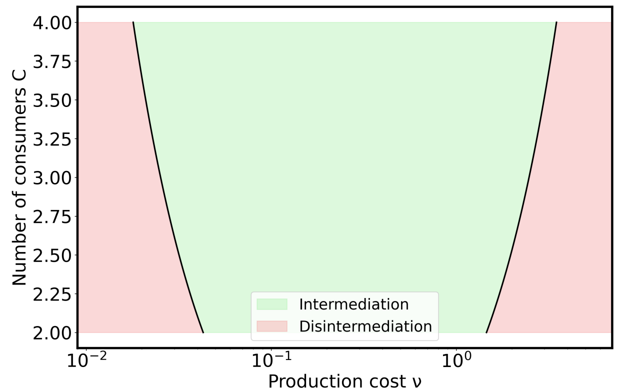
- In the disintermediation case, the intermediary chooses to produce at quality 0 (i.e., exits the market). Each of the C consumers then produces, directly and separately, their own content at a utility-maximizing level of quality, either by contracting with the minimum-index supplier or generating the content manually, whichever is cheapest.

- In the intermediation case, the intermediary produces at positive quality and all consumers choose to consume the content created by the intermediary.

For each choice of model parameters, exactly one of these two cases applies.



(a) Impact of transfer α



(b) Impact of number of consumers C

Figure 10.2: Production costs where disintermediation (red) vs. intermediation (green) occurs, for $g(w) = w^2$ (Theorem 74). We vary the transfer α (left) and number of consumers C (right). The intermediary only survives in the market when the production costs are at intermediate levels: the intermediary is driven out of the market when production costs are sufficiently low or sufficiently high. The range of values where intermediation occurs shifts lower when the fees α are higher, and the range expands when the number of consumers is larger. (We have generated these plots with a small number of consumers for ease of visualization, but our results apply for any number of consumers.)

10.3 Characterization of Disintermediation

In this section, we characterize when the disintermediation occurs: that is, when the intermediary does not survive in the market. To study this, we analyze the *intermediary usage* $\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]]$ which measures the number of consumers who consume content produced by the intermediary, at equilibrium. In Chapter 10.3.1, we characterize the intermediary usage in the special class of cost functions of the form $g(w) = w^\beta$. In Chapter 10.3.2, we extend this result to general cost functions g.

10.3.1 Special Class of Cost Functions: $g(w) = w^\beta$

To gain intuition, we first consider cost functions that are power functions of the form $g(w) = w^\beta$ for $\beta > 1$. This class of functions permits the following characterization.

Theorem 74. *Let $g(w) = w^\beta$ where $\beta > 1$. Fix $\alpha > 0$, $C > 1$. Suppose there are $P > 1$ providers. There exist thresholds $0 < T_L(C, \alpha, \beta) < T_U(C, \alpha, \beta) < \infty$ such that the* *intermediary usage at equilibrium satisfies:*

$$\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \textit{if } \min(\nu^* + \nu^H, \nu_0) < T_L(C, \alpha, \beta) \\ C & \textit{if } \min(\nu^* + \nu^H, \nu_0) \in [T_L(C, \alpha, \beta), T_U(C, \alpha, \beta)] \\ 0 & \textit{if } \min(\nu^* + \nu^H, \nu_0) > T_U(C, \alpha, \beta) \end{cases}$$

The thresholds $T_L(C, \alpha, \beta)$ and $T_U(C, \alpha, \beta)$ are the two unique solutions to:

$$\nu^{-\frac{1}{\beta(\beta-1)}} \cdot \left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right) \cdot \alpha^{-\frac{1}{\beta}} + \nu^{\frac{1}{\beta}} \cdot \alpha^{\frac{\beta-1}{\beta}} = C^{\frac{1}{\beta}}.$$

Moreover, the lower threshold $T_L(C, \alpha, \beta)$ is decreasing as a function of the number of consumers C, and the upper threshold $T_U(C, \alpha, \beta)$ is increasing as a function of C.

Theorem 74 (Figure 10.2) demonstrates that the equilibrium intermediary usage exhibits up to *three regimes of behavior* as a function of the production costs. In the first regime, which occurs when production costs are small, disintermediation occurs and consumers do not leverage the intermediary for content production. In the middle regime, intermediation instead occurs: all of the consumers rely on the intermediary for content production. In the last regime, where production costs are very large, disintermediation again occurs. Notably, the middle regime of intermediation is larger when the number of consumers C is large.

The intuition for Theorem 74 is as follows. The intermediary's main advantage is that they only need to incur the production cost once regardless of how many consumers consume their content, and yet their revenue still scales with the number of consumers that they attract. This economy of scale is more pronounced when the number of consumers is large. However, the intermediary's ability to leverage this advantage is constrained by the fact that the fee paid by consumers is fixed and might not align with production costs. This creates friction that can impede the intermediary's ability to generate utility. When costs are in the middle regime, the misalignment is outweighed by the intermediary's fixed-cost advantage: the intermediary can create much higher quality than what the consumer can afford to create for themselves, and the consumers are willing to pay the requisite fee for this additional quality.

When costs are sufficiently low, the consumers are incentivized to create even higher quality content than what the intermediary can afford with the fees they collect. When costs are sufficiently high, the fee is insufficient to cover the cost of producing quality that the consumers find acceptable for the price, so consumers are incentivized to create lower-quality content to reduce costs.

Taking a closer look at the structure of production costs, Theorem 74 also disentangles how two different axes of technological advances— reductions in supply-side costs and reductions in human-driven production costs—both affect whether disintermediation occurs. Specifically, production costs are captured by $\min(\nu^* + \nu^H, \nu_0)$. Assuming that manual production costs ν_0 exceed production costs $\nu^* + \nu^H$ from leveraging the technology, the first regime of disintermediation only occurs when technology improves to a sufficient degree along both of these axes. If supply-side costs ν^* are overly high, then even if the human-driven costs ν^H are pushed down to zero (i.e., fully automation is possible), this regime will not occur; similarly, if human-driven costs ν^H are overly high, then even if the supply-side costs ν^* are pushed down to zero (i.e., technology operation is free), then this regime will also not occur.

These findings have interesting implications for the integration of generative AI into content supply chains (Example 7). Prior to the generative AI era, the manual production costs likely placed the content creation ecosystem within the second regime, where all consumers rely on the intermediary (i.e., content creators) for production. If generative AI continues to reduce production costs, the ecosystem could move to the first regime where creators are cut out of the ecosystem. However, this shift would rely on technological advances along two axes: inference costs (which depend on the computational efficiency of querying these models), and human-driven production costs (which depends on the balance between automation vs. augmentation in content creation).

We provide a proof sketch of Theorem 74.

Proof sketch of Theorem 74. Disintermediation occurs when the intermediary can't afford to match the consumer utility from direct usage: that is, when the costs of producing content achieving that consumer utility level exceeds the intermediary's revenue from consumer usage. Lemma 234 shows that this occurs if and only if

$$(\nu^* + \nu^H) \cdot g(\alpha + \max_{w \geq 0}(w - (\nu^* + \nu^H)g(w))) > \alpha C. \tag{10.1}$$

The intuition for (10.1) is that $\max_{w \geq 0}(w - (\nu^* + \nu^H)g(w))$ is the utility that the consumer would have achieved from direct usage, so a content quality $\alpha + \max_{w \geq 0}(w - (\nu^* + \nu^H)g(w))$ is needed to match that utility level while also offsetting the fee that the consumer pays to the intermediary. The intermediary can't survive in the market if their production cost $(\nu^* + \nu^H) \cdot g(\alpha + \max_{w \geq 0}(w - (\nu^* + \nu^H)g(w)))$ exceeds their revenue αC from fees.

Using the fact that $g(w) = w^\beta$, we explicitly compute when this condition is satisfied:

$$\nu^{-\frac{1}{\beta(\beta-1)}} \cdot \left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right) \cdot \alpha^{-\frac{1}{\beta}} + \nu^{\frac{1}{\beta}} \cdot \alpha^{\frac{\beta-1}{\beta}} > C^{\frac{1}{\beta}}, \tag{10.2}$$

where $\nu = \nu^* + \nu^H$. The left-hand side of (10.2) is concave in ν (since it is the sum of two concave functions), and it approaches ∞ as $\nu \to \infty$ and as $\nu \to 0$. Moreover, the minimum value of the left-hand side across all $\nu \in (0, \infty)$ is 1 which violates (10.2). This establishes the existence of thresholds $0 < T_L(C, \alpha, \beta) < T_U(C, \alpha, \beta) < \infty$. Moreover, these properties, together with the fact that the right-hand of (10.2) is increasing in C, imply that the lower threshold is decreasing in C and the upper threshold is increasing in C. The full proof is deferred to Appendix G.3. □

10.3.2 General Cost Functions

We now move beyond the particular functional form in Chapter 10.3.1, and analyze disintermediation for more general cost functions g.

Specifically, we consider strictly increasing, continuously differentiable cost functions g which are (1) strictly convex, (2) satisfy $g(0) = g'(0) = 0$ and $\lim_{w\to\infty} g(w) = \lim_{w\to\infty} g'(w) = \infty$, and (3) strictly log-concave. Some examples of cost functions that satisfy assumptions (1)-(3) are $g(w) = w^\beta$ for $\beta > 1$, $g(w) = w^\beta \cdot e^w$ for $\beta \geq 1$, $g(w) = w^\beta \cdot e^{\sqrt{w}}$ for $\beta \geq 1$, and $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$ (Proposition 236). To elucidate the role of these assumptions, suppose that the consumer selects the direct creation option and optimally chooses their content quality level to maximally their utility. The first two assumptions imply that the consumer chooses content quality in the interior of $(0, \infty)$. Taken together with the first two assumptions, the third assumption implies that as production costs become cheaper, a consumer who directly creates their own content would expend more on content production.

Under these assumptions on the cost function g, we show a partial generalization of Theorem 74. The following result demonstrates that the intermediary usage exhibits up to three regimes of behavior as a function of the production costs.

Theorem 75. *Let g be a strictly increasing, continuously differentiable function which is strictly convex, satisfies $g(0) = g'(0) = 0$ and $\lim_{w\to\infty} g(w) = \lim_{w\to\infty} g'(w) = \infty$, and strictly log-concave. Fix $\alpha > 0$, $C > 1$. Suppose there are $P > 1$ providers. There exist thresholds $T_L(C, \alpha, g) < T_U(C, \alpha, g) \leq \infty$ such that the intermediary usage at equilibrium satisfies:*

$$\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) < T_L(C, \alpha, g) \\ C & \text{if } \min(\nu^* + \nu^H, \nu_0) \in [T_L(C, \alpha, g), T_U(C, \alpha, g)] \\ 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) > T_U(C, \alpha, g). \end{cases}$$

While Theorem 75 provides a partial generalization of Theorem 74, a key difference is that Theorem 75 only guarantees that there are *up to* three regimes rather than *exactly* three regimes. Crucially, Theorem 75 does not guarantee that disintermediation occurs when as production costs tend to zero. To help address this, we show a sufficient condition for having exactly three regimes.

Theorem 76. *Consider the setup of Theorem 75. Suppose also that*

$$\lim_{w\to\infty} \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g'(w)} = \infty.$$

Then it holds that $0 < T_L(C, \alpha, g) < T_U(C, \alpha, g) < \infty$. The lower threshold is decreasing as a function of the number of consumers C, and the upper threshold is increasing as a function of C. The thresholds are the two unique solutions to $\nu \cdot g(\alpha + \max_{w\geq 0}(w - \nu g(w))) = \alpha C$.

Theorem 76 provides a more complete generalization of Theorem 74 and guarantees that there are exactly three regimes. Specifically, disintermediation occurs when the production costs tend to zero and also occurs when production costs become sufficiently large.

The condition in Theorem 76 bears resemblance to requiring that the log derivative approaches ∞ in the limit as $w \to \infty$, which would instead take the form $\lim_{w\to\infty} \frac{g(w)}{g'(w)} = \infty$.

This simpler condition captures that the amount that the consumer expends on content production, if they directly use the technology, is unbounded in the limit as production costs become cheaper. However, the requirement in Theorem 76 is slightly stricter since the numerator is replaced with $g\left(w - \frac{g(w)}{g'(w)}\right)$. Some examples of functions g satisfying the conditions in Theorem 76 are $g(w) = w^\beta$ for $\beta > 1$, $g(w) = w^\beta \cdot e^{\sqrt{w}}$ for $\beta > 1$, and $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$ (Proposition 237).

We note that not all functions satisfy this condition. For example, functions of the form $g(w) = w^\beta \cdot e^w$ for $\beta > 1$ do not satisfy the conditions in Theorem 76, but do satisfy the conditions in Theorem 75. For this function class, this is not just an artifact of the analysis: we show that the intermediary survives even when production costs tend to zero.

Proposition 77. *Consider the setup of Theorem 75. Let $g = w^\beta \cdot e^w$ for $\beta > 1$. Fix $\alpha > 1$ and $C > 1$ satisfying $e^\alpha < \alpha \cdot C$. Then, $T_L(C, \alpha, g) \leq 0$. That is, the intermediary usage $\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = C$ even when the production costs $\min(\nu^* + \nu^H, \nu_0)$ are sufficiently small.*

10.4 Welfare Consequences

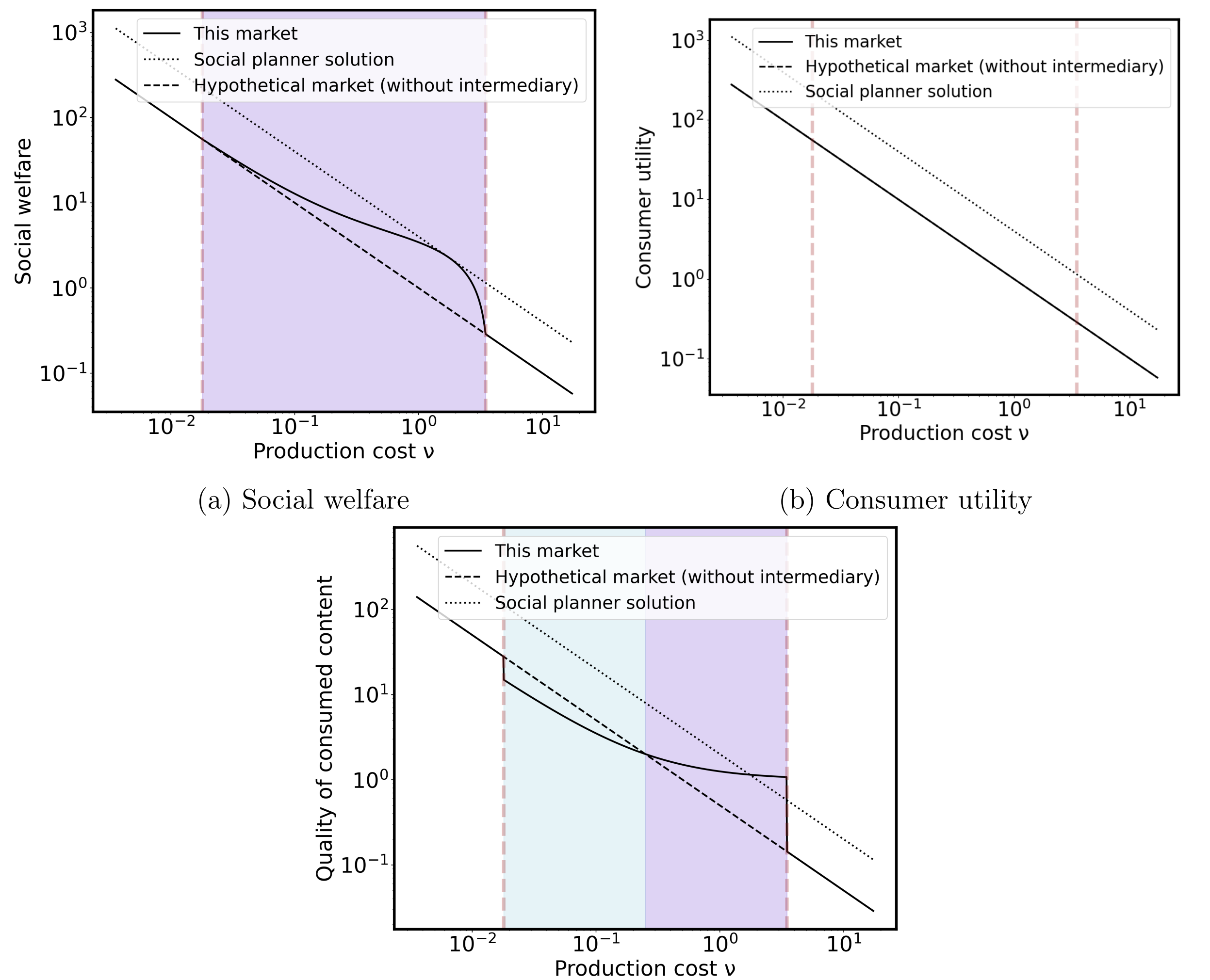
Having established when disintermediation occurs, we now turn to the consequences for social welfare and the overall digital economy. We study the impact on content quality (Chapter 10.4.1), the intermediary's utility (Chapter 10.4.2), consumer utility (Chapter 10.4.3), and social welfare (Chapter 10.4.4). We analyze how these metrics change with technology improvements, and to gain intuition for the impact of the intermediary, we also make comparisons to a hypothetical market where the intermediary does not exist (Figure 10.3). Throughout this section, we focus on the setup of Theorem 76 where disintermediation occurs exactly at the extreme values of production costs (i.e., where there are three regimes of behaviors).

10.4.1 Quality of content

We first examine how disintermediation impacts content quality. To gain intuition, we compute a closed-form characterization of the quality of the content consumed at equilibrium.

Proposition 78. *Consider the setup of Theorem 76. Let $\nu = \min(\nu^* + \nu^H, \nu_0)$. At equilibrium, the quality $w_{c,j}$ of the content consumed by any consumer $j \in [C]$ is:*

$$\begin{cases} \arg\max_{w \geq 0}(w - \nu \cdot g(w)) & \text{if } \nu < T_L(C, \alpha, g) \\ \alpha + \max_{w \geq 0}(w - \nu \cdot g(w)) & \text{if } \nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)] \\ \arg\max_{w \geq 0}(w - \nu \cdot g(w)) & \text{if } \nu > T_U(C, \alpha, g). \end{cases}$$

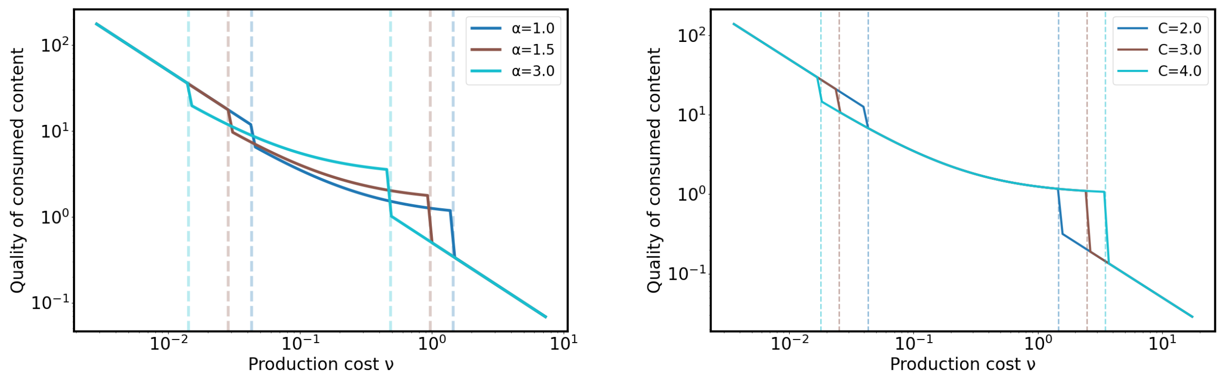


(a) Social welfare

(b) Consumer utility

(c) Content quality

Figure 10.3: Analysis of social welfare, consumer utility, and content quality in this market in comparison to a hypothetical market where the intermediary does not exist, for $g(w) = w^2$. We show how the the intermediary increases (purple), decreases (blue), or does not affect (white) each of these metrics. The intermediary is always welfare-improving (left; Theorem 86). However, the intermediary does not increase consumer utility (middle; Theorem 82), and instead extracts all of the surplus for themselves. The intermediary can increase content quality or decrease content quality (right; Theorem 79).



(a) Impact of transfer α (b) Impact of number of consumers C

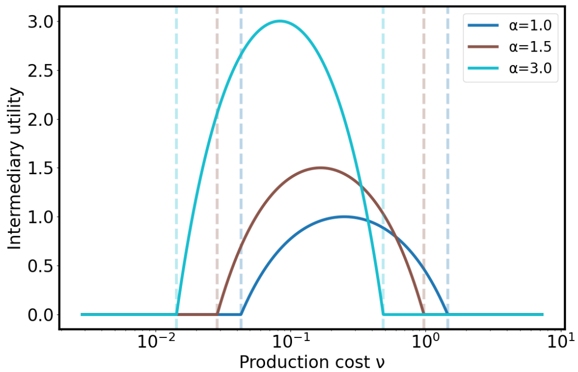
Figure 10.4: Quality of the content consumed at a pure strategy equilibrium as a function of production costs, for $g(w) = w^2$ (Proposition 78) We vary the transfers α (left), and the number of consumers C (right). The vertical dashed lines show the production costs at which disintermediation starts to occur. Observe that the quality is decreasing in production costs, and is discontinuous at the thresholds where disintermediation starts to occur (Theorem 79).

Proposition 78 demonstrates that the content quality has three regimes of behavior as a function of the production costs. These are conceptually the same three regimes as in Theorem 75. In the first and third regimes, the consumer consumes the content that they produce themselves; in the second regime, the intermediary survives in the market, and consumers consume the content that is produced by the intermediary.

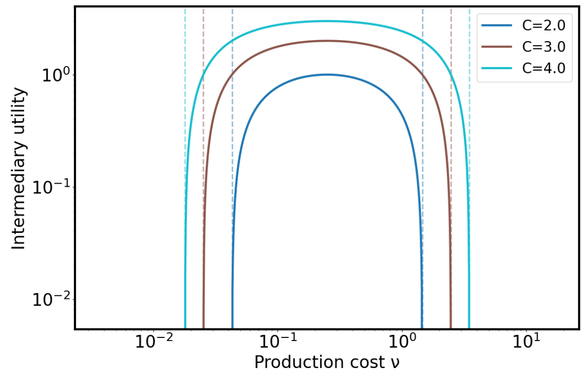
Using Proposition 78, we analyze how the content quality changes as production costs improve, and we also compare content quality to a hypothetical market where the intermediary does not exist, where the content quality would have been $\arg\max_{w \geq 0}(w - \nu \cdot g(w))$.

Theorem 79. *Consider the setup of Proposition 78. The quality of content consumed at equilibrium is decreasing in ν. Moreover, the quality is continuous in ν except for at the thresholds $T_L(C, \alpha, g)$ and $T_U(C, \alpha, g)$. The quality when $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$ can be higher or lower than $\arg\max_{w \geq 0}(w - \nu \cdot g(w))$ when ν is at the higher or lower end of the range, respectively.*

Theorem 79 (Figure 10.3c) illustrates how the intermediary distorts content quality. First, the shape of the curve qualitatively changes in the presence of the intermediary: specifically, the slope of the curve becomes flatter. This means that the intermediary reduces the responsiveness of quality to technology changes. Moreover, the intermediary can raise or lower content quality compared to a hypothetical market where the intermediary does not exist. Specifically, we see that when the production costs at the lower end of the regime where the intermediary survives, the content quality is lower than in this hypothetical market; when the production costs are at the upper end of the regime, then the content quality is



(a) Impact of transfer α



(b) Impact of number of consumers C

Figure 10.5: Intermediary utility at a pure strategy equilibrium as a function of production costs, for $g(w) = w^2$ (Proposition 80). We vary the transfers α (left), and the number of consumers C (right). The vertical dashed lines show the production costs at which disintermediation starts to occur. Observe that the intermediary utility is inverse U-shaped in production costs (Theorem 81).

higher than in this hypothetical market. A striking consequence is that disintermediation at the lower threshold *improves* content quality, even though the market no longer benefits from economies of scale from the intermediary.

Theorem 79 (Figure 10.4) also uncovers other global properties of the content quality in this market. Even though the market transitions between intermediation and disintermediation, content quality is decreasing with production costs: this illustrates how technological improvements consistently improves content quality. However, perhaps counterintuitively, increasing the number of consumers can lead to *lower* content quality for some production costs (Figure 10.4b). This comes as a side effect of intermediation, since the number of consumers impacts the range of production costs where intermediation occurs. The impact of the fees α is similarly ambiguous, since the fees also impact when intermediation occurs (Figure 10.4a).

10.4.2 Intermediary utility

We next turn to the intermediary's utility. To gain intuition, we first compute a closed-form characterization of the intermediary's utility at equilibrium.

Proposition 80. *Consider the setup of Theorem 76. Let* $\nu = \min(\nu^* + \nu^H, \nu_0)$. *The* *intermediary's utility at equilibrium is of the form:*

$$\begin{cases} 0 & \text{if } \nu < T_L(C, \alpha, g) \\ \alpha C - \nu g\left(\alpha + \max_{w \geq 0}\left(w - \nu \cdot g(w)\right)\right) & \text{if } \nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)] \\ 0 & \text{if } \nu > T_U(C, \alpha, g). \end{cases}$$

Proposition 80 lets us analyze how intermediary utility changes with production costs.

Theorem 81. *Consider the setup of Proposition 80. As a function of ν, the intermediary utility at equilibrium is continuous and inverse U-shaped. The maximum intermediary utility across all values $\nu > 0$ is equal to $\alpha(C-1)$ and occurs when $w_m = \arg\max_{w \geq 0}(w - \nu g(w))$.*

Theorem 81 (Figure 10.5) illustrates how the intermediary's utility is inverse U-shaped. This non-monotone behavior means that even though technology improvements first benefit the intermediary, the intermediary's utility later starts to fall until the intermediary is eventually driven out of the market. The intermediary's utility is maximized when production costs are in the middle of the range. At the optima, the intermediary expends one consumer's fee on content production, creating the same content that the consumer would have created if the intermediary did not exist. The intermediary retains the rest of the consumers' fees for themselves: in this sense, the intermediary extracts all the value from the economies of scale. The intermediary benefits from increasing the number of consumers (Figure 10.5b), but the impact of the fee α is ambiguous (Figure 10.5a).

10.4.3 Consumer utility

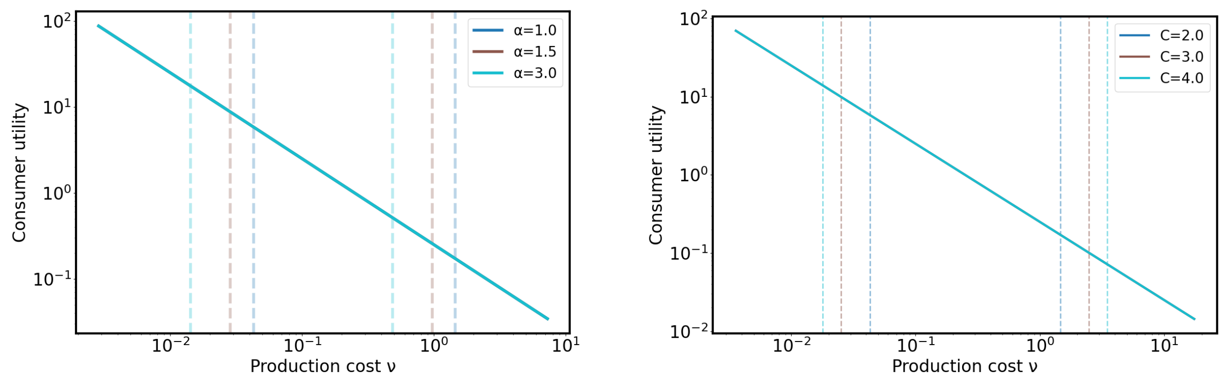
We next turn to consumer utility, which can be characterized at equilibrium in closed-form.

Theorem 82. *Consider the setup of Theorem 76. Let $\nu = \min(\nu^* + \nu^H, \nu_0)$. At equilibrium, the utility of any consumer $j \in [C]$ is equal to $\max_{w \geq 0}(w - \nu \cdot g(w))$.*

Theorem 82 (Figure 10.3b) illustrates how consumer utility is unaffected by the intermediary. Specifically, the consumer utility is the same as in a hypothetical market where the intermediary does not exist. This means that consumer utility is independent of the fees α (Figure 10.6a) as well as the number of other consumers in the market (Figure 10.6b). The intuition is that the intermediary is a monopolist, and is able to extract all of the value from the economies of scale for themselves. Interestingly, this occurs even though the intermediary can't influence the price ν: instead the intermediary extracts all of the surplus through the choice of quality produced.

Using Theorem 82, we show that consumer utility is decreasing is production costs, so consumers still do benefit from technological improvements.

Corollary 83. *Consider the setup of Theorem 82. As a function of ν, the the utility of each consumer $j \in [C]$ is continuous and decreasing.*



(a) Impact of transfer α

(b) Impact of number of consumers C

Figure 10.6: Consumer utility at a pure strategy equilibrium as a function of production costs, for $g(w) = w^2$ (Theorem 82). We vary the transfers α (left), and the number of consumers C (right). Observe that the consumer utility is continuous, decreasing in production costs, and independent of C and α (Corollary 83).

.

10.4.4 Social welfare

Finally, we turn to social welfare. We first analyze the social welfare at equilibrium in closed-form.

Proposition 84. *Consider the setup of Theorem 76. Let* $\nu = \min(\nu^* + \nu^H, \nu_0)$. *The equilibrium social welfare takes the form:*

$$\begin{cases} C \cdot \left(\max_{w \geq 0} \left(w - \nu \cdot g(w)\right)\right) & \text{if } \nu < T_L(C, \alpha, g) \\ C \cdot \left(\alpha + \max_{w \geq 0} \left(w - \nu \cdot g(w)\right)\right) - \nu g\left(\alpha + \max_{w \geq 0} \left(w - \nu \cdot g(w)\right)\right) & \text{if } \nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)] \\ C \cdot \left(\max_{w \geq 0} \left(w - \nu \cdot g(w)\right)\right) & \text{if } \nu > T_U(C, \alpha, g). \end{cases}$$

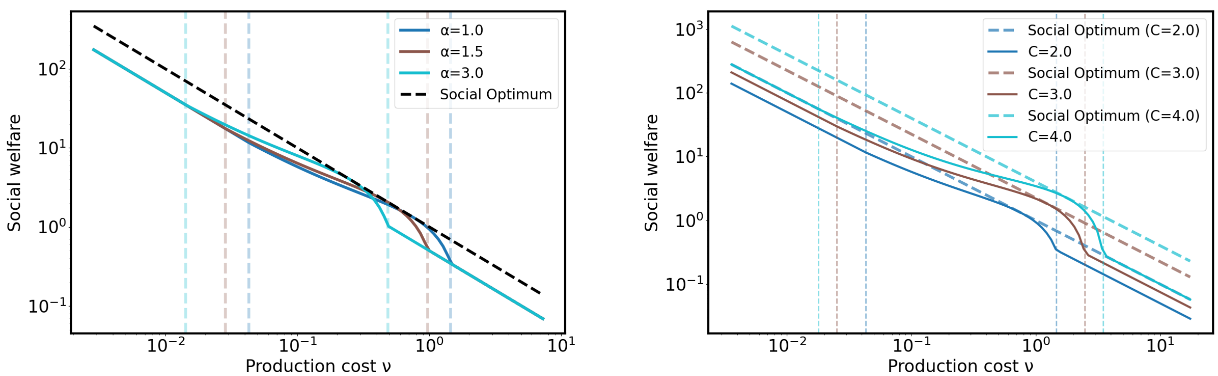
To interpret the social welfare achieved in this market, we consider a social planner whose goal is to maximize the total social welfare of the suppliers, intermediary, and consumers. We characterize the optimal social planner solution, both in the case where the intermediary exists and where the intermediary does not exist.

Proposition 85. *Let* $\nu = \min(\nu^* + \nu^H, \nu_0)$. *If the intermediary exists, then the social planner's solution achieves social welfare*

$$\max_{w \geq 0} \left(Cw - \nu \cdot g(w)\right).$$

If the intermediary does not exist, then the social planner's solution achieves social welfare

$$C \cdot \max_{w \geq 0} \left(w - \nu \cdot g(w)\right).$$



(a) Impact of transfer α (b) Impact of number of consumers C

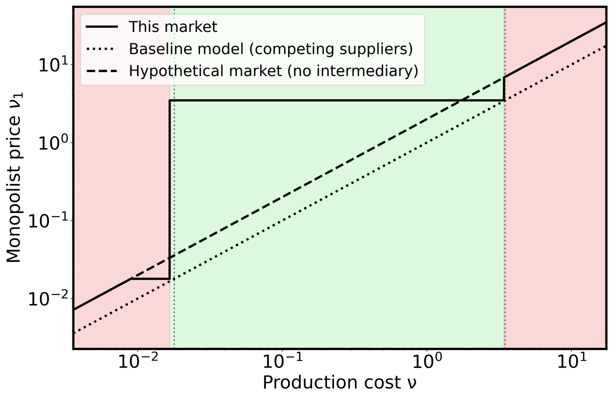
Figure 10.7: Social welfare at a pure strategy equilibrium as a function of production costs, for $g(w) = w^2$ (Proposition 84) We vary the transfers α (left), and the number of consumers C (right). The black line shows the social welfare of the optimal social planner solution. Observe that the social welfare utility is continuous, decreasing in production costs, and increasing in C (Theorem 86)

We now analyze how the social welfare achieved in the market changes with production costs and compares to the social planner's solutions.

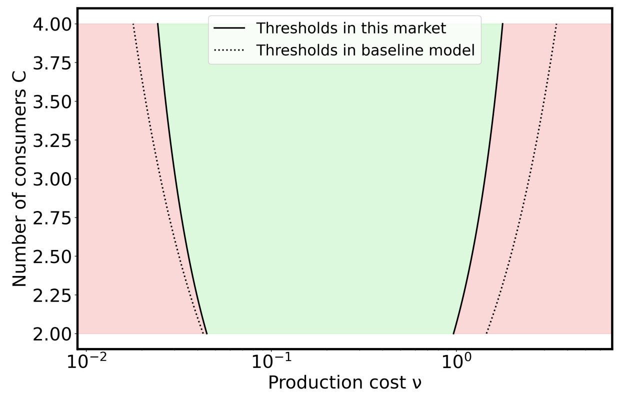
Theorem 86. *Consider the setup of Proposition 84. The social welfare is continuous and decreasing in production costs. It is strictly below the social planner's optimal except at at most one bliss point. Moreover, when $\nu \in (T_L(C, \alpha, g), T_U(C, \alpha, g))$, the social welfare is strictly greater than the social planner's optimal without the intermediary (i.e., $C \cdot (\max_{w \geq 0} (w - \nu \cdot g(w)))$).*

Theorem 86 (Figure 10.3a) shows that the intermediary is welfare-improving. Specifically, when the intermediary is present in the market, the social welfare at equilibrium is higher than the social planner solution in a hypothetical market where the intermediary does not exist. However, the social welfare almost always falls below the social planner solution which can take advantage of the intermediary, except at at most one bliss point. This bliss point always exists for costs of the form $g(w) = w^\beta$ for $\beta > 1$ (Proposition 240). The intuition is that the intermediary distorts content quality, producing too high-quality content when the production costs are above the bliss point and producing too low-quality content when the production costs are below the bliss point (Figure 10.3c). Figure 10.7 also suggests that the location of the bliss point appears to occur at higher production costs when the number of consumers is large (Figure 10.7b) and when the fees are smaller (Figure 10.7a).

Taken together with the earlier results in this section, this welfare analysis illustrates that while technology improvements lead to higher social welfare, the intermediary extracts all gains to social welfare. Specifically, any increase in social welfare over the hypothetical



(a) Monopolist supplier



(b) Nonzero marginal costs

Figure 10.8: Production costs where disintermediation (red) vs. intermediation (green) for $g(w) = w^2$. We consider extensions of the baseline model with a monopolist supplier (left; Theorem 87) and with nonzero marginal costs of production (right; Theorem 89). In both cases, disintermediation still occurs when production costs are sufficiently low or sufficiently high. However, relative to our baseline model, the range of technology levels that support intermediation changes: the range shifts to be lower with a monopolist supplier (though by a small amount) and shrinks in width with nonzero marginal costs.

market without the intermediary is captured by the monopolist intermediary themselves. Interestingly, this occurs even though the intermediary controls only the quality w_m, not the price ν.

10.5 Extensions

To check the robustness of our findings we consider extensions to our base model, focusing on the case $g(w) = w^\beta$ from Chapter 10.3.1 for simplicity. We find that our characterization of disintermediation from Chapter 10.3 readily generalizes to settings with a monopolist supplier (Figure 10.8a; Chapter 10.5.1) and where the intermediary faces nonzero marginal costs of production to serve each consumer (Figure 10.8b; Chapter 10.5.2): that is, disintermediation still occurs occurs at the extremes of production technology. However, we also show that disintermediation can be avoided entirely when users pay a fee that increases linearly with the quality of the content consumed: this demonstrates the importance of our modeling assumption in Chapter 10.3 that the intermediary does not have full control the fee structure offered to consumers in addition to content quality (Chapter 10.5.3).

10.5.1 Monopolist supplier

While Chapter 10.3 assumed competition between multiple suppliers, we now turn to the case of a single monopolist supplier. For simplicity, we focus on the case where manual production costs ν_0 are infinite, meaning that it is cheaper to produce content using the technology rather than without the technology. The following result characterizes when disintermediation occurs in this setting.

Theorem 87. *Let $g(w) = w^\beta$ where $\beta > 1$.*[11] *Fix $\nu_0 = \infty$, $\alpha > 0$, and assume that $C > \frac{\beta}{\beta-1}$.*[12] *Suppose there is a monopolist supplier (i.e., $P = 1$). There exist thresholds $0 < T_L^{mon}(C,\alpha,\beta) < T_U^{mon}(C,\alpha,\beta) < \infty$ such that the intermediary usage at equilibrium satisfies:*

$$\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) \leq T_L^{mon}(C,\alpha,\beta) \\ C & \text{if } \min(\nu^* + \nu^H, \nu_0) \in (T_L^{mon}(C,\alpha,\beta), T_U^{mon}(C,\alpha,\beta)] \\ 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) > T_U^{mon}(C,\alpha,\beta). \end{cases}$$

In comparison to the thresholds from Theorem 74, these thresholds satisfy $\beta^{-1} \cdot T_L(C,\alpha,\beta) < T_L^{mon}(C,\alpha,\beta) < T_L(C,\alpha,\beta)$ and $\beta^{-1} \cdot T_U(C,\alpha,\beta) < T_U^{mon}(C,\alpha,g) < T_U(C,\alpha,\beta)$.

Theorem 87 (Figure 10.8a) shows that the insights from Theorem 74 readily generalize to the case of a monopolist supplier, albeit with the intermediation range shifted to lower production costs. Disintermediation still occurs whenever production costs are sufficiently low or sufficiently high. However, the upper and lower thresholds for intermediation in Theorem 87 occur at lower production costs than the corresponding thresholds with competing suppliers. The intuition is that suppliers set prices above the marginal production costs, so the consumers and intermediary face higher prices, which shifts the thresholds downwards. The intermediary can more easily survive when technology costs are lower, but is less likely to enter when technology costs are high.

To prove Theorem 89, a key step is to analyze how the monopolist supplier sets prices. Unlike for competing suppliers, the price ν_1 is no longer driven down to the marginal production cost ν^*. The following lemma characterizes the optimal pricing decisions of the supplier.

[11] For this result we assume that each consumer j tiebreaks in favor of direct usage (i.e., $a_j = D$) rather than in favor of the intermediary (i.e., $a_j = M$) when $\nu < T_U(C,\alpha,\beta)$, but in favor of the intermediary when $\nu \geq T_U(C,\alpha,\beta)$.

[12] We assume that the number of consumers is sufficiently large (i.e., $C > \frac{\beta}{\beta-1}$) for technical convenience.

Lemma 88. *Consider the setup of Theorem 87, and let T_L and T_U be defined according to Theorem 74. Then the supplier's price ν_1 satisfies:*

$$\nu_1 = \begin{cases} \beta \cdot \nu^* & \text{if } \nu^* < \beta^{-1} \cdot T_L(C, \alpha, \beta) \\ T_L(C, \alpha, \beta) & \text{if } \nu^* \geq \beta^{-1} \cdot T_L(C, \alpha, \beta) \text{ and } \nu^* \leq T_L^{mon}(C, \alpha, \beta), \\ T_U(C, \alpha, \beta) & \text{if } \nu^* \in (T_L^{mon}(C, \alpha, \beta), T_U^{mon}(C, \alpha, \beta)] \\ \beta \cdot \nu^* & \text{if } \nu^* > T_U^{mon}(C, \alpha, \beta). \end{cases}$$

To interpret Lemma 88 (Figure 10.8a), consider a hypothetical market where the intermediary does not exist. In this hypothetical the monopolist provider would set $\nu_1 = \nu^* \cdot \beta$ (Lemma 243). By Lemma 88, when production costs are sufficiently low (or high), the monopolist sets the prices exactly as they would if the intermediary did not exist. However, for intermediate production costs, the monopolist supplier distorts prices to influence whether the intermediary survives in the market or not. To see why, first consider production costs at the lower end of this range, near (but higher than) the lower threshold $T_L(C, \alpha, \beta)$ (where intermediation would start to occur if the price were equal to marginal production costs ν^*). In this case, the monopolist supplier holds prices at $T_L(C, \alpha, \beta)$ in order to *avoid intermediation*. That is, the supplier suppresses their price to make direct usage by consumers more attractive and prevent the intermediary from entering the market. But if we grow the supplier's marginal production costs, these costs become sufficiently close to the supplier's price $T_L(C, \alpha, \beta)$ that the supplier's profit becomes too low. At this point, the supplier allows the intermediary to enter and discontinuously shifts to the maximal price $T_U(C, \alpha, \beta)$ that keeps the intermediary in the market. If we continue to increase costs, eventually the supplier's marginal production costs become sufficiently close to the price $T_U(C, \alpha, \beta)$, at which point they drive the intermediary out of the market and once again set prices as they would if the intermediary did not exist.

10.5.2 Marginal costs

In Chapter 10.3, we assumed that the intermediary faces no marginal costs for distributing content to additional consumers. We now relax this assumption and consider scenarios where the intermediary not only pays a fixed cost of $\nu \cdot g(w)$ to produce content w, but also pays a small additional cost of $\gamma \cdot \nu \cdot g(w)$ for every consumer to whom they serve the content. Here, we assume that $\gamma < 1$. The consumer likewise faces the same cost structure: they pay a total cost of $\nu(1+\gamma) \cdot g(w)$ to produce content w for themselves. The following result characterizes when disintermediation occurs.

Theorem 89. *Let $g(w) = w^\beta$ where $\beta > 1$. Fix $C > 1$, and fix $\gamma < 1$. Suppose there are $P > 1$ providers. Let $C' = \frac{C(1+\gamma)}{1+\gamma \cdot C}$. There exist thresholds $0 < T_L^{marg}(C, \alpha, \beta, \gamma) < T_U(C, \alpha, \beta, \gamma) < \infty$*

such that the intermediary usage at equilibrium satisfies:

$$\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) < T_L^{marg}(C, \alpha, \beta, \gamma) \\ C & \text{if } \min(\nu^* + \nu^H, \nu_0) \in [T_L^{marg}(C, \alpha, \beta, \gamma), T_U^{marg}(C, \alpha, \beta, \gamma)] \\ 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) > T_U^{marg}(C, \alpha, \beta, \gamma) \end{cases}$$

In fact, the thresholds are related to the thresholds from Theorem 74 as follows: $T_L^{marg}(C, \alpha, \beta, \gamma) = (1+\gamma)^{-1} \cdot T_L(C', \alpha, \beta)$ *and* $T_U^{marg}(C, \alpha, \beta, \gamma) = (1+\gamma)^{-1} \cdot T_U(C', \alpha, \beta)$.

Theorem 89 (Figure 10.8b) shows that the insights from Theorem 74 generalize to the case where the intermediary faces marginal costs, albeit with the intermediation range reduced in width. Intuitively, Theorem 89 captures how the market with marginal costs behaves the same as a market without marginal costs but with a smaller "effective" number of consumers $C' = \frac{C(1+\gamma)}{1+\gamma C} < C$ and also with a multiplicative reduction factor $(1+\gamma)^{-1}$. The effective number of consumers, which is decreasing in marginal costs, captures the extent to which the intermediary still enjoys economies of scale given that they face marginal costs for distribution; the multiplicative reduction factor captures how consumers also face higher costs of production relative to our baseline model. Given our prior finding that the range of thresholds supporting the intermediary shrinks in width as the number of consumers decreases in our baseline model (Theorem 74), this implies that marginal costs lead the intermediary to be supported on a more narrow range of technology levels.

10.5.3 Other fee structures

A key assumption in our baseline model is that the intermediary has no control over the fee structure: their marginal fee is α regardless of production costs. We verify the importance of this assumption by considering a model where the intermediary charges a fee that increases linearly with the quality of the content consumed. Specifically, instead of the consumer paying a fixed fee α to the intermediary, the consumer pays a *linear fee* $\alpha \cdot w_m$, which scales with the quality w_m of the content that the intermediary produces. In this setup, the value α captures the *fee scaling* rather than the fee itself. The following result characterizes when disintermediation occurs.

Theorem 90. *Let* $g(w) = w^\beta$ *where* $\beta > 1$*. Fix* $C > 1$ *and* $\alpha \in (0, 1)$*. Suppose there are* $P > 1$ *providers, and suppose that fees are linear. Then the intermediary usage* $\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]]$ *is independent of the production cost parameters* ν^*, ν^H, ν_0*. If the number of consumers* C *is sufficiently high, the intermediary usage at equilibrium is* $\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = C$*; if the number of consumers is sufficiently low, the intermediary usage is* $\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = 0$.

Theorem 89 shows that linear fees fundamentally change the nature of disintermediation: the intermediary usage at equilibrium is *independent* of the production costs $\nu = \min(\nu^* + \nu^H, \nu_0)$. This means that the intermediary always survives in the market when the number of consumers is sufficiently high, and never survives in the market when the number of consumers is sufficiently low. This result highlights the importance of our modeling assumption that the intermediary does not have full control over both the content quality and fee structure. We conclude that the this assumption—which is motivated by common practices in digital content recommendation (Example 7)—has a substantial impact on the market's sensitivity to technology improvements.[13]

10.6 Discussion

In this chapter, we investigate the relationship between production technology improvements and disintermediation. We focus on markets where the technology is available to both the intermediary and consumers, and where the intermediary's strategic choice is restricted to the level of production quality. We find that reduced production costs eventually drive the intermediary out of the market entirely. We also show that even at production cost levels where the intermediary does survive, the threat of disintermediation leads to striking implications for welfare and content quality. While the intermediary is welfare-improving, the intermediary extracts all gains to social welfare for themselves. Furthermore, the intermediary's utility is inverse U-shaped in production costs, and the presence of the intermediary can raise or lower content quality.

Our model and results open the door to several interesting avenues for future work. While our model focuses on a a single creator and a target audience of homogeneous consumers, it would be interesting to endogenize the audience-formation process and investigate differential impacts on different types of consumers and creators. For instance, one could consider multiple intermediaries who can differentiate horizontally as well as vertically. These could compete for heterogeneous consumers who might differ in horizontal taste, sensitivity to quality, and/or proclivity for niche content. In such an environment, which types of creators face the heaviest threat of disintermediation? Which types of consumers are better off in the world with new and improved production tools, and which (if any) are disadvantaged? How does the *type* of content available in the market vary as production technology improves?

[13]Specifically, digital content distribution platforms usually typically do not allow individual creators to individually set prices for their content. That being said, some platforms do reward creators based on engagement: when engagement is correlated with content quality (rather than other metrics, such as the length of videos), linear fees would capture online platforms that reward creators based on engagement rather than exposure.

Part IV

Repeated Human-AI Interactions

Chapter 11

Overview

When ML models are deployed into dynamic environments, these models often repeatedly interact with humans. For example, recommendation platforms such as Uber repeatedly match customers with service providers and learn about user preferences over time; recommendation platforms such as Google Maps repeatedly suggest routes to populations of pedestrians and drivers and learn about real-time traffic. As another example, deployed LLM agents repeatedly interact with humans over a chat session, and more recently, autonomous LLM agents have started interacting with the broader world by querying APIs.

When evaluating human-AI interactions, it is common to view agents as collaborating towards a shared goal (e.g., (Bansal et al., 2021)); however, these agents often have *competing preferences*. In some cases, the model-provider's preferences may be misaligned from human preferences: for example, recommendation platforms typically optimize for profit or aggregate user retention, whereas individual users may prefer outcomes that prioritize their preferences even at the cost of other users. When ML models act more autonomously, the ML model's learned preferences may themselves be misaligned from human preferences: for example, an LLM agent's learned preferences may fail to capture safety constraints (e.g., LLMs are susceptible to jailbreaks), or may reflect societal rather than individual preferences.

11.1 Our contributions

This part investigates the interplay between competing preferences and learning over repeated interactions. We develop evaluation metrics that account for competing preferences and are also achievable by learning algorithms. We then design learning algorithms that perform near-optimally against these evaluation metrics. We focus on two-sided matching platforms with monetary transfers, model-providers triggering distribution shifts, and decentralized two-agent systems, as detailed below:

- In Chapter 12, we show how a matching platform can simultaneously learn the preferences of customers and service-providers and also incentivize these users to stay on the platform. We develop an incentive-aware evaluation metric that captures the distance of a market outcome from equilibrium, and we design near-optimal algorithms with respect to cumulative performance against this evaluation metric.

- In Chapter 13, we illustrate how the model-provider can efficiently learn to steer the population of users towards distributions which are easier to predict. We evaluate performance with respect to the equilibrium loss, and we design learning algorithms where the excess loss over repeated interactions scales with the complexity of the distribution shifts.
- In Chapter 14, we investigate how repeated interactions between a human and AI agent impact the utilities of both agents. First, we show that the full-information equilibrium is fundamentally unachievable as a benchmark, since small errors can distort agent utilities due to competing preferences. We thus construct relaxed error-tolerant benchmarks, and design classes of decentralized learning algorithms which achieve these benchmarks.

11.2 Technical theme

In this part, a common technical theme is that we model repeated interactions by building on the *stochastic multi-armed bandits framework* (Lattimore and Szepesvári, 2020), but adapt this framework to capture the specifics of human-AI interactions. In all of these works, we build on standard bandit algorithms—specifically, ExploreThenCommit and UCB (Auer et al., 2002a). We carefully modify these algorithms to account for how different agents have competing preferences and to capture the learner's level of information about the actions of other agents. However, the specific approach that we take to instantiate each ecosystem within the bandits framework and to design learning algorithms varies across chapters. For example, we adjust the size of the confidence sets in UCB (Chapters 13 and 14), carefully design how the learner exploits its estimates of its preferences (Chapter 12 and 13), and use the learner's exploration process to steer the learning process of the other agent (Chapter 14).

More broadly, our work contributes to the rich literature on learning equilibria in games, as we outline in more detail in each of the chapters.

11.3 Other co-authored work

In other co-authored work which is not included in this thesis, we also further investigate repeated human-AI interactions. In Pan et al. (2024) (led by Alex Pan), we empirically deploy an LLM agent in simulated environments where the agent interacts with the external world: we show that these interactions enable the agent to optimize its learned preferences, but safety violations increase as a side effect. In Arunachaleswaran et al. (2025) (led by Natalie Collina), we investigate pricing games with many AI agents and humans, characterizing when algorithmic collusion persists in these mixed ecosystems.

Chapter 12

Two-Sided Matching Platforms

This chapter is based on *"Learning Equilibria in Matching Markets from Bandit Feedback"* (Jagadeesan et al., 2023d), which is joint work with Alex Wei, Yixin Wang, Michael I. Jordan, and Jacob Steinhardt.

12.1 Introduction

Data-driven marketplaces face the simultaneous challenges of learning agent preferences and aligning market outcomes with the incentives induced by these preferences. Consider, for instance, online platforms that match two sides of a market to each other (e.g., Lyft, TaskRabbit, and Airbnb). On these platforms, customers are matched to service providers and pay for the service they receive. If agents on either side are not offered desirable matches at fair prices, they would have an incentive to leave the platform and switch to a competing platform. Agent preferences, however, are often unknown to the platform and must be learned. When faced with uncertainty about agent preferences (and thus incentives), *when can a marketplace efficiently explore and learn market outcomes that align with agent incentives?*

We center our investigation around a model called *matching with transferable utilities*, proposed by Shapley and Shubik (Shapley and Shubik, 1971). In this model, there is a two-sided market of customers and service providers. Each customer has a utility that they derive from being matched to a given provider and vice versa. The platform selects a matching between the two sides and assigns a monetary transfer between each pair of matched agents. Transfers are a salient feature of most real-world matching markets: riders pay drivers on Lyft, clients pay freelancers on TaskRabbit, and guests pay hosts on Airbnb. An agent's net utility is their value for being matched to their partner plus the value of their transfer (either of which can be negative in the cases of costs and payments). In matching markets, the notion of *stability* captures alignment of a market outcome with agent incentives. Informally, a market outcome is *stable* if no pair of agents would rather match with each other than abide by the market outcome, and stable matchings can be computed when preferences are fully known.

In the context of large-scale matching platforms, however, the assumption that preferences are known breaks down. Platforms usually cannot have users report their complete preference profiles. Moreover, users may not even be aware of what their own preferences are. For example, a freelancer may not exactly know what types of projects they prefer until actually trying out specific ones. In reality, a data-driven platform is more likely to learn information about preferences from repeated feedback[1] over time. Two questions now emerge: In such marketplaces, how can stable matchings be learned? And what underlying structural assumptions are necessary for efficient learning to be possible?

To address these questions, we propose and investigate a model for learning stable matchings from noisy feedback. We model the platform's learning problem using stochastic multi-armed bandits, which lets us leverage the extensive body of work in the bandit literature to analyze the data efficiency of learning (see Lattimore and Szepesvári (2020) for a textbook treatment). More specifically, our three main contributions are: (i) We develop an incentive-aware learning objective—Subset Instability—that captures the distance of a market outcome from equilibrium. (ii) Using Subset Instability as a measure of regret, we show that any "UCB-based" algorithm from the classical bandit literature can be adapted to this incentive-aware setting. (iii) We instantiate this idea for several families of preference structures to design efficient algorithms for incentive-aware learning. This helps elucidate how preference structure affects the complexity of learning stable matchings.

Designing the learning objective. Since mistakes are inevitable while exploring and learning, achieving exact stability at every time step is an unattainable goal. To address this issue, we lean on approximation, focusing on learning market outcomes that are *approximately* stable. Thus, we need a metric that captures the distance of a market outcome from equilibrium.[2]

We introduce a notion for approximate stability that we call *Subset Instability*. Specifically, we define the Subset Instability of a market outcome to be the maximum difference, over all subsets $\mathcal{S}$ of agents, between the total utility of the maximum weight matching on $\mathcal{S}$ and the total utility of $\mathcal{S}$ under the market outcome.[3] We show that Subset Instability can be interpreted as the amount the platform would have to *subsidize* participants to keep them on the platform and make the resulting matching stable. We can also interpret Subset Instability as the platform's cost of learning when facing competing platforms with greater knowledge of user preferences. Finally, we show that Subset Instability is the maximum gain in utility that a coalition of agents could have derived from an alternate matching such that no agent in the coalition is worse off.

[1]Feedback might arise from explicit sources (e.g., riders rating drivers after a Lyft ride) or implicit sources (e.g., engagement metrics on an app); in either case, feedback is likely to be sparse and noisy.

[2]Previous work Das and Kamenica (2005); Liu et al. (2020a) has investigated utility difference (i.e. the difference between the total utility achieved by the selected matching and the utility achieved by a stable matching) as a measure of regret. However, this does not capture distance from equilibrium in matching markets with monetary transfers (see Chapter 12.4) or without monetary transfers (see Chapter 12.6.3).

[3]This formulation is inspired by the *strong ε-core* of Shapley and Shubik (1966).

Subset Instability also satisfies the following properties, which make it suitable for learning: (i) Subset Instability is equal to zero if and only if the market outcome is (exactly) stable; (ii) Subset Instability is robust to small perturbations to the utility functions of individual agents, which is essential for learning with noisy feedback; and (iii) Subset Instability upper bounds the utility difference of a market outcome from the socially optimal market outcome.

Designing algorithms for learning a stable matching. Using Subset Instability, we investigate the problem of learning a stable market outcome from noisy user feedback using the stochastic contextual bandit model (see, e.g., (Lattimore and Szepesvári, 2020)). In each round, the platform selects a market outcome (i.e., a matching along with transfers), with the goal of minimizing cumulative instability.

We develop a general approach for designing bandit algorithms within our framework. Our approach is based on a primal-dual formulation of matching with transfers Shapley and Shubik (1971), in which the primal variables correspond to the matching and the dual variables can be used to set the transfers. We find that "optimism in the face of uncertainty," the principle underlying many UCB-style bandit algorithms Auer et al. (2002a); Lattimore and Szepesvári (2020), can be adapted to this primal-dual setting. The resulting algorithm is simple: maintain upper confidence bounds on the agent utilities and compute, in each round, an optimal primal-dual pair in terms of these upper confidence bounds. The crux of the analysis is the following lemma, which bounds instability by the gap between the upper confidence bound and true utilities:

Lemma 91 (Informal, see Chapter 101 for a formal statement)**.** *Given confidence sets for each utility value such that each confidence set contains the true utility, let (X, τ) be a stable matching with transfers with respect to the utility functions given by the upper confidence bounds. The instability of (X, τ) is upper bounded by the sum of the sizes of the confidence sets of pairs in X.*

We can thus analyze our algorithms by combining Lemma 91 with the analyses of existing UCB-style algorithms. In particular, we can essentially inherit the bounds on the size of the confidence bounds from traditional analyses of multi-arm bandits.

Complexity of learning a stable matching. Our main technical result is a collection of regret bounds for different structural assumptions on agent preferences. These bounds resemble the classical stochastic multi-armed bandits bounds when rewards have related structural assumptions. We summarize these regret bounds in Table 12.1 and elaborate on them in more detail below.

Theorem 92 (Unstructured Preferences, Informal)**.** *For unstructured preferences, there exists a UCB-style algorithm that incurs $\widetilde{O}(N\sqrt{nT})$ regret according to Subset Instability after T rounds, where N is the number of agents on the platform and n is the number of agents that arrive in any round. (This bound is optimal up to logarithmic factors.)*

Theorem 93 (Typed Preferences, Informal)**.** *Consider preferences such that each agent a has a type $c_a \in \mathcal{C}$ and the utility of a when matched to another agent a' is given by a function*

	Regret bound
Unstructured preferences	$\widetilde{O}\big(N\sqrt{nT}\big)$
Typed preferences	$\widetilde{O}\big(\lvert\mathcal{C}\rvert\sqrt{nT}\big)$
Separable linear preferences	$\widetilde{O}\big(d\sqrt{N}\sqrt{nT}\big)$

Table 12.1: Regret bounds for different preference structures when there are N agents on the platform and no more than n agents arriving in each round.

of the types c_a and $c_{a'}$. There exists a UCB-style algorithm that incurs $\widetilde{O}(\lvert\mathcal{C}\rvert\sqrt{nT})$ regret according to Subset Instability after T rounds, where n is the maximum number of agents that arrive to the platform in any round.

Theorem 94 (Separable Linear Preferences, Informal)**.** *Consider preferences such that the utility of an agent a when matched to another agent a' is $\langle \phi(a), c_{a'} \rangle$, where $\phi(a) \in \mathbb{R}^d$ is unknown and $c_{a'} \in \mathbb{R}^d$ is known. There exists a UCB-style algorithm that incurs $\widetilde{O}(d\sqrt{N}\sqrt{nT})$ regret according to Subset Instability after T rounds, where N is the number of agents on the platform and n is the maximum number of agents that arrive in any round.*

These results elucidate the role of preference structure on the complexity of learning a stable matching. Our regret bounds scale with $N\sqrt{nT}$ for unstructured preferences (Chapter 92), $\lvert\mathcal{C}\rvert\sqrt{nT}$ for typed preferences (Chapter 93), and $d\sqrt{N}\sqrt{nT}$ for linear preferences (Chapter 94). To illustrate these differences in a simple setting, let's consider the case where all of the agents show up every round, so $n = N$. In this case, our regret bound for unstructured preferences is superlinear in N; in fact, this dependence on N is *necessary* as we demonstrate via a lower bound (see Chapter 102). On the other hand, the complexity of learning a stable matching changes substantially with preference structure assumptions. In particular, our regret bounds are sublinear / linear in N for typed preferences and separable linear preferences. This means that in large markets, a centralized platform can efficiently learn a stable matching with these preference structure assumptions.

Connections and extensions.

Key to our results and extensions is the primal-dual characterization of equilibria in matching markets with transfers. Specifically, equilibria are described by a linear program whose primal form maximizes total utility over matchings and whose dual variables correspond to transfers. This linear program inspires our definition of Subset Instability, connects Subset Instability to platform profit (see Chapter 12.6.2), and relates learning with Subset Instability to regret minimization in combinatorial bandits (see Chapter 12.5.4). We adapt ideas from combinatorial bandits to additionally obtain $O(\log T)$ instance-dependent regret bounds (see Chapter 12.6.1).

Our approach also offers a new perspective on learning stable matchings in markets with *non-transferable* utilities (Das and Kamenica, 2005; Liu et al., 2020a). Although this setting does not admit a linear program formulation, we show Subset Instability can be extended to what we call NTU Subset Instability (see Chapter 12.6.3), which turns out to have several advantages over the instability measures studied in previous work. Our algorithmic principles extend to NTU Subset Instability: we prove regret bounds commensurate with those for markets with transferable utilities.

12.1.1 Related work

In the machine learning literature, starting with Das and Kamenica (2005) and Liu et al. (2020a), several works Das and Kamenica (2005); Liu et al. (2020a); Sankararaman et al. (2021); Liu et al. (2021); Cen and Shah (2022); Basu et al. (2021) study learning stable matchings from bandit feedback in the Gale-Shapley stable marriage model Gale and Shapley (1962). A major difference between this setting and ours is the absence of monetary transfers between agents. These works focus on the *utility difference* rather than the instability measure that we consider. Cen and Shah (2022) extend this bandits model to incorporate fixed, predetermined cost/transfer rules. However, they do not allow the platform to set arbitrary transfers between agents. Moreover, they also consider a weaker notion of stability that does not consider agents negotiating arbitrary transfers: defecting agents must set their transfers according to a fixed, predetermined structure. In contrast, we follow the classical definition of stability Shapley and Shubik (1971).

Outside of the machine learning literature, several papers also consider the complexity of finding stable matchings in other feedback and cost models, e.g., communication complexity Gonczarowski et al. (2019); Ashlagi et al. (2020); Shi (2020) and query complexity Emamjomeh-Zadeh et al. (2020); Ashlagi et al. (2020). Of these works, Shi (2020), which studies the communication complexity of finding approximately stable matchings with transferable utilities, is perhaps most similar to ours. This work assumes agents know their preferences and focuses on the communication bottleneck, whereas we study the costs associated with learning preferences. Moreover, the approximate stability notion in Shi (2020) is the maximum unhappiness of any *pair* of agents, whereas Subset Instability is equivalent to the maximum unhappiness over any *subset* of agents. For learning stable matchings, Subset Instability has the advantages of being more fine-grained and having a primal view that motivates a clean UCB-based algorithm.

Our notion of instability connects to historical works in coalitional game theory: related are the concepts of the strong-ε core of Shapley and Shubik (1966) and the indirect function of MartÍnez-Legaz (1996), although each was introduced in a very different context than ours. Nonetheless, they reinforce the fact that our instability notion is a very natural one to consider.

A complementary line of work in economics Liu et al. (2014); Bikhchandani (2017); Alston (2020); Liu (2020) considers stable matchings under incomplete information. These works focus on defining stability when the agents have incomplete information about their own preferences, whereas we focus on the platform's problem of learning stable matchings from noisy feedback. As a result, these works relax the definition of stability to account for uncertainty in the preferences of agents, rather than the uncertainty experienced by the platform from noisy feedback.

Multi-armed bandits have also been applied to learning in other economic contexts. For example, learning a socially optimal matching (without learning transfers) is a standard application of combinatorial bandits Cesa-Bianchi and Lugosi (2012); Gai et al. (2012); Chen et al. (2013); Combes et al. (2015); Kveton et al. (2015). Other applications at the interface of bandit methodology and economics include dynamic pricing Rothschild (1974); Kleinberg and Leighton (2003); Badanidiyuru et al. (2018), incentivizing exploration Frazier et al. (2014); Mansour et al. (2015), learning under competition Aridor et al. (2025), and learning in matching markets without incentives Johari et al. (2016).

Finally, primal-dual methods have also been applied to other problems in the bandits literature (e.g., Immorlica et al. (2022); Tirinzoni et al. (2020); Li et al. (2021)).

12.2 Preliminaries

The foundation of our framework is the *matching with transfers* model of Shapley and Shubik (1971). In this section, we introduce this model along with the concept of stable matching.

12.2.1 Matching with transferable utilities

Consider a two-sided market that consists of a finite set $\mathcal{I}$ of customers on one side and a finite set $\mathcal{J}$ of providers on the other. Let $\mathcal{A} := \mathcal{I} \cup \mathcal{J}$ be the set of all agents. A *matching* $X \subseteq \mathcal{I} \times \mathcal{J}$ is a set of pairs (i, j) that are pairwise disjoint, representing the pairs of agents that are matched. Let $\mathscr{X}_\mathcal{A}$ denote the set of all matchings on $\mathcal{A}$. For notational convenience, we define for each matching $X \in \mathscr{X}_\mathcal{A}$ an equivalent functional representation $\mu_X : \mathcal{A} \to \mathcal{A}$, where $\mu_X(i) = j$ and $\mu_X(j) = i$ for all matched pairs $(i, j) \in X$, and $\mu_X(a) = a$ if $a \in \mathcal{A}$ is unmatched.

When a pair of agents $(i, j) \in \mathcal{I} \times \mathcal{J}$ matches, each experiences a utility gain. We denote these utilities by a global utility function $u : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$, where $u(a, a')$ denotes the utility that agent a gains from being matched to agent a'. (If a and a' are on the same side of the market, we take $u(a, a')$ to be zero by default.) We allow these utilities to be negative, if matching results in a net cost (e.g., if an agent is providing a service). We assume each agent $a \in \mathcal{A}$ receives zero utility if unmatched, i.e., $u(a, a) = 0$. When we wish to emphasize the role of an individual agent's utility function, we will use the equivalent notation $u_a(a') := u(a, a')$.

A *market outcome* consists of a matching $X \in \mathscr{X}_\mathcal{A}$ along with a vector $\tau \in \mathrm{R}^\mathcal{A}$ of transfers, where τ_a is the amount of money transferred from the platform to agent a for each $a \in \mathcal{A}$. These monetary transfers are a salient feature of most real-world matching markets: riders pay drivers on Lyft, clients pay freelancers on TaskRabbit, and guests pay hosts on Airbnb. Shapley and Shubik (1971) capture this aspect of matching markets by augmenting the classical two-sided matching model with transfers of utility between agents. Transfers are typically required to be *zero-sum*, meaning that $\tau_i + \tau_j = 0$ for all matched pairs $(i, j) \in X$

and $\tau_a = 0$ if a is unmatched. Here, X represents how agents are matched and τ_a represents the transfer that agent a receives (or pays). The net utility that an agent a derives from a matching with transfers (X, τ) is therefore $u(a, \mu_X(a)) + \tau_a$.

Stable matchings. In matching theory, stability captures when a market outcome aligns with individual agents' preferences. Roughly speaking, a market outcome (X, τ) is stable if: (i) no individual agent a would rather be unmatched, and (ii) no pair of agents (i, j) can agree on a transfer such that both would rather match with each other than abide by (X, τ). Formally:

Definition 9. *A market outcome (X, τ) is* stable *if: (i) it is* individually rational*, i.e.,*

$$u_a(\mu_X(a)) + \tau_a \geq 0 \tag{12.1}$$

for all agents $a \in \mathcal{A}$, and (ii) it has no blocking pairs*, i.e.,*

$$\big(u_i(\mu_X(i)) + \tau_i\big) + \big(u_j(\mu_X(j)) + \tau_j\big) \geq u_i(j) + u_j(i) \tag{12.2}$$

for all pairs of agents $(i, j) \in \mathcal{I} \times \mathcal{J}$.[4]

A fundamental property of the matching with transfers model is that if (X, τ) is stable, then X is a maximum weight matching, i.e., X maximizes $\sum_{a \in \mathcal{A}} u_a(\mu_X(a))$ over all matchings $X \in \mathscr{X}_{\mathcal{A}}$ (Shapley and Shubik, 1971). The same work shows that stable market outcomes coincide with Walrasian equilibria. (For completeness, we recapitulate the basic properties of this model in Chapter H.1.)

To make the matching with transfers model concrete, we use the simple market depicted in the center panel of Chapter 12.1 as a running example throughout the paper. This market consists of a customer Charlene and two providers Percy and Quinn, which we denote by $\mathcal{I} = \{C\}$ and $\mathcal{J} = \{P, Q\}$. If the agents' utilities are as given in Chapter 12.1, then Charlene would prefer Quinn, but Quinn's cost of providing the service is much higher. Thus, matching Charlene and Percy is necessary for a stable outcome. This matching is stable for any transfer from Charlene to Percy in the interval $[5, 7]$.

12.3 Learning Problem and Feedback Model

We instantiate the platform's learning problem in a stochastic contextual bandits framework. Matching takes place over the course of T rounds. We denote the set of all customers by $\mathcal{I}^*$, the set of all providers by $\mathcal{J}^*$, and the set of all agents on the platform by $\mathcal{A}^* = \mathcal{I}^* \cup \mathcal{J}^*$. Each agent $a \in \mathcal{A}^*$ has an associated context $c_a \in \mathcal{C}$, where $\mathcal{C}$ is the set of all possible contexts. This context represents the side information available to the platform about the agent, e.g.,

[4] We observe that (12.2) corresponds to no pair of agents (i, j) being able to agree on a transfer such that both would rather match with each other than abide by (X, τ). Notice that a pair (i, j) violates (12.2) if and only if they can find a transfer $\tau'_i = -\tau'_j$ such that $u_i(j) + \tau'_i > u_i(\mu_X(i)) + \tau_i$ and $u_j(i) + \tau'_j > u_j(\mu_X(j)) + \tau_j$.

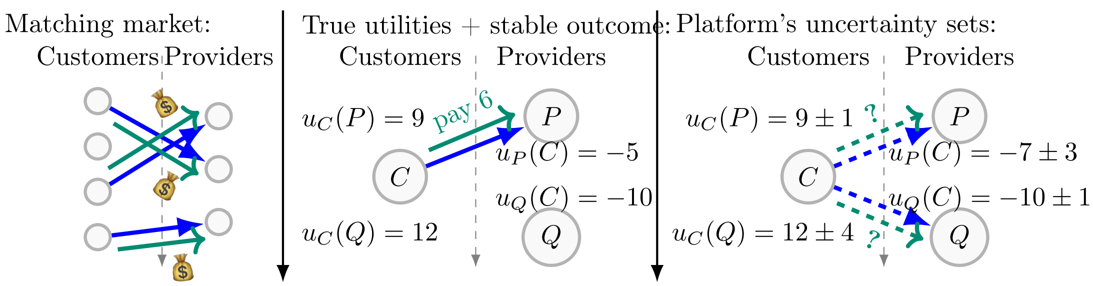


Figure 12.1: The left panel depicts a schematic of a matching (blue) with transfers (green). The center panel depicts a matching market with three agents and a stable matching with transfers for that market. (If the transfer 6 is replaced with any value between 5 and 7, the outcome remains stable.) The right panel depicts the same market, but with utilities replaced by uncertainty sets; note that no matching with transfers is stable for all realizations of utilities.

demographic, location, or platform usage information. Each round, a set of agents arrives to each side of the market. The platform then selects a market outcome and incurs a regret equal to the *instability* of the market outcome (which we introduce formally in Chapter 12.4). Finally, the platform receives noisy feedback about the utilities of each matched pair (i, j).

To interpret the noisy feedback, note that platforms in practice often receive feedback both explicitly (e.g., riders rating drivers after a Lyft ride) and implicitly (e.g., engagement metrics on an app). In either instance, feedback is likely to be sparse and noisy. For simplicity, we do not account for agents strategically manipulating their feedback to the platform and focus on the problem of learning preferences from unbiased reports.

We now describe this model more formally. In the t-th round:

1. A set $\mathcal{I}^t \subseteq \mathcal{I}^*$ of customers and a set $\mathcal{J}^t \subseteq \mathcal{J}^*$ of providers arrive to the market. Write $\mathcal{I}^t \cup \mathcal{J}^t =: \mathcal{A}^t$. The platform observes the identity a and the *context* $c_a \in \mathcal{C}$ of each agent $a \in \mathcal{A}^t$.
2. The platform selects a matching with *zero-sum* transfers (X^t, τ^t) between $\mathcal{I}^t$ and $\mathcal{J}^t$.
3. The platform observes noisy utilities $u_a(\mu_{X^t}(a)) + \varepsilon_{a,t}$ for each agent $a \in \mathcal{I}^t \cup \mathcal{J}^t$, where the $\varepsilon_{a,t}$ are independent, 1-subgaussian random variables.[5]
4. The platform incurs regret equal to the *instability* of the selected market outcome (X^t, τ^t). (We define instability formally in Chapter 12.4.)

The platform's total regret R_T is thus the cumulative instability incurred up through round T.

[5]Our feedback model corresponds to *semi-bandit* feedback, since the platform has (noisy) access to each agent's utility within the matching rather than the overall utility of the matching.

12.3.1 Preference structure

In this bandits framework, we can impose varying degrees of structure on agent preferences. We encode these preference structures via the functional form of agents' utility functions and their relation to agent contexts. More formally, let $\mathcal{U}$ be the set of functions $u\colon \mathcal{A}^* \times \mathcal{A}^* \to \mathbb{R}$, i.e., $\mathcal{U}$ is the set of all possible (global) utility functions. We now introduce several classes of preference structures as subsets of $\mathcal{U}$.

Unstructured preferences. The simplest setting we consider is one where the preferences are unstructured. Specifically, we consider the class of utility functions

$$\mathcal{U}_{\text{unstructured}} = \left\{u \in \mathcal{U} \mid u(a, a') \in [-1, 1]\right\}.$$

(Here, one can think of the context as being uninformative, i.e., $\mathcal{C}$ is the singleton set.) In this setup, the platform must learn each agent's utility function $u_a(\cdot) = u(a, \cdot)$.

Typed preferences. We next consider a market where each agent comes in one of finitely many *types*, with agents of the same type having identical preferences. Assuming typed preference structures is standard in theoretical models of markets (see, e.g., Debreu and Scarf (1963); Echenique et al. (2013); Azevedo and Hatfield (2018)). We can embed types into our framework by having each agent's context represent their type, with $|\mathcal{C}| < \infty$. The global utility function is then fully specified by agents' contexts:

$$\mathcal{U}_{\text{typed}} = \{u \in \mathcal{U} \mid u(a, a') = f(c_a, c_{a'}) \text{ for some } f\colon \mathcal{C} \times \mathcal{C} \to [-1, 1]\}.$$

Separable linear preferences. We next consider markets where each agent is associated with *known* information given by their context as well as *hidden* information that must be learned by the platform. (This differs from unstructured preferences, where all information was hidden, and typed preferences, where each agent's context encapsulated their full preferences.) We explore this setting under the assumption that agents' contexts and hidden information interact linearly.

We assume that all contexts belong to $\mathcal{B}^d$ (i.e., $\mathcal{C} = \mathcal{B}^d$) where $\mathcal{B}^d$ is the ℓ_2 unit ball in R^d. We also assume that there exists a function $\phi\colon \mathcal{A}^* \to \mathcal{B}^d$ mapping each agent to the hidden information associated to that agent. The preference class $\mathcal{U}^d_{\text{linear}}$ can then be defined as

$$\mathcal{U}^d_{\text{linear}} = \left\{u \in \mathcal{U} \;\middle|\; u(a, a') = \langle c_{a'}, \phi(a)\rangle \text{ for some } \phi\colon \mathcal{A}^* \to \mathcal{B}^d\right\}.$$

12.4 Measuring Approximate Stability

When learning stable matchings, we must settle for guarantees of approximate stability, since exact stability—a binary notion—is unattainable when preferences are uncertain. To see this, we return to the example from Chapter 12.1. Suppose that the platform has uncertainty sets given by the right panel. Recall that for the true utilities, all stable outcomes match

Charlene with Percy. If the true utilities were instead the upper bounds of each uncertainty set, then all stable outcomes would match Charlene and Quinn. Given only the uncertainty sets, it is impossible for the platform to find an (exactly) stable matching, so it is necessary to introduce a measure of approximate stability as a relaxed benchmark for the platform; we turn to this now.

Given the insights of Shapley and Shubik (1971)—that all stable outcomes maximize the sum of agents' utilities—it might seem natural to measure distance from stability simply in terms of the *utility difference*. To define this formally, let $\mathcal{A}$ be the set of agents participating in the market. (This corresponds to $\mathcal{A}^t$ at time step t in the bandits model.) The utility difference[6] of a market outcome (X, τ) is given by:

$$\left(\max_{X' \in \mathscr{X}_{\mathcal{A}}} \sum_{a \in \mathcal{A}} u_a(\mu_{X'}(a))\right) - \left(\sum_{a \in \mathcal{A}} u_a(\mu_X(a)) + \tau_a\right). \tag{12.3}$$

The first term $\max_{X' \in \mathscr{X}_{\mathcal{A}}} \sum_{a \in \mathcal{A}} u_a(\mu_{X'}(a))$ is the maximum total utility of any matching, and the second term $\sum_{a \in \mathcal{A}}(u_a(\mu_X(a)) + \tau_a)$ is the total utility of market outcome (X, τ). Since transfers are zero-sum, (12.3) can be equivalently written as

$$\left(\max_{X' \in \mathscr{X}_{\mathcal{A}}} \sum_{a \in \mathcal{A}} u_a(\mu_{X'}(a))\right) - \sum_{a \in \mathcal{A}} u_a(\mu_X(a)).$$

But this shows that utility difference actually ignores the transfers τ entirely! In fact, the utility difference can be zero even when the transfers lead to a market outcome that is far from stable (see Chapter H.2.1). Utility difference is therefore *not* incentive-aware, making it unsuitable as an objective for learning stable matchings with transfers.

In the remainder of this section, we propose a measure of instability—Subset Instability—which we will show serves as a suitable objective for learning stable matchings with transfers. Specifically, we show that Subset Instability captures the distance of a market outcome from equilibrium while reflecting both the platform's objective and the users' incentives. We additionally show that Subset Instability satisfies several structural properties that make it useful for learning.

12.4.1 Subset Instability

Subset Instability is based on utility difference, but rather than only looking at the market in aggregate, it takes a maximum ranging over all subsets of agents.

[6]Utility difference is standard as a measure of regret for learning a maximum weight matching in the combinatorial bandits literature (see, e.g., Gai et al. (2012)). However, we show that for learning stable matchings, a fundamentally different measure of regret is needed.

Definition 10. *Given utilities u, the* Subset Instability $I(X, \tau; u, \mathcal{A})$ *of a matching with transfers (X, τ) is*

$$\max_{\mathcal{S} \subseteq \mathcal{A}} \left[\left(\max_{X' \in \mathscr{X}_{\mathcal{S}}} \sum_{a \in \mathcal{S}} u_a(\mu_{X'}(a)) \right) - \left(\sum_{a \in \mathcal{S}} u_a(\mu_X(a)) + \tau_a \right) \right]. \qquad (*)$$

(The first term $\max_{X' \in \mathscr{X}_{\mathcal{S}}} \sum_{a \in \mathcal{S}} u_a(\mu_{X'}(a))$ is the maximum total utility of any matching over $\mathcal{S}$, and the second term $\sum_{a \in \mathcal{A}} (u_a(\mu_X(a)) + \tau_a)$ is the total utility of the agents in $\mathcal{S}$ under market outcome (X, τ).)

Intuitively, Subset Instability captures stability because it checks whether any subset of agents would prefer an alternate outcome. We provide a more extensive economic interpretation below; but before doing so, we first illustrate Definition 10 in the context of the example in Chapter 12.1.

Consider the matching $X = \{(C, Q)\}$ with transfers $\tau_C = -11$ and $\tau_Q = 11$. (This market outcome is stable for the upper bounds of the uncertainty sets of the platform in Chapter 12.1, but not stable for the true utilities.) It is not hard to see that the subset $\mathcal{S}$ that maximizes Subset Instability is $\mathcal{S} = \{C, P\}$, in which case $\max_{X' \in \mathscr{X}_{\mathcal{S}}} \sum_{a \in \mathcal{S}} u_a(\mu_{X'}(a)) = 4$ and $\sum_{a \in \mathcal{S}} (u_a(\mu_X(a)) + \tau_a) = 1$. Thus, the Subset Instability of (X, τ) is $I(X, \tau; u, \mathcal{A}) = 4 - 1 = 3$. In contrast, the utility difference of (X, τ) is 2.

We now discuss several interpretations of Subset Instability, which provide further insight into why Subset Instability serves as a meaningful notion of approximate stability in online marketplaces. In particular, Subset Instability can be interpreted as *the minimum stabilizing subsidy*, as *the platform's cost of learning*, as *a measure of user unhappiness*, and as a *distance from equilibrium.*

Subset Instability as the platform's minimum stabilizing subsidy. Subset Instability can be interpreted in terms of monetary subsidies from the platform to the agents. Specifically, the Subset Instability of a market outcome equals the minimum amount the platform could subsidize agents so that the subsidized market outcome is individually rational and has no blocking pairs.

More formally, let $s \in \mathrm{R}_{\geq 0}^{\mathcal{A}}$ denote subsidies made by the platform, where the variable $s_a \geq 0$ represents the subsidy provided to agent a.[7] For a market outcome (X, τ), the *minimum stabilizing subsidy* is

$$\min_{s \in \mathrm{R}_{\geq 0}^{\mathcal{A}}} \left\{ \sum_{a \in \mathcal{A}} s_a \;\middle|\; (X, \tau + s) \text{ is stable} \right\}, \qquad (12.4)$$

where we define stability in analogy to Chapter 9. Specifically, we say that a market outcome (X, τ) with subsidies s is *stable* if it is individually rational, i.e., $u_a(\mu_X(a)) + \tau_a + s_a \geq 0$ for

[7] The requirement that $s_a \geq 0$ enforces that all subsidies are nonnegative; without it, (12.5) would reduce to the utility difference, which is not incentive-aware.

all agents $a \in \mathcal{A}$, and has no blocking pairs, i.e., $(u_i(\mu_X(i)) + \tau_i + s_i) + (u_j(\mu_X(j)) + \tau_j + s_j) \geq u_i(j) + u_j(i)$ for all pairs of agents $(i, j) \in \mathcal{I} \times \mathcal{J}$.

Given this setup, we show the following equivalence:

Proposition 95. *Minimum stabilizing subsidy equals Subset Instability for any market outcome.*

The proof boils down to showing that the two definitions are "dual" to each other. To formalize this, we rewrite the minimum stabilizing subsidy as the solution to the following linear program:[8]:

$$
\begin{aligned}
&\min_{s \in \mathbb{R}^{|\mathcal{A}|}} \sum_{a \in \mathcal{A}} s_a && && (12.5)\\
&\text{s.t.} \quad \big(u_i(\mu_X(i)) + \tau_i + s_i\big) + \big(u_j(\mu_X(j)) + \tau_j + s_j\big) \geq u_i(j) + u_j(i) && \forall (i, j) \in \mathcal{I} \times \mathcal{J}\\
&\qquad\; u_a(\mu_X(a)) + \tau_a + s_a \geq 0 && \forall a \in \mathcal{A}\\
&\qquad\; s_a \geq 0 && \forall a \in \mathcal{A}.
\end{aligned}
$$

The crux of our argument is that the dual linear program to (12.5) maximizes the combinatorial objective ($*$). The equivalence of ($*$) and (12.5) then follows from strong duality.

With this alternate formulation of Subset Instability in mind, we revisit the example in Chapter 12.1. Again, consider the matching $X = \{(C, Q)\}$ with transfers $\tau_C = -11$ and $\tau_Q = 11$. (This is stable for the upper bounds of the uncertainty sets of the platform in Chapter 12.1, but not stable for the true utilities.) We have already shown above that the Subset Instability of this market outcome is 3. To see this via the subsidy formulation, note that the optimal subsidy s gives C and P a total of 3. (E.g., we give C a subsidy of $s_C = 2$ and P a subsidy of $s_P = 1$.) Indeed, if $s_C + s_P = 3$, then

$$\big(u_C(\mu_X(C)) + \tau_C + s_C\big) + \big(u_P(\mu_X(P)) + \tau_P + s_P\big) \geq u_C(P) + u_P(C)$$

holds (with equality), so the pair (C, P) could no longer gain by matching with each other.

The subsidy perspective turns out to be useful when designing learning algorithms. In particular, while the formulation in Chapter 10 involves a maximization over the $2^{|\mathcal{A}|}$ subsets of $\mathcal{A}$, the linear programming formulation (12.5) only involves $O(|\mathcal{A}|)$ variables and $O(|\mathcal{A}|^2)$ constraints.

Subset Instability as the platform's cost of learning. We next connect minimum stabilizing subsidies to the platform's *cost of learning*—how much the platform would have to pay to keep users on the platform in the presence of a worst-case (but budget-balanced) competitor with perfect knowledge of agent utilities.

Observe that (12.4) is the minimum amount the platform could subsidize agents so that no budget-balanced competitor could convince agents to leave. The way that we formalize

[8]In this linear program, the first set of constraints ensures there are no blocking pairs, while the second set of constraints ensures individual rationality.

"convincing agents to leave" is that: (a) an agent will leave the original platform if they prefer to be unmatched over being on the platform, or (b) a pair of agents who are matched on the competitor's platform will leave the original platform if they both prefer the new market outcome over their original market outcomes. Thus, if we imagine the platform as actually paying the subsidies, then the cumulative instability (i.e., our regret) can be realized as a "cost of learning": it is how much the platform pays the agents to learn a stable outcome while ensuring that no agent has the incentive to leave during the learning process. Later on, we will see that our algorithmic approach can be extended to efficiently compute feasible subsidies for (12.5) that are within a constant factor of our regret bound, meaning that subsidies can be implemented using only the information that the platform has. Moreover, in Chapter 12.6.2, we show that cost of learning can also be explicitly connected to the platform's revenue.

Subset Instability as a measure of user unhappiness. While the above interpretations focus on Subset Instability from the platform's perspective, we show that Subset Instability can also be interpreted as a measure of user unhappiness. Given a subset $\mathcal{S} \subseteq \mathcal{A}$ of agents, which we call a coalition, we define the *unhappiness* of $\mathcal{S}$ with respect to a market outcome (X, τ) to be the maximum gain (relative to (X, τ)) in total utility that the members of coalition $\mathcal{S}$ could achieve by matching only among themselves, such that no member is worse off than they were in (X, τ). (See Chapter H.2.3 for a formal definition.) The condition that no member is worse off ensures that all agents would actually want to participate in the coalition (i.e. they prefer it to the original market outcome).

User unhappiness differs from the original definition of Subset Instability in ($*$), because ($*$) does not require individuals to be better off in any alternative matching. However, we show that this difference is inconsequential:

Proposition 96. *The maximum unhappiness of any coalition $\mathcal{S} \subseteq \mathcal{A}$ with respect to (X, τ) equals the Subset Instability $I(X, \tau; u, \mathcal{A})$.*

See Chapter H.2.3 for a full proof. In the proof, we relate the maximum unhappiness of any coalition to the dual linear program to (12.5). To show this relation, we leverage the fact that optimal solutions to the dual program correspond to blocking pairs of agents as well as individual rationality violations.

The main takeaway from Chapter 96 is that Subset Instability not only measures costs to the platform, but also costs to users, in terms of the maximum amount they "leave on the table" by not negotiating an alternate arrangement amongst themselves.

Subset Instability as a distance from equilibrium. Finally, we connect Subset Instability to solution concepts for coalitional games, a general concept in game theory that includes matching with transfers as a special case. Coalitional games (also known as cooperative games) capture competition and cooperation amongst a group of agents. The *core* is the set of outcomes in a cooperative game such that no subset $\mathcal{S}$ of agents can achieve higher total utility among themselves than according to the given outcome. In games where the core is empty, a natural relaxation is the *strong ε-core* Shapley and Shubik (1966), which is the set of outcomes in a cooperative game such that no subset $\mathcal{S}$ of agents can achieve total utility among themselves that is at least ε greater than according to the given outcome.

Subset Instability can be seen as transporting the strong ε-core notion to a slightly different context. In particular, in the context of matching with transferable utilities, the core is exactly the set of stable matchings; since a stable matching always exists, the core is always nonempty. Even though the core is nonempty, we can nonetheless use the strong ε-core to measure *distance from the core*. More specifically, it is natural to consider the smallest ε such that (X, τ) is in the strong ε-core. This definition exactly aligns with Subset Instability, thus providing an alternate interpretation of Subset Instability within the context of coalitional game theory.

12.4.2 Properties of Subset Instability

We now describe additional properties of our instability measure that are important for learning. We show that Subset Instability is: (i) zero if and only if the matching with transfers is stable, (ii) Lipschitz in the true utility functions, and (iii) lower bounded by the utility difference.

Proposition 97. *Subset Instability satisfies the following properties:*

1. *Subset Instability is always nonnegative and is zero if and only if (X, τ) is stable.*

2. *Subset Instability is Lipschitz continuous with respect to agent utilities. That is, for any possible market outcome (X, τ), and any pair of utility functions u and uii it holds that:*

$$|I(X, \tau; u, \mathcal{A}) - I(X, \tau; uii, \mathcal{A})| \leq 2 \sum_{a \in \mathcal{A}} \|u_a - uii_a\|_\infty.$$

3. *Subset Instability is always at least the utility difference.*

We defer the proof to Chapter H.2.4.

These three properties show that Subset Instability is useful as a regret measure for learning stable matchings. The first property establishes that Subset Instability satisfies the basic desideratum of having zero instability coincide with exact stability. The second property shows that Subset Instability is robust to small perturbations to the utility functions of individual agents. The third property ensures that, when learning using Subset Instability as a loss function, the platform learns a socially optimal matching.

Note that the second property already implies the existence of an explore-then-commit algorithm that achieves $\widetilde{O}(N^{4/3}T^{2/3})$ regret in the simple setting where $\mathcal{A}^t = \mathcal{A}$ for some $\mathcal{A}$ of size N for all t.[9] In the next section, we will explore algorithms that improve the dependence on the number of rounds T to $\sqrt{T}$ and also work in more general settings.

12.5 Regret Bounds

In this section, we develop a general approach for designing algorithms that achieve near-optimal regret within our framework. To be precise, the platform's regret is defined to be

$$R_T = \sum_{t=1}^{T} I(X^t, \tau^t; u, \mathcal{A}^t).$$

While our framework bears some resemblance to the (incentive-free) combinatorial bandit problem of learning a maximum weight matching, two crucial differences differentiate our setting: (i) in each round, the platform must choose *transfers* in addition to a matching, and (ii) loss is measured with respect to *instability* rather than the utility difference. Nonetheless, we show that a suitable interpretation of "optimism in the face of uncertainty" can still apply.

Regret bounds for different preference structures. By instantiating this optimism-based approach, we derive regret bounds for the preference structures introduced in Chapter 12.3. We start with the simplest case of unstructured preferences, where we assume no structure on the utilities.

Theorem 98. *For preference class $\mathcal{U}_{\text{unstructured}}$ (see Chapter 12.3),* MatchUCB *(defined in Chapter 12.5.3) incurs expected regret $\mathbb{E}(R_T) = O\big(|\mathcal{A}|\sqrt{nT\log(|\mathcal{A}|T)}\big)$, where $n = \max_t |\mathcal{A}_t|$.*

In Chapter 12.5.4, we additionally give a matching (up to logarithmic factors) lower bound showing for $n = |\mathcal{A}|$ that such scaling in $|\mathcal{A}|$ is indeed necessary. This demonstrates that the regret scales with $|\mathcal{A}|\sqrt{n}$, which is superlinear in the size of the market. Roughly speaking, this bound means that the platform is required to learn a superconstant amount of information per agent in the marketplace. These results suggest that without preference structure, it is unlikely that a platform can efficiently learn a stable matching in large markets.

The next two bounds demonstrate that, with preference structure, efficient learning of a stable matching becomes possible. First, we consider typed preferences, which are purely specified by a function f mapping finitely many pairs of contexts to utilities.

Theorem 99. *For preference class $\mathcal{U}_{\text{typed}}$ (see Chapter 12.3),* MatchTypedUCB *(defined in Chapter 12.5.3) incurs expected regret $\mathbb{E}(R_T) = O\big(|\mathcal{C}|\sqrt{nT\log(|\mathcal{A}|T)}\big)$, where $n = \max_t |\mathcal{A}_t|$.*

[9] This bound can be achieved by adapting the explore-then-commit (ETC) approach where the platform explores by choosing each pair of agents $\widetilde{O}((T/N)^{2/3})$ times Lattimore and Szepesvári (2020). Thus, $\widetilde{O}(N^{1/3}T^{2/3})$ rounds are spent exploring, and the Subset Instability of the matching selected in the commit phase is $\widetilde{O}(N^{4/3}T^{2/3})$ with high probability. We omit further details since this analysis is a straightforward adaptation of the typical ETC analysis.

For a fixed type space $\mathcal{C}$, the regret bound in Chapter 99 scales sublinearly with the market size (captured by $|\mathcal{A}|$ and n). This demonstrates that the platform can efficiently learn a stable matching when preferences are determined by types. In fact, the regret bound only depends on the number of agents who arrive on the platform in any round; notably, it does not depend on the total number of agents on the platform (beyond logarithmic factors).

Finally, we consider separable linear preferences, where the platform needs to learn hidden information associated with each agent.

Theorem 100. *For preference class $\mathcal{U}_{\text{linear}}$ (see Chapter 12.3),* MatchLinUCB *(defined in Chapter 12.5.3) incurs expected regret $\mathbb{E}(R_T) = O\big(d\sqrt{|\mathcal{A}|}\sqrt{nT\log(|\mathcal{A}|T)}\big)$, where $n = \max_t |\mathcal{A}_t|$.*

When n is comparable to $|\mathcal{A}|$, the regret bound in Chapter 100 scales linearly with the market size (captured by $|\mathcal{A}|$) and linearly with the dimension d. Roughly speaking, this means that the platform learns (at most) a constant amount of information per agent in the marketplace. We interpret this as indicating that the platform can efficiently learn a stable matching in large markets for separable linear preferences, although learning in this setting is more demanding than for typed preferences.

12.5.1 Algorithm

Following the principle of optimism, our algorithm selects at each round a stable market outcome using upper confidence bounds as if they were the true agent utilities. To design and analyze this algorithm, we leverage the fact that, in the full-information setting, stable market outcomes are optimal solutions to a pair of primal-dual linear programs whose coefficients depend on agents' utility functions. This primal-dual perspective lets us compute a market outcome each round. A particular consequence is that any UCB-based algorithm for learning matchings in a semi-bandit setting can be transformed into an algorithm for learning *both* the matching and the prices.

Stable market outcomes via linear programming duality. Before proceeding with the details of our algorithm, we review how the primal-dual framework can be used to select a stable market outcome in the full information setting. Shapley and Shubik (1971) show that stable market outcomes (X, τ) correspond to optimal primal-dual solutions to the following pair of primal and dual linear programs (where we omit the round index t and consider matchings over $\mathcal{A} = \mathcal{I} \cup \mathcal{J}$):

Primal (P)

$$\begin{aligned}\max_{Z\in\mathbb{R}^{|\mathcal{I}|\times|\mathcal{J}|}} \quad & \sum_{(i,j)\in\mathcal{I}\times\mathcal{J}} Z_{i,j}(u_i(j)+u_j(i)) \\ \text{s.t.} \quad & \sum_{j\in\mathcal{J}} Z_{i,j}\le 1 \quad \forall i\in\mathcal{I} \\ & \sum_{i\in\mathcal{I}} Z_{i,j}\le 1 \quad \forall j\in\mathcal{J} \\ & Z_{i,j}\ge 0 \qquad \forall (i,j)\in\mathcal{I}\times\mathcal{J}\end{aligned}$$

Dual (D)

$$\begin{aligned}\min_{p\in\mathbb{R}^{|\mathcal{A}|}} \quad & \sum_{a\in\mathcal{A}} p_a \\ \text{s.t.} \quad & p_i+p_j\ge u_i(j)+u_j(i) \quad \forall (i,j)\in\mathcal{I}\times\mathcal{J} \\ & p_a\ge 0 \qquad \forall a\in\mathcal{A}\end{aligned}$$

The primal program (P) is a linear programming formulation of the maximum weight matching problem: the Birkhoff-von Neumann theorem states that its extreme points are exactly the indicator vectors for matchings between $\mathcal{I}$ and $\mathcal{J}$. Each dual variable p_a in (D) can be interpreted as a *price* that roughly corresponds to agent a's net utility. Specifically, given any optimal primal-dual pair (Z, p), one can recover a matching μ_X from the nonzero entries of Z and set transfers $\tau_a = p_a - u_a(\mu_X(a))$ to obtain a stable outcome (X, τ). Moreover, any stable outcome induces an optimal primal-dual pair (Z, p).

Algorithm overview. Leveraging the above primal-dual formulation of stability, we introduce a meta-algorithm MetaMatchUCB for learning stable outcomes (Chapter 1). In each round, we compute a matching with transfers by solving the primal-dual linear programs for our upper confidence bounds: Suppose we have a collection $\mathscr{C}$ of confidence sets $C_{i,j}, C_{j,i} \subseteq \mathbb{R}$ such that $u_i(j) \in C_{i,j}$ and $u_j(i) \in C_{j,i}$ for all $(i,j) \in \mathcal{I}\times\mathcal{J}$. Our algorithm uses $\mathscr{C}$ to get an upper confidence bound for each agent's utility function and then computes a stable matching with transfers as if these upper confidence bounds were the true utilities (see ComputeMatch in Chapter 2). This can be implemented efficiently if we use, e.g., the Hungarian algorithm Kuhn (1955) to solve (P) and (D).

Algorithm 1: MetaMatchUCB: A bandit meta-algorithm for matching with transferable utilities.

Input: Time horizon T

Initialize confidence intervals $\mathscr{C}$ over utilities;
for $1 \le t \le T$ **do**
  $(X^t, \tau^t) \leftarrow$ **ComputeMatch**($\mathscr{C}$);
  Update confidence intervals $\mathscr{C}$;

12.5.2 Main lemma

The key fact we need to analyze our algorithms is that Subset Instability is upper bounded by the sum of the sizes of the relevant confidence sets, assuming that the confidence sets over the utilities contain the true utilities. (In the following, we again omit the round index t.)

Lemma 101. *Suppose a collection of confidence sets $\mathscr{C} = \{C_{i,j}, C_{j,i} : (i,j) \in \mathcal{I} \times \mathcal{J}\}$ is such that $u_i(j) \in C_{i,j}$ and $u_j(i) \in C_{j,i}$ for all (i,j). Then the instability of $(X^{\text{UCB}}, \tau^{\text{UCB}}) \coloneqq$ ComputeMatch$(\mathscr{C})$ satisfies*

$$I(X^{\text{UCB}}, \tau^{\text{UCB}}; u, \mathcal{A}^t) \leq \sum_{a \in \mathcal{A}} \Big(\max\big(C_{a,\mu_{X^{\text{UCB}}}(a)}\big) - \min\big(C_{a,\mu_{X^{\text{UCB}}}(a)}\big)\Big). \tag{12.6}$$

Proof. Since $(X^{\text{UCB}}, \tau^{\text{UCB}})$ is stable with respect to u^{UCB}, we have $I(X^{\text{UCB}}, \tau^{\text{UCB}}; u^{\text{UCB}}, \mathcal{A}^t) = 0$. Thus, it is equivalent to bound the difference $I(X^{\text{UCB}}, \tau^{\text{UCB}}; u, \mathcal{A}^t) - I(X^{\text{UCB}}, \tau^{\text{UCB}}; u^{\text{UCB}}, \mathcal{A}^t)$.

At this stage, it might be tempting to bound this difference using the Lipschitz continuity of Subset Instability (see Chapter 97). However, this would only allow us to obtain an upper bound of the form $\sum_{a \in \mathcal{A}} \max_{a' \in \mathcal{A}} \big(\max\big(C_{a,a'}\big) - \min\big(C_{a,a'}\big)\big)$. The problem with this bound is that it depends on the sizes of the confidence sets for all pairs of agents, including those that are *not* matched in X^{UCB}, making it too weak to prove regret bounds for UCB-style algorithms.[10] Thus, we proceed with a more fine-grained analysis.

Define the function

$$f(\mathcal{S}, X, \tau; u) = \left(\max_{X' \in \mathscr{X}_{\mathcal{S}}} \sum_{a \in \mathcal{S}} u_a(\mu_{X'}(a)) \right) - \left(\sum_{a \in \mathcal{S}} u_a(\mu_X(a)) + \tau_a \right).$$

By definition, $I(X, \tau; u, \mathcal{A}) = \max_{\mathcal{S} \subseteq \mathcal{A}} f(\mathcal{S}, X, \tau; u)$. It follows that

$$\begin{aligned} I(X^{\text{UCB}}, \tau^{\text{UCB}}; u, \mathcal{A}^t) &- I(X^{\text{UCB}}, \tau^{\text{UCB}}; u^{\text{UCB}}, \mathcal{A}^t) \\ &\leq \max_{\mathcal{S} \subseteq \mathcal{A}} \left(f(\mathcal{S}, X^{\text{UCB}}, \tau^{\text{UCB}}; u) - f(\mathcal{S}, X^{\text{UCB}}, \tau^{\text{UCB}}; u^{\text{UCB}}) \right). \end{aligned}$$

To finish, we upper bound $f(\mathcal{S}, X^{\text{UCB}}, \tau^{\text{UCB}}; u) - f(\mathcal{S}, X^{\text{UCB}}, \tau^{\text{UCB}}; u^{\text{UCB}})$ for each $\mathcal{S} \subseteq \mathcal{A}$. We decompose this expression into two terms:

$$\begin{aligned} f(\mathcal{S}, X^{\text{UCB}}, \tau^{\text{UCB}}; u) &- f(\mathcal{S}, X^{\text{UCB}}, \tau^{\text{UCB}}; u^{\text{UCB}}) \\ &= \underbrace{\left(\max_{X' \in \mathscr{X}_{\mathcal{S}}} \sum_{a \in \mathcal{S}} u_a(\mu_{X'}(a)) - \max_{X' \in \mathscr{X}_{\mathcal{S}}} \sum_{a \in \mathcal{S}} u_a^{\text{UCB}}(\mu_{X'}(a)) \right)}_{\text{(A)}} \\ &\quad + \underbrace{\left(\sum_{a \in \mathcal{S}} \big(u_a^{\text{UCB}}(\mu_{X^{\text{UCB}}}(a)) + \tau_a^{\text{UCB}}\big) - \sum_{a \in \mathcal{S}} \big(u_a(\mu_{X^{\text{UCB}}}(a)) + \tau_a^{\text{UCB}}\big) \right)}_{\text{(B)}}. \end{aligned}$$

[10] For intuition, consider the classical stochastic multi-armed bandits setting and suppose that we could only guarantee that the loss incurred by an arm is bounded by the maximum of the sizes of the confidence sets over *all* arms. Then, we would only be able to obtain a weak bound on regret, since low-reward arms with large confidence sets may never be pulled.

Algorithm 2: COMPUTEMATCH: Compute matching with transfers from confidence sets of utilities

Input: Confidence sets $\mathscr{C}$

for $(i, j) \in \mathcal{I} \times \mathcal{J}$ **do**

  ; `// Instantiate UCB estimates of utilities.`

  $u_i^{\mathrm{UCB}}(j) \leftarrow \max\bigl(C_{i,j}\bigr)$;

  ;

$(X^*, p^*) \leftarrow$ optimal primal-dual pair for (P) and (D) given utilities u^{UCB};

for $a \in \mathcal{A}$ **do**

  ; `// Set transfers based on` (X^*, p^*) `and UCB utilities.`

  ;

return (X^*, τ);

To see that (A) is nonpositive, observe that the maximum weight matching of $\mathcal{S}$ with respect to u is no larger than the maximum weight matching of $\mathcal{S}$ with respect to u^{UCB}, since u^{UCB} pointwise upper bounds u. To upper bound (B), observe that the transfers cancel out, so the expression is equivalent to

$$\sum_{a \in \mathcal{S}} \bigl(u_a^{\mathrm{UCB}}(\mu_{X^{\mathrm{UCB}}}(a)) - u_a(\mu_{X^{\mathrm{UCB}}}(a))\bigr) \leq \sum_{a \in \mathcal{A}} \Bigl(\max\bigl(C_{a,\mu_{X^{\mathrm{UCB}}}(a)}\bigr) - \min\bigl(C_{a,\mu_{X^{\mathrm{UCB}}}(a)}\bigr)\Bigr). \quad \square$$

12.5.3 Instantiations of the meta-algorithm

As formalized in METAMATCHUCB, the regret bound of Chapter 101 suggests a simple approach: at each round, select the matching with transfers returned by COMPUTEMATCH and update confidence sets accordingly. To instantiate METAMATCHUCB, it remains to construct confidence intervals that contain the true utilities with high probability. This last step naturally depends on the assumptions made about the utilities and the noise.

Unstructured preferences. For this setting, we construct confidence intervals following the classical UCB approach: for each utility value involving the pair $(i, j) \in \mathcal{I} \times \mathcal{J}$, we take a confidence interval of length $O\bigl(\sqrt{\log(|\mathcal{A}|T)/n_{ij}}\bigr)$ centered at the empirical mean, where n_{ij} is the number of times the pair has been matched thus far. We describe this construction precisely in Chapter 3 (MATCHUCB).

To analyze MATCHUCB, recall that Chapter 101 bounds the regret at each step by the lengths of the confidence intervals of each pair in the selected matching. Bounding the lengths of the confidence intervals parallels the analysis of UCB for classical stochastic multi-armed bandits. We give the full proof of Chapter 98 in Chapter H.3.1.

Typed Preferences. For this setting, we construct our confidence intervals as follows: for each pair of types c_1 and c_2, we take a length $O\bigl(\sqrt{\log(|\mathcal{A}|T)/n_{c_1 c_2}}\bigr)$ confidence interval centered around the empirical mean, where $n_{c_1 c_2}$ is the number of times that an agent with

Algorithm 3: MatchUCB: A bandit algorithm for matching with transferable utilities for unstructured preferences.

Input: Time horizon T

for $(i, j) \in \mathcal{I} \times \mathcal{J}$ **do**

  ; `// Initialize confidence intervals.`

  $C_{i,j} \leftarrow [-1, 1]$;

  $C_{j,i} \leftarrow [-1, 1]$;

for $1 \leq t \leq T$ **do**

  $(X^t, \tau^t) \leftarrow$ **ComputeMatch**$(\mathscr{C})$;

  for $(i, j) \in X^t$ **do**

    ; `// Set confidence intervals and update means.`

    Update empirical means $\hat{u}_i(j)$ and $\hat{u}_j(i)$ from feedback; increment counter n_{ij};

    $C_{i,j} \leftarrow \left[\hat{u}_i(j) - 8\sqrt{\log(|\mathcal{A}|T)/n_{ij}}, \hat{u}_i(j) + 8\sqrt{\log(|\mathcal{A}|T)/n_{ij}}\,\right] \cap [-1, 1]$;

    $C_{j,i} \leftarrow \left[\hat{u}_j(i) - 8\sqrt{\log(|\mathcal{A}|T)/n_{ij}}, \hat{u}_j(i) + 8\sqrt{\log(|\mathcal{A}|T)/n_{ij}}\,\right] \cap [-1, 1]$;

type c_1 has been matched with an agent with type c_2. We describe this construction precisely in Chapter 4 (MatchTypedUCB). We give the full proof of Chapter 99 in Chapter H.3.2.

Algorithm 4: MatchTypedUCB: A bandit algorithm for matching with transferable utilities for typed preferences.

Input: Time horizon T

for $(c, c') \in \mathcal{C} \times \mathcal{C}$ **do**

  ; `// Initialize confidence intervals and empirical means.`

  $C_{c,c'} \leftarrow [-1, 1]$;

for $1 \leq t \leq T$ **do**

  $(X^t, \tau^t) \leftarrow$ **ComputeMatch**$(\mathscr{C})$;

  for $(i, j) \in X^t$ **do**

    ; `// Set confidence intervals and update means.`

    Update empirical means $\hat{f}(c_i, c_j)$ and $\hat{f}(c_i, c_j)$ from feedback; increment n_{c_i,c_j};

    $C_{c_i,c_j} \leftarrow \left[\hat{f}(c_i, c_j) - 8\sqrt{\log(|\mathcal{A}|T)/n_{c_i,c_j}}, \hat{f}(c_i, c_j) + 8\sqrt{\log(|\mathcal{A}|T)/n_{c_i,c_j}}\,\right] \cap [-1, 1]$;

    $C_{c_j,c_i} \leftarrow \left[\hat{f}(c_j, c_i) - 8\sqrt{\log(|\mathcal{A}|T)/n_{c_i,c_j}}, \hat{f}(c_i, c_j) + 8\sqrt{\log(|\mathcal{A}|T)/n_{c_i,c_j}}\,\right] \cap [-1, 1]$;

Separable Linear Preferences.

To build the confidence sets, we use a key idea from the design of LinUCB Russo and Van Roy (2013); Lattimore and Szepesvári (2020). The idea is to compute a confidence set for each hidden vector $\phi(a)$ using the least squares estimate and use that to construct confidence sets $C_{a,a'}$ for the utilities.

Algorithm 5: MatchLinUCB: A bandit algorithm for matching with transferable utilities for separable linear preferences.

Input: Time horizon T

for $(i, j) \in \mathcal{I} \times \mathcal{J}$ **do**
- ; `// Initialize confidence intervals.`
- $C_{i,j} \leftarrow [-1, 1]$;
- $C_{j,i} \leftarrow [-1, 1]$;

for $1 \leq t \leq T$ **do**
- $(X^t, \tau^t) \leftarrow$ **ComputeMatch**$(\mathscr{C})$;
- **for** $a \in \mathcal{A}^t$ **do**
  - ; `// Update confidence intervals.`
  - Increment the counter n_a;
  - $\beta \leftarrow O\left(d \log T + \frac{n_a \sqrt{\ln(n_a/(T|A|))}}{T^2}\right)$;
  - ; `// Parameter for width of confidence set.`
  - **if** $\mu_{X^t}(a) \neq a$ **then**
    - Add t to $\mathcal{T}_a$ (the set of rounds in which agent a has been matched);
    - Set $\mathcal{R}_{a,t}$ equal to the observed utility for agent a in round t;
    - $\phi^{\mathrm{LS}}(a) \leftarrow \mathrm{argmin}_{v \in \mathcal{B}^d} \left(\sum_{t' \in \mathcal{T}_a} \left(\langle v, c_{\mu_{X_{t'}}(a)} \rangle - \mathcal{R}_{a,t'}\right)^2\right)$;
    - ; `// Least squares estimate.`
    - $C_{\phi(a)} \leftarrow \left\{ v \mid \sum_{t' \in \mathcal{T}_a} \left(\langle v - \phi^{\mathrm{LS}}(a), c_{\mu_{X_{t'}}(a)} \rangle\right)^2 \leq \beta, \|v\|_2 \leq 1 \right\}$;
    - ; `// Conf. ellipsoid.`
    - **for** $a' \in \mathcal{A}$ **do**
      - $C_{a,a'} \leftarrow \left\{ \langle c_{a'}, v \rangle \mid v \in C_{\phi(a)} \right\} \cap [-1, 1]$;
      - ; `// Update confidence sets of agent` a.

More formally, let $\mathcal{T}_a$ be the set of rounds where agent a is matched on the platform thus far, and for $t' \in \mathcal{T}_a$, let $\mathcal{R}_{a,t'}$ be the observed utility at time t' for agent a. The center of the confidence set will be given by the least squares estimate

$$\phi^{\mathrm{LS}}(a) = \operatorname*{arg\,min}_{v \in \mathcal{B}^d} \left(\sum_{t' \in \mathcal{T}_a} (\langle v, c_{\mu_{X_{t'}}(a)} \rangle - \mathcal{R}_{a,t'} \right).$$

The confidence set for $\phi(a)$ is given by

$$C_{\phi(a)} := \left\{ v \;\middle|\; \sum_{t' \in \mathcal{T}_{a,t}} \left\langle v - \phi^{\mathrm{LS}}(a), c_{\mu_{X_{t'}}(a)} \right\rangle^2 \leq \beta \text{ and } \|v\|_2 \leq 1 \right\},$$

where $\beta = O\left(D \log T + \frac{n_a \sqrt{\ln(n_a/\delta)}}{T^2}\right)$ and n_a counts the number of times that a has appeared in selected matchings. The confidence set for $u(a, a')$ is given by

$$C_{a,a'} := \left\{ \langle c_{a'}, v \rangle \mid v \in C_{\phi(a)} \right\} \cap [-1, 1].$$

We describe this construction precisely in Chapter 5 (MatchLinUCB). We give the full proof of Chapter 100 in Chapter H.3.3.

12.5.4 Matching lower bound

For the case of unstructured preferences, we now show that MatchUCB achieves optimal regret (up to logarithmic factors) by showing a lower bound that (nearly) matches the upper bound in Theorem 98.

Lemma 102. *For any algorithm that learns a stable matching with respect to unstructured preferences, there exists an instance on which it has expected regret $\widetilde{\Omega}(|A|^{3/2}\sqrt{T})$ (where regret is given by Subset Instability).*

The idea behind this lemma is to show a lower bound for the easier problem of learning a maximum weight matching using utility difference as regret. By Proposition 97, this immediately implies a lower bound for learning a stable matching with regret measured by Subset Instability.

This lower bound illustrates the close connection between our setting and that of learning a maximum weight matching. Indeed, by applying MatchUCB and simply disregarding the transfers every round, we recover the classical UCB-based algorithm for learning the maximum weight matching Gai et al. (2012); Chen et al. (2013); Kveton et al. (2015). From this perspective, the contribution of MatchUCB is an approach to set the dual variables while asymptotically maintaining the same regret as the primal-only problem.

12.6 Extensions

In this section, we discuss several extensions of our results: instance-dependent regret bounds, connections between subset instability and platform revenue, and non-transferable utilities. These extensions illustrate the generality of our framework and also suggest several avenues for future research.

In Chapter 12.6.1, we derive instance-dependent regret bounds for Subset Instability, which allow us to improve the $O(\sqrt{T})$ convergence from Chapter 12.5 to $O(\log T)$ for any fixed instance. Achieving this logarithmic bound involves choosing "robust" dual solutions when setting transfers (rather than choosing an arbitrary optimal primal-dual pair as in ComputeMatch): we want our selected primal-dual pair to lead to stable outcomes even under perturbations of the transfers.

In Chapter 12.6.2, we connect the subsidy perspective of Subset Instability to platform revenue. We relate regret to platform revenue and show that, when there are search frictions, the platform can achieve substantial long-run profit despite starting with no knowledge of agent preferences.

In Chapter 12.6.3, we adapt our framework to matching with non-transferable utilities (where agents do not transfer money to other agents on the platform). We define an analogue of Subset Instability using the subsidy formulation and give an $\widetilde{O}(\sqrt{T})$ regret algorithm for learning stable matchings.

12.6.1 Instance-dependent regret bounds

While our analyses in Chapter 12.5.1 focused on bounds that hold uniformly for all problem instances, we now explore *instance-dependent* regret bounds. Instance-dependent bounds capture a different facet of bandit algorithms: how does the number of mistakes made by the algorithm scale on each instance with respect to T? Bounds of this nature have been explored in previous works Liu et al. (2020a); Basu et al. (2021); Sankararaman et al. (2021); Cen and Shah (2022); Liu et al. (2021) on learning stable matchings in the non-transferable utilities setting, and we show that they can be obtained within our framework as well.

Our instance-dependent regret bound depends on a gap $\Delta > 0$ determined by the true utility function u. We focus on the setting where agent utilities are unstructured (i.e., $u \in \mathcal{U}_{\text{unstructured}}$) and where the same set of agents $\mathcal{A}$ arrives in each round. As is common in analyses of combinatorial bandit problems (e.g., Kveton et al. (2015); Chen et al. (2013)), the gap Δ in the bound is global to the matching. Letting X^{opt} be a maximum weight matching with respect to u, we define the gap to Δ be the difference in utility between the optimal and second-best matchings[11]:

$$\Delta := \inf_{X \neq X^{\text{opt}}} \left\{ \sum_{a \in \mathcal{A}} u_a(\mu_{X^{\text{opt}}}(a)) - \sum_{a \in \mathcal{A}} u_a(\mu_X(a)) \right\}.$$

We prove the following regret bound:

Theorem 103 (Instance-Dependent Regret)**.** *Suppose that $\mathcal{A}_t = \mathcal{A}$ for all t. Let $u \in \mathcal{U}_{\text{unstructured}}$ be any utility function, and put*

$$\Delta := \inf_{X \neq X^*} \left\{ \sum_{a \in \mathcal{A}} u_a(\mu_{X^*}(a)) - \sum_{a \in \mathcal{A}} u_a(\mu_X(a)) \right\}.$$

Then MatchUCB′ *incurs expected regret $\mathbb{E}(R_T) = O(|\mathcal{A}|^5 \cdot \log(|\mathcal{A}|T)/\Delta^2)$.*

Remark. *MatchUCB′ is MatchUCB with a slight adjustment to ComputeMatch needed to prove Chapter 103. MatchUCB′, like MatchUCB, does not depend on the gap Δ and achieves the instance-independent regret bound in Chapter 98.*[12] *That is, MatchUCB′ achieves both our instance-independent and instance-dependent regret bounds.*

[11] Our bound is less fine-grained than the gap in (Chen et al., 2013), and in particular does not allow there to be multiple maximum weight matchings. We defer improving our definition of Δ to future work.

[12] The instance-independent regret bound can be shown using the same argument as the proof for Chapter 98.

Our starting point for proving Chapter 103 is to upper bound the number of "mistakes" that a platform makes while exploring and learning, i.e., the number of rounds where the chosen matching is suboptimal. That is, we bound the number of rounds where the chosen market outcome is not stable with respect to the true utilities u. This is similar in spirit to the analysis of the combinatorial bandits problem of learning a maximum weight matching in (Chen et al., 2013). However, a crucial difference is that a mistake can be incurred even when the selected matching is optimal, if the selected transfers do not result in a stable market outcome. Ensuring that the selected transfers result in a stable market outcome when the utility estimates are sufficiently accurate is the main technical hurdle in our analysis.

To make this argument work, we need to specify more precisely how the primal-dual solution is chosen in line 5 of ComputeMatch (which we previously did not specify). In particular, poor choices of the primal-dual solution can lead to many rounds where the chosen outcome is unstable, because the transfers violate the stability constraints. To see this, consider a market with a single customer C and a single provider P such that $u_C(P) = 2$ and $u_P(C) = -1$, and suppose we have nearly tight upper bounds $u_C^{\text{UCB}}(P) = 2 + \varepsilon$ and $u_P^{\text{UCB}}(C) = -1 + \varepsilon$ on the utilities. Then the market outcome with matching $\{(C, P)\}$ with $\tau_C = -2 - \varepsilon$ and $\tau_P = -\tau_C$ could be selected by ComputeMatch, since it corresponds to an optimal primal-dual pair for u^{UCB}. However, it is not stable with respect to the true utilities u (as individual rationality is violated for C), regardless of how small ε is. Thus, without assuming more about how the optimal primal-dual pair is chosen in ComputeMatch, we cannot hope to bound the number of unstable market outcomes selected.

We show that, by carefully selecting an optimal primal-dual pair each round, we can bound the number of mistakes. In particular, we design an algorithm ComputeMatch′ to find primal-dual pairs that satisfy the following property: if the confidence sets are small enough, then the selected matching will be stable with respect to the true utilities.

Lemma 104. *Suppose* ComputeMatch′ *is run on a collection $\mathscr{C}$ of confidence sets $C_{i,j}$ and $C_{j,i}$ over the agent utilities that satisfy*

$$\max(C_{i,j}) - \min(C_{i,j}) \leq 0.05 \frac{\Delta}{|\mathcal{A}|} \quad \textit{and} \quad \max(C_{j,i}) - \min(C_{j,i}) \leq 0.05 \frac{\Delta}{|\mathcal{A}|}$$

for all (i, j) in the matching returned by ComputeMatch′. *Suppose also that the confidence sets $\mathscr{C}$ contain the true utilities for all pairs of agents. Then the market outcome returned by* ComputeMatch′ *is stable with respect to the true utilities u.*

Remark. *Chapter 104 does not hold for* ComputeMatch*; its proof relies on the particular specification of the optimal primal-dual pair in* ComputeMatch′.

Using Lemma 104, we intuitively can bound the number of mistakes made by the algorithm by the number of samples needed to sufficiently reduce the size of the confidence sets. In Chapter H.4, we describe how we choose optimal primal-dual pairs in ComputeMatch′, prove Lemma 104, and provide a full proof of Chapter 103.

Chapter 103 opens the door to further exploring algorithmic properties of learning stable matchings. First, this result establishes fine-grained regret bounds, demonstrating the typical $O(\log T)$ regret bounds from the combinatorial bandits literature Chen et al. (2013) are achievable in our setting as well. Second, Chapter 103 provides insight into the number of mistakes made by the platform. In particular, we show within the proof of Chapter 103 that the platform fails to choose a matching that is stable with respect to u in at most $O(|\mathcal{A}|^4 \cdot \log(|\mathcal{A}|T)/\Delta^2)$ rounds.[13] This means that the platform selects a stable matching in at least $T - O(|\mathcal{A}|^5 \cdot \log(|\mathcal{A}|T)/\Delta^2) = T - O(\log T)$ of the rounds.

As we described, our bounds in Chapter 103 rely on choosing an appropriate primal-dual solution. An interesting direction for future work would be to provide further insight into how different methods for finding optimal primal-dual pairs affect both regret bounds and the trajectory of the selected market outcomes over time.

12.6.2 Search frictions and platform revenue

Next, we further ground Subset Instability by explicitly connecting it to the platform's revenue under a stylized economic model of search frictions. A major motivation for this is that it helps explain when an online platform can earn a profit in competitive settings, even when they start out with no information about agent preferences.

More specifically, we incorporate search frictions where an agent must lose utility ε in order to find an alternative to the given match (e.g., from the time spent finding an alternate partner, or from a cancellation fee). These search frictions weaken the requirements for stability: the platform now only needs matchings to be *ε-stable*:

$$u_i(j) + u_j(i) - 2\varepsilon \leq u_i(\mu_X(i)) + \tau_i + u_j(\mu_X(j)) + \tau_j$$

for all $(i, j) \in \mathcal{I} \times \mathcal{J}$ and $u_a(\mu_X(a)) + \tau_a \geq -\varepsilon$ for all $a \in \mathcal{A}$.[14]

To model revenue, we take the subsidy perspective on Subset Instability. Specifically, recall that Subset Instability is equal to the minimum subsidy needed to maintain stability (see Chapter 95). With search frictions, that subsidy can potentially be *negative*, thus allowing the platform to generate revenue. We are interested in analyzing the maximum revenue (minimum subsidy) the platform can generate while ensuring stability with high probability over all rounds. For realism, we also want this subsidy to be computed online using only

[13] The number of mistakes necessarily depends on the gap Δ because there exist utility functions u and $\tilde{u}$ where $\|u - \tilde{u}\|_\infty$ is arbitrary small, but where the stable market outcomes with respect to u and $\tilde{u}$ differ. To see this, consider a market where $\mathcal{I} = \{C\}$ and $\mathcal{J} = \{P\}$. Suppose that $u_C(P) = \tilde{u}_C(P) = 1$, while $u_P(C) = -1 + \varepsilon$ and $\tilde{u}_P(C) = -1 - \varepsilon$. Then, the maximum weight matchings under these utility functions differ: $\{(C, P)\}$ is the only maximum weight matching in the former, whereas $\emptyset$ is the only maximum weight matching in the latter.

[14] This definition corresponds to (X, τ) belonging to the weak ε-core of Shapley and Shubik (1966). We note that this definition also relaxes individual rationality. This formulation gives us the cleanest algorithmic results; while it can be extended to an analogue that does not relax individual rationality, it would involve bounds that (necessarily) depend on the specifics of agents' utilities.

information that the platform has access to, but it turns out we can do this with minimal modifications to our algorithm.

More formally, in this modified model, the platform must select an ε-stable matching in each round with high probability by choosing appropriate subsidies. That is, in round t, the platform selects a matching with transfers (X^t, τ^t) with the modification that the transfers need not be zero-sum. The transfers thus incorporate the amount that platform is subsidizing or charging agents for participation on the platform. The net profit of the platform is then $-\sum_{t=1}^{T} \sum_{a \in \mathcal{A}} \tau_a^t$. We impose the stability requirement that

$$\mathbb{P}[(X^t, \tau^t) \text{ is } \varepsilon\text{-stable for all } 1 \leq t \leq T] \geq 0.99.$$

Given this setup, we show the following:

Theorem 105. *For preference class $\mathcal{U}_{\text{unstructured}}$ (see Chapter 12.3), there exists an algorithm giving the platform*

$$\varepsilon T \sum_{t=1}^{T} |\mathcal{A}_t| - O\Big(|\mathcal{A}|\sqrt{nT}\sqrt{\log(|\mathcal{A}||T|)}\Big)$$

revenue in the presence of search frictions while maintaining stability with high probability.

Remark. *In particular if $\mathcal{A}_t = \mathcal{A}$ in every round, the platform will starting making a profit within $O(|\mathcal{A}|/\varepsilon^2 \cdot \log(|\mathcal{A}|/\varepsilon^2))$ rounds.*

We defer the proof of Chapter 105 to Chapter H.5.

Qualitatively, Theorem 105 captures that if the platform "pays to learn" in initial rounds, the information that it obtains will help it achieve a profit in the long run. We note that both the revenue objective and the model for search frictions that we consider in these preliminary results are stylized. An interesting direction for future work would be to integrate more realistic platform objectives and models for search frictions into the framework.

12.6.3 Matching with non-transferable utilities

While we have focused on matching with transferable utilities, utilities are not always transferable in practice, as in the cases of dating markets and college admissions (i.e., most people are not willing to date an undesirable partner in exchange for money, and a typical college admission slot is not sold for money). We can extend our findings to this setting following the model of matching with non-transferable utilities (NTU) Gale and Shapley (1962), which has also been studied in previous work Das and Kamenica (2005); Liu et al. (2020a); Cen and Shah (2022); Sankararaman et al. (2021). The definition of Subset Instability extends naturally and has advantages over the "utility difference" metric that is commonly used in prior work. Our algorithmic meta-approach also sheds new light on the convergence properties of the centralized UCB algorithm of Liu et al. (2020a).

The starting point of our instability measure is slightly different than in Chapter 12.4. Since stable matchings in the NTU model need not maximize total utility, we cannot define instability based on a maximum over all subsets of agents of the utility difference for that subset. On the other hand, the subsidy formulation of Subset Instability (see (12.4)) translates well to this setting. Our instability measure will correspond to the minimum amount the platform could subsidize agents so that individual rationality holds and no blocking pairs remain. For matching with NTU, we formalize this notion as follows:

Definition 11 (NTU Subset Instability)**.** *For utilities u and agents $\mathcal{A}$, the* NTU Subset Instability *$I(X; u, \mathcal{A})$ of a matching X is*

$$\begin{aligned} &\min_{s \in \mathrm{R}^{|\mathcal{A}|}} \sum_{a \in \mathcal{A}} s_a && (\dagger) \\ &s.t. \quad \min\big(u_i(j) - u_i(\mu_X(i)) - s_i, u_j(i) - u_j(\mu_X(j)) - s_j\big) \leq 0 && \forall (i,j) \in \mathcal{I} \times \mathcal{J} \\ &\qquad u_a(\mu_X(a)) + s_a \geq 0 && \forall a \in \mathcal{A} \\ &\qquad s_a \geq 0 && \forall a \in \mathcal{A}. \end{aligned}$$

NTU Subsidy Instability inherits some of the same appealing properties as Subsidy Instability.

Proposition 106 (Informal)**.** *NTU Subset Instability satisfies the following properties:*

1. *NTU Subset Instability is always nonnegative and is zero if and only if (X, τ) is stable.*

2. *NTU Subset Instability is Lipschitz continuous with respect to agent utilities. That is, for any matching X and any pair of utility functions u and $\tilde{u}$, it holds that:*

$$|I(X; u, \mathcal{A}) - I(X; uii, \mathcal{A})| \leq 2 \sum_{a \in \mathcal{A}} \|u_a - uii_a\|_\infty.$$

The proofs of this and subsequent results are deferred to Chapter H.6. Together, the preceding properties mean that NTU Subsidy Instability is useful as a regret measure for learning stable matchings.

As in the transferable utilities setting, Property 2 implies the existence of an explore-then-commit algorithm with $\widetilde{O}(|A|^{4/3}T^{2/3})$ regret. We show that this can be improved to a $\sqrt{T}$ dependence by adapting our approach from Chapter 12.5:

Theorem 107. *For matchings with non-transferable utilities, there exists an algorithm that for any utility function u incurs regret $R_T = O(|\mathcal{A}|^{3/2}\sqrt{T}\sqrt{\log(|\mathcal{A}|T)})$.*

While Theorem 107 illustrates that our approach easily generalizes to the NTU setting, we highlight two crucial differences between these settings. First, learning a stable matching is incomparable to learning a maximum weight matching because stable matchings do not maximize the sum of agents' utilities in the NTU setting. Next, the instability measure is *not* equivalent to the cumulative unhappiness of agents, unlike in the setting with transferable utilities. Intuitively, these definitions cease to be equivalent because non-transferable utilities render the problem more "discontinuous" and thus obstruct the duality results we applied earlier.

These results provide a preliminary application of our framework to the setting of matching with non-transferable utilities; an interesting direction for future inquiry would be to more thoroughly investigate notions of approximate stability and regret in this setting.

Comparison to the utility difference measure

It turns out that the algorithm underlying Chapter 107 is equivalent to the centralized UCB algorithm from previous work (Liu et al., 2020a; Cen and Shah, 2022), albeit derived from a different angle. However, an important difference is that Chapter 107 guarantees low regret relative to the incentive-aware NTU Subset Instability, as opposed to the incentive-unaware "utility difference" measure in prior work. In this section, we outline several properties that make our instability measure more suitable especially in the NTU setting. In particular, we show for utility difference that:

(a) There is no canonical formalization of utility difference when multiple stable matchings exist.

(b) The utility difference of a matching can be positive even if the matching is stable and negative even if the matching is unstable.

(c) Even when restricting to markets with unique stable matchings, the utility difference of a matching can be discontinuous in the true agent utilities. As a result, it does not allow for instance-independent regret bounds that are sublinear in T.

For (a), the utility difference requires specifying a stable matching to serve as a benchmark against which to measure relative utility. However, when multiple stable matchings exist, some ambiguity arises as to which one should be chosen as the benchmark. Because of this, previous works Das and Kamenica (2005); Liu et al. (2020a); Cen and Shah (2022); Sankararaman et al. (2021) study two different benchmarks. In particular, they assume providers' preferences are known and benchmark with respect to the customer-optimal and customer-pessimal stable matchings. For (b), notice that the utility difference for the maximum weight matching is negative, even though it is typically not stable in the NTU setting. Moreover, because of the ambiguity in the benchmark from (a), the utility difference may not be zero even when the matching is stable. For (c), to see that utility difference is not continuous as a function of the underlying agent utilities, consider the following example:

Example 8. *Consider a market where is a single customer i and two providers j_1 and j_2. Suppose their utility functions are given by $u_i(j_1) = \varepsilon$, $u_i(j_2) = 2\varepsilon$, $u_{j_1}(i) = 1$, and $u_{j_2}(i) = 0.5$. Then the unique stable matching $\{(i, j_2)\}$ has total utility $0.5 + 2\varepsilon$. Now, consider the perturbed utility function uii such that $uii_i(j_1) = 2\varepsilon$, $uii_i(j_2) = \varepsilon$, $uii_{j_1}(i) = 1$, and $uii_{j_2}(i) = 0.5$. For this perturbed utility function, the unique stable matching is $\{(i, j_1)\}$, which has total utility $1 + 2\varepsilon$. The utility difference (either optimal or pessimal) for matching*

$\{(i, j_2)\}$ is 0 for u and $0.5 + \varepsilon$ for uii. Since this holds for any $\varepsilon > 0$, taking $\varepsilon \to 0$ shows that utility difference is not continuous in the utility function.

That utility difference is discontinuous in agent utilities rules out the existence of bandit algorithms that achieve sublinear instance-independent regret when using utility difference as the regret measure. In particular, the analyses in previous work Liu et al. (2020a); Sankararaman et al. (2021); Cen and Shah (2022); Liu et al. (2021) focus entirely on *instance-dependent* regret bounds. They show that centralized UCB achieves logarithmic instance-dependent regret with respect to the utility difference relative to the customer-pessimal stable matching (but does not achieve sublinear regret with respect to the customer-optimal stable matching). Our insight here is that a new measure of instability can present a more appealing evaluation metric and paint a clearer picture of an algorithm's convergence to the *set* of stable matchings as a whole.

12.7 In what settings are equilibria learnable?

A core insight of our work is that, in a stochastic environment, "optimism in the face of uncertainty" can be effectively leveraged for the problem of learning stable matchings. This motivates us to ask: in what other settings, and with what other algorithmic methods, can equilibria be learned?

One interesting open direction is to understand when equilibria can be learned in *adversarial* environments where the utility functions can change between rounds. From an economic perspective, adversarial environments could capture evolving market conditions. In the adversarial bandit setting, most work relies on gradient-based algorithms instead of UCB-based algorithms to attain optimal regret bounds (see, e.g., Auer et al. (2002a); Abernethy et al. (2009)). Can these gradient-based algorithms similarly be adapted to Subset Instability?

Another interesting open direction is to consider more general market settings, even in stochastic environments. For example, within the context of matching markets, each agent might match to more than one agent on the other side of the market; and outside of matching markets, a buyer might purchase multiple units of multiple goods. In markets with transferable utilities, incentive-aligned outcomes can be captured by *Walrasian equilibria* (see, e.g., Bichler et al. (2021)). Can Subset Instability and our UCB-based algorithms be adapted to learning Walrasian equilibria in general?

Addressing these questions would provide a richer understanding of when and how large-scale, data-driven marketplaces can efficiently learn market equilibria.

Chapter 13

Model-Provider Triggering Distribution Shifts

This chapter is based on *"Regret Minimization with Performative Feedback"* (Jagadeesan et al., 2022), which is joint work with Tijana Zrnic and Celestine Mendler-Dünner.

13.1 Introduction

Perdomo et al. (2020) formalized this phenomenon under the name *performative prediction*. A key concept in this framework is the *distribution map*, which formalizes the dependence of the data distribution on the deployed predictive model. This object maps a model, encoded by a parameter vector θ, to a distribution $\mathcal{D}(\theta)$ over instances. Naturally, in a performative environment, a model's performance is measured on the distribution that results from its deployment. That is, given a loss function $\ell(z; \theta)$, which measures the learner's loss when they predict on instance z using model θ, we evaluate a model based on its *performative risk*, defined as

$$\mathrm{PR}(\theta) := \mathbb{E}_{z \sim \mathcal{D}(\theta)} \, \ell(z; \theta). \tag{13.1}$$

In contrast with the risk function studied in classical supervised learning, the performative risk takes an expectation over a model-dependent distribution. Importantly, this distribution is *unknown* ahead of time; for example, one can hardly anticipate the distribution of travel times induced by a traffic forecasting system without deploying the system first.

Due to this inherent uncertainty about $\mathcal{D}(\theta)$, it is not possible to find a model with low performative risk offline. The learner needs to interact with the environment and deploy models θ to explore the induced distributions $\mathcal{D}(\theta)$. Given the online nature of this task, we measure the loss incurred by deploying a sequence of models $\theta_1, \ldots, \theta_T$ by evaluating the *performative regret*:

$$\mathrm{Reg}(T) := \sum_{t=1}^{T} \left(\mathbb{E} \ \mathrm{PR}(\theta_t) - \min_{\theta} \mathrm{PR}(\theta) \right),$$

where the expectation is taken over the possible randomness in the choice of $\{\theta_t\}_{t=1}^T$. Performative regret measures the suboptimality of the deployed sequence of models relative to a *performative optimum* $\theta_{\mathrm{PO}} \in \arg\min_\theta \mathrm{PR}(\theta)$.

At first glance, performative regret minimization might seem equivalent to a classical bandit problem. Bandit solutions minimize regret while requiring only noisy zeroth-order access to the unknown reward function—in our case PR. The resulting regret bounds generally grow with some notion of complexity of the reward function.

However, a naive application of bandit baselines misses out on a crucial fact: performative regret minimization exhibits significantly richer feedback than bandit feedback. When deploying a model θ, the learner gains access to samples from the induced distribution $\mathcal{D}(\theta)$, rather than only a noisy estimate of the risk $\mathrm{PR}(\theta)$. We call this feedback model *performative feedback*. Together with the fact that the learner knows the loss $\ell(z;\theta)$, performative feedback can be used to inform the reward of unexplored arms. For instance, it allows the computation of an unbiased estimate of $\mathbb{E}_{z\sim\mathcal{D}(\theta)}\,\ell(z;\theta')$ for *any* point θ'.

To illustrate the power of this feedback model, consider the limiting case in which the performative effects entirely vanish and the distribution map is constant, i.e. $\mathcal{D}(\theta) \equiv \mathcal{D}_*$ for some fixed distribution $\mathcal{D}_*$ independent of θ. With zeroth-order feedback, the learner would still need to deploy different models to explore the landscape of PR and find a point with low risk. However, with performative feedback, a *single* deployment gives samples from $\mathcal{D}_*$, thus resolving all uncertainty in the objective (13.1) apart from finite-sample uncertainty. This raises the question: *with performative feedback, can one achieve regret bounds that scale only with the complexity of the distribution map, and not that of the performative risk?*

13.1.1 Our contribution

We study the problem of performative regret minimization based on performative feedback. Our main contribution is performative regret bounds that scale primarily with the complexity of the distribution map. The key conceptual idea is to apply bandits tools to carefully explore the distribution map, and then propagate this knowledge to the objective (13.1) in order to minimize performative regret.

Performative confidence bounds algorithm. Our main focus is on a setting where the distribution map is Lipschitz in an appropriate sense. We propose a new algorithm that takes advantage of performative feedback in order to construct non-trivial confidence bounds on the performative risk in unexplored regions of the parameter space and thus guide exploration. A crucial implication of these bounds is that the algorithm can discard highly suboptimal regions of the parameter space without ever deploying a model nearby. We summarize the regret guarantee of our *performative confidence bounds* algorithm:

Theorem 108 (Informal)**.** *Suppose that the distribution map $\mathcal{D}(\theta)$ is ε-Lipschitz and that the loss $\ell(z;\theta)$ is L_z-Lipschitz in z. Then, after T deployments, the performative confidence* *bounds algorithm achieves a regret bound of*

$$\mathrm{Reg}(T) = \tilde{\mathcal{O}}\left(\sqrt{T} + T^{\frac{d+1}{d+2}}(L_z\varepsilon)^{\frac{d}{d+2}}\right),$$

where d denotes the zooming dimension of the problem.

We compare the bound in Theorem 108 to a baseline Lipschitz bandits regret bound. The concept of zooming dimension stems from the work of Kleinberg et al. (2008) and serves as an instance-dependent notion of dimensionality. Kleinberg et al. (2008) showed that sublinear regret $\tilde{\mathcal{O}}\big(T^{\frac{d'+1}{d'+2}}L^{\frac{d'}{d'+2}}\big)$ can be achieved if the reward function is L-Lipschitz, where d' is a zooming dimension. The performative risk can be guaranteed to be Lipschitz if the distribution map is Lipschitz and the loss $\ell(z;\theta)$ is Lipschitz in *both* arguments.

The primary benefit of Theorem 108 is that our regret bound scales only with the Lipschitz constant of the distribution map, rather than the Lipschitz constant of the performative risk. In particular, our result allows $\mathrm{PR}(\theta)$ to be highly irregular as a function of θ, seeing that $\ell(z;\theta)$ as a function of θ is unconstrained. This difference becomes salient when $\varepsilon \to 0$, meaning that the performative effects vanish: our regret bound grows as $\tilde{\mathcal{O}}(\sqrt{T})$ in an essentially dimension-independent manner. More precisely, the dimension can only arise implicitly through a model class complexity term. On the other hand, the rate of classical Lipschitz bandits remains exponential in the dimension.

Another difference between our regret bound and that of Lipschitz bandits is in the zooming dimension. In particular, d' is a zooming dimension *no smaller than* the zooming dimension we obtain in Theorem 108. As we will elaborate on in later sections, the benefit we derive from the zooming dimension comes from the fact that it implicitly depends on the Lipschitz constant driving the objective, which is smaller when making full use of performative feedback.

Extension to location families. In addition, we study performative regret minimization for the special case where the distribution map has a *location family* form (Miller et al., 2021). We again prove regret bounds that scale only with the complexity of the distribution map, rather than the complexity of the performative risk. We adapt the LinUCB algorithm (Li et al., 2010) to learn the hidden parameters of the location family. This enables us to achieve a $\tilde{\mathcal{O}}(\sqrt{T})$ regret without placing any strong convexity assumptions on the performative risk that are required in (Miller et al., 2021). In particular, our result again allows $\mathrm{PR}(\theta)$ to be highly irregular as a function of θ.

Consequences for finding performative optima. While we have contextualized our work within online regret minimization, our performative confidence bounds algorithm has the additional property that it converges to the set of performative optima. Thus, if run for sufficiently many time steps, it generates a model with near-minimal performative risk: in particular, a model with risk at most $\tilde{\mathcal{O}}\left(T^{-\frac{1}{d+2}}(L_z\varepsilon)^{\frac{d}{d+2}}\right)$ greater than the minimum performative risk $\min_\theta \mathrm{PR}(\theta)$.

More broadly, our work establishes a connection between performative prediction and the bandits literature, which we believe is a worthwhile direction for future inquiry.

13.1.2 Related work

Performative prediction. Prior work on performative prediction has largely studied gradient-based optimization methods (Perdomo et al., 2020; Mendler-Dünner et al., 2020; Drusvyatskiy and Xiao, 2023; Brown et al., 2022a; Miller et al., 2021; Izzo et al., 2021; Maheshwari et al., 2022; Li and Wai, 2022; Ray et al., 2022; Dong et al., 2023). Many of the studied procedures only converge to *performatively stable* points, that is, points θ that satisfy the fixed-point condition $\theta \in \arg\min_{\theta'} \mathbb{E}_{z\sim\mathcal{D}(\theta)}\, \ell(z;\theta')$. In general, stable points are not minimizers of the performative risk (Perdomo et al., 2020; Miller et al., 2021), which implies that procedures converging to stable points do not achieve sublinear performative regret. There are exceptions in the literature that focus on finding performative optima (Miller et al., 2021; Izzo et al., 2021), but those algorithms rely on proving or assuming convexity of the performative risk; in this chapter we make no convexity assumptions. In fact, it is known that the performative risk can be nonconvex even when the loss $\ell(z;\theta)$ is convex and the performative effects are relatively weak (Perdomo et al., 2020; Miller et al., 2021). One other work that studies performative optimality, without imposing convexity, is that of Dong et al. (2023), but they focus on optimization heuristics that are not guaranteed to minimize performative regret.

Learning in Stackelberg games. Performative prediction is closely related to learning in *Stackelberg games*: if $\mathcal{D}(\theta)$ is thought of as a best response to the deployment of θ according to some unspecified utility function, then performative optima can be thought of as Stackelberg equilibria. There have been many works on learning dynamics in Stackelberg games in recent years (Balcan et al., 2015; Jin et al., 2020; Fiez et al., 2020; Fiez and Ratliff, 2021). Notably, Balcan et al. (2015) also study the benefit of a richer feedback model: they assume the agent's type is revealed after taking an action. When combined with a known agent-response model, this allows them to directly infer the loss of unexplored strategies. In contrast, performative feedback does not imply full-information feedback. One instance of performative prediction that has an explicit Stackelberg structure, meaning $\mathcal{D}(\theta)$ is defined as a best response, is *strategic classification* (Hardt et al., 2016). Several works have studied learning dynamics in strategic classification (Dong et al., 2018; Chen et al., 2020b; Bechavod et al., 2021; Zrnic et al., 2021); notably, Dong et al. (2018) and Chen et al. (2020b) provide solutions that minimize Stackelberg regret, of which performative regret is an analog in the performative prediction context. However, all of these works rely on strong structural assumptions, such as linearity of the predictor or convexity of the risk function, which significantly reduce the amount of necessary exploration compared to the mild Lipschitzness conditions we impose in our work.

Continuum-armed bandits. Particularly inspiring for our work is the literature on *continuum-armed bandits* (Agrawal, 1995; Kleinberg, 2004; Auer et al., 2007; Kleinberg et al., 2008; Podimata and Slivkins, 2021). As we will elaborate on in Chapter 13.2, performative prediction can be cast as a Lipschitz continuum-armed bandit problem. However, while this means that one can use an off-the-shelf Lipschitz bandit algorithm to minimize performative regret, this would generally be a conservative solution. After "pulling an arm" θ in performative prediction the learner observes samples from $\mathcal{D}(\theta)$. As explained earlier, in combination with the structure of our objective, this feedback model is more powerful than classical bandit feedback, where a noisy version of the mean reward at θ is observed. Moreover, it is fundamentally different from other partial-feedback and side-information models studied in the literature, e.g. (Mannor and Shamir, 2011; Kocák et al., 2014; Wu et al., 2015; Cohen et al., 2016).

13.1.3 Preliminaries

Performative prediction, set up as an online learning problem, can be formalized as follows. The learner chooses models θ in the parameter space $\Theta \subset \mathbb{R}^{d_\Theta}$. We assume[1] $\max\{\|\theta\| : \theta \in \Theta\} \leq 1$ for simplicity. The expected loss of model θ is given by $\mathrm{PR}(\theta) = \mathrm{E}_{z\sim\mathcal{D}(\theta)}\ell(z;\theta)$. We assume that the objective function is bounded so that $\ell(z;\theta) \in [0,1]$ for all z and θ.

At every time step t, the learner chooses a model θ_t and observes a constant number m_0 of i.i.d. samples,

$$\{z_t^{(i)}\}_{i\in[m_0]}, \text{ where } z_t^{(i)} \sim \mathcal{D}(\theta_t).$$

The regret incurred by choosing θ_t at time step t is $\Delta(\theta_t) := \mathrm{PR}(\theta_t) - \mathrm{PR}(\theta_{\mathrm{PO}})$, where θ_{PO} is the performative optimum.

The constant m_0 quantifies how many samples the learner can collect in a time window determined by how often they incur regret. For example, at the beginning of each week the learner might update the model, and thus at the end of each week they incur regret for the model they chose to deploy. In that case, m_0 is the number of samples the learner collects per week. Note that a learner with larger m_0 collects an empirical distribution that more accurately reflects $\mathcal{D}(\theta_t)$ and thus naturally minimizes regret at a faster rate.

To formally disentangle the effects of the parameter vector θ on the performative risk through the distribution map and the loss function, we use the notion of the *decoupled performative risk* (Perdomo et al., 2020):

$$\mathrm{R}(\theta,\theta') := \mathrm{E}_{z\sim\mathcal{D}(\theta)}\,\ell(z;\theta').$$

This object captures the risk incurred by a model θ' on the distribution $\mathcal{D}(\theta)$. Note that $\mathrm{PR}(\theta) = \mathrm{R}(\theta,\theta)$ by definition.

To measure the complexity of the distribution map we consider how much the distribution $\mathcal{D}(\theta)$ can change with changes in θ, as formalized by *ε-sensitivity*.

[1]Throughout we use $\|\cdot\|$ to denote the ℓ_2-norm for vectors and the operator norm for matrices.

Assumption 6 (ε-sensitivity (Perdomo et al., 2020))**.** *A distribution map $\mathcal{D}(\cdot)$ is ε-sensitive if for any pair $\theta, \theta' \in \Theta$ it holds that*

$$\mathcal{W}(\mathcal{D}(\theta), \mathcal{D}(\theta')) \leq \varepsilon \|\theta - \theta'\|,$$

where $\mathcal{W}$ denotes the Wasserstein-1 distance.

In the context of a traffic forecasting app, ε can be thought of as being proportional to the size of the user base of the app. When $\mathcal{D}(\theta)$ arises from the aggregate behavior of strategic agents manipulating their features in response to a model θ, the sensitivity ε grows when features are more easily manipulable.

13.2 A black-box bandits approach

Performative regret minimization can be set up as a continuum-armed bandits problem where an arm corresponds to a choice of model parameters θ. Performative feedback is sufficient to simulate noisy zeroth-order feedback about the reward function, as assumed in bandits. When we deploy θ_t, the samples from $\mathcal{D}(\theta_t)$ enable us to compute an unbiased estimate

$$\widehat{\mathrm{PR}}(\theta_t) = \frac{1}{m_0} \sum_{i=1}^{m_0} \ell\big(z_t^{(i)}; \theta_t\big)$$

of the risk $\mathrm{PR}(\theta_t)$. Moreover, since we assume the loss function is bounded, the noise in the estimate $\widehat{\mathrm{PR}}(\theta_t)$ is subgaussian, as typically required in bandits.

A standard condition that makes continuum-armed bandit problems tractable is a bound on how fast the reward can change when moving from one arm to a nearby arm. Formally, this regularity is ensured by assuming Lipschitzness of the reward function—in our case, Lipschitzness of the performative risk.

The dependence of $\mathrm{PR}(\theta)$ on θ is twofold, as seen in Equation (13.1). Thus, the most natural way to ensure that $\mathrm{PR}(\theta)$ is Lipschitz is to ensure that each of these two dependencies is Lipschitz. This yields the following bound:

Lemma 109 (Lipschitzness of PR)**.** *If the loss $\ell(z; \theta)$ is L_z-Lipschitz in z and L_θ-Lipschitz in θ and the distribution map is ε-sensitive, then the performative risk is $(L_\theta + \varepsilon L_z)$-Lipschitz.*

The intuition behind Lemma 109 is that $\mathrm{PR}(\theta)$ is guaranteed to be Lipschitz if $\mathrm{R}(\theta, \theta')$ is Lipschitz in each argument individually. Lipschitzness in the second argument follows from requiring that the loss be Lipschitz in θ. Lipschitzness in the first argument follows from combining Lipschitzness of the loss in z and ε-sensitivity of the distribution map.

13.2.1 Adaptive zooming

Once we have established Lipschitzness of the performative risk, we can apply techniques from the Lipschitz bandits literature. Kleinberg et al. (2008) proposed a bandit algorithm that adaptively discretizes promising regions of the space of arms, using Lipschitzness of the reward function to bound the additional loss due to discretization. Their method, called the *zooming algorithm*, will serve as a baseline for our problem. The algorithm enjoys an instance-dependent regret that takes advantage of nice problem instances, while maintaining tight guarantees in the worst case. The rate depends on the *zooming dimension*, which is upper bounded in the worst case by the dimension of the full space d_Θ.

Proposition 110 (Zooming algorithm (Kleinberg et al., 2008))**.** *Suppose $m_0 = o(\log T)$. Then, after T deployments, the zooming algorithm achieves a regret bound of*

$$\mathrm{Reg}(T) = \mathcal{O}\left(T^{\frac{d+1}{d+2}}\left(\frac{\log T}{m_0}\right)^{\frac{1}{d+2}}(L_\theta + \varepsilon L_z)^{\frac{d}{d+2}}\right),$$

where d denotes the $(L_\theta + \varepsilon L_z)$-zooming dimension.

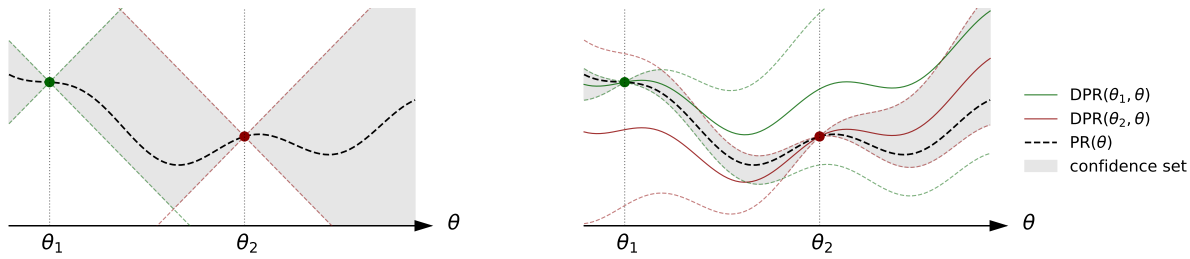
The zooming dimension quantifies the niceness of a problem instance by measuring the size of a covering of near-optimal arms, instead of the entire parameter space. Roughly speaking, if the reward function is very "flat" in that there are many near-optimal points, then the zooming dimension is close to the dimension d_Θ of the parameter space. However, if the reward has sufficient curvature, then the zooming dimension can be much smaller than d_Θ. The zooming dimension is defined formally as follows:

Definition 12 (α-zooming dimension)**.** *A performative prediction problem instance has α-zooming dimension equal to d if any minimal s-cover of any subset of $\{\theta : \Delta(\theta) \leq 16\alpha s\}$ includes at most a constant multiple of $(3/s)^d$ elements from $\{\theta : 16\alpha r \leq \Delta(\theta) < 32\alpha r\}$, for all $0 < r \leq s \leq 1$.*

For well-behaved instances, the definition intuitively requires every minimal s-cover of $\{\theta : 16\alpha r \leq \Delta(\theta) < 32\alpha r\}$ to have size at most of order $(3/s)^d$. Definition 12 slightly differs from the definition presented in (Kleinberg et al., 2008) and makes the dependence on the Lipschitz constant explicit; we use Definition 12 to later ease the comparison to our new algorithm. The differences between the two definitions are minor technicalities that we do not expect to alter the zooming dimension in a meaningful way, neither formally nor conceptually. See Appendix I.4.1 for a discussion.

13.3 Making use of performative feedback

In this section, we illustrate how we can take advantage of performative feedback beyond computing a point estimate of the deployed model's risk. For now, we ignore finite-sample



(a) Baseline confidence bounds

(b) Performative confidence bounds

Figure 13.1: Confidence bounds after deploying θ_1 and θ_2. (left) Confidence bounds via Lipschitzness, as stated in Equation (13.2). (right) Performative confidence bounds, as stated in Equation (13.3). The performative feedback model used for this illustration can be found in Appendix I.5.

considerations and assume access to the entire distribution $\mathcal{D}(\theta)$ after deploying a model θ. We will address finite-sample uncertainty when presenting our main algorithm in the next section.

13.3.1 Constructing performative confidence bounds

First, we demonstrate how performative feedback allows constructing tighter confidence bounds on the performative risk of unexplored models, compared to only relying on Lipschitzness of the risk function $\mathrm{PR}(\theta)$.

Suppose we deploy a set of models $\mathcal{S} \subseteq \Theta$ and for each $\theta \in \mathcal{S}$ we observe $\mathcal{D}(\theta)$. Then, under the regularity conditions of Lemma 109, we can bound the risk of any $\theta' \in \Theta$ as

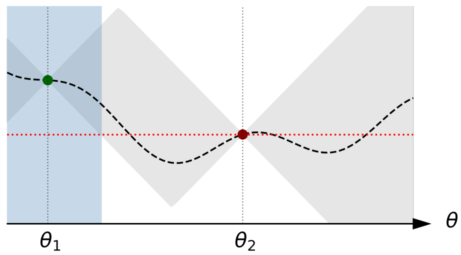
$$\max_{\theta \in \mathcal{S}} \mathrm{PR}(\theta) - (L_\theta + L_z\varepsilon)\|\theta - \theta'\| \leq \mathrm{PR}(\theta') \leq \min_{\theta \in \mathcal{S}} \mathrm{PR}(\theta) + (L_\theta + L_z\varepsilon)\|\theta - \theta'\|. \tag{13.2}$$

These confidence bounds only use $\mathcal{D}(\theta)$ for the purpose of computing $\mathrm{PR}(\theta)$ and rely on Lipschitzness to construct confidence sets around the risk of unexplored models. However, in light of the structure of our objective function (13.1), the bounds in Equation (13.2) do not make full use of performative feedback; in particular, access to $\mathcal{D}(\theta)$ actually allows us to evaluate $\mathrm{R}(\theta, \theta')$ for *any* θ'. Importantly, this information can further reduce our uncertainty about $\mathrm{PR}(\theta')$, and we can bound:

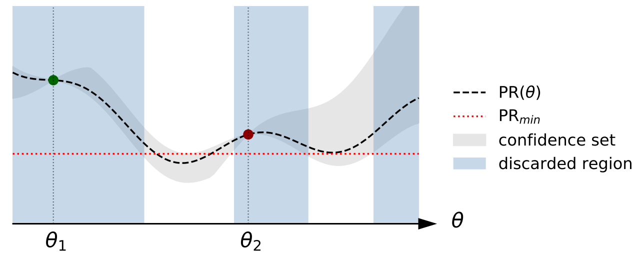
$$\begin{aligned} \mathrm{PR}(\theta') &= \mathrm{R}(\theta, \theta') + (\mathrm{R}(\theta', \theta') - \mathrm{R}(\theta, \theta')) \\ &\leq \mathrm{R}(\theta, \theta') + L_z\varepsilon\|\theta - \theta'\|. \end{aligned}$$

Thus we can get tighter bounds on the performative risk at an unexplored parameter θ':

$$\max_{\theta \in \mathcal{S}} \mathrm{R}(\theta, \theta') - L_z\varepsilon\|\theta - \theta'\| \leq \mathrm{PR}(\theta') \leq \min_{\theta \in \mathcal{S}} \mathrm{R}(\theta, \theta') + L_z\varepsilon\|\theta - \theta'\|. \tag{13.3}$$



(a) Baseline confidence bounds



(b) Performative confidence bounds

Figure 13.2: Performative feedback allows discarding unexplored suboptimal models even in regions that have not been explored. A model θ is discarded if $\mathrm{PR}_{\mathrm{LB}}(\theta) > \mathrm{PR}_{\min}$. The loss function and feedback model are the same as in Figure 13.1.

We call the confidence bounds computed in (13.3) *performative confidence bounds*. In Figure 13.1, we visualize and contrast these confidence bounds with the confidence bounds obtained via Lipschitzness. We observe that by computing R we can significantly tighten the confidence regions.

The tightness of the confidence bounds depends on the set $\mathcal{S}$ of deployed models. By choosing a cover of the parameter space, we can get an estimate of the performative risk that has low approximation error on the whole parameter space.

Proposition 111. *Let $\mathcal{S}_\gamma$ be a γ-cover of Θ and suppose we deploy all models $\theta \in \mathcal{S}_\gamma$. Then, using performative feedback we can compute an estimate of the performative risk $\widehat{\mathrm{PR}}(\theta)$ such that for any $\theta \in \Theta$ it holds that*

$$|\mathrm{PR}(\theta) - \widehat{\mathrm{PR}}(\theta)| \leq \gamma L_z \varepsilon.$$

Proposition 111 implies that after exploring the cover $\mathcal{S}_\gamma$, we can find a model whose suboptimality is at most $\mathcal{O}(\gamma L_z \varepsilon)$. To contextualize the bound in Proposition 111, consider an approach that uses the same cover $\mathcal{S}_\gamma$ but only relies on zeroth-order feedback, that is, $\{\mathrm{PR}(\theta) : \theta \in \mathcal{S}_\gamma\}$. Then, the only feasible estimate of PR over the whole space is $\widehat{\mathrm{PR}}(\theta) = \mathrm{PR}(\Pi_{\mathcal{S}_\gamma}(\theta))$, where $\Pi_{\mathcal{S}_\gamma}(\theta) = \arg\min_{\theta' \in \mathcal{S}_\gamma} \|\theta - \theta'\|$ is the projection onto the cover $\mathcal{S}_\gamma$. This zeroth-order approach only guarantees an accuracy of $|\mathrm{PR}(\theta) - \widehat{\mathrm{PR}}(\theta)| \leq (L_z\varepsilon + L_\theta)\gamma$, a strictly weaker approximation than the one in Proposition 111.

13.3.2 Sequential elimination of suboptimal models

Now we show how performative confidence bounds can guide exploration. Specifically, we show that every deployment informs the risk of unexplored models, which allows us to sequentially discard suboptimal regions of the parameter space.

To develop a formal procedure for discarding points, let $\mathrm{PR}_{\mathrm{LB}}(\theta)$ denote a lower confidence bound on $\mathrm{PR}(\theta)$ and $\mathrm{PR}_{\min}$ denote an upper confidence bound on $\mathrm{PR}(\theta_{\mathrm{PO}})$ based on the information from the models deployed so far:

$$\begin{aligned}\mathrm{PR}_{\mathrm{LB}}(\theta) &= \max_{\theta' \text{ already deployed}} \left(\mathrm{R}(\theta', \theta) - L_z \varepsilon \|\theta - \theta'\|\right), \\ \mathrm{PR}_{\min} &= \min_{\theta \in \Theta} \min_{\theta' \text{ already deployed}} \left(\mathrm{R}(\theta', \theta) + L_z \varepsilon \|\theta' - \theta\|\right).\end{aligned}$$

It is not difficult to see that the following lower bound on the suboptimality of model θ holds:

Proposition 112. *For all $\theta \in \Theta$, we have $\Delta(\theta) \geq \mathrm{PR}_{\mathrm{LB}}(\theta) - \mathrm{PR}_{\min}$.*

In particular, models θ with $\mathrm{PR}_{\mathrm{LB}}(\theta) > \mathrm{PR}_{\min}$ cannot be optimal. We recall our toy example from Figure 13.1 and illustrate in Figure 13.2 the parameter configurations we can discard after the deployment of two models, θ_1 and θ_2. We can see that access to R allows us to discard a large portion of the parameter space, and, in contrast to the baseline black-box approach, it is possible to discard regions of the space that have not been explored.

13.4 Performative confidence bounds algorithm

We introduce our main algorithm that builds on the two insights from the previous section. We furthermore provide a rigorous, finite-sample analysis of its guarantees.

13.4.1 Algorithm overview

Our *performative confidence bounds* algorithm, formally stated in Algorithm 6, takes advantage of performative feedback by assessing the risk of unexplored models and thus guiding exploration. We give an overview of the main steps.

Inspired by the successive elimination algorithm (Even-Dar et al., 2002), the algorithm keeps track of and refines an *active* set of models $\mathcal{A} \subseteq \Theta$. Roughly speaking, active models are those that are estimated to have low risk and only they are admissible to deploy. To deal with finite-sample uncertainty, the algorithm proceeds in phases which progressively refine the precision of the finite-sample risk estimates. More precisely, in phase p the algorithm chooses an error tolerance γ_p and deploys a model for n_p steps. In each step m_0 samples induced by the deployed model are collected, and n_p is chosen so that the inferred estimates of R are γ_p-accurate. Formally, if θ is deployed in phase p, we collect an empirical distribution $\widehat{\mathcal{D}}(\theta)$ of $n_p m_0$ samples so that $|\widehat{\mathrm{R}}(\theta, \theta') - \mathrm{R}(\theta, \theta')| \leq \gamma_p$ for all θ' with high probability, where

$$\widehat{\mathrm{R}}(\theta, \theta') := \mathop{\mathbb{E}}_{z \sim \widehat{\mathcal{D}}(\theta)} \ell(z; \theta').$$

These estimates of R are used to construct performative confidence bounds and refine $\mathcal{A}$.

Algorithm 6: Performative Confidence Bounds Algorithm

Input: time horizon T, number of samples collected per step m_0, sensitivity parameter ε, Lipschitz constant L_z, complexity bound $\mathfrak{C}$

Initialize $\mathcal{A} \leftarrow \Theta$;
for *phase* $p = 0, 1, \ldots$ **do**
  Set error tolerance $\gamma_p = 2^{-p}$ and net radius $r_p = \frac{\gamma_p}{L_z \varepsilon}$;
  Let $n_p = \left\lceil \frac{\left(2\mathfrak{C} + 3\sqrt{\log T}\right)^2}{\gamma_p^2 m_0} \right\rceil$;
  Initialize $\mathcal{S}_p \leftarrow \mathcal{N}_{r_p}(\mathcal{A})$; `// Initialize` $\mathcal{S}_p$ `to minimal` r_p`-cover of` $\mathcal{A}$
  Initialize $\mathcal{P}_p \leftarrow \emptyset$;
  while $\mathcal{S}_p \neq \emptyset$ **do**
    Draw $\theta_{\text{net}} \in \mathcal{S}_p$ uniformly at random;
    Deploy θ_t for n_p steps to form $\widehat{\mathrm{R}}(\theta_{\text{net}}, \cdot)$;
    $\mathcal{S}_p \leftarrow \mathcal{S}_p \setminus \theta_{\text{net}}$;
    $\mathcal{P}_p \leftarrow \mathcal{P}_p \cup \theta_{\text{net}}$; `// Update set of deployed models`
    $\mathrm{PR}_{\min} \leftarrow \min_{\theta \in \Theta} \min_{\theta' \in \mathcal{P}_p} \widehat{\mathrm{DPR}}(\theta', \theta) + L_z \varepsilon \|\theta' - \theta\|$; `// Update estimate of` $\mathrm{PR}(\theta_{\mathrm{PO}})$
    $\mathrm{PR}_{\mathrm{LB}}(\theta) \leftarrow \max_{\theta' \in \mathcal{P}_p} \left(\widehat{\mathrm{DPR}}(\theta', \theta) - L_z \varepsilon \|\theta' - \theta\| \right) \forall \theta \in \mathcal{A}$; `// Update LB for all models`
    $\mathcal{A} \leftarrow \mathcal{A} \setminus \{\theta \in \mathcal{A} : \mathrm{PR}_{\mathrm{LB}}(\theta) > \mathrm{PR}_{\min} + 2\gamma_p\}$; `// Update active region`
    $\mathcal{S}_p \leftarrow \mathcal{S}_p \setminus \{\theta \in \mathcal{S}_p : \mathrm{Ball}_{r_p}(\theta) \cap \mathcal{A} = \emptyset\}$; `// Remove net points in deactivated regions`

Each phase begins by constructing a net of the current active set $\mathcal{A}$. The points in the net are sequentially deployed in the phase, unless they are deemed to be suboptimal based on previous deployments in that phase and are in that case eliminated. During phase p, we denote by $\mathcal{P}_p$ the running set of deployed points and by $\mathcal{S}_p$ the running set of net points that have not been discarded. We initialize $\mathcal{S}_p$ to a minimal r_p-net of the current set of active points $\mathcal{A}$, denoted $\mathcal{N}_{r_p}(\mathcal{A})$, where r_p is proportional to γ_p. A net point θ gets eliminated from $\mathcal{S}_p$ if no point in $\mathrm{Ball}_{r_p}(\theta) := \{\theta' \in \Theta : \|\theta' - \theta\| \leq r_p\}$ is active. This means that we may deploy suboptimal points in the net if they help inform active points nearby.

13.4.2 Comparison with adaptive zooming algorithm

While we borrow the idea of an instance-dependent zooming dimension from Kleinberg et al. (2008), Algorithm 6 and its analysis are substantially different from prior work. In particular, Kleinberg et al. (2008) study an adaptive zooming algorithm which combines a UCB-based approach with an arm activation step. Adapting this method to our setting encounters several obstacles that we describe below.

First, a naive application of the adaptive zooming algorithm proposed by Kleinberg et al. (2008) does not lead to sublinear regret in our setting, unless we assume Lipschitzness of PR. Their rule for activating new arms requires that the reward of arms within a given radius in Euclidean distance of the pulled arm is similar. However, without Lipschitzness of PR, there is no radius that would ensure this property.

Given the shortcomings of this exploration strategy, one might imagine that selecting a better distance between arms, e.g. one based on performative confidence bounds, would result in a better algorithm. A natural distance function would be $d(\theta, \theta')$ taken as (an empirical estimate of) $\mathrm{PR}(\theta) - \mathrm{R}(\theta, \theta') + L_z\varepsilon\|\theta - \theta'\|$. The challenge is that the analysis in (Kleinberg et al., 2008) explicitly requires symmetry of the distance function, which $d(\theta, \theta')$ violates.

Therefore, to single out the $L_z\varepsilon$ dependence, it is necessary to disentangle learning the structure of the distribution map from the elimination of arms based on reward, which is in stark contrast with UCB-style adaptive zooming algorithms. Algorithm 6 achieves this by relying on a novel adaptation of successive elimination.

13.4.3 Regret bound

Before we state the regret bound for Algorithm 6, let us comment on an important component in the analysis. Recall that throughout the algorithm we operate with finite-sample estimates of the decoupled performative risk to bound the risk of unexplored models. Specifically, for any deployed θ, we make use of $\widehat{\mathrm{DPR}}(\theta, \theta')$ for *all* θ'. Since we need these estimates to be valid simultaneously for all θ', we rely on uniform convergence. As such, the Rademacher complexity of the loss function class naturally enters the bound.

Definition 13 (Rademacher complexity)**.** *Given a loss function $\ell(z;\theta)$, we define $\mathfrak{C}^*(\ell)$ to be:*

$$\mathfrak{C}^*(\ell) = \sup_{\theta\in\Theta} \sup_{n\in\mathbb{N}} \sqrt{n} \cdot \mathop{\mathbb{E}}_{\varepsilon, z^\theta} \left(\sup_{\theta'\in\Theta} \left| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \ell(z_j^\theta; \theta') \right| \right),$$

where $\varepsilon_j \sim$ Rademacher and $z_j^\theta \sim \mathcal{D}(\theta)$, $\forall j \in [n]$, which are all independent of each other.

Now we can state our regret guarantee for Algorithm 6.

Theorem 113 (Main regret bound)**.** *Assume the loss $\ell(z;\theta)$ is L_z-Lipschitz in z and let ε denote the sensitivity of the distribution map. Suppose that $\mathfrak{C}$ is any value such that $\mathfrak{C}^*(\ell) \leq \mathfrak{C}$ and $m_0 = o(\mathcal{B}^2_{\log T, \mathfrak{C}})$, where $\mathcal{B}_{\log T, \mathfrak{C}} := \sqrt{\log T} + \mathfrak{C}$. Then, after T time steps, Algorithm 6 achieves a regret bound of*

$$\mathrm{Reg}(T) = \mathcal{O}\left(T^{\frac{d+1}{d+2}} \left(\frac{(L_z\varepsilon)^d \mathcal{B}^2_{\log T, \mathfrak{C}}}{m_0} \right)^{\frac{1}{d+2}} + \sqrt{T}\, \frac{\mathcal{B}_{\log T, \mathfrak{C}}}{\sqrt{m_0}} \right),$$

where d is the $(L_z\varepsilon)$-sequential zooming dimension (see Definition 14).

Remark 2 (Consequences for finding performative optima). *Algorithm 6 has the additional property that it generates a model with near-minimal performative risk. In particular, an intermediate step in the proof of Theorem 113 shows if T is sufficiently large, the final iterate θ_T of Algorithm 6 satisfies:*

$$\mathbb{E}\left[\mathrm{PR}(\theta_T) - \min_{\theta\in\Theta} \mathrm{PR}(\theta)\right] \leq \mathcal{O}\left(T^{-\frac{1}{d+2}}\left(\frac{(L_z\varepsilon)^d \mathcal{B}^2_{\log T,\mathfrak{C}}}{m_0}\right)^{\frac{1}{d+2}}\right),$$

where d is the $(L_z\varepsilon)$-zooming dimension.

Notice that the regret in Theorem 113 depends on the sequential zooming dimension (formally defined in Definition 14). This sequential variant of zooming dimension accounts for the sequential elimination of models within each phase. We will show in the next section that the sequential zooming dimension is upper bounded by the usual zooming dimension (see Proposition 114).

The primary advantage of Theorem 113 over the Lipschitz bandit baseline can be seen by examining the first term in the regret bound. This term resembles the black-box regret bound from Chapter 13.2; however, the key difference is that that the bound of Theorem 113 depends on the complexity of the distribution map rather than that of the performative risk. In particular, the Lipschitz constant is $L_z\varepsilon$ and not $L_\theta + L_z\varepsilon$. The advantage is pronounced when $\varepsilon \to 0$, making the first term of the bound in Theorem 113 vanish so only the $\mathcal{O}(\sqrt{T})$ term remains. On the other hand, the bound in Proposition 114 maintains an exponential dimension dependence.

Taking the limit as $\varepsilon \to 0$ also reveals why the second term in the bound emerges. Even if the distribution map is constant, there is regret arising from finite-sample error. This is a key conceptual difference in the meaning of Lipschitzness of the distribution map versus that of the performative risk: $L_\theta + L_z\varepsilon$ being 0 implies that PR is flat and thus all models are optimal, while performative regret minimization is nontrivial even if $L_z\varepsilon = 0$. Unlike the first term, the second term due to finite samples is dimension-independent apart from any dependence implicit in the Rademacher complexity.

We note that the presence of the Rademacher complexity term $\mathfrak{C}^*(\ell)$ makes a direct comparison of the bound in Theorem 113 and the bound in Proposition 114 subtle. When the Rademacher complexity is very high, the regret bound in Theorem 113 may be worse. Nonetheless, for many natural function classes, the Rademacher complexity is polynomial in the dimension; in these cases, Theorem 113 can substantially outperform the regret bound in Proposition 114.

Another key feature of the regret bound in Theorem 113 worth highlighting is the zooming dimension. Definition 12 allows us to directly compare the dimension in Theorem 113 with the dimension in Proposition 114: the $(L_z\varepsilon)$-zooming dimension of Algorithm 6 is no larger than, and most likely smaller than, the $(L_\theta + L_z\varepsilon)$-zooming dimension in the black-box approach. Moreover, the sequential variant of zooming dimension in Theorem 113 can further reduce the dimension.

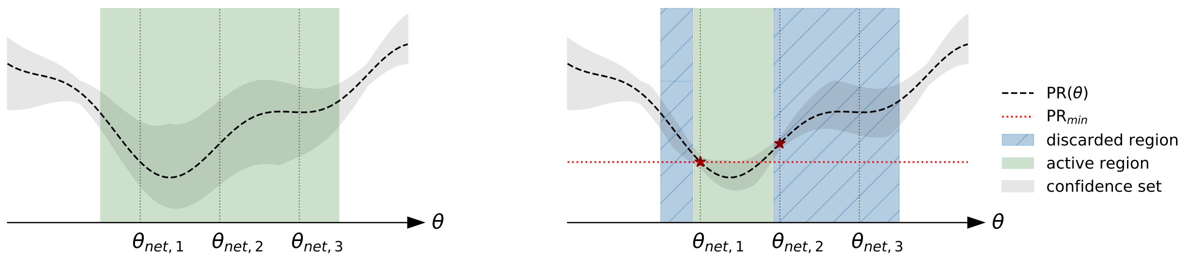


Figure 13.3: Sequential deployment of models allows Algorithm 6 to eliminate points from $\mathcal{S}_p$, reducing the number of deployments during the phase. We see how the deployment of $\theta_{\mathrm{net},1}$ and $\theta_{\mathrm{net},2}$ allows one to eliminate $\theta_{\mathrm{net},3}$.

Finally, the main assumption underpinning the bound in Theorem 113 is that R is $(L_z\varepsilon)$-Lipschitz in its first argument. Assumption 6 coupled with Lipschitzness of the loss in the data achieves this. However, this property can hold with different regularity assumptions on the distribution map and loss function; e.g., if the loss is bounded and the distribution map is Lipschitz in total variation distance.

13.4.4 Sequential zooming dimension

The zooming dimension of Definition 12 does not take into account that, using performative feedback, our algorithm can eliminate unexplored models *within* a phase. We illustrate the benefits of this sequential exploration strategy in Figure 13.3, where the deployment of two models is sufficient to eliminate the remaining model in the cover. This motivates a sequential definition of zooming dimension that captures the benefits of sequential exploration.

To set up the definition of *sequential zooming dimension*, we need to introduce some notation. For a set of points $\mathcal{S}$, enumeration $\pi : \mathcal{S} \to \{1, \ldots, |\mathcal{S}|\}$ that specifies an ordering on $\mathcal{S}$, and number $k \in \{1, \ldots, |\mathcal{S}|\}$, let

$$\begin{aligned}
\mathrm{PR}_{\mathrm{LB}}(\theta; k) &:= \max_{\theta' \in \mathcal{S}: \pi(\theta') < k} \left(\mathrm{R}(\theta', \theta) - L_z\varepsilon\|\theta - \theta'\|\right), \\
\mathrm{PR}^s_{\mathrm{LB}}(k) &:= \min_{\theta \in \mathrm{Ball}_s(\pi^{-1}(k))} \mathrm{PR}_{\mathrm{LB}}(\theta; k), \\
\mathrm{PR}_{\mathrm{min}}(k) &:= \min_{\theta} \min_{\theta' \in \mathcal{S}: \pi(\theta') < k} \left(\mathrm{R}(\theta', \theta) + L_z\varepsilon\|\theta' - \theta\|\right).
\end{aligned}$$

Here, $\mathrm{PR}_{\mathrm{LB}}(\theta; k)$ is a lower bound on $\mathrm{PR}(\theta)$ arising from the first $k-1$ deployments of the phase. Similarly, $\mathrm{PR}^s_{\mathrm{LB}}(k)$ captures the minimal lower confidence bound on the performative risk for any point in an s-ball around the k-th deployed model, $\pi^{-1}(k)$. Finally, $\mathrm{PR}_{\mathrm{min}}(k)$ captures an upper bound on $\mathrm{PR}(\theta_{\mathrm{PO}})$, estimated from the first $k-1$ deployments.

Using the above terms, we see that $\mathrm{PR}^s_{\mathrm{LB}}(k) \leq \mathrm{PR}_{\mathrm{min}}(k) + 4\alpha s$ is the population version of the condition that a model in the cover does not get discarded. The sequential zooming dimension captures the maximal number of models in each suboptimality band that can be deployed.

Definition 14 (Sequential zooming dimension)**.** *A performative prediction problem instance has α-zooming dimension equal to d if for any minimal s-cover $\mathcal{S}$ of any subset of $\{\theta : \Delta(\theta) \leq 16\alpha s\}$ and all $0 < r \leq s \leq 1$, the expected number of models $\theta \in \mathcal{S} \cap \{\theta : 16\alpha r \leq \Delta(\theta) < 32\alpha r\}$ with*

$$\mathrm{PR}^{s}_{\mathrm{LB}}(\pi(\theta)) \leq \mathrm{PR}_{\min}(\pi(\theta)) + 4\alpha s \tag{13.4}$$

is at most a constant multiple of $(3/s)^d$, where the expectation is taken over a uniformly sampled enumeration $\pi : \mathcal{S} \to \{1, \ldots, |\mathcal{S}|\}$.

The sequential zooming dimension is bounded by the zooming dimension in Definition 12.

Proposition 114. *For all $\alpha > 0$, the α-zooming dimension is at least as large as the α-sequential zooming dimension.*

The claim of Proposition 114 follows by definition. To see this, let d be the α-zooming dimension. This means that $\mathcal{S}$ includes at most a constant multiple of $(3/s)^d$ elements from $\{\theta : 16\alpha r \leq \Delta(\theta) < 32\alpha r\}$, for all $0 < r \leq s \leq 1$. This immediately guarantees that the subset of $\mathcal{S}$ characterized by (13.4) is at most a multiple of $(3/s)^d$, as desired.

In Appendix I.4.2, we provide an example where the sequential zooming dimension is *strictly* smaller than the zooming dimension.

13.5 Regret minimization for location families

In this section, we show how further knowledge about the structure of the distribution map can help reduce the complexity of performative regret minimization, without necessarily implying favorable structure of the performative risk. Once again, we apply our guiding principle of focusing exploration on learning the distribution map. Since the loss function is known, we can extrapolate knowledge about the distribution map to estimate the performative risk.

We focus on the setting of *location families* (Miller et al., 2021), which are distribution maps that depend on θ via a linear shift. More precisely, location families are distribution maps of the form $z \sim \mathcal{D}(\theta) \Leftrightarrow z \stackrel{d}{=} z_0 + \mu_*^\top \theta$, where $\mu_* \in \mathbb{R}^{d_\Theta \times m}$ is an unknown matrix and $z_0 \in \mathbb{R}^m$ is a zero-mean subgaussian sample from a base distribution $\mathcal{D}_0$.

Example 9 (Strategic classification)**.** *Location families arise in strategic classification (Hardt et al., 2016), where agents strategically manipulate their features in response to a deployed model. Suppose the learner uses a linear predictor $f_\theta(x) = \theta^T x$ and the agents incur quadratic cost for changing their original features x to manipulated features x', $c(x, x') = \frac{1}{2}(x - x')\Lambda(x - x')$. Then, the best response of an agent, typically modeled as $x_{\mathrm{BR}}(\theta) = \arg\max_{x'} f_\theta(x') -$*

$c(x, x')$, satisfies the location family structural assumption with z_0 being the agent's original features and $\mu^* = \Lambda^{-1}$.

At a high level, our algorithm can be described as follows: at every step t, the learner deploys a model θ_t and collects m_0 samples from $\mathcal{D}(\theta_t)$. We will write $\bar{z}_t := \frac{1}{m_0}\sum_{i=1}^{m_0} z_t^{(i)}$ for the corresponding sample average at time t. Then, based on all samples collected so far, the algorithm computes the least-squares estimate of μ_* along with a confidence region for μ_*. In the next step the algorithm picks the model that minimizes a lower confidence bound $\mathrm{PR}_{\mathrm{LB}}(\theta)$. See Algorithm 7 for details.

This algorithm is inspired by LinUCB (Li et al., 2010), a standard bandits algorithm for linear rewards whose regret scales as $\tilde{\mathcal{O}}(d\sqrt{T})$, where d is the dimension of the linear map. Importantly, unlike in the LinUCB analysis, our objective function $\mathrm{PR}(\theta)$ is *not* linear in θ. Still, the nature of performative feedback allows us to learn the hidden linear structure in the distribution map and apply this knowledge to obtain confidence bounds on the performative risk. Below we state our algorithm for performative regret minimization for location families together with its regret guarantees.

Theorem 115. *Suppose that $\ell(z;\theta)$ is L_z-Lipschitz in z, $\mathcal{D}_0$ is 1-subgaussian, and $m_0 = o(\log T)$. Then, after T time steps, Algorithm 7 achieves a regret bound of*

$$\mathrm{Reg}(T) = \tilde{\mathcal{O}}\left(\frac{1}{\sqrt{m_0}}\max\{L_z, 1\}\sqrt{T}\max\left\{d_\Theta, \sqrt{d_\Theta m}\right\}\right).$$

Remark 3. *For simplicity, we assume that $\mathcal{D}_0$ is known in Algorithm 7. This assumption is justified, for example, when we have plenty of historical data about a population, before any model deployment. We note that Theorem 115 can be extended to the case where we only have a finite data set from $\mathcal{D}_0$, by relying on a uniform convergence argument.*

Algorithm 7: Performative Regret Minimization for Location Families

---

Input: time horizon T, number of samples collected per step m_0, base distribution $\mathcal{D}_0$, bound M_* such that $\|\mu_*\| \le M_*$

Initialize confidence set $\mathcal{C}_1 \leftarrow \{\mu : \|\mu\| \le M_*\}$;

for *step* $t = 1, 2, \ldots$ **do**

  $\mathrm{PR}_{\mathrm{LB}}(\theta) \leftarrow \min_{\mu\in\mathcal{C}_t} \mathbb{E}_{z_0\sim\mathcal{D}_0}\, \ell(z_0 + \mu\theta; \theta)\ \forall\theta\in\Theta$; `// Update LB for all models`

  Deploy $\theta_t = \arg\min_\theta \mathrm{PR}_{\mathrm{LB}}(\theta)$; `// Deploy model with lowest LB`

  Compute $\bar{z}_t = \frac{1}{m_0}\sum_{i=1}^{m_0} z_t^{(i)}$ from collected samples;

  Let $\Sigma_t \leftarrow \sum_{i=1}^{t}\theta_i\theta_i^\top + \frac{1}{m_0}I$;

  $\hat{\mu}_t \leftarrow \Sigma_t^{-1}\left(\sum_{i=1}^{t}\theta_i\bar{z}_i^\top\right)$; `// Update estimate of` μ_*

  $\mathcal{C}_{t+1} \leftarrow \left\{\mu : \left\|\Sigma_t^{1/2}(\hat{\mu}_t - \mu)\right\| < \frac{M_* + \sqrt{8m_0 + 8\log T + 2d_\Theta \log\left(1 + \frac{Tm_0}{d_\Theta}\right)}}{\sqrt{m_0}}\right\}$; `// Update confidence set`

Theorem 115 shows that by leveraging the hidden linear structure of the distribution map, Algorithm 7 inherits the $\tilde{\mathcal{O}}(\sqrt{T})$ rate of LinUCB. This bears resemblance to the regret bound in Theorem 113 that also scaled primarily with the complexity of the distribution map. Furthermore, similarly to Algorithm 6, we see that the regret bound for Algorithm 7 holds while allowing the loss to have arbitrary dependence on θ. For example, the loss need not be convex and, as a result, the performative risk need not be convex either.

We conclude by comparing Theorem 115 to (Miller et al., 2021), which provided an algorithm for finding performative optima for location families in the special case when the performative risk is *strongly convex*. Converting their optimization error into a regret bound yields a bound of $\mathcal{O}(\sqrt{T}(d_\Theta + m))$. While this bears resemblance to Theorem 115, the rates are not directly comparable. The algorithm by Miller et al. (2021) does not assume knowledge of the base distribution $\mathcal{D}_0$, but rather deploys the model $\theta = 0$ in initial steps to collect samples from $\mathcal{D}_0$ (see Remark 3 for how to combine this strategy with our algorithm). In any case, the main benefit of Theorem 115 is that it applies to a more general setting, placing significantly fewer restrictions on the loss function and the performative risk.

13.6 Future directions

Having illuminated the connection between performative prediction and bandit problems, our work opens the door for interesting further investigations. We highlight several directions we consider promising.

Structural knowledge of the distribution map. Domain knowledge about performative distribution shifts is sometimes available: for example, a parametric approximation to the aggregate response (Miller et al., 2021; Izzo et al., 2021), a microfoundations model for individual behavior (Hardt et al., 2016; Jagadeesan et al., 2021), or basic constraints on the agents' action set (Chen et al., 2020b). For linear shifts, we demonstrated how such structural knowledge about the distribution map can help guide exploration. We expect this principle to apply to other structures of $\mathcal{D}(\theta)$.

Consequences of exploration. An important limitation of exploration in performative environments are social welfare concerns. Performative shifts can rarely be analyzed offline and every model deployment is consequential for the population the model acts upon. The ability to discard highly suboptimal regions of the parameter space without having to deploy a model within is highly appealing from a welfare perspective. Beyond this, we believe that incorporating constraints on what constitutes safe exploration (Wu et al., 2016; Turchetta et al., 2019; Kazerouni et al., 2017) is crucial for performative optimization in practice.

Costs of a new deployment. Our notion of regret quantifies the statistical complexity of regret minimization, but it does not differentiate between collecting more samples induced by the currently deployed model and deploying a new model. This difference has previously been studied by Mendler-Dünner et al. (2020) in the context of stochastic retraining methods. Due to the costs associated with a new deployment, collecting more samples from the same model typically comes at a reduced cost for the learner, and there may be a better notion of regret that reflects this.

Adapting to unknown sensitivity. Our algorithm relies on knowing $L_z\varepsilon$. While the Lipschitzness of a classifier in the data has been studied in the context of adversarial robustness (Szegedy et al., 2013; Cisse et al., 2017; Hein and Andriushchenko, 2017; Yang et al., 2020), which could help inform L_z, the sensitivity ε of an environment is generally unknown. Adapting the tools by Bubeck et al. (2011) could help relax the requirement of a known sensitivity.

"Best of both worlds" algorithm. When the Rademacher complexity of the function class is high, the Lipschitz bandit baseline may provide a better regret bound than Algorithm 6. It would be an interesting task for future work to design an algorithm that intersects the confidence sets of both algorithms and inherits the better of the two regret bounds.

Chapter 14

An AI Agent Interacting with a Human Agent

This is based on *"Impact of Decentralized Learning on Player Utilities in Stackelberg Games"* (Donahue et al., 2024b), which is joint work with Kate Donahue, Nicole Immorlica, Brendan Lucier, and Alex Slivkins.

14.1 Introduction

When learning agents such as recommender systems or chatbots are deployed into the world, these learning agents often repeatedly interact with other learning agents (such as humans). For example, a recommender system—through repeated interactions with a user—learns which content to suggest to the user, while the user simultaneously learns their own preferences over content (Example 11). As another example, a chatbot such as ChatGPT—during a chat session—can iteratively refine its generated content to the user's stylistic preferences, while the user (or prompt engineering agent) simultaneously learns how to best interact with the chatbot (Example 10).

These two-agent systems—among many others—share the following structural features. The environments are *decentralized* (both agents operate autonomously, without central coordination of their actions). Furthermore, the environments are *sequential* (one agent always chooses their action first[1]) and *misaligned* (the agents can obtain different utilities for the same pair of actions[2]). Finally, the environments exhibit *learning* (both agents learn

[1]E.g., a recommender system recommends a slate of items and the user picks among them; in a chatbot session, the user picks a prompt that the LLM responds to.

[2]Misalignment could arise from fundamental differences in agent preferences: the engagement metrics of recommender systems rarely align with user welfare (e.g., (Milli et al., 2021)); or a chatbot might be trained to optimize societal preferences or cultural norms (e.g., avoiding violent language) which conflict with individual user preferences (e.g., (Bakker et al., 2022)). Misalignment could also arise from misspecification, if the metrics that the AI system optimizes do not perfectly capture user preferences (e.g., (Zhuang and Hadfield-Menell, 2020)).

from repeated interactions about which actions to take). For such two-agent environments, core questions of interest include: *how quickly does this two-agent system learn, and what are the implications for each agent's objective?*

In the absence of learning, the interaction between misaligned agents taking sequential actions is formalized by Stackelberg games. In this setting, the leader chooses an action first and the follower chooses an action to respond. The two agents are allowed to have distinct utility functions over pairs of actions. The standard benchmark is the *Stackelberg equilibrium*, where the leader picks the best action they can, assuming that the follower will pick their own best response. However, this classical solution concept is tailored to the full-information setting where both players know their own utilities and the leader knows how the follower will best respond; in fact, the static Stackelberg game framework breaks down when agents instead must *learn* these utilities from noisy feedback.

Our focus is on a decentralized Stackelberg learning environment. In this setting, the leader and the follower repeatedly interact and each make decisions about which actions to take, where each agent only observes their own realized stochastic rewards. In this environment, it is natural to model each player's learning process as a multi-armed bandit algorithm[3] which learns over time which arms (actions) to pull. A unique feature of this sequential two-player learning environment is that agents must learn in two separate ways—first, both agents learn their own (fixed) preferences from stochastic observations, and secondly, the leader needs to learn and adapt to the follower's (evolving) responses to the leader's actions—which complicates the design of learning algorithms.

In this chapter, we initiate the study of how this learning environment impacts *both* the leader and follower's utility, motivated by how both objectives are of societal interest in natural real-world settings (see Example 10 and Example 11). Rather than only focusing on the regret of the leader as is typical in learning in Stackelberg games, we thus examine the *maximum regret* of the two agents. Our main contributions are to design appropriate benchmarks for each agent and to construct algorithms which achieve near-optimal regret against these benchmarks. Our results apply to the most general setting which allows for *arbitrary relationships between the two player's utilities.*

- **Linear regret for Stackelberg benchmarks:** We first show that the player utilities in the Stackelberg equilibrium are fundamentally unachievable and necessarily lead to linear regret for at least one agent (Theorem 116).

- **Relaxed benchmarks:** The possibility of linear regret motivates us to design relaxed benchmarks which are more tolerant to the other agent being suboptimal. We thus define *γ-tolerant benchmarks* (Definition 15), which account for incomplete learning: the benchmark captures an agent's worst-case utility if the other agent is up to γ-suboptimal.[4]

[3]See Slivkins (2019); Lattimore and Szepesvári (2020) for textbook treatments of multi-armed bandits.

[4]Chapter 14.4 describes our benchmark and γ in greater detail, and Chapter 14.1.1 compares our benchmarks to prior work.

- **Regret bounds:** Using the γ-tolerant benchmarks, we construct algorithms for the leader and follower that achieve $O(T^{2/3})$ regret (Theorem 119). Surprisingly, this dependence on T is unavoidable, and *any* pair of algorithms achieves $\Omega(T^{2/3})$ regret (Theorem 120). Nonetheless, under relaxed settings—either with a weaker benchmark or when players agree on which pairs of actions are meaningfully different[5]—we show that faster learning (i.e., $O(\sqrt{T})$ regret) is possible (Chapter 14.5).

From an algorithmic perspective, our results provide insight into which bandit algorithms for the leader allow for low regret for both players. Out-of-box stochastic algorithms do not provide this guarantee: for example, both agents choosing `ExploreThenCommit` can lead to linear regret even for the γ-tolerant benchmarks (Proposition 117). The intuition is that since the follower's actions can change between time steps, the leader is not operating in a stochastic environment; as a result, the follower's exploration phase can distort the leader's learning. This motivates us to design algorithms where *the leader waits for the follower to partially converge before starting to learn*: `ExploreThenCommitThrowOut` (Algorithm 9) and `ExploreThenUCB` (Algorithm 10). The more sophisticated of these two algorithms, `ExploreThenUCB` (Algorithm 10), guarantees a $T^{2/3}$ regret bound when the follower applies any algorithm with certain properties (i.e., high-probability instantaneous regret bounds) (Theorem 119). We then consider two relaxed environments where the leader no longer needs to worry about being overly distorted by the follower; in these environments, *the leader can start learning before the follower has partially converged*, which enables $O(\sqrt{T})$ regret bounds (Theorems 121 and 122).

More broadly, our work takes a step towards assessing the utility of *both learning agents* in decentralized, misaligned environments. Our model and results capture the general setting where the player utilities are arbitrarily related, where players might not even agree upon which pairs of actions give similar or different rewards. This motivated us to design benchmarks which are tolerant to small errors in the other player. We hope that our benchmarks and algorithms serve as a starting point for assessing when two-agent learning systems in misaligned environments can ensure high utility for both agents.

14.1.1 Related Work

Most closely related is the work on learning in Stackelberg games (SGs) where both players incur stochastic rewards. Bai et al. (2021); Gan et al. (2023) focus on the centralized setting where the learner controls the actions and observes the rewards of both players; in contrast, we study a decentralized setting where each player controls their own actions and only observes their own rewards. Nonetheless, the benchmarks proposed in these papers are related to the γ-tolerant benchmarks that we consider, but with some key differences. For the leader's utility, their benchmark is equivalent to our γ-tolerant benchmark with a fixed value of ε

[5]We formalize this as a continuity requirement on the utilities (Chapter 14.5). This requirement allows players to be misaligned (e.g. different preferences), but requires them to agree on which outcomes are *substantially different* from each other.

(rather than an inf over $\varepsilon \leq \gamma$ with a regularizer). For the follower's utility, their benchmark only ensures ε-optimality with respect to the leader's selected action; in contrast, we consider a different style of follower benchmark that is more conceptually similar to the benchmark for the leader. Also, we study regret, whereas they study the speed of convergence.

Several papers study *decentralized* online learning in SGs. Camara et al. (2020); Collina et al. (2024) posit that the follower runs a no-counterfactual-internal-regret algorithm and design no-regret algorithms for the leader. However, they assume *strong alignment* between the players' rewards: Camara et al. (2020) requires that a follower choosing an ε-suboptimal action only results in an $O(\varepsilon)$ utility loss for the leader.[6] Collina et al. (2024) partially relax this assumption, but still require the existence of *stable* actions for the leader. In contrast, we do not place any alignment conditions: in fact, misalignement is the driver of our linear regret result for the original Stackelberg benchmarks (Theorem 116). Other differences are that we focus on stochastic, rather than adversarial, rewards, and our benchmark is independent of the follower's choice of learning algorithm.[7] Haghtalab et al. (2023) takes a different perspective and considers the follower running a *calibrated* algorithm. They design a leader algorithm which waits for the follower to partially converge, and show that the Stackelberg value is obtained in the limit as $T \to \infty$. In contrast, we focus on *instance-independent regret bounds* for a fixed time horizon T, which requires us to relax the benchmark. Other differences are we focus on stochastic, rather than deterministic, rewards, we assume the follower observes the leader's action, and we consider the follower's utility in addition to the leader's utility.

The literature on learning in SGs is vast and includes many other variations. Many works (e.g., (Letchford et al., 2009; Balcan et al., 2015; Zhao et al., 2023; Lauffer et al., 2023)) consider the leader performing (offline or) online learning and followers myopically best-responding. Other model variants studied include the leader strategizing against a follower who is running a no-regret learning algorithm (Braverman et al., 2018; Deng et al., 2019; Guruganesh et al., 2024; Brown et al., 2023; Lin and Chen, 2024), the leader and follower both running gradient-based algorithms (Fiez et al., 2019; Goktas et al., 2022), non-myopic followers who best-respond to a discounted utility over future time steps (Haghtalab et al., 2022b; Hajiaghayi et al., 2024), repeated game formulations under complete information (Zuo and Tang, 2015; Collina et al., 2023) the leader offering a menu of actions to the follower (Han et al., 2024), the (human) follower having cognitive biases in responding (Pita et al., 2009), and both players having side information (Harris et al., 2024). Other works have studied learning in structured SGs, including delegated choice (e.g., (Kleinberg and Kleinberg, 2018; Hajiaghayi et al., 2024)), strategic classification (e.g., (Dong et al., 2018; Chen et al., 2020a; Zrnic et al., 2021; Ahmadi et al., 2021)), performative prediction (e.g., (Perdomo et al., 2020; Miller et al., 2021)), pricing under buyer and seller uncertainty (e.g., (Guo et al.,

[6]See Assumption 2 in Camara et al. (2020). Appendix D therein considers some relaxations, but they lead to $\Omega(T)$ worst-case regret.

[7]However, our regret bounds assume that the follower's algorithm gracefully improves over time, see Chapter 14.2.4.

2024)), contract theory (e.g., (Zhu et al., 2022)), cake cutting (e.g., (Brânzei et al., 2024)), and aligned utilities (e.g., (Kao et al., 2022)).

Our work also connects to a broader literature on interacting learners. This literature examines interactions between *multiple bandit learners*, studying aspects such as the convergence of systems of no-regret learners to coarse correlated and correlated equilibrium (e.g. (Daskalakis et al., 2011; 2021; Anagnostides et al., 2022)), multiple bandit learners competing for market share (e.g., (Aridor et al., 2025; Jagadeesan et al., 2023c)), and multiple autobidding algorithms competing in an auction (e.g., (Borgs et al., 2007; Balseiro and Gur, 2019; Lucier et al., 2024)). Most closely related to this chapter is *corralling bandit algorithms* (e.g., (Agarwal et al., 2017; Pacchiano et al., 2020)), where a "master algorithm" dynamically chooses among several "base algorithms": our decentralized learning environment in the case of aligned player utilities is essentially an instance of corralling bandits, with the "base algorithms" corresponding to different leader actions. The interacting learner literature also examines *human-algorithm collaboration* studying aspects such as misalignment between engagement metrics and user welfare (e.g., (Ekstrand and Willemsen, 2016; Milli et al., 2021; Stray et al., 2021; Kleinberg et al., 2024)), impact of underspecification on human-AI misalignment (e.g., (Zhuang and Hadfield-Menell, 2020)), and "assistive" algorithmic tools (e.g. (Chan et al., 2019)). Most closely related to this chapter is work on online learning in subset selection and conformal prediction, where goals often revolve around selecting a subset of items to present to a learning user (Straitouri and Rodriguez, 2023; Corvelo Benz and Rodriguez, 2023; Straitouri et al., 2023; Wang et al., 2022; Donahue et al., 2024a; Brown and Agarwal, 2024; Yao et al., 2022), often with the goal of achieving complementarity (Bansal et al., 2021). The related area of *human-AI interaction* (see (Preece et al., 1994; Kim, 2015; MacKenzie, 2024; Lazar et al., 2017) for textbook treatments) studies similar questions, often from a more behavioral angle. More broadly, the interacting learner literature also studies applied domains including *multi-agent reinforcement learning* (see Zhang et al. (2021) for a survey) and *federated learning* (see Yang et al. (2019) for a survey).

14.2 Model and assumptions

In this section we describe our formal model. We first define an instance $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ in our setup, which captures the setup of the underlying static Stackelberg game. Let $\mathcal{A}$ be a finite action set for the leader (Player 1) and let $\mathcal{B}$ be a finite action set for the follower (Player 2). Let $v_i(a, b) \in [0, 1]$ denote Player i's value (i.e., mean reward) for the leader choosing a and the follower choosing b. The Stackelberg equilibrium takes the following form. Let $b^*(a)$ be the best-response with respect to the follower's rewards:[8] $b^*(a) = \arg\max_{b \in \mathcal{B}} v_2(a, b)$. The Stackelberg equilibrium (a^*, b^*) is defined to be the best action the leader can take, assuming that the follower will exactly best-respond:

$$a^* = \arg\max_{a \in \mathcal{A}} v_1(a, b^*(a)) \text{ and } b^* = b^*(a^*).$$

[8] In case of ties in follower utility, $b^*(a)$ is the best-response with lowest leader utility.

Note that for simplicity, we restrict both players to pure strategies.[9]

In this chapter, we move from the static Stackelberg Equilibrium environment to a repeated dynamic environment which we call a *decentralized Stackelberg game (DSG)*. A DSG operates over T time steps where each player selects actions using a multi-armed bandit algorithm. A DSG is *sequential*: at each time step t, the leader chooses an action $a_t \in \mathcal{A}$ and then the follower chooses an action $b_t \in \mathcal{B}$. A DSG is also *decentralized*: each player i can observe their own stochastic rewards but not the stochastic rewards of the other player. We measure *regret* for each player i by their cumulative reward across all time steps relative to a benchmark.

14.2.1 Interaction between players

In a DSG, the interaction between the leader and follower proceeds as follows. The leader chooses an algorithm $\texttt{ALG}_1$ mapping their history (formalized below) of observed actions and rewards to a distributions over actions $\mathcal{A}$, and the follower similarly chooses an algorithm $\texttt{ALG}_2$ mapping the leader's action and the follower's history to a distribution over actions $\mathcal{B}$. After the players select algorithms, the interaction between the leader and the follower proceeds as follows at each time step t:

1. The leader chooses action $a_t \sim \texttt{ALG}_1(H_{1,t})$ as a function of their history $H_{1,t}$ and reveals a_t to the follower.

2. After observing a_t, the follower chooses action $b_t \sim \texttt{ALG}_2(a_t, H_{2,t})$ as a function of their history $H_{2,t}$.

3. Players 1 and 2 incur stochastic rewards $r_{1,t}(a_t, b_t) \sim \mathcal{N}(v_1(a_t, b_t), 1)$ and $r_{2,t}(a_t, b_t) \sim \mathcal{N}(v_2(a_t, b_t), 1)$. The noise distribution is Gaussian with unit variance[10].

This interaction captures that the players are *dynamic* in their learning: in particular, this framework is sufficiently general to capture a wide range of learning strategies. However, we do not study the *meta-game* where agents strategically pick learning algorithms (e.g., see Kolumbus and Nisan (2022) for an example of a work studying the meta-game). We believe that our model captures many real-world environments such as user-chatbot interactions and recommender system-user interactions, where agents learn about their incurred rewards from past interactions even if they do not actively optimize the higher-order manner in which they learn. See Chapter 14.2.3 (Example 10 and Example 11) for more details of how these real-world examples are captured by our model.

Information structures. Having described how the players interact, we next discuss the players' history, which further enforces decentralization. Each player i can only observe their

[9]Other works (e.g., Bai et al. (2021)) similarly restricts both players to pure strategies.

[10]We assume the reward distributions are independent across time steps and players. We make the Gaussian assumption for simplicity, and we expect that our results would likely extend to subgaussian Bernoulli distributions.

own reward $r_{i,t}(a_t, b_t)$ (and cannot observe the reward of the other player). In a *strongly decentralized Stackelberg game (StrongDSG)*, the follower can observe the leader's action a_t, but the leader cannot observe the follower's action b_t, whereas in a *weakly decentralized Stackelberg game (WeakDSG)*, the leader can additionally observe the follower's action b_t. Notation for each player's histories is presented in Chapter J.1.4.

Note that a strongly decentralized Stackelberg game (StrongDSG) restricts what information the leader has access to, and is thus *more challenging* than a weakly decentralized Stackelberg game (WeakDSG). One motivation for a strongly decentralized Stackelberg game is interpretability: the leader and follower may be taking actions in spaces that are not mutually understandable (e.g. a chatbot's representation of human preferences may not be interpretable). Most of our positive results (i.e., algorithm constructions) focus on StrongDSGs, whereas our negative results apply to both StrongDSGs and WeakDSGs.

14.2.2 Measuring regret

As is typical in multi-armed bandits, we measure performance by the *regret* of each player with respect to a benchmark, where higher benchmarks make learning more challenging (further detail on benchmarks is in Chapters 14.3 and 14.4). For each player $i \in \{1, 2\}$, given a benchmark β_i, we define the (expected) regret of Player i on instance $\mathcal{I}$ to be:

$$R_i(T; \mathcal{I}) = \beta_i \cdot T - \left(\sum_{t=1}^{T} \mathbb{E}[r_{i,t}(a_t, b_t)] \right)$$

where the expectation is over randomness in the algorithm and in the stochastic rewards. Given action sets $\mathcal{A}$ and $\mathcal{B}$, we let $R_i(T)$ denote the worst-case regret across all value functions v_1 and v_2 on instances of the form $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$.

Our goal is to obtain sublinear worst-case regret for *both* players: that is, we will assess $\max(R_1(T), R_2(T))$. We note that this challenging goal is a departure from previous work which has typically focused solely on sublinear regret for the leader (see Chapter 14.1.1). Our motivation for selecting this objective is that (a) a human could be either the leader or the follower, and (b) societal welfare may demand that we care about the utility of both the leader and the follower (discussed further in Chapter 14.2.3).

14.2.3 Real-world examples

We describe two real-world examples which fit into our framework.

Example 10 (User-chatbot interaction)**.** *Consider user-chatbot interactions where the user (e.g., a human or a prompt engineer) selects a prompt and the chatbot (e.g., an LLM-based application such as chatGPT) selects a response. We model the user as the leader and the chatbot as the follower: the user picks a (perhaps high-level) prompt or prompt engineering technique $a \in \mathcal{A}$, and the chatbot picks a response or style of response $b \in \mathcal{B}$. Repeated* *interactions may occur within a single chatbot session, such as with ChatGPT, where sessions can be resumed when the user logs in at a later time. An example of such an interaction is where the user repeatedly asks for help with similar queries (e.g. content generation or help with technical tasks) and learns better prompt engineering techniques (Chen et al., 2023), while the chatbot learns how to best respond to this user by using the session history as its context (Hong et al., 2023; Pan et al., 2024). The user and chatbot may have misaligned rewards for each prompt-output pair: this misalignment could arise from fundamental differences in preferences if the chatbot is trained to optimize societal preferences or cultural norms (e.g., avoiding violent language) which conflict with individual user preferences (Bakker et al., 2022). Misalignment could also arise from unintentional misspecification if chatbot optimizes a metric which does not fully capture user preferences (e.g., if the user has an imperfect ability to communicate preferences (Zhuang and Hadfield-Menell, 2020)).*

Example 11 (User-recommender system interaction)**.** *Consider interactions between a recommender system and a user, where the recommender system gives a slate (or subset) of items $a \in \mathcal{A}$ to the user, and the user picks an action $b \in \mathcal{B}$ from the slate. When the user returns to the same content recommendation system (e.g. a Netflix/Hulu user with a profile) many times, this becomes a repeated game where both the recommendation system and user learn about their preferences (Hajiaghayi et al., 2024). Again, misalignment could occur from the engagement metric being misaligned with user welfare (Milli et al., 2021) or for unintentional reasons (e.g., misspecification due to discrete thumbs up/thumbs down user feedback, since true preferences are more nuanced).*

Examples 10-11 motivate why our objective is to minimize regret for both the leader and the follower. First, we may inherently care about utility for the human, who could be either the leader (Example 10) or the follower (Example 11). Secondly, we may also care about utility for the algorithmic tools: for example, a recommendation system that fails to make money may go out of business, or in certain cases, the chatbot/recommender system may better capture societal objectives than certain humans.

14.2.4 Assumptions on the follower's algorithm $\texttt{ALG}_2$

Finally, we present some technical assumptions on the follower's algorithm. When we analyze regret in Chapters 14.4-14.5, many of our constructions do not require that the follower run a particular algorithm, but instead allow the follower to run any algorithm that has sufficiently good performance along certain fine-grained performance metrics that capture the extent to which an algorithm's performance gracefully improves over time.

These fine-grained performance metrics capture the follower's errors while learning. These errors are captured by the difference between $v_2(a_t, b_t)$ (the follower's realized mean reward) and the $\max_{b \in \mathcal{B}} v_2(a_t, b)$ (the best mean reward that the follower could achieve for the leader's action a_t). Note that this measure of suboptimality $\max_{b \in \mathcal{B}} v_2(a_t, b) - v_2(a_t, b_t)$ captures how well the follower is best-responding to the leader's action. This differs from our main notion of regret in Chapter 14.2.2, which captures the follower's level of discontent relative to a fixed benchmark.

For intuition, we first describe these performance metrics—*high-probability instantaneous regret* and *high-probability anytime regret*—for a typical single-bandit learner which acts in isolation. For the single learner setting, instances $\mathcal{I} = (\mathcal{C}, v)$ capture a single action set and a single value function. *High-probability instantaneous regret* measures the suboptimality of the arms that the algorithm pick arms at each time step. More formally, an algorithm acting over an action space $\mathcal{C}$ satisfies high-probability instantaneous regret g if for any instance $\mathcal{I} = (\mathcal{C}, v)$:

$$\mathbb{P}\left[\forall t \in [T] \mid v(c_t) \geq \max_{c \in \mathcal{C}} v(c) - g(t, T, \mathcal{C})\right] \geq 1 - T^{-3},$$

where the probability captures randomness in the algorithm and in the stochastic rewards. A *high-probability anytime regret* bound guarantees that the regret bound for the algorithm holds with high probability at every time step t. More formally, an algorithm acting over an action space $\mathcal{C}$ satisfies high-probability anytime regret bound h if for any instance $\mathcal{I} = (\mathcal{C}, v)$, it holds that:

$$\mathbb{P}\left[\forall t \in [T] \mid \sum_{t' \leq t} v(c_{t'}) \geq \sum_{t' \leq t} \max_{c \in \mathcal{C}} v(c) - h(t, T, \mathcal{C})\right] \geq 1 - T^{-3},$$

where the randomness is over the algorithm.[11]

In a DSG, we will require similar properties to hold for the follower's algorithm $\texttt{ALG}_2$, but we take account how the algorithm $\texttt{ALG}_2$ depends on the action a_t which is selected by the leader's algorithm $\texttt{ALG}_1$. Let $n_{t+1}(a)$ be the number of times that arm a has been pulled prior to the $(t+1)$th time step. An algorithm $\texttt{ALG}_2$ satisfies a high-probability instantaneous regret bound of g if for any $\texttt{ALG}_1$ chosen by the leader and any $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$, it holds that:

$$\mathbb{P}\left[\forall t \in [T], a \in \mathcal{A} \mid v_2(a_t, b_t) \geq \max_{b \in \mathcal{B}} v_2(a_t, b) - g(n_{t+1}(a), T, \mathcal{B})\right] \geq 1 - |\mathcal{A}| \cdot T^{-3}.$$

An algorithm $\texttt{ALG}_2$ satisfies a high-probability anytime regret bound of h if for any $\texttt{ALG}_1$ chosen by the leader and any instance $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$, it holds that:

$$\mathbb{P}\left[\forall t \in [T], a \in \mathcal{A} \mid \sum_{t' \leq t \mid a_{t'} = a} v_2(a, b_{t'}) \geq \left(\sum_{t' \leq t \mid a_{t'} = a} \max_{b \in \mathcal{B}} v_2(a, b)\right) - h(n_{t+1}(a), T, \mathcal{B})\right] \geq 1 - |\mathcal{A}| \cdot T^{-3}.$$

In Chapter 14.7, we discuss the relationship between these metrics, the performance of standard algorithms for the follower on these metrics, and algorithms for the leader for more general g and h. As an example, if the follower runs a separate instantiation of `ActiveArmElimination` (Algorithm 14) on every arm $a \in \mathcal{A}$, this satisfies high-probability instantaneous regret $g(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot \log T/t})$ and high-probability anytime regret $h(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot t \cdot \log T})$ (Chapter 126).

[11] Compared with standard definitions of instaneous and anytime regret, we require a high-probability bound (rather than expectation). For anytime regret, we also require the bound for all t for a given T (rather than for all T).

14.3 Stackelberg value is unachievable

In this section, we show that the natural benchmark given by the players' utilities in the underlying static Stackelberg game is unachievable. More formally, given an instance $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$, let (a^*, b^*) be the Stackelberg equilibrium. We define the *Stackelberg benchmarks* to be each player's utility at (a^*, b^*), that is: $\beta_1^{\text{orig}} = v_1(a^*, b^*)$ and $\beta_2^{\text{orig}} = v_2(a^*, b^*)$ (where the superscript "orig" denotes that this is the benchmark for original offline Stackelberg games). The following result illustrates that it is not possible to simultaneously achieve sublinear regret with respect to both players' regret.[12]

Theorem 116. *Consider StrongDSGs or WeakDSGs. For any algorithms* $\texttt{ALG}_1$ *and* $\texttt{ALG}_2$*, there exists an instance* $\mathcal{I}^*$ *with* $|\mathcal{A}| = |\mathcal{B}| = 2$ *where at least one of the players incurs linear regret with respect to the Stackleberg benchmarks* β_1^{orig} *and* β_2^{orig}*. That is, it holds that* $\max(R_1(T; \mathcal{I}^*), R_2(T; \mathcal{I}^*)) = \Omega(T)$.

Proof sketch of Theorem 116. It suffices to prove this lower bound in a *centralized* environment where a single learner can choose action pairs (a, b) and observes rewards for both players (Lemma 270). We show that on the instances $\mathcal{I}$ and $\tilde{\mathcal{I}}$ in Table 14.1 (with $\delta = O(1/\sqrt{T})$), at least one of the players incurs linear regret on at least one of the instances. The small value of δ means that the algorithm fails to distinguish between these instances with constant probability. Nonetheless, the benchmarks are very different: on instance $\mathcal{I}$, $(a^*, b^*) = (a_1, b_1)$, $\beta_1^{\text{orig}} = 0.6$ and $\beta_2^{\text{orig}} = \delta > 0$; on instance $\tilde{\mathcal{I}}$, $(a^*, b^*) = (a_2, b_1)$, $\beta_1^{\text{orig}} = 0.5$, and $\beta_2^{\text{orig}} = 0.6$. Intuitively, when the algorithm fails to distinguish between these instances, then it must choose the same distribution over $\mathcal{A} \times \mathcal{B}$ on both $\mathcal{I}$ and $\tilde{\mathcal{I}}$. However, any such distribution either incurs constant loss for the leader on $\mathcal{I}$ or constant loss for the follower on $\tilde{\mathcal{I}}$. We formalize this proof in Chapter J.2.4. $\square$

The linear regret in Theorem 116 is driven by *misalignment* between the leader's utilities and the follower's utilities: small differences in the follower's utilities can lead to arbitrarily large differences in the leader's utilities. As a result, the suboptimal actions that the follower takes while learning are amplified in the leader's regret. This motivates the design of relaxed benchmarks that take into account the suboptimal actions players inevitably take while learning.

14.4 γ-tolerant benchmark and regret bounds

Having shown that the Stackelberg equilibrium is unattainable, we next propose a novel benchmark and give tight sublinear regret bounds with respect to it.

[12] There exists a simple algorithm in the centralized environment that achieves sublinear regret for Player i: run a sublinear regret multi-armed bandit algorithm on the arms (a, b) using Player i's stochastic rewards (ignoring the rewards of the other player).

	b_1	b_2
a_1	$(0.6, \delta)$	$(0.2, \mathbf{0})$
a_2	$(0.5, 0.6)$	$(0.4, 0.4)$

(a) Mean rewards $(v_1(a,b), v_2(a,b))$ for $\mathcal{I}$

	b_1	b_2
a_1	$(0.6, \delta)$	$(0.2, \boldsymbol{2\delta})$
a_2	$(0.5, 0.6)$	$(0.4, 0.4)$

(b) Mean rewards $(\tilde{v}_1(a,b), \tilde{v}_2(a,b))$ for $\tilde{\mathcal{I}}$

Table 14.1: Two instances $\mathcal{I}$ (left) and $\tilde{\mathcal{I}}$ (right), which differ solely in the follower's reward for (a_1, b_2) (shown in **bold**). For δ sufficiently small, the instances $\mathcal{I}$ and $\tilde{\mathcal{I}}$ are hard to distinguish and turn out to imply a $\Omega(T)$ lower bound on regret with respect to the original Stackelberg benchmarks (Theorem 116).

14.4.1 γ-tolerant benchmark

Our benchmark is related to the Stackelberg Equilibrium, but adapted to account for the fact that both players are learning and cannot be counted on to exactly best respond. This benchmark is a function of the instance $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ but *independent* of the learning algorithms for either player. At a high-level, we construct a set of *approximate best responses* for each player and use these sets to construct more realistic benchmarks; within these sets, our benchmark will be *tolerant* to suboptimality with respect to the other player.

If the leader takes action a, then we define the follower's ε-best-response set $\mathcal{B}_\varepsilon(a)$ as:

$$\mathcal{B}_\varepsilon(a) := \{b \in \mathcal{B} \mid v_2(a,b) \geq \max_{b' \in \mathcal{B}} v_2(a,b') - \varepsilon\}.$$

Defining the ε-best response set $\mathcal{A}_\varepsilon$ for the leader is more subtle. Informally, we define this set to include any action $a \in \mathcal{A}$ which has "any chance" of doing at least as well as the the leader's best action if the follower is ε-best responding. Specifically, this includes actions a where *some* action $b \in \mathcal{B}_\varepsilon(a)$ achieves utility close to $\max_{a' \in \mathcal{A}} \min_{b' \in \mathcal{B}_\varepsilon(a)} v_1(a', b')$ for the leader:[13]

$$\mathcal{A}_\varepsilon = \{a \in \mathcal{A} \mid \max_{b \in \mathcal{B}_\varepsilon(a)} v_1(a,b) \geq \max_{a' \in \mathcal{A}} \min_{b' \in \mathcal{B}_\varepsilon(a')} v_1(a', b') - \varepsilon\}.$$

Observe that the set $\mathcal{B}_\varepsilon(a)$ always contains the follower's best-response set $\{b \in \arg\max_{b' \in \mathcal{B}} v_1(a, b')\}$, and furthermore approaches this best-response set in the limit as $\varepsilon \to 0$; similarly, the set $\mathcal{A}_\varepsilon$ always contains the leader's best-response set $\{a \in \arg\max_{a' \in \mathcal{A}} v_1(a', b^*(a'))\}$ where ties are broken in favor of the follower, and furthermore approaches this best-response set in the limit as $\varepsilon \to 0$.

We use these ε-best-response sets to create the relaxed benchmarks for the leader and follower. In these benchmarks, we add an *ε-relaxed Stackelberg utility* term with a *ε-regularizer*

[13] At first glance, it might appear more natural to instead define $\mathcal{A}_\varepsilon$ to be all actions where the follower's worst-case approximate best-response yields high utility for the leader, that is: $\{a \in \mathcal{A} \mid \mathbf{min}_{b \in \mathcal{B}_\varepsilon(a)} v_1(a,b) \geq \max_{a' \in \mathcal{A}} \min_{b' \in \mathcal{B}_\varepsilon(a')} v_1(a', b') - \varepsilon\}$. However, this set does not necessarily contain the leader's best-response set $\{a \in \arg\max_{a' \in \mathcal{A}} v_1(a', b^*(a'))\}$, which makes it a less natural definition of an approximate best-response set.

term, and then take an infimum over all possible values $\varepsilon \leq \gamma$. In particular, the ε-relaxed Stackelberg utility takes a max over the player's actions and a min over the other player's ε-best response set; the regularizer adds a ε penalty for errors made by the other player.

Definition 15. *Given a maximum tolerance $\gamma > 0$, we define the γ-tolerant benchmarks β_1^{tol} and β_2^{tol} to be:*

$$\beta_1^{tol} = \inf_{\varepsilon \leq \gamma} \Bigg(\underbrace{\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_\varepsilon(a)} v_1(a,b)}_{\varepsilon\text{-relaxed Stackelberg utility}} + \underbrace{\varepsilon}_{\varepsilon\text{-regularizer}} \Bigg)$$

$$\beta_2^{tol} = \inf_{\varepsilon \leq \gamma} \Bigg(\underbrace{\min_{a \in \mathcal{A}_\varepsilon} \max_{b \in \mathcal{B}} v_2(a,b)}_{\varepsilon\text{-relaxed Stackelberg utility}} + \underbrace{\varepsilon}_{\varepsilon\text{-regularizer}} \Bigg).$$

We provide some high-level intuition for why our benchmark might be appropriate for learning environments. For small values of ε, the ε-best-response sets describe actions that (from the player's perspective) are similar and difficult to distinguish between while learning. The ε-relaxed Stackelberg utility takes a worst-case perspective and takes a minimum over the other player's ε-best-response set, since the player's choices within this set can be unpredictable while learning.[14] The ε-regularizer captures the player's intolerance of suboptimality of the other player (see Chapter 14.6 for a discussion of regularizers and γ).

We illustrate these benchmarks in the following example.

Example 12. *Consider the example in Table 14.2 (with $0.4 > \gamma \geq 4\delta$). In this case, the leader's benchmark is equal to the Stackelberg utility ($\beta_1^{orig} = \beta_1^{tol} = 0.5 + \delta$), while the follower's benchmark is weaker ($\beta_2^{orig} = 0.4 > 3\delta + \delta = \beta_2^{tol}$), where the second δ comes from the regularizer. The intuition is that the leader's ε-best-response set $\mathcal{A}_\delta = \{a_1, a_2\}$ contains both actions, even though a_2 is not a Stackelberg equilibrium, which noticeably lowers the follower's ε-relaxed Stackelberg utility. In Chapter J.1, we provide more detailed discussions of examples.*

	b_1	b_2
a_1	$(0.5+\delta, 0.4)$	$(0.2, 0)$
a_2	$(0.5, 3\delta)$	$(0.4, 2\delta)$

Table 14.2: A single instance, illustrating the γ-tolerant benchmark (Example 12).

[14]For the leader, Bai et al. (2021) also takes a similar worst-case perspective over the ε-best-response set, but does not introduce a regularizer or take a minimum over ε.

14.4.2 Linear regret for ExploreThenCommit

We first show that out-of-box stochastic bandit algorithms do not directly provide sublinear regret against the γ-tolerant benchmark, where the challenge is that the leader's learning gets distorted when both players simultaneously learn. To demonstrate this, we consider `ExploreThenCommit` (Algorithm 8) and show that if both the leader and follower are running this algorithm, the regret could be linear for *both players*.

`ExploreThenCommit`$(E, \mathcal{C})$ (Algorithm 8). The algorithm `ALG` = `ExploreThenCommit`$(E, \mathcal{C})$ takes as input $E \in [T]$ and a set of arms $\mathcal{C}$.[15] When `ALG` is applied to an instance, for the first $|\mathcal{C}| \cdot E$ rounds, the algorithm `ALG` pulls each arm in $\mathcal{C}$ a total of E times in a round-robin fashion. For the remaining $T - |\mathcal{C}| \cdot E$ rounds, the algorithm commits to the optimal empirical mean from the first $|\mathcal{C}| \cdot E$ rounds. This is a standard algorithm for multi-armed bandits (Slivkins, 2019; Lattimore and Szepesvári, 2020).

Algorithm 8: `ExploreThenCommit`$(E, \mathcal{C})$ applied to history H (see e.g., (Slivkins, 2019; Lattimore and Szepesvári, 2020))

Fix an arbitrary ordering $\mathcal{C} = \{c^1, \ldots, c^{|\mathcal{C}|}\}$.
Let $t = |H|$.
`/* Explore for the first` $E \cdot |\mathcal{C}|$ `rounds */`
if $t \leq E \cdot |\mathcal{C}|$ **then**
  Let $i = t \pmod{|\mathcal{C}|} + 1$ be the index of the action that should be pulled.
  return *point mass at* c^i
`/* Commit for the remaining rounds */`
if $t > E \cdot |\mathcal{C}|$ **then**
  `/* Discard history all but the first` $E \cdot |\mathcal{C}|$ `rounds. */`
  $H^* = \{(t', c_{t'}, r) \mid \exists (t', b_{t'}, r) \in H \text{ s.t. } t' \leq E \cdot |\mathcal{C}|\}$
  `/* Choose highest empirical mean. */`
  for $c \in \mathcal{C}$ **do**
    Set $S(c) := \{r \mid \exists (t', c_{t'}, r) \in H^* \text{ s.t. } c = c_{t'}\}$ `// observed rewards`
    $\hat{v}(c) \leftarrow (\sum_{r \in S(c)} r)/|S(c)|$ `// compute empirical mean`
  return *point mass at* $\arg\max_{c \in \mathcal{C}} \hat{v}(c)$

When both players run `ExploreThenCommit`, we show that if the leader's exploration phase ends before the follower's exploration phase, then both players can incur linear regret. This lower bound holds for any maximum tolerance $\gamma \leq 1$.

Proposition 117. *Consider StrongDSGs where the follower runs a separate instantiation of `ExploreThenCommit`$(E, \mathcal{B})$ for every $a \in \mathcal{A}$. Moreover, suppose that the leader runs `ExploreThenCommit`$(E' \cdot |\mathcal{B}|, \mathcal{A})$ for any $E' \leq E$ (i.e., the leader's exploration phase ends*

[15] By setting $\mathcal{C} = \mathcal{A}$, this algorithm can be instantiated as $\texttt{ALG}_1$ for the leader; by setting $\mathcal{C} = \mathcal{B}$, this algorithm can be instantiated as $\texttt{ALG}_2$ for the follower.

before the follower's exploration phase). Then, there exists an instance $\mathcal{I}^*$ such that both players incur linear regret with respect to the γ-tolerant benchmarks β_1^{tol} and β_2^{tol}: that is, $\min(R_1(T;\mathcal{I}^*), R_2(T;\mathcal{I}^*)) = \Omega(T)$.

Proof sketch of Proposition 117. The intuition is that in the leader's exploration phase, the follower alternates uniformly between all actions $\mathcal{B}$. This distorts the leader's learning during the leader's exploration phase, and as a result, the leader can choose a highly suboptimal arm during the commit phase. This can lead to linear regret for both players. We construct a single instance (Table 14.2, with $\delta = 0.1$) where both players incur linear regret. The full proof is deferred to Chapter J.3.1. □

14.4.3 Warmup Algorithm

The constant regret in Proposition 117 motivates the design of more sophisticated algorithms where the leader waits for the follower to partially converge before starting to learn. As a warmup, we show that a simple modification of the setup of Proposition 117 guarantees sublinear regret (i.e., $O\left(|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3}T^{2/3}\right)$ regret for both players). In this algorithm, both players run `ExploreThenCommit`-based algorithms, but the leader waits for the follower to finish exploring before starting to explore. More specifically, the leader runs `ExploreThenCommitThrowOut`, which acts similar to `ExploreThenCommit`, but with an extra exploration phase at the start, after which all rewards are thrown out. This initial phase is to allow the follower to partially converge.

`ExploreThenCommitThrowOut`$(E, E', \mathcal{C})$ (Algorithm 9). The algorithm $\texttt{ALG}_1$ equal to `ExploreThenCommitThrowOut`$(E, E', \mathcal{C})$ takes as input $E, E' \in [T]$ and a set of arms $\mathcal{C}$. It throws out the first $E' \cdot |\mathcal{C}|$ rounds and then runs `ExploreThenCommit`$(E, \mathcal{C})$.

Algorithm 9: `ExploreThenCommitThrowOut`$(E, E', \mathcal{C})$ applied to history H

Fix an arbitrary ordering $\mathcal{C} = \left\{c^1, \ldots, c^{|\mathcal{C}|}\right\}$. `/* Explore for first` $E' \cdot |\mathcal{C}|$ `rounds */`

if $t \leq E' \cdot |\mathcal{C}|$ **then**

  Let $i = t \pmod{|\mathcal{C}|} + 1$ be the index of the action that should be pulled.

  return *point mass at* c^i

`/* Run ETC for the remainder of time, throwing out first` $E' \cdot |\mathcal{C}|$ `rounds */`

if $t > E' \cdot |\mathcal{C}|$ **then**

  $H^* = \{(t' - E' \cdot |\mathcal{C}|, c_{t'}, r) \mid \exists (t', c_{t'}, r) \in H \text{ s.t. } t' > E' \cdot |\mathcal{C}|\}$ `// Throw out first` $E' \cdot |\mathcal{C}|$ `rounds of history`

  return *`ExploreThenCommit`*$(E, \mathcal{C})$ *applied to* H^*

We show that if the follower runs `ExploreThenCommit` and the leader runs `ExploreThenCommitThrowOut`, then both players achieve sublinear regret. For this result, we require that γ is not *too small*: $\gamma = \omega\left(T^{-1/3}\left(|\mathcal{A}| \cdot |\mathcal{B}|\right)^{1/3}\right)$ (see Chapter 14.6 for a discussion).

Theorem 118. *Consider a StrongDSG where the follower runs a separate instantiation of* `ExploreThenCommit`$(E_2, \mathcal{B})$ *for every* $a \in \mathcal{A}$*, and where the leader runs* `ExploreThenCommitThrowOut`$(E_1, E_2 \cdot |\mathcal{B}|, \mathcal{A})$*. If* $E_2 = \Theta(|\mathcal{A}|^{-2/3}|\mathcal{B}|^{-2/3} \cdot (\log T)^{1/3}T^{2/3})$*, and* $E_1 = \Theta(|\mathcal{A}|^{-2/3} \cdot (\log T)^{1/3}T^{2/3})$*, then, the regret with respect to the* γ*-tolerant benchmarks is bounded as:*

$$\max(R_1(T), R_2(T)) = O\left(|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3}T^{2/3}\right).$$

Proof sketch for Theorem 118. The "throw out" phase for $\texttt{ALG}_1$ enables $\texttt{ALG}_2$ to learn and commit to near-optimal actions. The meaningful exploration for $\texttt{ALG}_1$ thus begins *after* the follower has committed to actions. This enables $\texttt{ALG}_1$ to identify a near-optimal action given the arms that $\texttt{ALG}_2$ has already committed to after the first phase of exploration. Returning to our γ-tolerant benchmarks, for each player, we can upper bound regret by setting ε to be the suboptimality of the other player and achieve the desired regret bound. The full proof is deferred to Appendix J.3.2. □

One drawback of Theorem 118 is that requiring the follower to run a single algorithm is relatively restrictive. In the next subsection, we allow for a rich class of follower algorithms.

14.4.4 Main Algorithm

Our main result in this section is an adaptive algorithm for the leader (`ExploreThenUCB`, Algorithm 10) that achieves the same regret bounds while permitting greater flexibility for the follower (Theorem 119). Specifically, we only require that the follower converges to ε-optimal responses quickly, which we formalize through high-probability instantaneous regret (Chapter 14.2.4). Since the leader's algorithm needs to be robust to a broader range of follower behaviors, we replace the commit phase of `ExploreThenCommit` with an adaptive algorithm. This motivates `ExploreThenUCB`, which explores in the first phase, and then runs a version of UCB on the arms $\mathcal{A}$. The initial exploration phase in `ExploreThenUCB`, similar to the initial exploration phase in `ExploreThenCommitThrowOut`, ensures that the leader waits for the follower to partially converge before starting to learn.

`ExploreThenUCB`(E) **(Algorithm 10).** The algorithm $\texttt{ALG}_1 = \texttt{ExploreThenUCB}(E)$ takes as input $E \in [T/|\mathcal{A}|]$. When $\texttt{ALG}_1$ applied to an instance, for the first $|\mathcal{A}| \cdot E$ rounds, the algorithm $\texttt{ALG}_1$ pulls each arm in $\mathcal{A}$ a total of E times, and then discards all history from these rounds. For the remaining $T - |\mathcal{A}| \cdot E$ rounds, the algorithm runs UCB, computing the upper confidence bound $v_1^{\text{UCB}}(a) = \hat{v}_{1,t}(a) + \alpha_t(a)$ using confidence bound $\alpha_t(a) = \Theta\left(\sqrt{(\log T)/n_{E\cdot|\mathcal{A}|,t}(a)}\right)$, where $\hat{v}_{1,t}(a)$ is the empirical mean and $n_{E\cdot|\mathcal{A}|,t}(a)$ is the number of times that action a is chosen in the UCB phase after time step $E \cdot |\mathcal{A}|$ and prior to time step t. The algorithm then chooses the arm with maximum upper confidence bound.

Algorithm 10: `ExploreThenUCB`(E) applied to H

Fix an arbitrary ordering $\mathcal{A} = \left\{a^1, \ldots, a^{|\mathcal{A}|}\right\}$.
Let $t = |H|$.
`/* Explore for the first` $E \cdot |\mathcal{A}|$ `rounds */`
if $t \leq E \cdot |\mathcal{A}|$ **then**
  Let $i = \lceil \frac{t}{E} \rceil$ be the index of the action that should be pulled.
  return *point mass at* a^i
if $t > E \cdot |\mathcal{A}|$ **then**
  $H^* = \{(t' - E \cdot |\mathcal{A}|, a_{t'}, r) \mid \exists (t', a_{t'}, r) \in H \text{ s.t. } t' > E \cdot |\mathcal{A}|\}$ `// Throw out first` $E \cdot |\mathcal{A}|$ `rounds of history`
  Initialize $\hat{v}_1(a) = 1$ for $a \in \mathcal{A}$. `// Initialize empirical means.`
  Initialize $v_1^{\text{UCB}}(a) = 1$ for $a \in \mathcal{A}$. `// Initialize UCB.`
  for $a \in \mathcal{A}$ **do**
    Set $S(a) := \{r \mid \exists (t', a_{t'}, r_{1,t'}(a_{t'}, b_{t'})) \in H^* \text{ s.t. } a = a_{t'}, r_{1,t'}(a_{t'}, b_{t'}) = r\}$ `// Observed rewards`
    if $S(a) \neq \emptyset$ **then**
      $\hat{v}_1(a) \leftarrow (\sum_{r \in S(a)} r)/|S(a)|$ `// Empirical mean`
      $\alpha(a) \leftarrow 10 \cdot \frac{\sqrt{\log T}}{\sqrt{|S(a)|}}$ `// confidence bound width`
      $v_1^{\text{UCB}}(a) \leftarrow \min(1, \hat{v}_1(a) + \alpha(a))$
  Let $a^* = \arg\max_{a \in \mathcal{A}} \left(v_1^{\text{UCB}}(a)\right)$. `// arm with max upper confidence bound`
  return *point mass at* a^*

Even though the rewards observed by the leader are *not* stochastic (since the follower can pick different arms over time), we show if the leader runs `ExploreThenUCB` and the follower runs algorithms with sufficiently low high-probability instantaneous regret, then both players achieve $O\left(|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3}T^{2/3}\right)$ regret. The assumptions on the follower's algorithm are satisfied by standard algorithms such as `ActiveArmElimination` (Algorithm 14; Proposition 126) and `ExploreThenCommit` (Algorithm 8; Proposition 127). For this result, we require that the maximum tolerance γ is not *too small*: $\gamma = \omega\left(T^{-1/3}\left(|\mathcal{A}| \cdot |\mathcal{B}|\right)^{1/3}\right)$ (see Chapter 14.6 for a discussion).

Theorem 119. *Let $E = \Theta(|\mathcal{A}|^{-2/3}(|\mathcal{B}|\log T)^{1/3}T^{2/3})$. Consider a StrongDSG, where* `ALG`$_2$ *is any algorithm with high-probability instantaneous regret $g(t, T, \mathcal{B}) = O\left((|\mathcal{A}||\mathcal{B}|\log T)^{1/3}T^{-1/3}\right)$ for $t > E$ and $g(t, T, \mathcal{B}) = 1$ for $t \leq E$, and where* `ALG`$_1$ = `ExploreThenUCB`(E)*. Then, it holds that the regret with respect to the γ-tolerant benchmarks β_1^{tol} and β_2^{tol} is bounded as:*

$$\max(R_1(T), R_2(T)) = O\left(|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3}T^{2/3}\right).$$

Proof sketch of Theorem 119. The intuition is that the exploration phase of `ExploreThenUCB` ensures that all of the follower's actions have bounded suboptimality, and the UCB phase accounts for the follower changing which action they choose over time. In more detail, high-probability instantaneous regret guarantees that after the explore phase, all actions that the follower's chooses are within the ε-best-response set $\mathcal{B}_{\varepsilon^*}(a)$ for $\varepsilon^* = \Theta((|\mathcal{A}| \cdot |\mathcal{B}| \cdot \log T)^{1/3} T^{-1/3})$. For the UCB phase, the main lemma (Lemma 277) is that if an arm $a \in \mathcal{A}$ is pulled, the empirical mean is at least $\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) - \Theta\left(\sqrt{\log T / n_{E \cdot |\mathcal{A}|, t}(a)}\right)$ (the optimal utility for the leader when the follower worst-case ε-best-responds minus the confidence set size). Lemma 277 enables us to analyze the leader's cumulative reward from each arm $a \in \mathcal{A}$ and thus bound the leader's regret. For the follower's regret, Lemma 277 enables us to bound the number of times that arms outside of $\mathcal{A}_\varepsilon$ are chosen, which enables us to bound the follower's regret. We defer the full proof to Chapter J.3.3. □

14.4.5 Regret lower bound

A natural question is whether the regret bound in Theorem 119 can be improved from $\tilde{O}(T^{2/3})$ to $\tilde{O}(\sqrt{T})$, given that such dependence is possible in single-player bandit problems. Interestingly, we show a *lower bound* of $T^{2/3}$ with respect to the γ-tolerant benchmarks, thus demonstrating that the dependence on T in Theorem 120 is near-optimal. This lower bound holds for any maximum tolerance $\gamma \leq 1$.

Theorem 120. *Consider StrongDSGs or WeakDSGs with action sets $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}| \geq 2$ and $|\mathcal{B}| \geq 2$. For any algorithms $\texttt{ALG}_1$ and $\texttt{ALG}_2$, there exists an instance $\mathcal{I}^* = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ such that at least one of the players incurs $\Omega(T^{2/3} \cdot (|\mathcal{B}|)^{1/3})$ regret with respect to the γ-tolerant benchmarks β_1^{tol} and β_2^{tol}:*

$$\max(R_1(T; \mathcal{I}^*), R_2(T; \mathcal{I}^*)) = \Omega(T^{2/3} \cdot (|\mathcal{B}|)^{1/3}).$$

Proof sketch. In this sketch, we give intuition for a weaker bound of $\Omega(T^{2/3})$, deferring the strengthening to $\Omega(T^{2/3} \cdot (|\mathcal{B}|)^{1/3})$ to Chapter J.2.5. Like in the proof of Theorem 116, it suffices to consider a centralized environment (Lemma 270). We show that on the $\mathcal{I}$ and $\tilde{\mathcal{I}}$ in Table 14.3 (with $\delta = \Theta(T^{-1/3})$), at least one player incurs $\Omega(T^{2/3})$ regret on at least one of these instances. The only way to distinguish the instances is to pull (a_1, b_2) at least $\Omega(T^{2/3})$ times, which gives low utility for both players. Intuitively, when the algorithm fails to distinguish $\mathcal{I}$ and $\tilde{\mathcal{I}}$, the algorithm must choose the same distribution over $\mathcal{A} \times \mathcal{B}$, but this gives $\Theta(T^{-1/3})$ loss for the leader on $\mathcal{I}$ or $\Theta(T^{-1/3})$ loss for the follower on $\tilde{\mathcal{I}}$. The full proof, which relies on a KL-divergence argument, is deferred to Chapter J.2.5. □

At a high-level, the $T^{2/3}$ regret bound in Theorem 120 is driven by the need to obtain precise estimates of *highly suboptimal action pairs* in order to learn to distinguish between two instances. This is fundamentally different from single learner environments, where the learner only needs to obtain precise estimates of *near-optimal* arms. Our regret upper bound (Theorem 119) and lower bound (Theorem 120) have near-matching dependence on T and $|\mathcal{B}|$, but a gap in dependence on $|\mathcal{A}|$ (the upper bound scales with $|\mathcal{A}|^{1/3}$ while the lower bound is independent of $|\mathcal{A}|^{1/3}$). An interesting direction for future work is to close this gap.

	b_1	b_2
a_1	$(0.5+\delta,\delta)$	$(0,\mathbf{0})$
a_2	$(0.5,3\delta)$	$(0.5,3\delta)$

(a) Mean rewards $(v_1(a,b), v_2(a,b))$ for $\mathcal{I}$

	b_1	b_2
a_1	$(0.5+\delta,\delta)$	$(0,\boldsymbol{2\delta})$
a_2	$(0.5,3\delta)$	$(0.5,3\delta)$

(b) Mean rewards $(\tilde{v}_1(a,b), \tilde{v}_2(a,b))$ for $\tilde{\mathcal{I}}$

Table 14.3: Two instances $\mathcal{I}$ (left) and $\tilde{\mathcal{I}}$ (right), which differ solely in the follower's reward for (a_1, b_2) (shown in **bold**). For δ sufficiently small, the instances $\mathcal{I}$ and $\tilde{\mathcal{I}}$ are hard to distinguish and turn out to imply a $\Omega(T^{2/3})$ lower bound on regret with respect to the γ-tolerant benchmark (Theorem 120).

14.5 Relaxed Settings with Faster Learning

The lower bound in the previous section showed that $\Theta(T^{2/3})$ regret is optimal for the benchmarks β_1^{tol} and β_2^{tol} for general instances. Since a $T^{2/3}$ lower bound is atypical for K-armed bandits problems, we next consider relaxed environments under which faster learning—-i.e., $O(\sqrt{T})$ regret—is possible. In the first environment, we consider well-behaved instances (Chapter 14.5.1) and in the second environment, we weaken the benchmarks (Chapter 14.5.2). In both environments, we show that the learner does not need to worry about their learning being overly distorted by the follower; thus, the leader can start learning immediately, even before the follower's actions have partially converged, which leads to improved regret bounds. The algorithms that we design for the leader are variants of UCB.

14.5.1 Continuity condition on utilities

We first show that improved regret bounds are possible with a continuity condition on the player utilities. For intuition, the example in Table 14.1 gave a "hard" example resulting in linear regret in Theorem 116 and the related example in Table 14.3 resulted in $\Omega(T^{2/3})$ regret in Theorem 120. These examples relied on two outcomes with nearly identical utilities for the follower having significantly different utilities for the leader, which could be viewed as a violation of continuity. This suggests that if arms that are *sufficiently different* for the leader were also *sufficiently different* for the follower, then it might be possible to beat the regret lower bound from Theorem 116 and Theorem 120.

We formalize continuity as follows: given an instance $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$, we define the Lipschitz constant L^* by[16]:

$$L^* = \sup_{i \neq j \in \{1,2\}} \sup_{(a,b) \neq (a',b')} \frac{|v_i(a,b) - v_i(a',b')|}{|v_j(a,b) - v_j(a',b')|}.$$

[16]In the case of ties in rewards, if the numerator and denominator are both 0, we define $\frac{|v_i(a,b)-v_i(a',b')|}{|v_j(a,b)-v_j(a',b')|}$ to be 1 (because both players agree the elements are equivalent). If the denominator is 0 and numerator is nonzero, we define this fraction to be ∞ (because the items are indistinguishable to one player, while they give different rewards to the other).

For example, when the two players have the same utilities (i.e., $v_1 = v_2$), then $L^* = 1$. More generally, our continuity condition captures the extent to which players agree on which outcomes are different from each other (a more detailed discussion is given in Chapter J.4.1). Returning to the examples in Tables 14.1, 14.3, the "hard" instances yielding linear regret for the original Stackelberg benchmarks (Theorem 116; Table 14.1) have $L^* = \Theta(T^{-1/2})$ and the corresponding "hard" instances for $T^{2/3}$ regret for the γ-tolerant benchmarks (Theorem 120; Table 14.3) require that $L^* = \Theta(T^{-1/3})$; in contrast, we focus on utility functions where L^* is a constant.

When L^* is bounded, we show that it is possible for both players to achieve $\tilde{O}(\sqrt{T})$ regret even with respect to the *original Stackelberg benchmarks*. The follower can run any algorithm $\texttt{ALG}_2$ with sufficiently low high-probability anytime regret (e.g., `ActiveArmElimination` as in Chapter 126 or UCB as in Chapter 128). We construct another UCB-based algorithm `LipschitzUCB` (Algorithm 11) for the leader, which expands the confidence sets based on the Lipschitz constant L^*.

`LipschitzUCB`(L, C) **(Algorithm 11).** The algorithm $\texttt{ALG}_1 = \texttt{LipschitzUCB}(L, C)$ takes as inputs parameters L and C. (The parameter L is intended to be an upper bound on the Lipschitz constant L^*, and the parameter C' is intended to be such that $\texttt{ALG}_2$ satisfies anytime regret bound $h(t, T, \mathcal{B}) = \sqrt{Ct \log T}$, where $C = C' \cdot \sqrt{|\mathcal{B}|}$ for a constant C'.) For each arm $a \in \mathcal{A}$, the algorithm computes UCB estimates $v_1^{\mathrm{UCB}}(a)$ of the quantity $\max_{b \in \mathcal{B}} v_1(a, b)$ using the high-probability anytime regret bounds of $\texttt{ALG}_2$ as well as the upper bound on the Lipschitz constant. The algorithm then chooses the arm $a_t = \arg\max_{a \in \mathcal{A}} \max_{b \in \mathcal{B}'(a)} v_1^{\mathrm{UCB}}(a)$.

Algorithm 11: `LipschitzUCB`(L, C) applied to H

Initialize $\hat{v}_1(a) = 1$ for $a \in \mathcal{A}$. `// Initialize empirical means for` $\max_{b \in \mathcal{B}} v_1(a, b)$.
Initialize $v_1^{\mathrm{UCB}}(a) = 1$ for $a \in \mathcal{A}$. `// Initialize UCB for` $\max_{b \in \mathcal{B}} v_1(a, b)$
for $a \in \mathcal{A}$ **do**
  Set $S(a) := \{r \mid \exists (t', a_{t'}, r) \in H \text{ s.t. } a = a_{t'}\}$ `// Observed rewards`
  if $S(a) \neq \emptyset$ **then**
    $\hat{v}_1(a) \leftarrow (\sum_{r \in S(a)} r)/|S(a)|$ `// Empirical mean`
    $\alpha(a) \leftarrow \frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{|S(a)|}} + C \cdot L \cdot \frac{\sqrt{\log T}}{\sqrt{|S(a)|}}$ `// confidence bound width`
    $v_1^{\mathrm{UCB}}(a) \leftarrow \min(1, \hat{v}_1(a) + \alpha(a))$
Let $a^* = \arg\max_{a \in \mathcal{A}} \left(v_1^{\mathrm{UCB}}(a)\right)$. `// arm with max upper confidence bound`
return *point mass at* a^*.

We obtain the following regret bound with respect to the Stackelberg benchmark, our strongest benchmark.

Theorem 121. *Consider a StrongDSG where $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ has Lipschitz constant L^*. Let $\texttt{ALG}_2$ be any algorithm satisfying high-probability anytime regret $h(t, T, \mathcal{B}) = C'\sqrt{|\mathcal{B}| t \log T}$*

where C' is a constant, and let $\texttt{ALG}_1 = \texttt{LipschitzUCB}(L, C'\sqrt{|\mathcal{B}|})$ for any $L \geq L^*$. Then both players achieve the following regret bounds with respect to the original Stackelberg benchmarks β_1^{orig} and β_2^{orig}: that is, $R_1(T; \mathcal{I}) = O\left(L\sqrt{T|\mathcal{A}||\mathcal{B}|\log T}\right)$ and $R_2(T; \mathcal{I}) = O\left(L^2\sqrt{T|\mathcal{A}| \cdot |\mathcal{B}|\log T}\right)$.

Proof sketch for Theorem 121. The intuition is the continuity conditions imply that small errors by the follower translate into bounded suboptimality for the leader (and vice versa); moreover, the high-probability anytime regret requirements bound the follower's errors. Together, these properties guarantee that the leader's empirical mean $\hat{v}_1(a)$ for each arm $a \in \mathcal{A}$ is close to the mean reward $v_1(a, b^*(a))$ that they would receive if the follower best-responded: in more detail, the main lemma (Lemma 279) is that the empirical mean $\hat{v}_1(a)$ is at least $v_1(a, b^*(a)) - \Theta(L\sqrt{\log T}/\sqrt{n_t(a)})$, where $n_t(a)$ is the number of times that arm a has been pulled prior to time step t. Using Lemma 279 to bound the suboptimality of the leader's choice of actions $a_t \in \mathcal{A}$ and using the anytime regret requirements to bound the follower's suboptimality, we can bound both the leader's regret and the follower's regret. We defer the full proof to Chapter J.4.2. □

Finally, we compare our continuity condition and results with those in other works. Our continuity condition bears resemblance to the restrictions on utilities in Camara et al. (2020); Collina et al. (2024): in fact, our conditions are conceptually stronger since we require Lipschitz continuity across *all* pairs of actions rather only for near-optimal actions. However, Theorem 121 is not directly comparable with the results in Camara et al. (2020); Collina et al. (2024) since we consider a stronger benchmark (the original Stackelberg benchmark) and also restrict to stochastic rewards. An interesting direction for future work would be to relax the Lipschitz continuity assumptions in our work, perhaps borrowing intuition from the stable action requirement of Collina et al. (2024).

14.5.2 Weaker benchmark

Finally, we will consider the case where utilities are allowed to be arbitrary (L^* can be unbounded), but where we compete with weakened benchmarks, which we call *self-γ-tolerant*. These benchmarks capture the case where the player is not only tolerant of suboptimality the other player, but also tolerant of their own suboptimality. We thus take a min over the ε-best-response sets of *both* players.

Definition 16. *Given a maximum tolerance $\gamma > 0$, we define the self-γ-tolerant benchmarks, $\beta_1^{self\text{-}tol}$ and $\beta_2^{self\text{-}tol}$, to be:*

$$\beta_1^{self\text{-}tol} = \inf_{\varepsilon \leq \gamma} \left(\min_{a \in \mathcal{A}_\varepsilon} \min_{b \in \mathcal{B}_\varepsilon(a)} v_1(a, b) + \varepsilon \right)$$

$$\beta_2^{self\text{-}tol} = \inf_{\varepsilon \leq \gamma} \left(\min_{a \in \mathcal{A}_\varepsilon} \min_{b \in \mathcal{B}_\varepsilon(a)} v_2(a, b) + \varepsilon \right).$$

The tolerance of a player to their own suboptimality is the key difference from the γ-tolerant benchmarks from Chapter 14.4. For the follower, the benchmarks behave similarly: for a given value of ε, moving from $\max_{b\in\mathcal{B}} v_2(a,b)$ to $\min_{b\in\mathcal{B}_\varepsilon(a)} v_2(a,b)$ differs by only an additive value of ε. However for the leader, there is a conceptual difference: the value $\mathbf{min}_{a\in\mathcal{A}_\varepsilon} \min_{b\in\mathcal{B}_\varepsilon(a)} v_1(a,b)+\varepsilon$ is *not* necessarily within ε of $\mathbf{max}_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_\varepsilon(a)} v_1(a,b)+\varepsilon$. This is because $\mathcal{A}_\varepsilon$ includes any action a that achieves high reward for *some* (near-optimal) actions by the follower, even if the worst-case (near-optimal) action by the follower yields arbitrarily low reward for the leader. As an illustration, the "hard" instances specified in Table 14.3 with $\delta=\Theta(T^{-1/3})$ led to the $T^{2/3}$ regret bound. The self-tolerant benchmark $\beta_1^{\text{self-tol}}$ reduces to $0.2+\delta$ (rather than 0.5), so choosing (a_1,b_2) no longer results in constant loss for the leader.

Example 4.2 [Continued]. *Let's again consider $\mathcal{I}$ in Table 14.2, which we also used to illustrate the γ-tolerant benchmark. The minimum is again attained at $\varepsilon=\delta$, but the benchmark values change to $\beta_1^{\textit{self-tol}}=0.4+\delta$ and $\beta_2^{\textit{self-tol}}=2\cdot\delta+\delta$. The intuition is that the self-γ-tolerance benchmark only requires each agent to compete with the* worst *element within the product set $\mathcal{A}_\delta\times\mathcal{B}_\delta(a)$. Note that the resulting benchmark differs from the γ-tolerant benchmark for the follower only by δ, but differs by* 0.4 *(a constant) for the leader. We provide a detailed derivation of this example along with diagrams illustrating richer examples in Chapter J.1.*

For the self-tolerant benchmarks, we show it is possible to achieve $\tilde{O}(\sqrt{T|\mathcal{A}||\mathcal{B}|})$ regret for both players (Theorem 122), which outperforms the $T^{2/3}$ lower bound for the stronger benchmark shown in Theorem 120. To demonstrate this is feasible, we focus on WeakDSGs, and we construct a specific pair of algorithms that achieve a $O(\sqrt{T})$ regret upper bound. For the follower, we take the algorithm $\texttt{ALG}_2$ to be `ActiveArmElimination` (Algorithm 14), which cycles through phases of exploration, after which all *sufficiently suboptimal arms* are eliminated. For the leader, we construct a UCB-based algorithm `PhasedUCB` (Algorithm 12) which constructs confidence bounds for every pair of actions (a,b).

`PhasedUCB`$(M_1,\ldots,M_P)$ **(Algorithm 12).** The algorithm $\texttt{ALG}_1=$ `PhasedUCB`$(M_1,\ldots,M_P)$ takes as input the parameters $M_1,\ldots,M_P\geq 0$. (The parameter M_i is intended to capture the number of times that an arm is pulled in phase i by the instantiation of `ActiveArmElimination` specified by $\texttt{ALG}_2$.) The algorithm $\texttt{ALG}_1$ computes UCB estimates $v_1^{\text{UCB}}(a,b)$ for $v_1(a,b)$, computes the set of active arms $\mathcal{B}'(a)$ in the previous phase of $\texttt{ALG}_2$'s instantiation of `ActiveArmElimination` for each arm a (`ComputeActiveArms`, Algorithm 13), and chooses the arm with maximum UCB: $a_t=\arg\max_{a\in\mathcal{A}}\max_{b\in\mathcal{B}'(a)} v_1^{\text{UCB}}(a,b)$. `ComputeActiveArms` computes the active arms $\mathcal{B}'(a)$ by iterating through H and keeping track of whenever a new phase is entered using the parameters $M_1,\ldots,M_P$.

We show that both players achieve $O(\sqrt{T})$ regret. For this result, we require that the γ is not *too small*: $\gamma=\Omega(T^{-1/4}(|\mathcal{A}|\cdot|\mathcal{B}|\cdot\log T)^{1/2}))$ (see Chapter 14.6 for a discussion).

Theorem 122. *Consider a WeakDSG, where for each $a\in\mathcal{A}$, the algorithm $\texttt{ALG}_2$ runs a separate instantiation of `ActiveArmElimination` with parameters $M_1,\ldots,M_P$ (where $M_i=$*

Algorithm 12: `PhasedUCB`$(M_1, \ldots, M_P)$ applied to H

Let $\hat{v}_1(a,b) = 0$ for $a \in \mathcal{A}$ and $b \in \mathcal{B}$. `// initialize empirical mean of` $v_1(a,b)$
Let $v_1^{\mathrm{UCB}}(a,b) = 1$ for $a \in \mathcal{A}$ and $b \in \mathcal{B}$. `// initialize UCB for` $v_1(a,b)$
Let $\mathcal{B}'(a) =$ `ComputeActiveArms`$(M_1, \ldots, M_P, H)$. `// active arms in previous phase for` $\texttt{ALG}_2$
for $a \in \mathcal{A}$ **do**
  for $b \in \mathcal{B}$ **do**
    Set $S(a,b) := \{r \mid \exists (t', a_{t'}, b_{t'}, r) \in H \text{ s.t. } a = a_{t'}, b = b_{t'}\}$ `// observed rewards`
    if $S(a,b) \neq \emptyset$ **then**
      $\hat{v}_1(a,b) \leftarrow (\sum_{r \in S(a,b)} r)/|S(a,b)|$ `// compute empirical mean`
      $\alpha(a,b) := 10 \cdot \sqrt{\frac{\log T}{|S(a,b)|}}$ `// confidence bound width`
      $v_1^{\mathrm{UCB}}(a,b) \leftarrow \min(1, \hat{v}_1(a,b) + \alpha(a,b))$ `// compute UCB`
Let $a^* = \arg\max_{a \in \mathcal{A}} \max_{b \in \mathcal{B}'(a)} \left(v_1^{\mathrm{UCB}}(a,b)\right)$. `// arm with max upper confidence bound for any valid` b
return *point mass at* a^i

Algorithm 13: `ComputeActiveArms`$(M_1, \ldots, M_P, H)$

Initialize $s'(a) = 0$ for $a \in \mathcal{A}$. `// Index of the last completed phase for` $\texttt{ALG}_2$ `on arm` a.
Initialize $t'(a) = 1$ for $a \in \mathcal{A}$. `// Time step marking beginning of phase` $s'+1$ `for` $\texttt{ALG}_2$ `on arm` a.
Initialize $\mathcal{B}'(a) = \mathcal{B}$. `// Active arms in phase` s' `for` $\texttt{ALG}_2$ `on arm` a.
Initialize $newphase_a =$ False for $a \in \mathcal{A}$. `// Boolean for first time step in phase for` $\texttt{ALG}_2$ `on` a.
Let $t = |H|$.
for $t'' = 1$ **to** t **do**
  for $a \in \mathcal{A}$ **do**
    for $b \in \mathcal{B}$ **do**
      Let $n(a,b) := |\{(t'', a_{t''}, b_{t''}, r_{1,t''}(a_{t''}, b_{t''})) \in H \mid a_{t''} = a, b_{t''} = b, t'' \geq t'_a\}|$.
      if $n(a,b) > M_{s'_a+1}$ **then**
        $newphase_a =$ True.
    if $newphase_a = True$ **then**
      Update $\mathcal{B}'(a) \leftarrow$
      $\{b \in \mathcal{B} \mid \exists (t'', a_{t''}, b_{t''}, r_{1,t''}(a_{t''}, b_{t''})) \in H \text{ s.t. } t'_a \leq t'' < t, a_{t''} = a, b_{t''} = b\}$
      Update $s'(a) \leftarrow s'(a) + 1$.
      Update $t'(a) \leftarrow t$.
      $newphase_a =$ False.
return $\{\mathcal{B}'(a)\}_{a \in \mathcal{A}}$.

$\Theta(\log T \cdot 2^{2i})$ denotes the number of times that each arm is pulled in phase i). Let $\texttt{ALG}_1 =$ $\texttt{PhasedUCB}(M_1, \ldots, M_P)$. Then it holds that the regret with respect to the self-γ-tolerant benchmarks $\beta_1^{\text{self-tol}}$ and $\beta_2^{\text{self-tol}}$ is bounded as:

$$\max(R_1(T), R_2(T)) = O\left(\sqrt{|\mathcal{A}| \cdot |\mathcal{B}| \cdot T \cdot \log T}\right).$$

Proof sketch for Theorem 122. The intuition is that the benchmark allows the leader to choose any $a \in \mathcal{A}_\varepsilon$. The definition of $\mathcal{A}_\varepsilon$ means that the leader can take an optimistic perspective on the follower's choice of action $\mathcal{B}_\varepsilon(a)$ (and not have to prepare for the worst-case action in $\mathcal{B}_\varepsilon(a)$). This optimistic perspective surfaces in `PhasedUCB` in terms of how the leader evaluates an action a based on the maximum UCB $\max_{b \in \mathcal{B}'(a)} v_1^{\text{UCB}}(a, b)$ across all active arms $b \in \mathcal{B}'(a)$. To analyze this pair of algorithms, we show a bound ε_t for each time step t such that a_t is an ε_t-best-response for the leader and b_t is an ε_t-best-response for the follower: the main lemma (Lemma 283) shows that we can set ε_t to be $\Theta(\sqrt{|\mathcal{B} \cdot \log T / n_t(a_t)})$ where $n_t(a_t)$ is the number of times that arm a_t has been pulled prior to time step t. The full proof is deferred to Chapter J.4.3. □

The regret bound in Theorem 122 is nearly optimal, as we show in the following $\Omega(\sqrt{T|\mathcal{A}| \cdot |\mathcal{B}|})$ lower bound for self-γ-tolerant benchmarks, which holds for any maximum tolerance $\gamma \leq 1$.

Proposition 123. *Consider StrongDSGs or WeakDSGs with actions sets $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}| \geq 2$ and $|\mathcal{B}| \geq 2$. For any algorithms $\texttt{ALG}_1$ and $\texttt{ALG}_2$, there exists an instance $\mathcal{I}^* =$ $(\mathcal{A}, \mathcal{B}, v_1, v_2)$ such that at least one of the players incurs $\Omega(\sqrt{T \cdot (|\mathcal{A}| - 1) \cdot |\mathcal{B}|})$ regret with respect to the self-γ-tolerant benchmarks $\beta_1^{\text{self-tol}}$ and $\beta_2^{\text{self-tol}}$, that is:* $\max(R_1(T; \mathcal{I}^*), R_2(T; \mathcal{I}^*)) =$ $\Omega(\sqrt{T \cdot (|\mathcal{A}| - 1) \cdot |\mathcal{B}|})$.

Taken together, Theorem 122 and Proposition 123 demonstrate the self-γ-tolerant benchmarks lead to $\tilde{\Theta}(\sqrt{T|\mathcal{A}|\mathcal{B}|})$ regret bounds for each player.

We note that Theorem 122 requires the follower to run a specific algorithm: this contrasts with our results for the γ-tolerant benchmark (Theorem 119) and the Lipschitz benchmark (Theorem 121) which allowed for greater flexibility in the follower algorithm. An interesting direction for future work would be to design a leader algorithm for the self-γ-tolerant benchmark that permits a richer family of follower behaviors.

14.6 Discussion of Benchmark Parameters

Our relaxed benchmarks—the γ-tolerant benchmarks (Definition 15) and the self-γ-tolerant benchmarks (Definition 16)—depend on two parameters: (1) the maximum tolerance γ and (2) the ε-regularizer. In this section, we discuss the role of each parameter and describe extensions of our results to alternate settings of these parameters.

	b_1	b_2
a_1	(0.6, 0.05)	(0.2, 0.1)
a_2	(0.5, 0.2)	(0.4, 0.15)

Table 14.4: Taking γ to be too small makes the benchmark too easy: for $\gamma = 0$, we have $\beta_1^{\text{tol}} = 0.5, \beta_2^{\text{tol}} = 0.2$, but for $\gamma = 0.05$ we have $\beta_1^{\text{tol}} = 0.5$ and $\beta_2^{\text{tol}} = 0.15$ (see Chapter 14.6.1)

14.6.1 Maximum Tolerance γ

The value γ intuitively captures the players' maximum tolerance for suboptimality. Taking γ to be small makes our benchmarks more challenging, because it reduces the space of permissible suboptimality levels ε over which the infimum is taken. In contrast, taking γ to be large can make our benchmarks *too easy*: for example, consider Table 14.4, which shows a case where setting $\gamma = 0.05$ reduces the benchmark for the follower, but the instance has rewards that are sufficiently far apart that for large T the Stackelberg equilibrium should intuitively be learnable.

We briefly discuss how our results extend to different maximum tolerances γ. First, we prove our lower bounds (Theorem 120, Proposition 123) for the "hardest case" of $\gamma = 1$, which means that these lower bounds hold for *all* maximum tolerances γ.

On the other hand, our upper bounds require sufficiently large γ. For some intuition, all of our analyses require that $\gamma = \omega(1/\sqrt{T})$, since followers with high-probability instantaneous regret rates of $\Theta(\sqrt{|\mathcal{B}| \cdot \log(T)/t})$ require $\Omega(T)$ rounds to find a $O(1/\sqrt{T})$-optimal solution. As to what specific values of γ that each result requires, Theorems 118 and 119 hold for any $\gamma = \omega\left(T^{-1/3} |\mathcal{A}|^{1/3} |\mathcal{B}|^{1/3} \cdot (\log(T)^{1/3})\right)$, while Theorem 122 assumes that $\gamma = \Omega\left(T^{-1/4}\sqrt{|\mathcal{A}||\mathcal{B}| \cdot \log T}\right)$.

14.6.2 ε-Regularizer

Since the ε-regularizer adds an implicit penalty for increasing ε in the benchmark, a natural question is how our benchmark would change if we changed the regularizer from ε to other functional forms $f(\varepsilon)$. To provide some preliminary intuition for this, we consider $f(\varepsilon) = c \cdot \varepsilon^d$ regularizer, which leads to the following generalized γ-tolerant benchmarks.

Definition 17 (Generalization of Definition 15)**.** *Given a maximum tolerance $\gamma > 0$ and parameters $c > 0$, and $d > 0$, we define the generalized (c, d, γ)-tolerant benchmarks β_1^{tol} and β_2^{tol} to be:*

$$\beta_1^{tol} = \inf_{\varepsilon \leq \gamma} \Bigg(\underbrace{\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_\varepsilon(a)} v_1(a, b)}_{\varepsilon\text{-relaxed Stackelberg utility}} + \underbrace{c \cdot \varepsilon^d}_{\varepsilon\text{-regularizer}} \Bigg)$$

$$\beta_2^{tol} = \inf_{\varepsilon \leq \gamma} \Bigg(\underbrace{\min_{a \in \mathcal{A}_\varepsilon} \max_{b \in \mathcal{B}} v_2(a, b)}_{\varepsilon\text{-relaxed Stackelberg utility}} + \underbrace{c \cdot \varepsilon^d}_{\varepsilon\text{-regularizer}} \Bigg).$$

At a conceptual level, different settings of c and d capture different levels of tolerance that a player has for sub-optimality in the other player. Higher values of c and smaller values of d capture greater intolerance, and thus lead to harsher penalties. The resulting changes in the benchmarks capture that if a player is less tolerant, we might expect them to experience a higher regret for a given suboptimality level of the other player.

We show how our two main upper bounds in Chapter 14.5 generalize to these new benchmarks, focusing on the case of $c \geq 1$ and $d \leq 1$ (where the benchmark becomes harder). We first show the following generalization of Theorem 118 by adjusting the explore phase length to depend on c and d.

Theorem 124. *Suppose that $c \geq 1$ and $d \leq 1$, and let $\eta := 2/(2+d)$. Consider a StrongDSG, where the follower runs a separate instantiation of* `ExploreThenCommit`$(E_2, \mathcal{B})$ *for every $a \in \mathcal{A}$, and the leader runs* `ExploreThenCommitThrowOut`$(E_1, E_2 \cdot |\mathcal{B}|, \mathcal{A})$*. If $E_2 = \Theta(|\mathcal{A}|^{-\eta} |\mathcal{B}|^{-\eta} \cdot (\log T)^{1-\eta} (c \cdot T)^\eta)$, and $E_1 = \Theta(|\mathcal{A}|^{-\eta} \cdot (\log T)^{1-\eta} (c \cdot T)^\eta)$, then the leader and follower regret with respect to the generalized (c, d, γ)-tolerant benchmarks are both at most:*

$$\max(R_1(T), R_2(T)) = O\big((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta} \cdot (c \cdot T)^\eta\big).$$

We next show the following generalizations of Theorem 119 by again adjusting the explore phase length to depend on c and d. Like in Theorem 119, the assumptions on the follower's algorithm in this result are satisfied by standard algorithms such as `ActiveArmElimination` (Algorithm 14; Proposition 126) and `ExploreThenCommit` (Algorithm 8; Proposition 127).

Theorem 125. *Suppose that $c \geq 1$ and $d \leq 1$, and let $\eta := 2/(2+d)$. Let $E = \Theta(|\mathcal{A}|^{-\eta} (|\mathcal{B}| \log T)^{1-\eta} (c \cdot T)^\eta)$. Consider a StrongDSG where* `ALG`$_2$ *is any algorithm with high-probability instantaneous regret*
$g(t, T, \mathcal{B}) = O\left((|\mathcal{A}| \cdot |\mathcal{B}| \cdot \log T)^{\eta/2} \cdot (c \cdot T)^{-\eta/2}\right)$ for $t > E$ and $g(t, T, \mathcal{B}) = 1$ for $t \leq E$, and where `ALG`$_1$ = `ExploreThenUCB`(E)*. Then, then the leader and follower regret with respect to the generalized (c, d, γ)-tolerant benchmarks are both bounded as:*

$$\max(R_1(T), R_2(T)) = O((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta} \cdot (c \cdot T)^\eta).$$

The proofs of Theorem 124 and Theorem 125 follows from the same arguments as the proof of Theorem 118 and Theorem 119, respectively, but with the values of E_1, E_2 modified (full proofs are deferred to Appendix J.5). Note that as d decreases, the regret bound worsens: this aligns with the intuition that smaller values of d capture greater intolerance. Similarly, the regret increases with c.

We defer a more extensive treatment of these generalized benchmarks to future work. Moreover, another interesting for future work would be to extend our model and results to more general functions $f(\varepsilon)$ and also allow the two players to have different regularizers.

14.7 Discussion of Assumptions on the Follower's Algorithm

Our algorithms for the leader placed assumptions on the fine-grained performance of the follower's algorithm. More specifically, the regret bound for `ExploreThenUCB` required an *high-probability instantaneous regret bound* g for the follower (Theorem 119), and the regret bound for `LipschitzUCB` required an *high-probability anytime regret bound* h for the follower (Theorem 121).

In this section, we examine these two conditions in more detail. First, we relate these two conditions and show that many standard algorithms satisfy the conditions on g and h in Theorem 119 and Theorem 121 (Chapter 14.7.1). Then, we extend our analysis of `ExploreThenUCB` and `LipschitzUCB` to more general conditions on g and h, respectively (Chapter 14.7.2).

14.7.1 Algorithms Satisfying These Fine-Grained Regret Guarantees

As a warmup, we first observe that high probability instantaneous regret bounds immediately translate to high-probability anytime regret bounds.

Observation 14.7.1. *Suppose that $\texttt{ALG}_2$ satisfies a high-probability instantaneous regret bound of g. Then it holds that $\texttt{ALG}_2$ satisfies an anytime regret bound of h defined as $h(t,T) := \sum_{t'=1}^{t} g(t',T)$.*

As a consequence, if $\texttt{ALG}_2$ satisfies the high-probability instantaneous regret bound in Theorem 119 (i.e., $g(t,T,\mathcal{B}) = O\left((|\mathcal{A}||\mathcal{B}|\log T)^{1/3}T^{-1/3}\right)$ for $t > E := \Theta(|\mathcal{A}|^{-2/3}(|\mathcal{B}|\log T)^{1/3}T^{2/3})$ and $g(t,T,\mathcal{B}) = 1$ for $t \leq E$), then $\texttt{ALG}_2$ also satisfies an anytime regret bound of h defined for $t > E$ as:

$$h(t,T) := \sum_{t'=1}^{t} g(t',T) = E + \sum_{t'=E+1}^{t} O\left((|\mathcal{A}||\mathcal{B}|\log T)^{1/3}T^{-1/3}\right) = O\left((|\mathcal{A}||\mathcal{B}|\log T)^{1/3}T^{2/3}\right).$$

However, this naive high-probability anytime regret bound is not strong enough for Theorem 121. We can nonetheless achieve the desired regret bound with additional assumptions on $\texttt{ALG}_2$ as we describe below.

By leveraging the structural properties of specific algorithms, we show that many standard algorithms achieve high-probability instantaneous regret g and/or high-probability anytime regret h, where g and h are specified according to the functional forms in Theorem 119 and Theorem 121. Proofs of these results are deferred to Chapter J.6.1.

First, we show that `ActiveArmElimination` (Even-Dar et al., 2002) (Algorithm 14, see Lattimore and Szepesvári (2020) for a textbook treatment) satisfies both the high-probability instantaneous regret bound required for Theorem 119 and the high-probability anytime regret bound required for Theorem 121.

Algorithm 14: `ActiveArmElimination`$(M_1, \ldots, M_P)$ applied to (a, H) (adapted from (Even-Dar et al., 2002; Lattimore and Szepesvári, 2020))

Initialize $s' = 0$, $t' = 1$, $\mathcal{B}' = \mathcal{B}$ `// Index of the last completed phase, time step marking beginning of phase` $s'+1$`, active arms in phase` s'`.`
Initialize *newphase* = False. `// Boolean for first time step in phase.`
Let $t = |H|$.
for $t'' = 1$ **to** t **do**
  for $b \in \mathcal{B}'$ **do**
    Let $n(a, b) := |\{(t'', a_{t''}, b_{t''}, r) \in H \mid a_{t''} = a, b_{t''} = b, t'' \geq t'\}|$.
  if $n(a, b) = M_{s'+1} \forall b \in \mathcal{B}'$ **then**
    newphase = True.
  if $newphase = True$ **then**
    for $b \in \mathcal{B}'$ **do**
      Set $S(a, b) := \{r \mid \exists (t'', a_{t''}, b_{t''}, r) \in H \text{ s.t. } a_{t''} = a, b = b_{t''}, t'' \geq t'\}$ `// observed rewards`
      $\hat{v}_2(a, b) \leftarrow (\sum_{r \in S(a,b)} r)/|S(a, b)|$ `// compute empirical mean`
    Update $\mathcal{B}' \leftarrow \{b \mid \hat{v}_2(a, b) + \frac{20 \cdot \sqrt{\log T}}{\sqrt{M_{s'}}} \geq \max_{b \in \mathcal{B}'} \hat{v}_2(a, b)\}$.
    Update $s' \leftarrow s' + 1$.
    Update $t' \leftarrow t$.
    newphase = False.
$i = ((t - t') \mod (|\mathcal{B}'|)) + 1$. `// Calculate next arm to be pulled`
return *point mass at* b_i.

Proposition 126. *Suppose that for every $a \in \mathcal{A}$, the follower runs a separate instantiation of* `ActiveArmElimination`$(M_1, \ldots, M_P)$ *(Algorithm 14) with $M_i = \Theta(\log T \cdot 2^{2i})$. Then the follower satisfies high-probability instantaneous regret $g(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot \log(T)/t}$, which implies $g(t, T, \mathcal{B}) = O\left((|\mathcal{A}||\mathcal{B}| \log T)^{1/3} T^{-1/3}\right)$ for $t \geq \Theta(|\mathcal{A}|^{-2/3}(|\mathcal{B}| \log T)^{1/3} T^{2/3})$. Moreover, the follower satisfies high-probability anytime regret $h(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot \log(T) \cdot t})$.*

Next, we show that `ExploreThenCommit` (Algorithm 8, see Slivkins (2019); Lattimore and Szepesvári (2020) for a textbook treatment) satisfies the high-probability instantaneous regret bound required for Theorem 119.

Proposition 127. *Suppose that the follower runs a separate instantiation of* `ExploreThenCommit`$(E, \mathcal{B})$ *(Algorithm 8) for every $a \in \mathcal{A}$. Then, the follower satisfies high-probability instantaneous regret $g(t, T, \mathcal{B}) = \mathcal{O}(\sqrt{\log T/E})$ for all time steps $t \geq E \cdot |\mathcal{B}|$. If $E = \Theta((|\mathcal{A} \cdot |\mathcal{B}|)^{-2/3}(\log T)^{1/3} T^{2/3})$, then $g(t, T, \mathcal{B}) = O\left((|\mathcal{A}||\mathcal{B}| \log T)^{1/3} T^{-1/3}\right)$ for $t \geq \Theta(|\mathcal{A}|^{-2/3}(|\mathcal{B}| \log T)^{1/3} T^{2/3})$.*

Note that `ExploreThenCommit` does *not* satisfy the high-probability anytime regret bound required for Theorem 121 due to the uniform exploration phase at the beginning of the algorithm.

Finally, we show that UCB (Auer et al., 2002a) (see Slivkins (2019); Lattimore and Szepesvári (2020) for a textbook treatment) satisfies the high-probability anytime regret bound required in Theorem 121.

Proposition 128. *Suppose that the follower runs a separate instantiation of UCB for every $a \in \mathcal{A}$. Then, the follower satisfies high-probability anytime regret bound $h(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot t \cdot \log(T)})$.*

We do not expect that UCB satisfies the high-probability instantaneous regret bound required for Theorem 119, using the intuition that UCB does not provide final-iterate convergence guarantees.

14.7.2 Generalized Analysis of `ExploreThenUCB` and `LipschitzUCB`

While the specific instantations of g and h in Theorem 119 and Theorem 121 are tailored to standard algorithms (Chapter 14.7.1), we generalize our analysis of `ExploreThenUCB` and `LipschitzUCB` to a richer class of functions g and h, respectively.

We generalize Theorem 119 to functions $g(t, T) = O(E^{-c_1}|\mathcal{B}|^{c_2}(\log T)^{c_3})$ for $t > E$, where $c_1 \in (0, 1)$ and $c_2, c_3 > 0$ are arbitrary parameters and where E is equal to $\Theta(|\mathcal{A}|^{-1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)} \log(T)^{c_2/(1+c_1)} \cdot T^{1/(1+c_1)})$.

Theorem 129. *Let $c_1 \in (0, 1)$ and $c_2, c_3 > 0$. Let $E = \Theta(|\mathcal{A}|^{-1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)}(\log T)^{c_3/(1+c_1)} \cdot T^{1/(1+c_1)})$. Consider a StrongDSG where $\texttt{ALG}_2$ is any algorithm with high-probability instantaneous regret $g(t, T, \mathcal{B}) = O\left(E^{-c_1}|\mathcal{B}|^{c_2}(\log T)^{c_3}\right)$ for $t > E$ and $g(t, T, \mathcal{B}) = 1$ for $t \le E$, and where $\texttt{ALG}_1 = \texttt{ExploreThenUCB}(E)$. Then, it holds that the regret $\max(R_1(T), R_2(T))$ with respect to the γ-tolerant benchmarks β_1^{tol} and β_2^{tol} is bounded as:*

$$O\left(T^{1/(1+c_1)} \cdot |\mathcal{A}|^{c_1/(1+c_1)} \cdot |\mathcal{B}|^{c_2/(1+c_1)} \cdot (\log T)^{c_3/(1+c_1)}\right) + \Theta\left(\sqrt{T|\mathcal{A}| \log T}\right).$$

Note that the special case of $c_1 = c_2 = c_3 = 1/2$ recovers the functional form of g in Theorem 119. The proof, which follows similarly to the proof of Theorem 119, is deferred to Appendix J.6.2.

We similarly generalize Theorem 121 to functions $h(t, T) = C' \cdot t^{c_1} \cdot |\mathcal{B}|^{c_2} \cdot (\log(T))^{c_3}$ for $t > E$, where $c_1, c_2, c_3 \in (0, 1)$ are arbitrary parameters. This result requires the leader to instead run `LipschitzUCBGen` (Algorithm 15), a generalized version of `LipschitzUCB` which adjusts the confidence set size based on the parameters c_1, c_2, and c_3.

Theorem 130. *Let $c_1 \in (0, 1)$, $c_2, c_3 > 0$, and $C' > 0$ be arbitrary constants. Consider a StrongDSG where $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ has Lipschitz constant L^*. Let $\texttt{ALG}_2$ be any algorithm satisfying high-probability anytime regret $h(t, T, \mathcal{B}) = C' \cdot t^{c_1} \cdot |\mathcal{B}|^{c_2} \cdot (\log(T))^{c_3}$. Let $\texttt{ALG}_1 = \texttt{LipschitzUCBGen}(L, C'B^{c_2}, c_1, c_3)$ for any $L \ge L^*$. Then both players achieve* *the following regret bounds with respect to the original Stackelberg benchmarks β_1^{orig} and β_2^{orig}: that is, $R_1(T;\mathcal{I}) = O\left(\sqrt{T|\mathcal{A}||\mathcal{B}|\log T} + L|\mathcal{A}|^{1-c_1}|\mathcal{B}|^{c_2}(\log T)^{c_3}T^{c_1}\right)$ and $R_2(T;\mathcal{I}) = O\left(L\sqrt{T|\mathcal{A}||\mathcal{B}|\log T} + L^2|\mathcal{A}|^{1-c_1}|\mathcal{B}|^{c_2}T^{c_1}(\log T)^{c_3}\right)$.*

Again, note that the special case of $c_1 = c_2 = c_3 = 1/2$ recovers the functional form of g in Theorem 121. The proof, which follows similarly to the proof of Theorem 121, is deferred to Appendix J.6.2.

Algorithm 15: `LipschitzUCBGen`(L, C, c_1, c_3) applied to H

Initialize $\hat{v}_1(a) = 1$ for $a \in \mathcal{A}$. `// Initialize empirical means for` $\max_{b\in\mathcal{B}} v_1(a,b)$.
Initialize $v_1^{\mathrm{UCB}}(a) = 1$ for $a \in \mathcal{A}$. `// Initialize UCB for` $\max_{b\in\mathcal{B}} v_1(a,b)$
for $a \in \mathcal{A}$ **do**
  Set $S(a) := \{r \mid \exists(t', a_{t'}, r) \in H \text{ s.t. } a = a_{t'}\}$ `// Observed rewards`
  if $S(a) \neq \emptyset$ **then**
    $\hat{v}_1(a) \leftarrow (\sum_{r\in S(a)} r)/|S(a)|$ `// Empirical mean`
    $\alpha(a) \leftarrow \frac{10\sqrt{\mathcal{B}\log T}}{\sqrt{|S(a)|}} + C \cdot L \cdot (\log T)^{c_3}T^{c_1-1}$ `// confidence bound width`
    $v_1^{\mathrm{UCB}}(a) \leftarrow \min(1, \hat{v}_1(a) + \alpha(a))$
Let $a^* = \arg\max_{a\in\mathcal{A}} \left(v_1^{\mathrm{UCB}}(a)\right)$. `// arm with max upper confidence bound`
return *point mass at* a^*.

14.8 Discussion

In this chapter, we studied two-agent environments where interactions are *sequential*, utilities are *misaligned*, and each agent *learns* their utilities over time. We modeled these environments as decentralized Stackelberg games where both agents are bandit learners who only observe their own utilities, and we investigated the implications for each agent's cumulative utility over time. Motivated by the offline Stackelberg equilibrium benchmarks being infeasible (Theorem 116), we designed γ-tolerant benchmarks which allow for approximate best responses by the other agent.

We proved that both players can achieve $\tilde{\Theta}(T^{2/3})$ regret with respect to the γ-tolerant benchmarks. To achieve this regret bound, we designed an algorithm (i.e., `ExploreThenUCB`; Algorithm 10) where the leader waits for the follower to partially converge before starting to learn; this algorithm achieves $\tilde{\Theta}(T^{2/3})$ regret for both players under a rich class of follower learning algorithms (Theorem 119). We further show that $\tilde{\Theta}(T^{2/3})$ regret is unavoidable for any pair of algorithms (Theorem 120). Furthermore, we showed that $O(\sqrt{T})$ regret is possible in two relaxed environments: i.e., under a relaxed benchmark that is (self-)tolerant of a player's own mistakes (Theorem 122) or when players agree on which pair of actions are different (Theorem 121)

Our results have broader implications for *designing* two-agent environments to achieve favorable utility for both agents. For example, given that our results illustrate that certain properties for the follower (such as high-probability instantaneous regret or high-probability anytime regret bounds) and certain properties for the leader (such as waiting for the follower to partially converge) are conducive to low regret, it may be helpful for a designer to engineer or encourage agents to follow these algorithmic principles. As another example, our continuity results in Chapter 14.5.1 illustrate the importance of reducing *near-ties* in utilities between different items, which could be achieved by allowing agents to express preferences between items in a nuanced fashion.

More broadly, our benchmarks and regret analysis suggest several interesting avenues for future work. For example, while Theorem 119 offered flexibility in the follower's choice of algorithm, we required that the leader follow a particular algorithm: it would be interesting to explore richer classes of leader algorithms which maintain low regret. Additionally, while our framework captures a range of real-world applications including chatbots (Example 10 in Chapter 14.2.3) and recommender systems (Example 11 in Chapter 14.2.3), an interesting future direction would be to focus on a particular application and incorporate application-specific nuances (e.g., bidder learning rates in advertising auctions (Nekipelov et al., 2015; Noti and Syrgkanis, 2021; Nisan and Noti, 2017)). Finally, while we study the role of continuity requirements that reflect alignment (Chapter 14.5.1), it would be interesting to consider other structured bandit environments such as linear utility functions and generalize our benchmarks and results accordingly.

Bibliography

Jacob D Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Conference on Learning Theory*, number 110, 2009.

Daron Acemoglu. The simple macroeconomics of ai. *Economic Policy*, 40(121):13–58, 2025.

Daron Acemoglu and Pablo D Azar. Endogenous production networks. *Econometrica*, 88(1): 33–82, 2020.

Daron Acemoglu and Alireza Tahbaz-Salehi. The macroeconomics of supply chain disruptions. *Review of Economic Studies*, page rdae038, 2024.

Josef Adalian. Inside the binge factory, 2018. URL https://www.vulture.com/2018/06/how-netflix-swallowed-tv-industry.html.

Gediminas Adomavicius, Jesse C Bockstedt, Shawn P Curley, and Jingjing Zhang. Do recommender systems manipulate consumer preferences? a study of anchoring effects. *Information Systems Research*, 24(4):956–975, 2013.

Alekh Agarwal, Haipeng Luo, Behnam Neyshabur, and Robert E Schapire. Corralling a band of bandit algorithms. In *Conference on Learning Theory*, pages 12–38. PMLR, 2017.

Rajeev Agrawal. The continuum-armed bandit problem. *SIAM journal on control and optimization*, 33(6):1926–1951, 1995.

Saba Ahmadi, Hedyeh Beyhaghi, Avrim Blum, and Keziah Naggita. The strategic perceptron. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 6–25, 2021.

Saba Ahmadi, Hedyeh Beyhaghi, Avrim Blum, and Keziah Naggita. On classification of strategic agents who can both game and improve. *arXiv preprint arXiv:2203.00124*, 2022.

S Nageeb Ali, Nicole Immorlica, Meena Jagadeesan, and Brendan Lucier. Flattening supply chains: When do technology improvements lead to disintermediation? *arXiv preprint arXiv:2502.20783*, 2025.

Max Alston. On the non-existence of stable matches with incomplete information. *Games and Economic Behavior*, 120:336–344, 2020.

Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 736–749, 2022.

Simon P Anderson, Andre De Palma, and Jacques-Francois Thisse. *Discrete choice theory of product differentiation*. MIT press, 1992.

Guy Aridor and Duarte Gonçalves. Recommenders' originals: The welfare effects of the dual role of platforms as producers and recommender systems. *International Journal of Industrial Organization*, 83:102845, 2022.

Guy Aridor, Yishay Mansour, Aleksandrs Slivkins, and Steven Wu. Competing bandits: The perils of exploration under competition. *ACM Transactions on Economics and Computation*, 13(1):1–47, 2025.

Eshwar Ram Arunachaleswaran, Natalie Collina, and Meena Jagadeesan. Breaking algorithmic collusion via simple defections. 2025.

Itai Ashlagi, Mark Braverman, Yash Kanoria, and Peng Shi. Clearing matching markets efficiently: informative signals and match recommendations. *Management Science*, 66(5): 2163–2193, 2020.

Alexander Atanasov, Jacob A Zavatone-Veth, and Cengiz Pehlevan. Scaling and renormalization in high-dimensional regression. *arXiv preprint arXiv:2405.00592*, 2024.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002a.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002b.

Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *International Conference on Computational Learning Theory*, pages 454–468. Springer, 2007.

José Azar, Martin C Schmalz, and Isabel Tecu. Anticompetitive effects of common ownership. *The Journal of Finance*, 73(4):1513–1565, 2018.

Eduardo M Azevedo and John William Hatfield. Existence of equilibrium in large matching markets with complementarities. *Available at SSRN 3268884*, 2018.

Francis Bach. High-dimensional analysis of double descent for linear regression with random projections. *SIAM Journal on Mathematics of Data Science*, 6(1):26–50, 2024.

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. *Journal of the ACM (JACM)*, 65(3):1–55, 2018.

Yasaman Bahri, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma. Explaining neural scaling laws. *Proceedings of the National Academy of Sciences*, 121(27):e2311878121, 2024.

Yu Bai, Chi Jin, Huan Wang, and Caiming Xiong. Sample-efficient learning of stackelberg equilibria in general-sum games. *Advances in Neural Information Processing Systems*, 34: 25799–25811, 2021.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

Jonathan B Baker. Beyond schumpeter vs. arrow: How antitrust fosters innovation'(2007). *Antitrust Law Journal*, 74:575.

Michiel Bakker, Martin Chadwick, Hannah Sheahan, Michael Tessler, Lucy Campbell-Gillingham, Jan Balaguer, Nat McAleese, Amelia Glaese, John Aslanides, Matt Botvinick, et al. Fine-tuning language models to find agreement among humans with diverse preferences. *Advances in Neural Information Processing Systems*, 35:38176–38189, 2022.

Maria-Florina Balcan, Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Commitment without regrets: Online learning in stackelberg security games. In *Proceedings of the sixteenth ACM conference on economics and computation*, pages 61–78, 2015.

Ian Ball. Scoring strategic agents. *American Economic Journal: Microeconomics*, 17(1): 97–129, 2025.

Santiago R Balseiro and Yonatan Gur. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–16, 2021.

Ran Ben Basat, Moshe Tennenholtz, and Oren Kurland. A game theoretic analysis of the adversarial retrieval setting. *Journal of Artificial Intelligence Research*, 60:1127–1164, 2017.

Soumya Basu, Karthik Abinav Sankararaman, and Abishek Sankararaman. Beyond log2(t) regret for decentralized bandits in matching markets. In *International Conference on Machine Learning*, pages 705–715. PMLR, 2021.

Michael R. Baye and Dan Kovenock. *Bertrand competition.* Palgrave Macmillan UK, London, 2008.

Yahav Bechavod, Katrina Ligett, Steven Wu, and Juba Ziani. Gaming helps! learning from strategic interactions in natural dynamics. In *International Conference on Artificial Intelligence and Statistics*, pages 1234–1242. PMLR, 2021.

Yahav Bechavod, Chara Podimata, Steven Wu, and Juba Ziani. Information discrepancy in strategic learning. In *International Conference on Machine Learning*, pages 1691–1715. PMLR, 2022.

Omer Ben-Porat and Moshe Tennenholtz. Best response regression. *Advances in Neural Information Processing Systems*, 30, 2017.

Omer Ben-Porat and Moshe Tennenholtz. A game-theoretic approach to recommendation systems with strategic content providers. *Advances in Neural Information Processing Systems*, 31, 2018.

Omer Ben-Porat and Moshe Tennenholtz. Regression equilibrium. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 173–191, 2019.

Omer Ben-Porat and Rotem Torkan. Learning with exposure constraints in recommendation systems. In *Proceedings of the ACM Web Conference 2023*, pages 3456–3466, 2023.

Omer Ben-Porat, Itay Rosenberg, and Moshe Tennenholtz. Content provider dynamics and coordination in recommendation ecosystems. *Advances in Neural Information Processing Systems*, 33:18931–18941, 2020.

Priyanjana Bengani, Jonathan Stray, and Luke Thorburn. What's right and what's wrong with optimizing for engagement. *Understanding Recommenders*, Apr 2022. URL https://medium.com/understanding-recommenders/whats-right-and-what-s-wrong-with-optimizing-for-engagement-5abaac021851.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8): 1798–1828, 2013.

Benjamin Bennett, Rene M. Stulz, and Zexi Wang. Does greater public scrutiny hurt a firm's performance? Available at SSRN: https://ssrn.com/abstract=4321191, 2023.

Dirk Bergemann and Alessandro Bonatti. Data, competition, and digital platforms. *American Economic Review*, 114(8):2553–2595, 2024.

Dirk Bergemann and Juuso Välimäki. Experimentation in markets. *The Review of Economic Studies*, 67(2):213–234, 2000.

Steven T Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

Martin Bichler, Maximilian Fichtl, and Gregor Schwarz. Walrasian equilibria from an optimization perspective: A guide to the literature. *Naval Research Logistics (NRL)*, 68 (4):496–513, 2021.

Sushil Bikhchandani. Stability with one-sided incomplete information. *Journal of Economic Theory*, 168:372–399, 2017.

Kostas Bimpikis, Ozan Candogan, and Shayan Ehsani. Supply disruptions and optimal network structures. *Management Science*, 65(12):5504–5517, 2019.

Avrim Blum, Nika Haghtalab, Ariel D Procaccia, and Mingda Qiao. Collaborative pac learning. *Advances in Neural Information Processing Systems*, 30, 2017.

Lawrence Blume, David Easley, Jon Kleinberg, Robert Kleinberg, and Éva Tardos. Network formation in the presence of contagious risk. *ACM Transactions on Economics and Computation (TEAC)*, 1(2):1–20, 2013.

Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2): 349–374, 1999.

Patrick Bolton and Christopher Harris. Strategic experimentation: the undiscounted case. *Incentives, Organizations and Public Economics – Papers in Honour of Sir James Mirrlees*, pages 53–68, 2000a.

Patrick Bolton and Christopher Harris. Strategic experimentation : The undiscounted case. 2000b. URL https://www0.gsb.columbia.edu/faculty/pbolton/PDFS/strategi.pdf.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Rishi Bommasani, Kathleen A Creel, Ananya Kumar, Dan Jurafsky, and Percy S Liang. Picking on the same person: Does algorithmic monoculture lead to outcome homogenization? *Advances in Neural Information Processing Systems*, 35:3663–3678, 2022.

Blake Bordelon, Abdulkadir Canatar, and Cengiz Pehlevan. Spectrum dependent learning curves in kernel regression and wide neural networks. In *International Conference on Machine Learning*, pages 1024–1034. PMLR, 2020.

Blake Bordelon, Alexander Atanasov, and Cengiz Pehlevan. A dynamical model of neural scaling laws. *arXiv preprint arXiv:2402.01092*, 2024.

Christian Borgs, Jennifer Chayes, Nicole Immorlica, Kamal Jain, Omid Etesami, and Mohammad Mahdian. Dynamics of bid optimization in online advertisement auctions. In *Proceedings of the 16th international conference on World Wide Web*, pages 531–540, 2007.

Simina Brânzei and Yuval Peres. Multiplayer bandit learning, from competition to cooperation. In *Conference on Learning Theory*, pages 679–723. PMLR, 2021.

Simina Brânzei, MohammadTaghi Hajiaghayi, Reed Phillips, Suho Shin, and Kun Wang. Dueling over dessert, mastering the art of repeated cake cutting. *Advances in Neural Information Processing Systems*, 37:97699–97765, 2024.

Mark Braverman, Jieming Mao, Jon Schneider, and Matt Weinberg. Selling to a no-regret buyer. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 523–538, 2018.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *International conference on artificial intelligence and statistics*, pages 6045–6061. PMLR, 2022a.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *International conference on artificial intelligence and statistics*, pages 6045–6061. PMLR, 2022b.

William Brown and Arpit Agarwal. Online recommendations for agents with discounted adaptive preferences. In *International Conference on Algorithmic Learning Theory*, pages 244–281. PMLR, 2024.

William Brown, Jon Schneider, and Kiran Vodrahalli. Is learning in games good for the learners? *Advances in Neural Information Processing Systems*, 36:54228–54249, 2023.

Michael Brückner, Christian Kanzow, and Tobias Scheffer. Static prediction games for adversarial learning problems. *The Journal of Machine Learning Research*, 13(1):2617–2654, 2012.

Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. Generative ai at work. *The Quarterly Journal of Economics*, page qjae044, 2025.

Sébastien Bubeck, Gilles Stoltz, and Jia Yuan Yu. Lipschitz bandits without the lipschitz constant. In *Algorithmic Learning Theory: 22nd International Conference, ALT 2011, Espoo, Finland, October 5-7, 2011. Proceedings 22*, pages 144–158. Springer, 2011.

Thomas Kleine Buening, Aadirupa Saha, Christos Dimitrakakis, and Haifen Xu. Bandits meet mechanism design to combat clickbait in online recommendation. In *The Twelfth International Conference on Learning Representations*, pages 1–29, 2024.

Gordon Burtch, Dokyun Lee, and Zhichen Chen. The consequences of generative ai for online knowledge communities. *Scientific Reports*, 14(1):10413, 2024.

California Legislature. California senate bill no. 1047 (2023-2024). https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1047, 2024.

Emilio Calvano and Michele Polo. Market power, competition and innovation in digital markets: A survey. *Information Economics and Policy*, 54:100853, 2021a.

Emilio Calvano and Michele Polo. Market power, competition and innovation in digital markets: A survey. *Information Economics and Policy*, 54:100853, 2021b.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. Artificial intelligence, algorithmic recommendations and competition. *Social Science Research Network*, May 2023.

Modibo K Camara, Jason D Hartline, and Aleck Johnsen. Mechanisms for a no-regret agent: Beyond the common prior. In *2020 ieee 61st annual symposium on foundations of computer science (focs)*, pages 259–270. IEEE, 2020.

Micah D Carroll, Anca Dragan, Stuart Russell, and Dylan Hadfield-Menell. Estimating and penalizing induced preference shifts in recommender systems. In *International Conference on Machine Learning*, pages 2686–2708. PMLR, 2022.

Jacopo Castellini, Amelia Fletcher, Peter L. Ormosi, and Rahul Savani. Recommender systems and competition on subscription-based platforms. *Social Science Research Network*, April 2023.

Sarah H Cen and Devavrat Shah. Regret, stability & fairness in matching markets with bandit learners. In *International Conference on Artificial Intelligence and Statistics*, pages 8938–8968. PMLR, 2022.

Nicolo Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

Lawrence Chan, Dylan Hadfield-Menell, Siddhartha Srinivasa, and Anca Dragan. The assistive multi-armed bandit. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 354–363. IEEE, 2019.

Chandra R Chegireddy and Horst W Hamacher. Algorithms for finding k-best perfect matchings. *Discrete applied mathematics*, 18(2):155–165, 1987.

Banghao Chen, Zhaofeng Zhang, Nicolas Langrené, and Shengxin Zhu. Unleashing the potential of prompt engineering in large language models: a comprehensive review. *arXiv preprint arXiv:2310.14735*, 2023.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International conference on machine learning*, pages 151–159. PMLR, 2013.

Yiling Chen, Yang Liu, and Chara Podimata. Learning strategy-aware linear classifiers. *Advances in Neural Information Processing Systems*, 33:15265–15276, 2020a.

Yiling Chen, Yang Liu, and Chara Podimata. Learning strategy-aware linear classifiers. *Advances in Neural Information Processing Systems*, 33:15265–15276, 2020b.

Shih-fen Cheng, Daniel M. Reeves, Yevgeniy Vorobeychik, and Michael P. Wellman. Notes on equilibria in symmetric games. In *Proceedings of the 6th International Workshop on Game Theoretic and Decision Theoretic Agents (GTDT)*, pages 71–78, 2004.

Moustapha Cisse, Piotr Bojanowski, Edouard Grave, Yann Dauphin, and Nicolas Usunier. Parseval networks: Improving robustness to adversarial examples. In *International conference on machine learning*, pages 854–863. PMLR, 2017.

Alon Cohen, Tamir Hazan, and Tomer Koren. Online learning with feedback graphs without the graphs. In *International Conference on Machine Learning*, pages 811–819. PMLR, 2016.

Natalie Collina, Eshwar Ram Arunachaleswaran, and Michael Kearns. Efficient stackelberg strategies for finitely repeated games. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2023, London, United Kingdom, 29 May 2023 - 2 June 2023*, pages 643–651. ACM, 2023.

Natalie Collina, Aaron Roth, and Han Shao. Efficient prior-free mechanisms for no-regret agents. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 511–541, 2024.

Richard Combes, Mohammad Sadegh Talebi Mazraeh Shahi, Alexandre Proutiere, et al. Combinatorial bandits revisited. *Advances in neural information processing systems*, 28, 2015.

Competition and Markets Authority. AI foundation models: Technical update report. Technical report, UK Government, 2024. URL `https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf`.

Jodie Cook. ChatGPT, Claude, Gemini or another: The AI tool entrepreneurs prefer. *Forbes*, 2024. URL `https://www.forbes.com/sites/jodiecook/2024/05/07/chatgpt-claude-gemini-or-another-the-ai-tool-entrepreneurs-prefer/`.

Nina Corvelo Benz and Manuel Rodriguez. Human-aligned calibration for ai-assisted decision making. *Advances in Neural Information Processing Systems*, 36:14609–14636, 2023.

Ian Covert, Wenlong Ji, Tatsunori Hashimoto, and James Zou. Scaling laws for the value of individual data points in machine learning. *arXiv preprint arXiv:2405.20456*, 2024.

Jacques Crémer, Yves-Alexandre de Montjoye, and Heike Schweitzer. *Competition Policy for the digital era : Final report*. Publications Office of the European Union, 2019.

Hugo Cui, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Generalization error rates in kernel regression: The crossover from the noiseless to noisy regime. *Advances in Neural Information Processing Systems*, 34:10131–10143, 2021.

Zheyuan Kevin Cui, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. The effects of generative ai on high skilled work: Evidence from three field experiments with software developers. *Available at SSRN 4945566*, 2024.

Mihaela Curmei, Andreas A Haupt, Benjamin Recht, and Dylan Hadfield-Menell. Towards psychologically-grounded dynamic preference models. In *Proceedings of the 16th ACM Conference on Recommender Systems*, pages 35–48, 2022.

Jessica Dai, Bailey Flanigan, Nika Haghtalab, Meena Jagadeesan, and Chara Podimata. Can probabilistic feedback drive user impacts in online platforms? In *International Conference on Artificial Intelligence and Statistics*, pages 2512–2520. PMLR, 2024.

Sanmay Das and Emir Kamenica. Two-sided bandits and the dating market. In *IJCAI*, volume 5, page 19, 2005.

Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 235–254. SIAM, 2011.

Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems*, 34:27604–27616, 2021.

Claude d'Aspremont, J Jaskold Gabszewicz, and J-F Thisse. On hotelling's" stability in competition". *Econometrica: Journal of the Econometric Society*, pages 1145–1150, 1979.

Jean Magnan de Bornier. The 'cournot-bertrand debate': A historical perspective. *History of Political Economy*, 24(3):623–656, 1992.

Sarah Dean and Jamie Morgenstern. Preference dynamics under personalized recommendations. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 795–816, 2022.

Sarah Dean, Sarah Rich, and Benjamin Recht. Recommendations and user agency: the reachability of collaboratively-filtered information. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 436–445, 2020.

Sarah Dean, Mihaela Curmei, Lillian Ratliff, Jamie Morgenstern, and Maryam Fazel. Emergent specialization from participation dynamics and multi-learner retraining. In *International Conference on Artificial Intelligence and Statistics*, pages 343–351. PMLR, 2024a.

Sarah Dean, Evan Dong, Meena Jagadeesan, and Liu Leqi. Accounting for ai and users shaping one another: The role of mathematical models. *arXiv preprint arXiv:2404.12366*, 2024b.

Gerard Debreu and Herbert Scarf. A limit theorem on the core of an economy. *International Economic Review*, 4(3):235–246, 1963.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, pages 248–255. IEEE Computer Society, 2009.

Yuan Deng, Jon Schneider, and Balasubramanian Sivan. Strategizing against no-regret learners. *Advances in neural information processing systems*, 32, 2019.

Elvis Dohmatob, Yunzhen Feng, Pu Yang, Francois Charton, and Julia Kempe. A tale of tails: Model collapse as a change of scaling laws. *arXiv preprint arXiv:2402.07043*, 2024.

Kate Donahue, Sreenivas Gollapudi, and Kostas Kollias. When are two lists better than one?: Benefits and harms in joint decision-making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 10030–10038, 2024a.

Kate Donahue, Nicole Immorlica, Meena Jagadeesan, Brendan Lucier, and Aleksandrs Slivkins. Impact of decentralized learning on player utilities in stackelberg games. In *International Conference on Machine Learning*, pages 11253–11310. PMLR, 2024b.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

Jinshuo Dong, Hadi Elzayn, Shahin Jabbari, Michael Kearns, and Zachary Schutzman. Equilibrium characterization for data acquisition games. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 252–258, 2019.

Roy Dong, Heling Zhang, and Lillian Ratliff. Approximate regions of attraction in learning with decision-dependent distributions. In *International Conference on Artificial Intelligence and Statistics*, pages 11172–11184. PMLR, 2023.

Dmitriy Drusvyatskiy and Lin Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 48(2):954–998, 2023.

Federico Echenique, Sangmok Lee, Matthew Shum, and M Bumin Yenmez. The revealed preference theory of stable and extremal stable matchings. *Econometrica*, 81(1):153–171, 2013.

Michael D Ekstrand and Martijn C Willemsen. Behaviorism is not enough: better recommendations through listening to users. In *Proceedings of the 10th ACM conference on recommender systems*, pages 221–224, 2016.

Matthew Elliott, Benjamin Golub, and Matthew V Leduc. Supply network formation and fragility. *American Economic Review*, 112(8):2701–2747, 2022.

Ehsan Emamjomeh-Zadeh, Yannai A Gonczarowski, and David Kempe. The complexity of interactively learning a stable matching by trial and error. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 599–599, 2020.

Seyed A Esmaeili, Kevin Lim, Kshipra Bhawalkar, Zhe Feng, Di Wang, and Haifeng Xu. How to strategize human content creation in the era of genai? *arXiv preprint arXiv:2406.05187*, 2024.

European Union. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and Amending Directive 2000/31/EC (Digital Services Act). Official Journal of the European Union, 2022a. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065.

European Union. Regulation (EU) 2022/1925 of the European parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and Amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), 2022b. URL https://eur-lex.europa.eu/eli/reg/2022/1925/oj.

European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 March 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (AI Act). AI Act Explorer, 2024. URL https://artificialintelligenceact.eu/ai-act-explorer/. Accessed: 2025-05-11.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Pac bounds for multi-armed bandit and markov decision processes. In *Computational Learning Theory: 15th Annual Conference on Computational Learning Theory, COLT 2002 Sydney, Australia, July 8–10, 2002 Proceedings 15*, pages 255–270. Springer, 2002.

Alireza Fallah and Michael Jordan. Contract design with safety inspections. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 616–638, 2024.

Alireza Fallah, Michael I Jordan, Ali Makhdoumi, and Azarakhsh Malekian. On three-layer data markets. *arXiv preprint arXiv:2402.09697*, 2024.

Yiding Feng, Ronen Gradwohl, Jason Hartline, Aleck Johnsen, and Denis Nekipelov. Bias-variance games. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 328–329, 2022.

Tanner Fiez and Lillian J Ratliff. Local convergence analysis of gradient descent ascent with finite timescale separation. In *Proceedings of the International Conference on Learning Representation*, 2021.

Tanner Fiez, Benjamin Chasnov, and Lillian J Ratliff. Convergence of learning dynamics in stackelberg games. *arXiv preprint arXiv:1906.01217*, 2019.

Tanner Fiez, Benjamin Chasnov, and Lillian Ratliff. Implicit learning dynamics in stackelberg games: Equilibria characterization, convergence analysis, and empirical study. In *International Conference on Machine Learning*, pages 3133–3144. PMLR, 2020.

Seth Flaxman, Sharad Goel, and Justin M Rao. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320, 2016.

Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):2, 2002.

Alex Frankel and Navin Kartik. Muddled information. *Journal of Political Economy*, 127(4): 1739–1776, 2019.

Alex Frankel and Navin Kartik. Improving information from manipulable data. *Journal of the European Economic Association*, 20(1):79–115, 2022.

Peter Frazier, David Kempe, Jon Kleinberg, and Robert Kleinberg. Incentivizing exploration. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 5–22, 2014.

Yi Gai, Bhaskar Krishnamachari, and Rahul Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. *IEEE/ACM Transactions on Networking*, 20(5):1466–1478, 2012.

Michal S Gal and Oshrit Aviv. The competitive effects of the gdpr. *Journal of Competition Law & Economics*, 16(3):349–391, 2020.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American mathematical monthly*, 69(1):9–15, 1962.

Jiarui Gan, Minbiao Han, Jibang Wu, and Haifeng Xu. Robust stackelberg equilibria. *arXiv preprint arXiv:2304.14990*, 2023.

Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning*, pages 10835–10866. PMLR, 2023.

Ernest Gellhorn. An introduction to antitrust economics. *Duke LJ*, page 1, 1975.

Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, et al. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. *arXiv preprint arXiv:2404.01413*, 2024.

Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, and Nir Rosenfeld. Strategic classification in the dark. In *International Conference on Machine Learning*, pages 3672–3681. PMLR, 2021.

Arpita Ghosh and Patrick Hummel. Learning and incentives in user-generated content: Multi-armed bandits with endogenous arms. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 233–246, 2013.

Arpita Ghosh and Preston McAfee. Incentivizing high-quality user-generated content. In *Proceedings of the 20th international conference on World wide web*, pages 137–146, 2011.

Tony Ginart, Eva Zhang, Yongchan Kwon, and James Zou. Competing ai: How does competition feedback affect machine learning? In *International Conference on Artificial Intelligence and Statistics*, pages 1693–1701. PMLR, 2021.

John C Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(2):148–164, 1979.

John C Gittins and David M Jones. A dynamic allocation index for the discounted multiarmed bandit problem. *Biometrika*, 66(3):561–565, 1979.

Denizalp Goktas, Jiayi Zhao, and Amy Greenwald. Robust no-regret learning in min-max stackelberg games. *arXiv preprint arXiv:2203.14126*, 2022.

Yannai A Gonczarowski, Noam Nisan, Rafail Ostrovsky, and Will Rosenbaum. A stable marriage requires communication. *Games and Economic Behavior*, 118:626–647, 2019.

Sachin Goyal, Pratyush Maini, Zachary C Lipton, Aditi Raghunathan, and J Zico Kolter. Scaling laws for data filtering–data curation cannot be compute agnostic. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22702–22711, 2024.

Ronen Gradwohl and Moshe Tennenholtz. Coopetition against an amazon. *Journal of Artificial Intelligence Research*, 76:1077–1116, 2023.

Sanford J Grossman. Nash equilibrium and the industrial organization of markets with large fixed costs. *Econometrica: Journal of the Econometric Society*, pages 1149–1172, 1981.

Wenshuo Guo, Karl Krauth, Michael Jordan, and Nikhil Garg. The stereotyping problem in collaboratively filtered recommender systems. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–10, 2021.

Wenshuo Guo, Nika Haghtalab, Kirthevasan Kandasamy, and Ellen Vitercik. Leveraging reviews: Learning to price with buyer and seller uncertainty. *ACM SIGecom Exchanges*, 22(1):74–82, 2024.

Guru Guruganesh, Yoav Kolumbus, Jon Schneider, Inbal Talgam-Cohen, Emmanouil-Vasileios Vlatakis-Gkaragkounis, Joshua Wang, and S Weinberg. Contracting with a learning agent. *Advances in Neural Information Processing Systems*, 37:77366–77408, 2024.

C. W. Ha. A non-compact minimax theorem. *Pacific Journal of Mathematics*, 97:115–117, 1981.

Walid Hachem, Philippe Loubaton, and Jamal Najim. Deterministic equivalents for certain functionals of large random matrices. 2007.

Nika Haghtalab, Nicole Immorlica, Brendan Lucier, and Jack Z Wang. Maximizing welfare with incentive-aware evaluation mechanisms. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 160–166, 2021.

Nika Haghtalab, Michael Jordan, and Eric Zhao. On-demand sampling: Learning optimally from multiple distributions. *Advances in Neural Information Processing Systems*, 35: 406–419, 2022a.

Nika Haghtalab, Thodoris Lykouris, Sloan Nietert, and Alexander Wei. Learning in stackelberg games with non-myopic agents. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 917–918, 2022b.

Nika Haghtalab, Chara Podimata, and Kunhe Yang. Calibrated stackelberg games: Learning optimal commitments against calibrated agents. *Advances in Neural Information Processing Systems*, 36:61645–61677, 2023.

Mohammad Hajiaghayi, Mohammad Mahdavi, Keivan Rezaei, and Suho Shin. Regret analysis of repeated delegated choice. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9757–9764, 2024.

Minbiao Han, Michael Albert, and Haifeng Xu. Learning in online principal-agent interactions: The power of menus. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17426–17434, 2024.

Tinashe Handina and Eric Mazumdar. Rethinking scaling laws for learning in strategic environments. *arXiv e-prints*, pages arXiv–2402, 2024.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122, 2016.

Moritz Hardt, Meena Jagadeesan, and Celestine Mendler-Dünner. Performative power. *Advances in Neural Information Processing Systems*, 35:22969–22981, 2022.

F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TIIS)*, 5(4):1–19, 2015.

Winston Harrington. Enforcement leverage when penalties are restricted. *Journal of Public Economics*, 37(1):29–53, 1988.

Keegan Harris, Steven Z Wu, and Maria-Florina F Balcan. Regret minimization in stackelberg games with side information. *Advances in Neural Information Processing Systems*, 37: 12944–12976, 2024.

Md Rajibul Hasan, Ashish Kumar Jha, and Yi Liu. Excessive use of online video streaming services: Impact of recommender system use, psychological factors, and motives. *Computers in Human Behavior*, 80:220–228, 2018.

Tatsunori Hashimoto. Model performance scaling with multiple data sources. In *International Conference on Machine Learning*, pages 4107–4116. PMLR, 2021.

Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, and Percy Liang. Fairness without demographics in repeated loss minimization. In *International Conference on Machine Learning*, pages 1929–1938. PMLR, 2018.

Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J Tibshirani. Surprises in high-dimensional ridgeless least squares interpolation. *Annals of statistics*, 50(2):949, 2022.

Andreas Haupt, Dylan Hadfield-Menell, and Chara Podimata. Recommending to strategic users. *arXiv preprint arXiv:2302.06559*, 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016.

Matthias Hein and Maksym Andriushchenko. Formal guarantees on the robustness of a classifier against adversarial manipulation. *Advances in neural information processing systems*, 30, 2017.

Christopher A Hennessy and Charles AE Goodhart. Goodhart's law and machine learning: a structural perspective. *International Economic Review*, 64(3):1075–1086, 2023.

Danny Hernandez, Jared Kaplan, Tom Henighan, and Sam McCandlish. Scaling laws for transfer. *arXiv preprint arXiv:2102.01293*, 2021.

Thomas Hodgson. Spotify and the democratisation of music. *Popular Music*, 40(1):1–17, 2021.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022.

Joey Hong, Sergey Levine, and Anca Dragan. Zero-shot goal-directed dialogue via rl on imagined conversations. *arXiv preprint arXiv:2311.05584*, 2023.

Aspen Hopkins, Sarah H Cen, Andrew Ilyas, Isabella Struckman, Luis Videgaray, and Aleksander Mądry. Ai supply chains: An emerging ecosystem of ai actors, products, and services. *arXiv preprint arXiv:2504.20185*, 2025.

Johannes Hörner and Andrzej Skrzypacz. Learning, experimentation and information design. *Advances in Economics and Econometrics*, 1:63–98, 2017.

Harold Hotelling. Stability in competition. *Economic Journal*, 39(153):41–57, 1929.

Jiri Hron, Karl Krauth, Michael I Jordan, Niki Kilbertus, and Sarah Dean. Modeling content creator incentives on algorithm-curated platforms. *arXiv preprint arXiv:2206.13102*, 2022.

Xinyan Hu, Meena Jagadeesan, Michael I Jordan, and Jacob Steinhardt. Incentivizing high-quality content in online recommender systems. *arXiv preprint arXiv:2306.07479*, 2023.

Daniel Huttenlocher, Hannah Li, Liang Lyu, Asuman Ozdaglar, and James Siderius. Matching of users and creators in two-sided markets with departures. *arXiv preprint arXiv:2401.00313*, 2023.

Enrique Ide and Eduard Talamas. Artificial intelligence in the knowledge economy. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 834–836, 2024.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.

Nicole Immorlica, Adam Tauman Kalai, Brendan Lucier, Ankur Moitra, Andrew Postlewaite, and Moshe Tennenholtz. Dueling algorithms. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 215–224, 2011.

Nicole Immorlica, Greg Stoddard, and Vasilis Syrgkanis. Social status and badge design. In *Proceedings of the 24th international conference on World Wide Web*, pages 473–483, 2015.

Nicole Immorlica, Karthik Sankararaman, Robert Schapire, and Aleksandrs Slivkins. Adversarial bandits with knapsacks. *Journal of the ACM*, 69(6):1–47, 2022.

Nicole Immorlica, Meena Jagadeesan, and Brendan Lucier. Clickbait vs. quality: How engagement-based optimization shapes the content landscape in online platforms. In *Proceedings of the ACM Web Conference 2024*, pages 36–45, 2024.

David Ingram. Fewer people using Elon Musk's X as platform struggles to keep users. *NBC News*, 2024. URL https://www.nbcnews.com/tech/tech-news/fewer-people-using-elon-musks-x-struggles-keep-users-rcna144115.

Ganesh Iyer and T Tony Ke. Competitive algorithmic targeting and model selection. *Available at SSRN 4214973*, 2022.

Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: performative gradient descent. In *International Conference on Machine Learning*, pages 4641–4650. PMLR, 2021.

Meena Jagadeesan, Celestine Mendler-Dünner, and Moritz Hardt. Alternative microfoundations for strategic classification. In *International Conference on Machine Learning*, pages 4687–4697. PMLR, 2021.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-Dünner. Regret minimization with performative feedback. In *International Conference on Machine Learning*, pages 9760–9785. PMLR, 2022.

Meena Jagadeesan, Nikhil Garg, and Jacob Steinhardt. Supply-side equilibria in recommender systems. *Advances in Neural Information Processing Systems*, 36:14597–14608, 2023a.

Meena Jagadeesan, Michael Jordan, Jacob Steinhardt, and Nika Haghtalab. Improved bayes risk can yield reduced social welfare under competition. *Advances in Neural Information Processing Systems*, 36:66940–66952, 2023b.

Meena Jagadeesan, Michael I Jordan, and Nika Haghtalab. Competition, alignment, and equilibria in digital marketplaces. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5689–5696, 2023c.

Meena Jagadeesan, Alexander Wei, Yixin Wang, Michael I Jordan, and Jacob Steinhardt. Learning equilibria in matching markets with bandit feedback. *Journal of the ACM*, 70(3): 1–46, 2023d.

Meena Jagadeesan, Michael I Jordan, and Jacob Steinhardt. Safety vs. performance: How multi-objective learning reduces barriers to market entry. *arXiv preprint arXiv:2409.03734*, 2024.

Ayush Jain, Andrea Montanari, and Eren Sasoglu. Scaling laws for learning with real and surrogate data. *arXiv preprint arXiv:2402.04376*, 2024.

Chi Jin, Praneeth Netrapalli, and Michael Jordan. What is local optimality in nonconvex-nonconcave minimax optimization? In *International conference on machine learning*, pages 4880–4889. PMLR, 2020.

Ramesh Johari, Vijay Kamble, and Yash Kanoria. Matching while learning. *arXiv preprint arXiv:1603.04549*, 2016.

Bruno Jullien and Wilfried Sand-Zantman. The economics of platforms: A theory guide for competition policy. *Information Economics and Policy*, 54:100880, 2021.

Hsu Kao, Chen-Yu Wei, and Vijay Subramanian. Decentralized cooperative reinforcement learning with hierarchical information structure. In *International Conference on Algorithmic Learning Theory*, pages 573–605. PMLR, 2022.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Abbas Kazerouni, Mohammad Ghavamzadeh, Yasin Abbasi Yadkori, and Benjamin Van Roy. Conservative contextual linear bandits. *Advances in Neural Information Processing Systems*, 30, 2017.

Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

Gerard Jounghyun Kim. *Human-computer interaction: fundamentals and practice*. CRC press, 2015.

Jon Kleinberg and Robert Kleinberg. Delegated search approximates efficient search. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 287–302, 2018.

Jon Kleinberg and Manish Raghavan. How do classifiers induce agents to invest effort strategically? *ACM Transactions on Economics and Computation (TEAC)*, 8(4):1–23, 2020.

Jon Kleinberg and Manish Raghavan. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22):e2018340118, 2021.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. The challenge of understanding what users want: Inconsistent preferences and engagement optimization. *Management science*, 70(9):6336–6355, 2024.

Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 17, 2004.

Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605. IEEE, 2003.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 681–690, 2008.

Tomáš Kocák, Gergely Neu, Michal Valko, and Rémi Munos. Efficient learning by implicit exploration in bandit problems with side observations. *Advances in Neural Information Processing Systems*, 27, 2014.

Yoav Kolumbus and Noam Nisan. How and why to manipulate your own agent: On the incentives of users of learning agents. *Advances in Neural Information Processing Systems*, 35:28080–28094, 2022.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pages 1106–1114, 2012.

Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Tight regret bounds for stochastic combinatorial semi-bandits. In *Artificial Intelligence and Statistics*, pages 535–543. PMLR, 2015.

Yongchan Kwon, Antonio Ginart, and James Zou. Competition over data: how does data purchase affect users? *arXiv preprint arXiv:2201.10774*, 2022.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Benjamin Laufer, Jon Kleinberg, and Hoda Heidari. Fine-tuning games: Bargaining and adaptation for general-purpose models. In *Proceedings of the ACM Web Conference 2024*, pages 66–76, 2024.

Niklas Lauffer, Mahsa Ghasemi, Abolfazl Hashemi, Yagiz Savas, and Ufuk Topcu. No-regret learning in dynamic stackelberg games. *IEEE Transactions on Automatic Control*, 69(3): 1418–1431, 2023.

Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

A. P. Lerner. The concept of monopoly and the measurement of monopoly power. *The Review of Economic Studies*, 1(3):157–175, 1934.

Joshua Letchford, Vincent Conitzer, and Kamesh Munagala. Learning and approximating the optimal strategy to commit to. In *International symposium on algorithmic game theory*, pages 250–262. Springer, 2009.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. In *International Conference on Artificial Intelligence and Statistics*, pages 3164–3186. PMLR, 2022.

Xiaocheng Li, Chunlin Sun, and Yinyu Ye. The symmetry between arms and knapsacks: A primal-dual approach for bandits with knapsacks. In *International Conference on Machine Learning*, pages 6483–6492. PMLR, 2021.

Licong Lin, Jingfeng Wu, Sham M Kakade, Peter L Bartlett, and Jason D Lee. Scaling laws in linear regression: Compute, parameters, and data. *arXiv preprint arXiv:2406.08466*, 2024.

Tao Lin and Yiling Chen. Persuading a learning agent. *arXiv preprint arXiv:2402.09721*, 2024.

Lydia T Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. In *International Conference on Machine Learning*, pages 3150–3158. PMLR, 2018.

Lydia T Liu, Horia Mania, and Michael Jordan. Competing bandits in matching markets. In *International Conference on Artificial Intelligence and Statistics*, pages 1618–1628. PMLR, 2020a.

Lydia T Liu, Ashia Wilson, Nika Haghtalab, Adam Tauman Kalai, Christian Borgs, and Jennifer Chayes. The disparate equilibria of algorithmic decision making when individuals invest rationally. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 381–391, 2020b.

Lydia T Liu, Feng Ruan, Horia Mania, and Michael I Jordan. Bandit learning in decentralized matching markets. *Journal of Machine Learning Research*, 22(211):1–34, 2021.

Lydia T Liu, Nikhil Garg, and Christian Borgs. Strategic ranking. In *International Conference on Artificial Intelligence and Statistics*, pages 2489–2518. PMLR, 2022.

Qingmin Liu. Stability and bayesian consistency in two-sided markets. *American Economic Review*, 110(8):2625–2666, 2020.

Qingmin Liu, George J Mailath, Andrew Postlewaite, and Larry Samuelson. Stable matching with incomplete information. *Econometrica*, 82(2):541–587, 2014.

Yang Liu and Chien-Ju Ho. Incentivizing high quality user contributions: New arm generation in bandit learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Brendan Lucier, Sarath Pattathil, Aleksandrs Slivkins, and Mengxiao Zhang. Autobidders with budget and roi constraints: Efficiency, regret, and pacing dynamics. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 3642–3643. PMLR, 2024.

I Scott MacKenzie. Human-computer interaction: An empirical research perspective. 2024.

Chinmay Maheshwari, Chih-Yuan Chiu, Eric Mazumdar, Shankar Sastry, and Lillian Ratliff. Zeroth-order methods for convex-concave min-max problems: Applications to decision-dependent risk minimization. In *International Conference on Artificial Intelligence and Statistics*, pages 6702–6734. PMLR, 2022.

Neil Mallinar, James Simon, Amirhesam Abedsoltan, Parthe Pandit, Misha Belkin, and Preetum Nakkiran. Benign, tempered, or catastrophic: Toward a refined taxonomy of overfitting. *Advances in neural information processing systems*, 35:1182–1195, 2022.

Neil Mallinar, Austin Zane, Spencer Frei, and Bin Yu. Minimum-norm interpolation under covariate shift. *arXiv preprint arXiv:2404.00522*, 2024.

Shie Mannor and Ohad Shamir. From bandits to experts: On the value of side-observations. *Advances in neural information processing systems*, 24, 2011.

Yishay Mansour, Aleksandrs Slivkins, and Vasilis Syrgkanis. Bayesian incentive-compatible bandit exploration. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 565–582, 2015.

Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games. *Operations Research*, 70(2): 1105–1127, 2022.

VA Marchenko and Leonid A Pastur. Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.(NS)*, 72(114):4, 1967.

JE MartÍnez-Legaz. Dual representation of cooperative games based on fenchel-moreau conjugation. *Optimization*, 36(4):291–319, 1996.

Glenn McDonald. On netflix and spotify, algorithms hold the power. but there's a way to get it back., 2019. URL https://expmag.com/2019/11/endless-loops-of-like-the-future-of-algorithmic-entertainment/.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems*, 33:4929–4939, 2020.

Celestine Mendler-Dünner, Gabriele Carovano, and Moritz Hardt. An engine not a camera: Measuring performative power of online search. *arXiv preprint arXiv:2405.19073*, 2024.

Eric Meyerson. Youtube now: Why we focus on watch time, 2012. URL https://blog.youtube/news-and-events/youtube-now-why-we-focus-on-watch-time/.

John P Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *International Conference on Machine Learning*, pages 7710–7720. PMLR, 2021.

Smitha Milli, John Miller, Anca D Dragan, and Moritz Hardt. The social cost of strategic classification. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 230–239, 2019.

Smitha Milli, Luca Belli, and Moritz Hardt. From optimizing engagement to measuring value. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 714–722, 2021.

Smitha Milli, Emma Pierson, and Nikhil Garg. Choosing the right weights: Balancing value, strategy, and noise in recommender systems. *arXiv preprint arXiv:2305.17428*, 2023.

Smitha Milli, Micah Carroll, Yike Wang, Sashrika Pandey, Sebastian Zhao, and Anca D Dragan. Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS nexus*, 4(3):pgaf062, 2025.

Martin Mladenov, Elliot Creager, Omer Ben-Porat, Kevin Swersky, Richard Zemel, and Craig Boutilier. Optimizing long-term social welfare in recommender systems: A constrained matching approach. In *International Conference on Machine Learning*, pages 6987–6998. PMLR, 2020.

Mehryar Mohri, Gary Sivek, and Ananda Theertha Suresh. Agnostic federated learning. In *International conference on machine learning*, pages 4615–4625. PMLR, 2019.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1): 124–143, 1996.

Niklas Muennighoff, Alexander Rush, Boaz Barak, Teven Le Scao, Nouamane Tazi, Aleksandra Piktus, Sampo Pyysalo, Thomas Wolf, and Colin A Raffel. Scaling data-constrained language models. *Advances in Neural Information Processing Systems*, 36:50358–50376, 2023.

Luke Munn. Angry by design: toxic communication and technical architectures. *Humanities and Social Sciences Communications*, 7(1):1–11, 2020.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J Ratliff. Multiplayer performative prediction: Learning in decision-dependent games. *Journal of Machine Learning Research*, 24(202):1–56, 2023.

Sridhar Narayanan and Kirthi Kalyanam. Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science*, 34(3):388–407, 2015.

John F Nash et al. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

Denis Nekipelov, Vasilis Syrgkanis, and Eva Tardos. Econometrics for learning agents. In *Proceedings of the sixteenth acm conference on economics and computation*, pages 1–18, 2015.

Noam Nisan and Gali Noti. An experimental evaluation of regret-based econometrics. In *Proceedings of the 26th International Conference on World Wide Web*, pages 73–81, 2017.

Gali Noti and Vasilis Syrgkanis. Bid prediction in repeated auctions with learning. In *Proceedings of the Web Conference 2021*, pages 3953–3964, 2021.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

Aldo Pacchiano, My Phan, Yasin Abbasi Yadkori, Anup Rao, Julian Zimmert, Tor Lattimore, and Csaba Szepesvari. Model selection in contextual stochastic bandit problems. *Advances in Neural Information Processing Systems*, 33:10328–10337, 2020.

Alexander Pan, Erik Jones, Meena Jagadeesan, and Jacob Steinhardt. Feedback loops with language models drive in-context reward hacking. In *Proceedings of the 41st International Conference on Machine Learning*, pages 39154–39200, 2024.

Geoffrey Parker, Georgios Petropoulos, and Marshall W Van Alstyne. Digital platforms and antitrust. 2019.

Pratik Patil, Jin-Hong Du, and Ryan J Tibshirani. Optimal ridge regularization for out-of-distribution prediction. *arXiv preprint arXiv:2404.01233*, 2024.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.

Jacopo Perego and Sevgi Yuksel. Media competition and social disagreement. *Econometrica*, 90(1):223–265, 2022.

Billy Perrigo. The new AI-powered Bing is threatening users. that's no laughing matter. *Time Magazine*, 2023. URL https://time.com/6256529/bing-openai-chatgpt-danger-alignment/.

Georgios Piliouras and Fang-Yi Yu. Multi-agent performative prediction: From global stability and optimality to chaos. In *Proceedings of the 24th ACM Conference on Economics and Computation*, pages 1047–1074, 2023.

James Pita, Manish Jain, Fernando Ordóñez, Milind Tambe, Sarit Kraus, and Reuma Magori-Cohen. Effective solutions for real-world stackelberg games: When agents must deal with human uncertainties. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 369–376, 2009.

Chara Podimata and Alex Slivkins. Adaptive discretization for adversarial lipschitz bandits. In *Conference on Learning Theory*, pages 3788–3805. PMLR, 2021.

Siddharth Prasad, Martin Mladenov, and Craig Boutilier. Content prompting: Modeling content provider dynamics to improve user welfare in recommender ecosystems. *arXiv preprint arXiv:2309.00940*, 2023.

Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. *Human-computer interaction*. Addison-Wesley Longman Ltd., 1994.

Jens Prüfer and Christoph Schottmüller. Competing with big data. *The Journal of Industrial Economics*, 69(4):967–1008, 2021.

Kun Qian and Sanjay Jain. Digital content creation: An analysis of the impact of recommendation systems. *Management Science*, 70(12):8668–8684, 2024.

Manish Raghavan. Competition and diversity in generative ai. *arXiv preprint arXiv:2412.08610*, 2024.

Aditi Raghunathan, Sang Michael Xie, Fanny Yang, John Duchi, and Percy Liang. Understanding and mitigating the tradeoff between robustness and accuracy. *arXiv preprint arXiv:2002.10716*, 2020.

Steve Rathje, Jay J Van Bavel, and Sander Van Der Linden. Out-group animosity drives engagement on social media. *Proceedings of the national academy of sciences*, 118(26): e2024292118, 2021.

Mitas Ray, Lillian J Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8081–8088, 2022.

Philip J Reny. On the existence of pure and mixed strategy nash equilibria in discontinuous games. *Econometrica*, 67(5):1029–1056, 1999.

Charles S ReVelle and Horst A Eiselt. Location analysis: A synthesis and survey. *European journal of operational research*, 165(1):1–19, 2005.

Esther Rolf, Theodora T Worledge, Benjamin Recht, and Michael Jordan. Representation matters: Assessing the importance of subgroup allocations in training data. In *International Conference on Machine Learning*, pages 9040–9051. PMLR, 2021.

Dinah Rosenberg, Eilon Solan, and Nicolas Vieille. Social learning in one-arm bandit problems. *Econometrica*, 75(6):1591–1611, 2007.

Robert W Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202, 1974.

Daniel Russo and Benjamin Van Roy. Eluder dimension and the sample complexity of optimistic exploration. *Advances in Neural Information Processing Systems*, 26, 2013.

Marc Rysman. The economics of two-sided markets. *Journal of economic perspectives*, 23(3): 125–143, 2009.

Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731*, 2019.

Steven C Salop. Monopolistic competition with outside goods. *The Bell Journal of Economics*, pages 141–156, 1979.

Abishek Sankararaman, Soumya Basu, and Karthik Abinav Sankararaman. Dominate or delete: Decentralized competing bandits in serial dictatorship. In *International Conference on Artificial Intelligence and Statistics*, pages 1252–1260. PMLR, 2021.

Ilya Segal and Michael D Whinston. Antitrust in innovative industries. *American Economic Review*, 97(5):1703–1730, 2007.

Mark Sellke and Aleksandrs Slivkins. The price of incentivizing exploration: A characterization via thompson sampling and sample complexity. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 795–796, 2021.

Lloyd S Shapley and Martin Shubik. Quasi-cores in a monetary economy with nonconvex preferences. *Econometrica: Journal of the Econometric Society*, pages 805–827, 1966.

Lloyd S Shapley and Martin Shubik. The assignment game i: The core. *International Journal of game theory*, 1(1):111–130, 1971.

Utkarsh Sharma and Jared Kaplan. A neural scaling law from the dimension of the data manifold. *arXiv preprint arXiv:2004.10802*, 2020.

Eliot Shekhtman and Sarah Dean. Strategic usage in a multi-learner setting. In *International Conference on Artificial Intelligence and Statistics*, pages 2665–2673. PMLR, 2024.

Judy Hanwen Shen, Inioluwa Deborah Raji, and Irene Y Chen. The data addition dilemma. *arXiv preprint arXiv:2408.04154*, 2024.

Peng Shi. Efficient matchmaking in assignment games with application to online platforms. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 601–602, 2020.

Galit Shmueli and Ali Tafti. Improving" prediction of human behavior using behavior modification. *arXiv preprint arXiv:2008.12138*, 2020.

W Sierpinski. Un théoreme sur les continus. *Tohoku Mathematical Journal, First Series*, 13: 300–303, 1918.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *Found. Trends Mach. Learn.*, 12 (1-2):1–286, 2019.

Ben Smith. How tiktok reads your mind. *The New York Times*, Dec 2021. URL `https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html`.

Yanke Song, Sohom Bhattacharya, and Pragya Sur. Generalization error of min-norm interpolators in transfer learning. *arXiv preprint arXiv:2406.13944*, 2024.

Statista. Leading countries based on facebook audience size as of january 2024, 2024. URL https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/#:~:text=With%20around%202.9%20billion%20monthly,most%20popular%20social%20media%20worldwide.

Stigler Committee. Final report: Stigler committee on digital platforms. https://www.chicagobooth.edu/-/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf, 2019. Accessed: 2023-4-26.

Eleni Straitouri and Manuel Gomez Rodriguez. Designing decision support systems using counterfactual prediction sets. *arXiv preprint arXiv:2306.03928*, 2023.

Eleni Straitouri, Lequn Wang, Nastaran Okati, and Manuel Gomez Rodriguez. Improving expert predictions with conformal prediction. In *International Conference on Machine Learning*, pages 32633–32653. PMLR, 2023.

Jonathan Stray, Ivan Vendrov, Jeremy Nixon, Steven Adler, and Dylan Hadfield-Menell. What are you optimizing for? aligning recommender systems with human values. *arXiv preprint arXiv:2107.10939*, 2021.

Jinyan Su and Sarah Dean. Learning from streaming data when users choose. *arXiv preprint arXiv:2406.01481*, 2024.

Chad Syverson. Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, 33(3):23–43, 2019.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

Boaz Taitler and Omer Ben-Porat. Braess's paradox of generative ai. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14139–14147, 2025a.

Boaz Taitler and Omer Ben-Porat. Selective response strategies for genai. *arXiv preprint arXiv:2502.00729*, 2025b.

Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, 2008.

The White House. Executive order on the safe, secure, and trustworthy development and use of Artificial Intelligence, 2023.

The YouTube Team. Strengthening enforcement against egregious clickbait on youtube, 2024. URL https://blog.google/intl/en-in/products/platforms/strengthening-enforcement-against-egregious-clickbait-on-youtube/.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Andrea Tirinzoni, Matteo Pirotta, Marcello Restelli, and Alessandro Lazaric. An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits. *Advances in Neural Information Processing Systems*, 33:1417–1427, 2020.

Jean Tirole. *The theory of industrial organization*. MIT press, 1988.

Aidan Toner-Rodgers. Artificial intelligence, scientific discovery, and product innovation. *arXiv preprint arXiv:2412.17866*, 2024.

Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

Matteo Turchetta, Felix Berkenkamp, and Andreas Krause. Safe exploration for interactive machine learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Twitter. Twitter's recommendation algorithm. https://github.com/twitter/the-algorithm-ml/tree/main, 2023.

Raluca M Ursu. The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552, 2018.

Jai Vipra and Anton Korinek. Market concentration implications of foundation models. *arXiv preprint arXiv:2311.01550*, 2023.

Jonathan Vogel. Spatial competition with heterogeneous firms. *Journal of Political Economy*, 116(3):423–466, 2008.

Lequn Wang, Thorsten Joachims, and Manuel Gomez Rodriguez. Improving screening processes via calibrated subset selection. In *International Conference on Machine Learning*, pages 22702–22726. PMLR, 2022.

Jacob Ward. *The loop: How technology is creating a world without choices and how to fight back*. Grand Central Publishing, 2022.

Alexander Wei. *Learning and Decision-Making in Complex Environments*. PhD thesis, EECS Department, University of California, Berkeley, May 2024.

Alexander Wei, Wei Hu, and Jacob Steinhardt. More than a toy: Random matrix models predict how real-world neural representations generalize. In *International conference on machine learning*, pages 23549–23588. PMLR, 2022.

Alexander Wei, Nika Haghtalab, and Jacob Steinhardt. Jailbroken: How does llm safety training fail? *Advances in Neural Information Processing Systems*, 36:80079–80110, 2023.

E Glen Weyl and Alexander White. Let the right'one'win: Policy lessons from the new economics of platforms. *Competition Policy International*, 10(2):29–51, 2014.

Thomas G Wollmann. Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, 108(6):1364–1406, 2018.

Killian Wood, Gianluca Bianchin, and Emiliano Dall'Anese. Online projected gradient descent for stochastic optimization with decision-dependent distributions. *IEEE Control Systems Letters*, 6:1646–1651, 2021.

Yifan Wu, András György, and Csaba Szepesvári. Online learning with gaussian payoffs and side observations. *Advances in Neural Information Processing Systems*, 28, 2015.

Yifan Wu, Roshan Shariff, Tor Lattimore, and Csaba Szepesvári. Conservative bandits. In *International Conference on Machine Learning*, pages 1254–1262. PMLR, 2016.

Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy S Liang, Quoc V Le, Tengyu Ma, and Adams Wei Yu. Doremi: Optimizing data mixtures speeds up language model pretraining. *Advances in Neural Information Processing Systems*, 36:69798–69818, 2023.

Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

Yao-Yuan Yang, Cyrus Rashtchian, Hongyang Zhang, Russ R Salakhutdinov, and Kamalika Chaudhuri. A closer look at accuracy vs. robustness. *Advances in neural information processing systems*, 33:8588–8601, 2020.

Fan Yao, Chuanhao Li, Denis Nekipelov, Hongning Wang, and Haifeng Xu. Learning the optimal recommendation from explorative users. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 9457–9465, 2022.

Fan Yao, Chuanhao Li, Denis Nekipelov, Hongning Wang, and Haifeng Xu. How bad is top-k recommendation under competing content creators? In *International Conference on Machine Learning*, pages 39674–39701. PMLR, 2023a.

Fan Yao, Chuanhao Li, Karthik Abinav Sankararaman, Yiming Liao, Yan Zhu, Qifan Wang, Hongning Wang, and Haifeng Xu. Rethinking incentives in recommender systems: are monotone rewards always beneficial? *Advances in Neural Information Processing Systems*, 36:74582–74601, 2023b.

Fan Yao, Chuanhao Li, Denis Nekipelov, Hongning Wang, and Haifeng Xu. Human vs. generative ai in content creation competition: symbiosis or conflict? *arXiv preprint arXiv:2402.15467*, 2024.

Doron Yeverechyahu, Raveesh Mayya, and Gal Oestreicher-Singer. The impact of large language models on open-source innovation: Evidence from github copilot. *arXiv preprint arXiv:2409.08379*, 2024.

Xinyang Yi, Ji Yang, Lichan Hong, Derek Zhiyuan Cheng, Lukasz Heldt, Aditee Kumthekar, Zhe Zhao, Li Wei, and Ed Chi. Sampling-bias-corrected neural modeling for large corpus item recommendations. In *Proceedings of the 13th ACM conference on recommender systems*, pages 269–277, 2019.

YouTube. Continuing our work to improve recommendations on youtube, 2019. URL https://blog.youtube/news-and-events/continuing-our-work-to-improve/.

Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, pages 321–384, 2021.

Geng Zhao, Banghua Zhu, Jiantao Jiao, and Michael Jordan. Online learning in stackelberg games with an omniscient follower. In *International Conference on Machine Learning*, pages 42304–42316. PMLR, 2023.

Eric Zhou and Dokyun Lee. Generative artificial intelligence, human creativity, and art. *PNAS nexus*, 3(3):pgae052, 2024.

Banghua Zhu, Stephen Bates, Zhuoran Yang, Yixin Wang, Jiantao Jiao, and Michael I Jordan. The sample complexity of online contract design. *arXiv preprint arXiv:2211.05732*, 2022.

Banghua Zhu, Sai Praneeth Karimireddy, Jiantao Jiao, and Michael I Jordan. Online learning in a creator economy. *arXiv preprint arXiv:2305.11381*, 2023.

Simon Zhuang and Dylan Hadfield-Menell. Consequences of misaligned ai. *Advances in Neural Information Processing Systems*, 33:15763–15773, 2020.

Tijana Zrnic, Eric Mazumdar, Shankar Sastry, and Michael Jordan. Who leads and who follows in strategic classification? *Advances in Neural Information Processing Systems*, 34: 15257–15269, 2021.

Song Zuo and Pingzhong Tang. Optimal machine strategies to commit to in two-person repeated games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

Appendix A

Appendix for Chapter 3

A.1 Additional Details of Simulations

*Hyperparameters. We introduce a temperature parameter τ within our loss function, defining the loss $\ell(f_{w,b}(x), y)$ to be $|\,\mathrm{sigmoid}((\langle w, x\rangle + b)/\tau) - 1|$. This reparameterizes, but does not change, the model family.

When we run the best-response dynamics, we always initialize the model parameters as mean-zero Gaussians with standard deviation σ. When we reinitialize model parameters, we again initialize them as mean-zero Gaussians with standard deviation σ. For Chapter 3.4.2, we set $I = 5000$, $\tau = 0.1$, $\varepsilon = 0.001$, $\eta = 0.1$, $\sigma = 0.1$, and $\rho = \rho' = 1.0$. For Chapter 3.4.3 and Chapter 3.4.4, we set $I = 2000$, $\sigma = 0.5$, $\tau = 1.0$, $\varepsilon = 0.001$, and η with the following learning rate schedule to expedite convergence: $\eta = 1.0$ if the risk $\mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f_{w_j,b_j}(x), y)]$ is at least 0.5, $\eta = 5.0$ if the risk is in $[0.4, 0.5)$, $\eta = 15$ if the risk is in $[0.3, 0.4)$, and $\eta = 20$ if the risk is less than 0.3. We set $\rho = \rho' = 0.3$ for Chapter 3.4.3 and we set $\rho = 0.7$ and $\rho' = 1$ for Chapter 3.4.4.

For Chapter 3.4.3 and Chapter 3.4.4, we ran over several trials for each data point and the error bars show two standard errors from the mean. For binary classification, the number of trials was 20 for $m = 3$ and $m = 4$ and 8 for $m = 5$, $m = 6$, and $m = 8$. For 10-class classification, the number of trials was 40 for $m = 3$ and $m = 4$ and 8 for $m = 5$, $m = 6$, and $m = 8$.

In addition to computing the equilibria, we also approximate the optimal Bayes risk. For Chapter 3.4.2, we run gradient descent for $10,000$ iterations with learning rate equal to one and parameters initialized to independent Gaussians with zero mean and standard deviation 0.1. For Chapter 3.4.3, we run gradient descent for $50,000$ iterations with learning rate equal to 0.1 and parameters initialized to independent Gaussians with zero mean and standard deviation 0.005. For Chapter 3.4.4, we run gradient descent for $70,000$ iterations with learning rate equal to 0.1 and parameters initialized to independent Gaussians with zero mean and standard deviation 0.005.

*Generation of the synthetic dataset. In Setting 1 (Figures 3.2a, 3.3a, and 3.2d), we consider a zero-dimensional population where $Y \mid X$ is distributed as a Bernoulli with probability α. In Figure 3.2d, the meaning of a zero-dimensional representation is that the only parameter is the bias.

In Setting 2 (Figures 3.2b, 3.3b, and 3.2e), we consider a one-dimensional population given by a mixture of Gaussians. In particular, the Gaussian $X \mid Y = 0$ is distributed as $N(-\mu, \sigma^2)$ and the Gaussian $X \mid Y = 1$ is distributed as $N(\mu, \sigma^2)$. The mean μ is taken to be 1. The distribution of the labels is given by $\mathbb{P}[Y = 1] = 0.4$ and $\mathbb{P}[Y = 1] = 0.6$.

In Setting 3 (Figures 3.2c, 3.3c, and 3.2f), let $D_{\text{base}} = 4$. The distribution over (X^{all}, Y) consists of D_{base} subpopulations. We define the distribution of (X^{all}, Y) as follows: each subpopulation $1 \leq i \leq D_{\text{base}}$ has a different mean vector $\mu_i \in \mathbb{R}^{D_{\text{base}}}$ and is distributed as $X^{\text{all}} \sim Y = 0 \sim N(-\mu_i, \sigma^2)$, let $X^{\text{all}} \sim Y = 1 \sim N(\mu_i, \sigma^2)$, and let $\mathbb{P}[Y = 0] = \mathbb{P}[Y = 1] = 1/2$.We define $(\mu_i)_d = 0$ for $1 \leq d \leq i - 1$ and $(\mu_i)_d = 1$ for $i \leq d \leq D_{\text{base}}$, and we let $\sigma = 1$. If the representation dimension is D, then we define X to consist of the first D coordinates of X^{all}. When $D = 0$, the model-provider is not given representations and thus must assign all users to the same output. (Our setup captures that the dimension D must be at least i to see any nontrivial features about subpopulation i.) The distribution across the 4 subpopulations is 0.7, 0.15, 0.1, and 0.05.

In each case, we draw 10,000 samples and take the resulting empirical distribution to be $\mathcal{D}$.

*Generation of the CIFAR-10 task. We consider a binary classification task consisting of the first 10,000 images in the training set of CIFAR-10. The class 0 is defined to be {airplane, bird, automobile, ship, horse, truck} and class 1 is defined to be {cat, deer, dog, frog}. To generate representations, we use the pretrained models from the Pytorch `torchvision.models` package; these models were pretrained on ImageNet.

*Compute details. We run our simulations on a single A100 GPU.

A.2 Additional Results and Proofs for Chapter 3.3

In Chapter A.2.1, we show a decomposition lemma and prove existence of equilibrium (Proposition 1). We prove the results from Chapter 3.3.2 in Chapter A.2.2, prove the results from Chapter 3.3.4 in Chapter A.2.3, and prove the results from Chapter 3.3.5 in Chapter A.2.4.

A.2.1 Decomposition lemma and existence of equilibrium

We first show that we can decompose model-provider actions into independent decisions about each representation x. To formalize this, let $\mathcal{D}$ be the data distribution, and let $\mathcal{D}_x$ be the conditional distribution over $(X, Y) \mid X = x$ where $(X, Y) \sim \mathcal{D}$. Let $(\mathcal{F}_{\text{all}}^{\text{multi-class}})^x := \{f^0, f^1, \ldots, f^{K-1}\}$ be the class of K functions from a single representation x to $\{0, 1, \ldots, K - 1\}$, where $f^i(x) = i$.

Lemma 131. *Let X be a finite set of representations, let $\mathcal{F} = \mathcal{F}_{all}^{multi\text{-}class}$, and let $\mathcal{D}$ be the distribution over (X, Y). For each $x \in X$, let $\mathcal{D}_x$ be the conditional distribution over $(X, Y) \mid X = x$ where $(X, Y) \sim \mathcal{D}$, and let $(\mathcal{F}_{all}^{multi\text{-}class})^x := \left\{f^0, f^1, \ldots, f^{K-1}\right\}$ be the class of the K functions from a single representation x to $\{0, 1\}$, where $f^i(x) = i$. Suppose that user decisions are noiseless (i.e., $c \to 0$, so user decisions are given by (3.3)). A market outcome $f_1, \ldots, f_m$ is a pure-strategy equilibrium if and only if for every $x \in X$, the market outcome $(f^{f_1(x)}, \ldots, f^{f_m(x)})$ is a pure-strategy equilibrium for $(\mathcal{F}_{all}^{multi\text{-}class})^x$ with data distribution $\mathcal{D}_x$.*

The intuition is that since $\mathcal{F}_{\text{all}}^{\text{multi-class}}$ is all possible functions, model-providers make independent decisions for each data representation.

Proof. Let $\mathcal{D}^R$ be the marginal distribution of X with respect to the distribution $(X, Y) \sim \mathcal{D}$. First, we write model-provider j's utility as:

$$u(f_j; \mathbf{f}_{-j}) = \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}} [\mathbb{P}[j^*(x, y) = j]] = \mathop{\mathbb{E}}_{x'\sim\mathcal{D}^R} \left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{x'}} [\mathbb{P}[j^*(x, y) = j]] \right], \tag{A.1}$$

where $\mathbf{f}_{-j}$ denotes the predictors chosen by the other model-providers. The key intuition for the proof will be that the predictions $[f_1(x''), \ldots, f_m(x'')]$ affect $\mathbb{E}_{(x,y)\sim\mathcal{D}_{x'}} [\mathbb{P}[j^*(x, y) = j]]$ if and only if $x' = x''$.

First we show that if $f_1, \ldots, f_m$ is a pure-strategy equilibrium, then $(f^{f_1(x')}, \ldots, f^{f_m(x')})$ is a pure-strategy equilibrium for $(\mathcal{F}_{\text{all}}^{\text{multi-class}})^{x'}$ with data distribution $\mathcal{D}_{x'}$. Assume for sake of contradiction that $(f^{f_1(x')}, \ldots, f^{f_m(x')})$ is not an equilibrium. Then there exists $j' \in [m]$ such that model-provider j' would achieve higher utility if they switched from $f^{f_{j'}(x')}$ to f^l for some $l \neq f_{j'}(x')$. Let $f'_{j'}$ be the predictor given by $f'_{j'}(x) = f_{j'}(x)$ if $x \neq x'$ and $f'_{j'}(x') = l$. By equation (A.1), this would mean that $u(f'_{j'}; \mathbf{f}_{-j'})$ is strictly higher than $u(f_{j'}; \mathbf{f}_{-j'})$ which is a contradiction.

Next, we show that if $(f^{f_1(x')}, \ldots, f^{f_m(x')})$ is a pure-strategy equilibrium for $(\mathcal{F}_{\text{all}}^{\text{binary}})^{x'}$ with data distribution $\mathcal{D}_{x'}$ for all $x' \in X$ then $f_1, \ldots, f_m$ is a pure-strategy equilibrium. Assume for sake of contradiction that there exists j' such that $u(f'_{j'}; \mathbf{f}_{-j'}) > u(f_j; \mathbf{f}_{-j'})$. By equation (A.1), there must exist x' such that $\mathbb{E}_{(x,y)\sim\mathcal{D}_{x'}} [\mathbb{P}[j^*(x, y) = j']]$ is higher for $f'_{j'}$ than $f_{j'}$. This means that $(f^{f_1(x')}, \ldots, f^{f_m(x')})$ is not an equilibrium (since f^l would be a better response for model-provider j') which is a contradiction. □

We next prove Proposition 1, showing that a pure-strategy equilibrium exists by applying the proof technique of Lemma 3.7 of Ben-Porat and Tennenholtz (2019).

Proof of Proposition 1. By Lemma 131, it suffices to show that there exists a pure-strategy equilibrium whenever there is a single data representation $X = \{x\}$. In this case, the function class $\mathcal{F}_{\text{all}}^{\text{multi-class}}$ consists of K predictors $\left\{f^0, f^1 \ldots, f^{K-1}\right\}$ given by $f^i(x) = i$. For each class i, let $\mathbb{P}[Y = i \mid X] = p_i$.

For the special case of $K = 2$ (binary classification), the game between model-providers is thus a 2-action game with symmetric utility functions. Thus, it must possess a (possibly asymmetric) pure Nash equilibrium (Cheng et al., 2004).

For the general case of $K \geq 2$, we can no longer apply the result in (Cheng et al., 2004) since there can be more than 2 actions. We instead show that the game is a potential game, following a similar argument to Ben-Porat and Tennenholtz (2019). We define the potential function $\Phi(\cdot)$ as follows. For each $i \in \{0, 1, \ldots, K-1\}$, we define the function $G_i : \left\{f^0, f^1 \ldots, f^{K-1}\right\}^m \rightarrow \mathbb{R}_{\geq 0}$ to be:

$$G_i(f_1, \ldots, f_m) := \begin{cases} \frac{1}{m} & \text{if } |\{j \in [m] \mid f_j = f^i\}| = 0 \\ \sum_{l=1}^{|\{j \in [m] | f_j = f^i\}|} \frac{1}{l} & \text{if } |\{j \in [m] \mid f_j = f^i\}| \geq 1. \end{cases}$$

We let

$$\Phi(f_1, \ldots, f_m) := \sum_{i=1}^{K} p_i \cdot G_i(f_1, \ldots, f_m).$$

We show that Φ is a potential function for this game. Suppose that model-provider j switches from $f_j := f^i$ to $f'_j = f^{i'}$ for $i' \neq i$. For each $i \in \{0, 1, \ldots, K-1\}$, let $N_i = |\{j \in [m] \mid f_j = f^i\}|$ be the number of model-providers who choose f^i on the original outcome $[f_1, \ldots, f_m]$. We observe that:

$$u(f_j; \mathbf{f}_{-j}) - u(f'_j; \mathbf{f}_{-j}) = \begin{cases} p_i \cdot \frac{1}{N_i} - p_{i'} \cdot \frac{1}{N_{i'}+1} \text{ if } N_i > 1, N_{i'} > 0 \\ p_i \cdot \left(1 - \frac{1}{m}\right) - p_{i'} \cdot \frac{1}{N_{i'}+1} \text{ if } N_i = 1, N_{i'} > 0 \\ p_i \cdot \frac{1}{N_i} - p_{i'} \cdot \left(1 - \frac{1}{m}\right) \text{ if } N_i > 1, N_{i'} = 0 \\ p_i \cdot \left(1 - \frac{1}{m}\right) - p_{i'} \cdot \left(1 - \frac{1}{m}\right) \text{ if } N_i = 1, N_{i'} = 0 \end{cases}$$

Moreover, we see that:

$$\begin{aligned} &\Phi(f_1, \ldots, f_m) - \Phi(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m) \\ &= \sum_{i''=1}^{K} p_{i''} \cdot G_{i''}(f_1, \ldots, f_m) - \sum_{i''=1}^{K} p_{i''} \cdot G_{i''}(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m) \\ &= p_i \cdot \left(G_i(f_1, \ldots, f_m) - G_i(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m)\right) \\ &+ p_{i'} \left(G_{i'}(f_1, \ldots, f_m) - G_{i'}(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m)\right). \end{aligned}$$

If $N_i > 1$, then:

$$G_i(f_1, \ldots, f_m) - G_i(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m) = \frac{1}{N^i}$$

and if $N_i = 1$, then

$$G_i(f_1, \ldots, f_m) - G_i(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m) = 1 - \frac{1}{m}.$$

Similarly, if $N_{i'} > 0$, then:

$$G_{i'}(f_1, \ldots, f_m) - G_{i'}(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m) = -\frac{1}{N^{i'} + 1}$$

and if $N_{i'} = 0$, then

$$G_{i'}(f_1, \ldots, f_m) - G_i(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m) = -\left(1 - \frac{1}{m}\right).$$

Altogether, this implies that:

$$\Phi(f_1, \ldots, f_m) - \Phi(f_1, f_2, \ldots, f_{j-1}, f'_j, f_{j+1}, \ldots, f_m) = u(f_j; \mathbf{f}_{-j}) - u(f'_j; \mathbf{f}_{-j}),$$

which shows that Φ is a potential function of the game. Since pure strategy equilibria exist in potential games (Rosenthal, 1973; Monderer and Shapley, 1996), a pure strategy equilibrium must exist in the game. □

A.2.2 Proofs for Chapter 3.3.2

We next prove Proposition 2. The high-level intuition of the proof is as follows. By Lemma 131, we can focus on one data representation x at a time. Let $y^* = \arg\max_y \mathbb{P}[y \mid x]$ be the Bayes optimal label of x. The proof boils down to characterizing when the market outcome, $f_j(x) = y^*$ for $j \in [m]$, is an equilibrium, and the equilibrium social loss is determined by whether this market outcome is an equilibrium or not.

Proof of Proposition 2. Let $\mathcal{D}^R$ be the marginal distribution of X with respect to the distribution $(X, Y) \sim \mathcal{D}$. Let $f_1^*, \ldots, f_m^*$ be a pure-strategy equilibrium. The social loss is equal to:

$$\begin{aligned}
\mathtt{SL}(f_1^*, \ldots, f_m^*) &= \mathbb{E}[\ell(f^*_{j^*(x,y)}(x), y)] \\
&= \mathbb{E}_{x' \sim \mathcal{D}^R}\left[\mathbb{E}_{(x,y)\sim\mathcal{D}}[\ell(f^*_{j^*(x,y)}(x), y) \mid x = x']\right] \\
&= \mathbb{E}_{x' \sim \mathcal{D}^R}\left[\mathbb{E}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)]\right],
\end{aligned}$$

where $\mathcal{D}_{x'}$ denotes the conditional distribution $(X, Y) \mid X = x'$ where $(X, Y) \sim \mathcal{D}$. Thus, to analyze the overall social loss, we can separately analyze the social loss on each distribution $\mathcal{D}_{x'}$ and then average across distributions. It suffices to show that $\mathbb{E}_{\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)] = \alpha(x')$ if $\alpha(x') < 1/m$ and zero if $\alpha(x') > 1/m$.

To compute the social loss on $\mathcal{D}_{x'}$, we first apply Lemma 131. This means that $(f_1^*(x'), \ldots, f_m^*(x'))$ is pure-strategy equilibrium with $\mathcal{D}_{x'}$. We characterize the equilibrium structure for $\mathcal{D}_{x'}$ and use this characterization to compute the equilibrium social loss.

*Equilibrium structure for $\mathcal{D}_{x'}$. For notational convenience, let $y_i := f_i^*(x')$ denote the label chosen by model-provider i and let let $y^* = \arg\max_y \mathbb{P}[y \mid x']$ be the Bayes optimal label for x'. We also abuse notation slightly and let $u(y_1; y_{-j})$ be model-provider 1's utility if they choose the label y_1 for x' and the other model-provider's choose y_{-j}.

We first show that all model-providers choosing y^* is an equilibrium if and only if $\alpha(x') \leq 1/m$. Let's fix $y_j = y^*$ for all $j \geq 2$ and look at model-provider 1's utility. We see that $u(y^*; y_{-j}) = 1/m$ and $u(1-y^*; y_{-j}) = \alpha(x')$. This means that y^* is a best-response (i.e., $y^* \in \arg\max_y u(y; y_{-j})$) if and only if $\alpha(x') \leq 1/m$.

We next show that if $\alpha(x') < 1/m$, then the market outcome $y_i = y^*$ for all $i \in [m]$ is the only pure-strategy equilibrium. Let $y_1, \ldots, y_m$ be a pure-strategy equilibrium. It suffices to show that y^* is the unique best response to y_{-j}; that is, that $\{y^*\} = \arg\max_y u(y; y_{-j})$. To show this, let m' denote the size of the set $\{2 \leq i \leq m \mid y_i = y^*\}$. First, if $m' = 0$, then we have that

$$u(y^*; y_{-j}) = 1 - \alpha(x') > 1/m = u(1 - y^*; y_{-j}),$$

where $1 - \alpha(x') > 1/m$ follows from the fact that $1 - \alpha(x') \geq 1/2 \geq 1/m$ along with our assumption that $\alpha(x') \neq 1/m$. This demonstrates that y^* is indeed the unique best response. If $m' = m - 1$, then we have that:

$$u(y^*; y_{-j}) = 1/m > \alpha(x') = u(1 - y^*; y_{-j}),$$

as desired. Finally, if $1 \leq m' \leq m - 2$, then:

$$u(y^*; y_{-j}) = \frac{1 - \alpha(x')}{m' + 1} \geq \frac{1 - \alpha(x')}{m - 1} > \frac{1}{m} > \alpha(x') > \frac{\alpha(x')}{m - m'} = u(1 - y^*; y_{-j}),$$

as desired.

Finally, we show that all model-providers choosing $1 - y^*$ is never an equilibrium. Let's fix $y_j = 1 - y^*$ and look at model-provider 1's utility. We see that:

$$u(y^*; y_{-j}) = 1 - \alpha(x') > \frac{\alpha(x')}{m} = u(1 - y^*; y_{-j}),$$

which shows that y^* is the unique best response as desired.

*Characterization of equilibrium social loss. It follows from (3.5) that the equilibrium social loss $\mathbb{E}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)]$ is $\alpha(x')$ if all of the model-providers choose $y_i = y^*$, it is zero if a nonzero number of model-providers choose y^* and a nonzero number of model-providers choose $1 - y^*$, and it is $1 - \alpha(x')$ if all of the model-providers choose $1 - y^*$.

Let's combine this with our equilibrium characterization results. If $\alpha(x') < 1/m$, then the unique equilibrium is at $y_i = y^*$ so the equilibrium social loss is $\alpha(x)$ as desired. If $\alpha(x') > 1/m$, then neither $y_i = y^*$ for all $i \in [m]$ nor $y_i = 1 - y^*$ for all $i \in [m]$ is an equilibrium. Since there exists a pure strategy equilibrium by Proposition 1, there must be a pure strategy equilibrium where a nonzero number of model-providers choose y^* and a nonzero number of model-providers choose $1 - y^*$. The equilibrium social loss is thus zero.

Note when $\alpha(x') = 1 - 1/m$, there is actually an equilibrium where all of the model-providers choose $y_i = y^*$, 0 and an equilibrium where a nonzero number of model-providers choose y^* and a nonzero number of model-providers choose $1 - y^*$; thus, the equilibrium social loss can be zero or $1/m$.

$\square$

A.2.3 Proofs for Chapter 3.3.4

We prove Proposition 3.

Proof of Proposition 3. Let $\mathcal{D}^R$ be the marginal distribution of X with respect to the distribution $(X, Y) \sim \mathcal{D}$. Let $f_1^*, \ldots, f_m^*$ be a pure-strategy equilibrium. The social loss is equal to:

$$\begin{aligned}\texttt{SL}(f_1, \ldots, f_m) &= \mathbb{E}[\ell(f^*_{j^*(x,y)}(x), y)] \\ &= \mathop{\mathbb{E}}_{x' \sim \mathcal{D}^R}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(f^*_{j^*(x,y)}(x), y) \mid x = x']\right] \\ &= \mathop{\mathbb{E}}_{x' \sim \mathcal{D}^R}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)]\right],\end{aligned}$$

where $\mathcal{D}_{x'}$ denotes the conditional distribution $(X, Y) \mid X = x'$ where $(X, Y) \sim \mathcal{D}$. Thus, to analyze the overall social loss, we can separately analyze the social loss on each distribution $\mathcal{D}_{x'}$ and then average across distributions. It suffices to show that

$$\mathop{\mathbb{E}}_{\mathcal{D}_{x'}}\left[\sum_{i=1}^{K} \alpha^i(x) \cdot \mathbb{1}\left[\alpha^i(x) < \frac{c}{m}\right]\right] \le \mathop{\mathbb{E}}_{\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)] \le \mathop{\mathbb{E}}_{\mathcal{D}_{x'}}\left[\sum_{i=1}^{K} \alpha^i(x) \cdot \mathbb{1}\left[\alpha^i(x) \le \frac{1}{m}\right]\right].$$

To compute the social loss on $\mathcal{D}_{x'}$, we first apply Lemma 131. This means that $(f_1^*(x'), \ldots, f_m^*(x'))$ is pure-strategy equilibrium with $\mathcal{D}_{x'}$. We then prove properties of the equilibrium structure for $\mathcal{D}_{x'}$ and use these properties to bound the equilibrium social loss. For notational convenience, let $y_i := f_i^*(x')$ denote the label chosen by model-provider i and let let $y^* = \arg\max_y \mathbb{P}[y \mid x']$ be the Bayes optimal label for x'. We also abuse notation slightly and let $u(y_1; y_{-j})$ be model-provider 1's utility if they choose the label y_1 for x' and the other model-provider's choose y_{-j}. We can rewrite:

$$\mathop{\mathbb{E}}_{\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)] = \mathop{\mathbb{E}}_{\mathcal{D}_{x'}}\left[\sum_{i=1}^{K} \alpha^i(x) \cdot \mathbb{1}\left[y_j \ne i \text{ for all } j \in [m]\right]\right].$$

We first prove the lower bound on $\mathbb{E}_{\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)]$ and then we prove the upper bound on $\mathbb{E}_{\mathcal{D}_{x'}}[\ell(f^*_{j^*(x,y)}(x), y)]$.

*Proof of lower bound. Let $y_1, \ldots, y_m$ be a pure strategy equilibrium. To prove the lower bound, it suffices to show that if $\alpha^i(x) < c/m$, then $y_j \ne i$ for all $j \in [m]$.

Assume for sake of contradiction that $\alpha^i(x) < c/m$ and $y_j = i$ for some $j \in [m]$. Let $i' = \text{argmax}_{i'' \in \{0,1,\ldots,K-1\}} \alpha^{i''}(x)$ be the class with maximal conditional probability. By the definition of c, we see that $\alpha^{i'}(x) \ge c > c/m$ which also implies that $i' \ne i$. We split into two cases—(1) $y_{j'} \ne i'$ for all $j' \in \{0, 1, \ldots, K-1\}$, and (2) $y_{j'} = i'$ for some $j' \in \{0, 1, \ldots, K-1\}$—and derive a contradiction in each case.

Consider the first case where $y_{j'} \neq i'$ for all $j' \in \{0, 1, \ldots, K-1\}$. Then if model-provider j switched from y_j to i', the difference in their utility would be bounded as:

$$\begin{aligned} u(i'; y_{-j}) - u(y_j; y_{-j}) &\geq \alpha^{i'}(x) - \left(\frac{\alpha^{i'}(x)}{m} + \alpha^i(x)\right) \\ &= \alpha^{i'}(x)\left(1 - \frac{1}{m}\right) - \alpha^i(x) \\ &> c\left(1 - \frac{1}{m}\right) - \frac{c}{m} \\ &= c\left(1 - \frac{2}{m}\right) \\ &\geq 0, \end{aligned}$$

so y_j is not a best-response for model-provider j, which is a contradiction.

Now, consider the second case, where $y_{j'} = i'$ for some $j' \in \{0, 1, \ldots, K-1\}$. If we compare the utility when model-provider j chooses i' versus y_j as their action, the difference is utility can be bounded as:

$$u(i'; y_{-j}) - u(y_j; y_{-j}) \geq \frac{\alpha^{i'}(x)}{m} - \alpha^i(x) > \frac{c}{m} - \frac{c}{m} = 0.$$

so y_j is not a best-response for model-provider j, which is a contradiction.

This proves the lower bound as desired.

*Proof of upper bound. Let $y_1, \ldots, y_m$ be a pure strategy equilibrium. To prove the upper bound, it suffices to show if $\alpha^i(x) > 1/m$, then $y_j = i$ for some $j \in [m]$. Assume for sake of contradiction that $\alpha^i(x) > 1/m$ and $y_j \neq i$ for all $j \in [m]$. For any set of actions $y_1, \ldots, y_m$, the total utility $\sum_{j=1}^m u(y_j; y_{-j}) = 1$ sums to 1. Thus, some model provider $j \in [m]$ must have utility satisfying $u(y_j; y_{-j}) \leq 1/m$. However, if model-provider j instead chose action i, then they would achieve utility:

$$u(i; y_{-j}) \geq \alpha^i(x) > \frac{1}{m} \geq u(y_j; y_{-j}),$$

so y_j is not a best-response for model-provider j, which is a contradiction. This proves the upper bound as desired.

□

A.2.4 Proofs for Chapter 3.3.5

A useful lemma is the following calculation of the game matrix when there is a single representation $X = \{x\}$.

	$y_2 = 1 - y^*$	$y_2 = y^*$
$y_1 = 1 - y^*$	$(w_{\max}, w_{\min})$	$(\alpha(x), 1 - \alpha(x))$
$y_1 = y^*$	$(1 - \alpha(x), \alpha(x))$	$(w_{\max}, w_{\min})$

Table A.1: Let $X = \{x\}$, $\mathcal{F} = \mathcal{F}_{\text{all}}^{\text{binary}}$, user decisions are noiseless, and user decisions are noiseless (i.e., $c \to 0$, so user decisions are given by (3.8)). Suppose that there are $m = 2$ model-providers with market reputations $w_{\min}$ and $w_{\max}$, where $w_{\max} \geq w_{\min}$ and $w_{\max} + w_{\min} = 1$. Let $y^* = \arg\max_y \mathbb{P}[y \mid x]$ be the Bayes optimal label for x'. The table shows the game matrix when model-provider 1 chooses the label y_1 and model provider 2 chooses the label y_2.

Lemma 132. *Let $X = \{x\}$, and let $\mathcal{F} = \mathcal{F}_{all}^{binary}$. Suppose that there are $m = 2$ model-providers with market reputations w_{min} and w_{max}, where $w_{max} \geq w_{min}$ and $w_{max} + w_{min} = 1$. Suppose that user decisions are noiseless (i.e., $c \to 0$, so user decisions are given by (3.8)). Then the game matrix is specified by Table A.1.*

Proof. This follows from applying (3.8) and using the fact that $\ell(y, y') = \mathbb{1}[y \neq y']$. □

We show that pure strategy equilibria are no longer guaranteed to exist when model-providers have unequal market reputations, even when there is a single representation $X = \{x\}$.

Lemma 133. *Let $X = \{x\}$ let $\mathcal{F} = \mathcal{F}_{all}^{binary}$. Suppose that there are $m = 2$ model-providers with market reputations w_{min} and w_{max}, where $w_{max} \geq w_{min}$ and $w_{max} + w_{min} = 1$. Suppose that user decisions are noiseless (i.e., $c \to 0$, so user decisions are given by (3.8)). If $\alpha(x) > w_{min}$, then a pure strategy equilibrium does not exist.*

Proof. For notational convenience, let $y_i := f_i(x')$ denote the label chosen by model-provider i and let $y^* = \arg\max_y \mathbb{P}[y \mid x']$ be the Bayes optimal label for x'. We also abuse notation slightly and let $u_i(y; y')$ be model-provider i's utility if they choose the label y for x and the other model-providers choose y'. The proof follows from the game matrix show in Table A.1 (Lemma 132). Using the fact that model-provider 1 must best-respond to model-provider 2's action, this leaves $y_1 = 1 - y^*$, $y_2 = 1 - y^*$ and $y_1 = y^*$, $y_2 = y^*$. However, neither of these market outcomes captures a best-response for model-provider 2: if $y_1 = 1 - y^*$, then model-provider 2's unique best response is y^*; if $y_1 = y^*$, then model-provider 2's unique best response is $1 - y^*$. This rules out the existence of a symmetric or asymmetric pure strategy equilibrium. □

Given the lack of existence of pure strategy equilibria, we must turn to mixed strategies. A mixed strategy equilibrium is guaranteed to exist since the game has finitely many actions $\mathcal{F}_{\text{all}}^{\text{binary}}$ and finitely many players m. Let $(\mu_1, \mu_2, \ldots, \mu_m)$ denote a mixed strategy profile over

$\mathcal{F}_{\text{all}}^{\text{binary}}$. We show the following analogue of Lemma 131 that allows us to again decompose model-provider actions into independent decisions about each representation x. To formalize this, let $\mathcal{D}$ be the data distribution, and again let $\mathcal{D}_x$ be the conditional distribution of (X, Y) when $X = x$, where $(X, Y) \sim \mathcal{D}$. Again, let $(\mathcal{F}_{\text{all}}^{\text{binary}})^x := \{f_0, f_1\}$ be the class of the (two) functions from a single representation x to $\{0, 1\}$, where $f_0(x) = 0$ and $f_1(x) = 1$. Given a mixed strategy profile μ and a representation x, we define the conditional mixed strategy μ^x over $(\mathcal{F}_{\text{all}}^{\text{binary}})^x := \{f_0, f_1\}$ to be defined so $\mathbb{P}_{\mu^x}[f_i] := \mathbb{P}_{f\sim\mu}[f(x) = i]$ for $i \in \{0, 1\}$.

Lemma 134. *Let X be a finite set of representations, let $\mathcal{F} = (\mathcal{F}_{all}^{binary})$, and let $\mathcal{D}$ be the distribution over (X, Y). For each $x \in X$, let $\mathcal{D}_x$ be the conditional distribution of (X, Y) given $X = x$, where $(X, Y) \sim \mathcal{D}$, and let $(\mathcal{F}_{all}^{binary})^x := \{f_0, f_1\}$ be the class of the (two) functions from a single representation x to $\{0, 1\}$, where $f_0(x) = 0$ and $f_1(x) = 1$. Suppose that user decisions are noiseless (i.e., $c \to 0$, so user decisions are given by (3.3)). A strategy profile $(\mu_1, \mu_2, \ldots, \mu_m)$ is an equilibrium if and only if for every $x \in X$, the market outcome $(\mu_1^x, \mu_2^x, \ldots, \mu_m^x)$ (where μ_1^x, ..., μ_m^x are the conditional mixed strategies defined above) is an equilibrium for $(\mathcal{F}_{all}^{binary})^x$ with data distribution $\mathcal{D}_x$.*

Proof. The proof follows similarly to the proof of Lemma 134, but some minor generalizations to account for mixed strategy equilibria. Let $\mathcal{D}^R$ be the marginal distribution of X with respect to the distribution $(X, Y) \sim \mathcal{D}$. Let $\mathcal{D}^R$ be the marginal distribution of X with respect to the distribution $(X, Y) \sim \mathcal{D}$. First, we write model-provider j's utility $\mathbb{E}_{\substack{f_j\sim\mu_j \\ \mathbf{f}_{-j}\sim\mu_{-j}}}[u(f_j; \mathbf{f}_{-j})]$ as:

$$\mathbb{E}_{\substack{f_j\sim\mu_j \\ \mathbf{f}_{-j}\sim\mu_{-j}}}\left[\mathbb{E}_{(x,y)\sim\mathcal{D}}[\mathbb{P}[j^*(x, y) = j]]\right] = \mathbb{E}_{x'\sim\mathcal{D}^R}\left[\mathbb{E}_{\substack{f_j\sim\mu_j^{x'} \\ \mathbf{f}_{-j}\sim\mu_{-j}^{x'}}}\left[\mathbb{E}_{(x,y)\sim\mathcal{D}_{x'}}[\mathbb{P}[j^*(x, y) = j]]\right]\right]. \tag{A.2}$$

where μ_{-j} denotes the mixed strategies chosen by the other model-providers.

First we show that if $\mu_1, \mu_2, \ldots, \mu_m$ is an equilibrium, then $(\mu_1^{x'}, \ldots, \mu_m^{x'})$ is an equilibrium for $(\mathcal{F}_{\text{all}}^{\text{binary}})^{x'}$ with data distribution $\mathcal{D}_{x'}$. Let f_j be in $\text{supp}(\mu_{j'})$. Assume for sake of contradiction that $(\mu_1^{x'}, \ldots, \mu_m^{x'})$ is not an equilibrium. Then there exists $j' \in [m]$ such that model-provider j' would achieve higher utility on $f^{1-f_{j'}(x')}$ than $f^{f_{j'}(x')}$. Let $f'_{j'}$ be the predictor given by $f'_{j'}(x) = f_{j'}(x)$ if $x \neq x'$ and $f'_{j'}(x') = 1 - f_{j'}(x')$. By equation (A.2), this would mean that $u(f'_{j'}; \mu_{-j'})$ is strictly higher than $u(f_{j'}; \mu_{-j'})$ which is a contradiction.

Next, we show that if $(\mu_1^{x'}, \ldots, \mu_m^{x'})$ is an equilibrium for $(\mathcal{F}_{\text{all}}^{\text{binary}})^{x'}$ with data distribution $\mathcal{D}_{x'}$ for all $x' \in X$ then $\mu_1, \ldots, \mu_m$ is an equilibrium. Let f_j be in $\text{supp}(\mu_{j'})$. Assume for sake of contradiction that there exists j' such that $u(f'_{j'}; \mu_{-j'}) > u(f_j; \mu_{-j'})$. By equation (A.1), there must exist x' such that

$$\mathbb{E}_{\mathbf{f}_{-j'}\sim\mu_{-j'}^{x'}}\left[\mathbb{E}_{(x,y)\sim\mathcal{D}_{x'}}[\mathbb{P}[j^*(x, y) = j']]\right]$$

is higher for $f'_{j'}$ than $f_{j'}$. This means that $(\mu_1^{x'}, \ldots, \mu_m^{x'})$ is not an equilibrium, which is a contradiction. $\square$

We now prove Proposition 4.

Proof of Proposition 4. Let $\mathcal{D}^R$ be the marginal distribution of x with respect to the distribution $(x, y) \sim \mathcal{D}$. Let μ_1, μ_2 be a mixed strategy equilibrium. The social loss is equal to:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\substack{f_1 \sim \mu_1 \\ f_2 \sim \mu_2}}[\texttt{SL}(f_1, f_2)] &= \mathbb{E}[\ell(f_{j^*(x,y)}(x), y)] \\
&= \mathop{\mathbb{E}}_{\substack{f_1 \sim \mu_1 \\ f_2 \sim \mu_2}}\left[\mathop{\mathbb{E}}_{x' \sim \mathcal{D}^R}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}}[\ell(f_{j^*(x,y)}(x), y) \mid x = x']\right]\right] \\
&= \mathop{\mathbb{E}}_{\substack{f_1 \sim \mu_1 \\ f_2 \sim \mu_2}}\left[\mathop{\mathbb{E}}_{x' \sim \mathcal{D}^R}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f_{j^*(x,y)}(x), y)]\right]\right] \\
&= \mathop{\mathbb{E}}_{x' \sim \mathcal{D}^R}\left[\mathop{\mathbb{E}}_{\substack{f_1 \sim \mu_1^* \\ f_2 \sim \mu_2^*}}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f_{j^*(x,y)}(x), y)]\right]\right] \\
&= \mathop{\mathbb{E}}_{x' \sim \mathcal{D}_X}\left[\mathop{\mathbb{E}}_{\substack{f_1 \sim \mu_1^{x'} \\ f_2 \sim \mu_2^{x'}}}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f_{j^*(x,y)}(x), y)]\right]\right]
\end{aligned}$$

where $\mathcal{D}_{x'}$ denotes the conditional distribution $(X, Y) \mid X = x'$ where $(X, Y) \sim \mathcal{D}$ and where μ^x denotes the conditional mixed strategy $(\mathcal{F}_{\text{all}}^{\text{binary}})^x := \{f^0, f^1\}$ to be defined so $\mathbb{P}_{\mu^x}[f^i] := \mathbb{P}_{f\sim\mu}[f(x) = i]$ for $i \in \{0, 1\}$ Thus, to analyze the overall social loss, we can separately analyze the social loss on each distribution $\mathcal{D}_{x'}$ and then average across distributions. It suffices to show that:

$$\mathop{\mathbb{E}}_{\substack{f_1 \sim \mu_1^{x'} \\ f_2 \sim \mu_2^{x'}}}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f_{j^*(x,y)}(x), y)]\right] = \begin{cases} \alpha(x') & \text{if } \alpha(x') < w_{\min} \\ \frac{2(\alpha(x') - w_{\min})\cdot(w_{\max} - \alpha(x))}{(1 - 2\cdot w_{\min})^2} & \text{if } \alpha(x') > w_{\min}. \end{cases}$$

To compute the social loss on $\mathcal{D}_{x'}$, we first apply Lemma 134. This means that $(\mu_1^{x'}, \mu_2^{x'})$ is mixed-strategy equilibrium with $\mathcal{D}_{x'}$. We characterize the equilibrium structure for $\mathcal{D}_{x'}$ and use this characterization to compute the equilibrium social loss.

Our main technical ingredient is the game matrix in Table A.1 (Lemma 132). We will slightly abuse notation and view choosing the label y as the strategy of the model-provider. Accordingly, we view a mixed strategy as a distribution over $\{0, 1\}$. For notational convenience, let $y_i := f_i(x')$ denote the label chosen by model-provider i and let $y^* = \arg\max_y \mathbb{P}[y \mid x']$ be the Bayes optimal label for x'. We split into two cases: $\alpha(x') < w_{\min}$ and $\alpha(x') > w_{\min}$.

*Case 1: $\alpha(x') < w_{\min}$. We claim that the unique equilibrium is a pure strategy equilibrium where $y_1 = y_2 = y^*$. First, if $\alpha(x) < w_{\min}$, we show that choosing y^* is a strictly dominant strategy for model-provider 1. This follows from the fact that $1 - \alpha(x) > w_{\max}$ and $w_{\max} \geq w_{\min} > \alpha(x)$. Thus, model-provider 1 must play a pure strategy where they always choose $y_1 = y^*$. When model-provider 1 chooses y^*, then the unique best response for model-provider 2 is also to choose y^* since $\alpha(x') < w_{\min}$. This establishes that $y_1 = y_2 = y^*$ is the unique equilibrium. This also implies that the equilibrium social loss satisfies:

$$\mathop{\mathbb{E}}_{\substack{f_1 \sim \mu_1^{x'} \\ f_2 \sim \mu_2^{x'}}} \left[\mathop{\mathbb{E}}_{(x,y) \sim \mathcal{D}_{x'}} [\ell(f_{j^*(x,y)}(x), y)] \right] = \alpha(x')$$

as desired.

*Case 2: $\alpha(x') > w_{\min}$. Let $p_1 = \mathbb{P}_{\mu_1^{x'}}[y_1 = y^*]$ and let $p_2 = \mathbb{P}_{\mu_2^{x'}}[y_2 = y^*]$. By Lemma 133, a pure strategy equilibrium does not exist. Thus, we consider mixed strategies. Since pure strategy equilibria do not exist, at least one of p_1 and p_2 must be strictly between zero and one. We compute p_1 and p_2, splitting into two cases: (1) $p_1 > 0$ and (2) $p_2 > 0$.

If $p_1 > 0$, then we know that model-provider 1 must be indifferent between choosing y^* and $1 - y^*$. This means that:

$$p_2 \alpha(x') + (1 - p_2) w_{\max} = (1 - p_2)(1 - \alpha(x')) + p_2 w_{\max}.$$

Solving for p_2, we obtain:

$$p_2 = \frac{w_{\max} - (1 - \alpha(x'))}{2 w_{\max} - 1} = \frac{\alpha(x') - w_{\min}}{1 - 2 w_{\min}} > 0.$$

If $p_2 > 0$, then we know that model-provider 2 must be indifferent between choosing y^* and $1 - y^*$. This means that:

$$p_1 \alpha(x') + (1 - p_1) w_{\min} = (1 - p_1)(1 - \alpha(x')) + p_1 w_{\min}.$$

Solving for p_1, we obtain:

$$p_1 = \frac{(1 - \alpha(x')) - w_{\min}}{1 - 2 w_{\min}} = \frac{w_{\max} - \alpha(x)}{1 - 2 w_{\min}} > 0.$$

Putting this all together, we see that:

$$p_1 = \frac{w_{\max} - \alpha(x')}{1 - 2 w_{\min}}$$
$$p_2 = \frac{\alpha(x') - w_{\min}}{1 - 2 w_{\min}},$$

and in fact $p_1 + p_2 = 1$.

Using this characterization of p_1 and p_2, we see that the equilibrium social loss is equal to:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{f_1\sim\mu_1^{x'}\\ f_2\sim\mu_2^{x'}}}\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{x'}}[\ell(f_{j^*(x,y)}(x),y)]\right]\\
&=\alpha(x')\mathbb{P}[y_1=y^*]\mathbb{P}[y_2=y^*]+(1-\alpha(x'))\mathbb{P}[y_1=1-y^*]\mathbb{P}[y_2=1-y^*]\\
&=\alpha(x')p_1p_2+(1-\alpha(x))(1-p_1)(1-p_2)\\
&=\alpha(x')p_1p_2+(1-\alpha(x))p_1p_2\\
&=p_1p_2\\
&=\frac{(\alpha(x')-w_{\min})\cdot(w_{\max}-\alpha(x))}{(1-2\cdot w_{\min})^2},
\end{aligned}$$

as desired.

$\square$

Appendix B

Appendix for Chapter 4

B.1 Proofs for Chapter 4.3

In this section, we prove Theorem 5. First, we state relevant facts (Appendix B.1.1) and prove intermediate lemmas (Appendix B.1.2), and then we use these ingredients to prove Theorem 5 (Appendix B.1.3). Throughout this section, we let

$$L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)\Sigma(\beta_1 - \beta_2)^T].$$

Moreover, let

$$\beta(\alpha, \lambda) = \arg\min_{\beta} \left(\alpha \cdot \mathbb{E}_{X \sim \mathcal{D}_F}[(\langle \beta - \beta_1, X \rangle)^2] + (1-\alpha) \cdot \mathbb{E}_{X \sim \mathcal{D}_F}[(\langle \beta - \beta_2, X \rangle)^2] + \lambda \|\beta\|_2^2\right)$$

be the infinite-data ridge regression predictor.

B.1.1 Facts

We can explicitly solve for the infinite-data ridge regression predictor

$$\begin{aligned}\beta(\alpha, \lambda) &= \arg\min_{\beta} \left(\alpha \cdot \mathbb{E}_{x \sim \mathcal{D}_F}[\langle \beta - \beta_1, x \rangle^2] + (1-\alpha) \cdot \mathbb{E}_{x \sim \mathcal{D}_F}[\langle \beta - \beta_2, x \rangle^2] + \lambda ||\beta||_2^2\right) \\ &= \Sigma(\Sigma + \lambda I)^{-1}(\alpha\beta_1 + (1-\alpha)\beta_2).\end{aligned}$$

A simple calculation shows that $\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, 0))] = (1-\alpha)^2 L^*(\rho)$ and $\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, 0))] = \alpha^2 L^*(\rho)$. Thus, it holds that:

$$\alpha \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, 0))] + (1-\alpha)\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, 0))] = \alpha(1-\alpha)L^*(\rho).$$

Moreover, by the definition of the ridge regression objective, we see that:

$$\alpha \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, \lambda))] + (1-\alpha)\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, \lambda))] \geq \alpha \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, 0))] + (1-\alpha)\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, 0))].$$

B.1.2 Lemmas

The first lemma upper bounds the performance loss when there is regularization.

Lemma 135. *Suppose that power-law scaling holds for the eigenvalues and alignment coefficients with scaling exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$ and suppose that $P = \infty$. Let $L^*(\rho) = (\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)^T$. Let*

$$\beta(\alpha, \lambda) = \arg\min_{\beta} \left(\alpha \cdot \mathbb{E}_{X \sim \mathcal{D}_F}[(\langle \beta - \beta_1, X \rangle)^2] + (1 - \alpha) \cdot \mathbb{E}_{X \sim \mathcal{D}_F}[(\langle \beta - \beta_2, X \rangle)^2] + \lambda \|\beta\|_2^2\right)$$

be the infinite-data ridge regression predictor. Assume that $\alpha \geq 1/2$. Then it holds that

$$\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, \lambda))] \geq (1 - \alpha)^2 L^*(\rho)$$

and

$$\frac{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, \lambda))]}{\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, \lambda))]} \geq \frac{(1 - \alpha)^2}{\alpha^2}.$$

Proof. We define the quantities:

$$A := \lambda^2 \sum_{i=1}^{P} \frac{\lambda_i}{(\lambda_i + \lambda)^2} i^{-\delta}$$

$$B := (1 - \alpha)^2 (1 - \rho)^2 \sum_{i=1}^{P} \frac{\lambda_i^3}{(\lambda_i + \lambda)^2} i^{-\delta}$$

$$C := \lambda (1 - \rho) \sum_{i=1}^{P} \frac{\lambda_i^2}{(\lambda_i + \lambda)^2} i^{-\delta}.$$

We compute the performance loss as follows:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, \lambda))] \\
&= \mathbb{E}_{\mathcal{D}_W}[\mathrm{Tr}(\Sigma(\beta_1 - \beta(\alpha, \lambda))(\beta_1 - \beta(\alpha, \lambda))^T)] \\
&= \mathbb{E}_{\mathcal{D}_W}\left[\mathrm{Tr}\left((\Sigma + \lambda I)^{-2} \Sigma \left(\lambda \beta_1 + \Sigma \cdot (1 - \alpha)(\beta_1 - \beta_2)\right) \left(\lambda \beta_1 + \Sigma \cdot (1 - \alpha)(\beta_1 - \beta_2)\right)^T\right)\right] \\
&= \mathbb{E}_{\mathcal{D}_W}\left[\mathrm{Tr}\left((\Sigma + \lambda I)^{-2} \Sigma \cdot \left(\lambda \beta_1 + \Sigma \cdot (1 - \alpha)(\beta_1 - \beta_2)\right) \left(\lambda \beta_1 + \Sigma \cdot (1 - \alpha)(\beta_1 - \beta_2)\right)^T\right)\right] \\
&= \lambda^2 \mathbb{E}_{\mathcal{D}_W}\left[\mathrm{Tr}\left((\Sigma + \lambda I)^{-2} \Sigma \cdot \beta_1 \beta_1^T\right)\right] + (1 - \alpha)^2 \mathbb{E}_{\mathcal{D}_W}\left[\mathrm{Tr}\left((\Sigma + \lambda I)^{-2} \Sigma^3 \cdot (\beta_1 - \beta_2)(\beta_1 - \beta_2)^T\right)\right] \\
&\quad + \lambda (1 - \alpha) \mathbb{E}_{\mathcal{D}_W}\left[\mathrm{Tr}\left((\Sigma + \lambda I)^{-2} \Sigma^2 \cdot \beta_1 (\beta_1 - \beta_2)^T\right)\right] \\
&= \lambda^2 \sum_{i=1}^{P} \frac{\lambda_i}{(\lambda_i + \lambda)^2} \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle^2] + (1 - \alpha)^2 \sum_{i=1}^{P} \frac{\lambda_i^3}{(\lambda_i + \lambda)^2} \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1 - \beta_2, v_i \rangle^2] \\
&\quad + \lambda (1 - \alpha) \sum_{i=1}^{P} \frac{\lambda_i^2}{(\lambda_i + \lambda)^2} \mathbb{E}_{\mathcal{D}_W}\left[\langle \beta_1, v_i \rangle \langle \beta_1 - \beta_2, v_i \rangle\right] \\
&= \lambda^2 \sum_{i=1}^{P} \frac{\lambda_i}{(\lambda_i + \lambda)^2} i^{-\delta} + (1 - \alpha)^2 (1 - \rho)^2 \sum_{i=1}^{P} \frac{\lambda_i^3}{(\lambda_i + \lambda)^2} i^{-\delta} + \lambda (1 - \alpha)(1 - \rho) \sum_{i=1}^{P} \frac{\lambda_i^2}{(\lambda_i + \lambda)^2} i^{-\delta} \\
&= A + (1 - \alpha)^2 B + (1 - \alpha) C.
\end{aligned}$$

An analogous calculation shows that the safety violation can be written as:

$$\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha,\lambda))] = A + \alpha^2 B + \alpha C$$

Since $\alpha \geq 1/2$, then it holds that:

$$\frac{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha,\lambda))]}{\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha,\lambda))]} = \frac{A + (1-\alpha)^2 B + (1-\alpha) C}{A + \alpha B + \alpha C} \geq \frac{(1-\alpha)^2}{\alpha^2}.$$

Combining this with the facts from Appendix B.1.1—which imply that $\alpha \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha,\lambda))] + (1-\alpha)\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha,\lambda))] \geq \alpha(1-\alpha)L^*(\rho)$—we have that $\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha,\lambda))] \geq (1-\alpha)^2 L^*(\rho)$ as desired. □

The following lemma computes the optimal values of α and λ for the incumbent.

Lemma 136. *Suppose that power-law scaling holds for the eigenvalues and alignment coefficients with scaling exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$ and suppose that $P = \infty$. Let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)^T]$. Suppose that $N_I = \infty$, and suppose that the safety constraint τ_I satisfies (4.1). Then it holds that $\alpha_I = \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}}$, and $\lambda_I = 0$ is optimal for the incumbent. Moreover, it holds that:*

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_I, \infty, \alpha_O)] = (\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))})^2.$$

Proof. First, we apply Lemma 150 with $N = \infty$ to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, \infty, \alpha)] = \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha,\lambda))]$$

and apply the definition of L_2^* to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_2^*(\beta_1, \beta_2, \mathcal{D}_F, \alpha)] = \mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, 0))].$$

Let $\alpha^* = \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}}$. By the assumption in the lemma statement, we know that:

$$\alpha^* \geq \sqrt{\frac{\mathbb{E}_{\mathcal{D}_W}[L_2^*(\beta_1, \beta_2, \mathcal{D}_F, 0.5)]}{L^*(\rho)}} = 0.5.$$

We show that $(\alpha_I, \lambda_I) = (\alpha^*, 0)$. Assume for sake of contradiction that $(\alpha, \lambda) \neq (\alpha^*, 0)$ satisfies the safety constraint $\mathbb{E}_{\mathcal{D}_W}[L_2^*(\beta_1, \beta_2, \mathcal{D}_F, \alpha)] \leq \tau_I$ and achieves strictly better performance loss:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, \infty, \alpha)] < \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, 0, \infty, \alpha^*)].$$

We split into two cases: $\alpha^* = \alpha, \lambda \neq 0$ and $\alpha^* \neq \alpha$.

Case 1: $\alpha^* = \alpha$**,** $\lambda \neq 0$**.** By Lemma 135, we know that

$$\mathbb{E}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, \infty, \alpha^*)] = \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha^*, \lambda))] \geq (1-\alpha^*)^2 L^*(\rho).$$

Equality is obtained at $\lambda = 0$, which is a contradiction.

Case 2: $\alpha \neq \alpha^*$**.** By Lemma 135, it must hold that $\alpha > \alpha^*$ in order for the performance to beat that of $(\alpha^*, 0)$. However, this means that the safety constraint

$$\mathbb{E}_{\mathcal{D}_W}[L_2^*(\beta_1, \beta_2, \mathcal{D}_F, \alpha)] = \alpha^2 L^*(\rho) > (\alpha^*)^2 L^*(\rho) = \tau_I$$

is violated, which is a contradiction.

Concluding the statement. This means that$(\alpha_I, \lambda_I) = (\alpha^*, 0)$, which also means that:

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_I, \infty, \alpha_I)] &= \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha_I, \lambda_I))] \\
&= (1-\alpha_I)^2 \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)] \\
&= \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2.
\end{aligned}$$

□

The following claim calculates $\mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)]$.

Claim 137. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, suppose that $P = \infty$. Then it holds that:*

$$\mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)] = 2(1-\rho)\left(\sum_{i=1}^{P} i^{-\delta-1-\gamma}\right) = \Theta(1-\rho).$$

Proof. Let $\Sigma = V \Lambda V^T$ be the eigendecomposition of Σ, where Λ is a diagonal matrix consisting of the eigenvalues. We observe that

$$\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1 - \beta_2, v_i \rangle^2] = \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle^2] + \mathbb{E}_{\mathcal{D}_W}[\langle \beta_2, v_i \rangle^2] - 2\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle \langle \beta_2, v_i \rangle]$$
$$* = i^{-\delta} + i^{-\delta} - 2\rho i^{-\delta} = 2(1-\rho) i^{-\delta}.$$

This means that:

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)] &= \mathrm{Tr}(\Sigma \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)(\beta_1 - \beta_2)^T]) \\
&= \mathrm{Tr}(\Lambda \mathbb{E}_{\mathcal{D}_W}[V^T(\beta_1 - \beta_2)(\beta_1 - \beta_2)^T V]) \\
&= \sum_{i=1}^{P} i^{-1-\gamma} \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1 - \beta_2, v_i \rangle^2] \\
&= 2(1-\rho) \sum_{i=1}^{P} i^{-\delta-1-\gamma} \\
&= \Theta(1-\rho).
\end{aligned}$$

□

B.1.3 Proof of Theorem 5

We prove Theorem 5 using the above lemmas along with Corollary 12 (the proof of which we defer to Appendix B.3).

Proof of Theorem 5. We analyze (α_C, λ_C) first for the incumbent $C = I$ and then for the entrant $C = E$.

Analysis of the incumbent $C = I$. To compute α_I and λ_I, we apply Lemma 136. By Lemma 136, we see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_I, \infty, \alpha_I)] = \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2.$$

Analysis of the entrant $C = E$. Since the entrant faces no safety constraint, the entrant can choose any $\alpha \in [0.5, 1]$. We apply Corollary 12 to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] = \inf_{\alpha \in [0.5,1]} \inf_{\lambda > 0} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)] = \Theta\left(N^{-\nu}\right),$$

which means that:

$$N_E^*(\infty, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F) = \Theta\left(\left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^{-2/\nu}\right)$$

as desired. We can further apply Claim 137 to see that $L^*(\rho) = \Theta(1 - \rho)$. □

B.2 Proofs for Chapter 4.4

B.2.1 Proofs for Chapter 4.4.2

We prove Theorem 8. The main technical tool is Theorem 11, the proof of which we defer to Appendix B.3.

Proof of Theorem 8. We analyze (α_C, λ_C) first for the incumbent $C = I$ and then for the entrant $C = E$. Like in the theorem statement, let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta_2)] = \Theta(1 - \rho)$ (Claim 137) and $G_I := (\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))})^2$, and $\nu = \min(2(1 + \gamma), \delta + \gamma)$.

Analysis of the incumbent $C = I$. Recall from the facts in Appendix B.1.1 that:

$$L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \alpha) = \alpha^2 L^*(\rho).$$

This means that the safety constraint is satisfied if and only if $\alpha_I \leq \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}} =: \alpha^*$. The bound in Corollary 12 implies that:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_I, N_I, \alpha_I)] \\
&= \inf_{\alpha \in [0.5, \alpha^*]} \inf_{\lambda > 0} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N_I, \alpha)] \\
&= \Theta\left(\inf_{\lambda>0} \mathbb{E}_{\mathcal{D}_W}\left[L_1^*\left(\beta_1, \beta_2, \Sigma, \lambda, N_I, \alpha^*\right)\right]\right) \\
&= \begin{cases} \Theta\left(N_I^{-\nu}\right) & \text{if } N_I \leq (1-\alpha^*)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \Theta\left(\left(\frac{N_I}{(1-\alpha^*)(1-\rho)}\right)^{-\frac{\nu}{\nu+1}}\right) & \text{if } (1-\alpha^*)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N_I \leq (1-\alpha^*)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \Theta((1-\alpha^*)^2(1-\rho)) & \text{if } N_I \geq (1-\alpha^*)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}, \end{cases} \\
&= \begin{cases} \Theta\left(N_I^{-\nu}\right) & \text{if } N_I \leq G_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ \Theta\left(N_I^{-\frac{\nu}{\nu+1}} \cdot G_I^{\frac{\nu}{2(\nu+1)}}(1-\rho)^{\frac{\nu}{2(\nu+1)}}\right) & \text{if } G_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq G_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(G_I\right) & \text{if } N_I \geq G_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \end{cases}.
\end{aligned}$$

Analysis of the entrant $C = E$. Since the entrant faces no safety constraint, the entrant can choose any $\alpha \in [0.5, 1]$. We apply Corollary 11 to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] = \inf_{\alpha \in [0.5, 1]} \inf_{\lambda > 0} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)] = \Theta\left(N^{-\nu}\right),$$

which means that $N_E^*(N_I, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F)$ equals:

$$\begin{cases} \Theta\left(N_I\right) & \text{if } N_I \leq G_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ \Theta\left(N_I^{\frac{1}{\nu+1}} \cdot G_I^{-\frac{1}{2(\nu+1)}}(1-\rho)^{-\frac{1}{2(\nu+1)}}\right) & \text{if } G_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq G_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(G_I^{-\frac{1}{\nu}}\right) & \text{if } N_I \geq G_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}}. \end{cases}$$

as desired.

□

B.2.2 Proofs for Chapter 4.4.3

We prove Theorem 9. When the the safety constraints of the two firms are sufficiently close, it no longer suffices to analyze the loss up to constants for the entrant, and we require a more fine-grained analysis of the error terms than is provided in the scaling laws in Corollary 12. In this case, we turn to scaling laws for the *excess loss* as given by Corollary 14.

Proof of Theorem 9. We analyze (α_C, λ_C) first for the incumbent $C = I$ and then for the entrant $C = E$. Like in the theorem statement, let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)] = \Theta(1-$

ρ) (Claim 137), $G_I = (\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))})^2$, $G_E = (\sqrt{L^*(\rho)} - \sqrt{\min(\tau_E, L^*(\rho))})^2$, $D = G_I - G_E$, and $\nu = \min(2(1+\gamma), \delta + \gamma)$.

Analysis of the incumbent $C = I$. Since the incumbent has infinite data, we apply Lemma 136 to see that:

$$\begin{aligned}\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_I, \infty, \alpha_I)] &= \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2 \\ &= D + G_E.\end{aligned}$$

Analysis of the entrant $C = E$. Recall from the facts in Appendix B.1.1 that:

$$L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \alpha) = \alpha^2 L^*(\rho).$$

This means that the safety constraint is satisfied if and only if $\alpha_E \leq \sqrt{\frac{\min(\tau_E, L^*(\rho))}{L^*(\rho)}} =: \alpha^*$. The bound in Corollary 14 implies that:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] \\
&= \inf_{\alpha \in [0.5, \alpha^*]} \inf_{\lambda > 0} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)] \\
&= \inf_{\alpha \in [0.5, \alpha^*]} \left(\inf_{\lambda > 0} (\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) - L_1(\beta(\alpha, 0))]) + \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, 0))] \right) \\
&= \inf_{\alpha \in [0.5, \alpha^*]} \left(\inf_{\lambda > 0} (\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) - L_1(\beta(\alpha, 0))]) + (1-\alpha)^2 L^*(\rho) \right) \\
&= \Theta\left(\inf_{\lambda > 0} (\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) - L_1(\beta(\alpha^*, 0))]) \right) + (1-\alpha^*)^2 L^*(\rho) \\
&= \begin{cases} (1-\alpha^*)^2 L^*(\rho) + \Theta\left(N^{-\nu}\right) & \text{if } N \leq (1-\alpha^*)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ (1-\alpha^*)^2 L^*(\rho) + \Theta\left(\left(\frac{N}{(1-\alpha^*)(1-\rho)}\right)^{-\frac{\nu}{\nu+1}}\right) & \text{if } ((1-\alpha^*)(1-\rho))^{-\frac{1}{\nu}} \leq N \leq ((1-\alpha^*)(1-\rho))^{-\frac{\nu'+1}{\nu-\nu'}} \\ (1-\alpha^*)^2 L^*(\rho) + \Theta\left((1-\alpha^*)(1-\rho) N^{-\frac{\nu'}{\nu'+1}}\right) & \text{if } N \geq (1-\alpha^*)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}, \end{cases} \\
&= \begin{cases} G_E + \Theta\left(N^{-\nu}\right) & \text{if } N \leq (1-\alpha^*)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ G_E + \Theta\left(\left(\frac{N}{(1-\alpha^*)(1-\rho)}\right)^{-\frac{\nu}{\nu+1}}\right) & \text{if } (1-\alpha^*)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha^*)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}} \\ G_E + \Theta\left((1-\alpha^*)(1-\rho) N^{-\frac{\nu'}{\nu'+1}}\right) & \text{if } N \geq (1-\alpha^*)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}, \end{cases} .
\end{aligned}$$

Using this, we can compute the market-entry threshold as follows:

$$N_E^*(\infty, \tau_I, \tau_E, \mathcal{D}_W, \mathcal{D}_F)$$
$$= \begin{cases} \Theta(D^{-\frac{1}{\nu}}) & \text{if } D \geq (1-\alpha^*)(1-\rho) \\ \Theta\left(D^{-\frac{\nu+1}{\nu}}(1-\alpha^*)(1-\rho)\right) & \text{if } (1-\alpha^*)^{\frac{\nu}{\nu-\nu'}}(1-\rho)^{\frac{\nu}{\nu-\nu'}} \leq D \leq (1-\alpha^*)(1-\rho) \\ \Theta\left(\left(\frac{D}{(1-\alpha^*)(1-\rho)}\right)^{-\frac{\nu'+1}{\nu'}}\right) & \text{if } D \leq (1-\alpha^*)^{\frac{\nu}{\nu-\nu'}}(1-\rho)^{\frac{\nu}{\nu-\nu'}} \end{cases}$$
$$= \begin{cases} \Theta(D^{-\frac{1}{\nu}}) & \text{if } D \geq G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(D^{-\frac{\nu+1}{\nu}}G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}}\right) & \text{if } G_E^{\frac{\nu}{2(\nu-\nu')}}(1-\rho)^{\frac{\nu}{2(\nu-\nu')}} \leq D \leq G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(\left(\frac{D}{G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}}}\right)^{-\frac{\nu'+1}{\nu'}}\right) & \text{if } D \leq G_E^{\frac{\nu}{2(\nu-\nu')}}(1-\rho)^{\frac{\nu}{2(\nu-\nu')}} \end{cases}$$

□

B.3 Proofs for Chapter 4.5

In this section, we derive a deterministic equivalent and scaling laws for high-dimensional multi-objective linear regression. Before diving into this, we introduce notation, derive a basic decomposition, and give an outline for the remainder of the section.

Notation. Recall that (X_i, Y_i) denotes the labelled training dataset. Let the sample covariance be:

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} X_i X_i^T.$$

We also consider the following reparameterization where we group together inputs according to how they are labelled. For $j \in \{1,2\}$, we let $X_{1,j}, \ldots, X_{N_j,j}$ be the inputs labelled by β_j. We let

$$\hat{\Sigma}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} X_{i,1} X_{i,1}^T$$
$$\hat{\Sigma}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} X_{i,2} X_{i,2}^T.$$

It is easy to see that $\Sigma = \alpha\hat{\Sigma}_1 + (1-\alpha)\hat{\Sigma}_2$. Moreover, $\mathbb{E}[\hat{\Sigma}] = \mathbb{E}[\hat{\Sigma}_1] = \mathbb{E}[\hat{\Sigma}_2] = \Sigma$. Furthermore, $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are fully independent. We let $\sim$ denote asymptotic equivalence following Bach (2024).

Basic decomposition. A simple calculation shows that the solution and population-level loss of ridge regression takes the following form.

Claim 138. *Assume the notation above. Let $B^{\mathtt{sn}} = \beta_1\beta_1^T$, let $B^{\mathtt{df}} = (\beta_1 - \beta_2)(\beta_1 - \beta_2)^T$, and let $B^{\mathtt{mx}} = (\beta_1 - \beta_2)\beta_1^T$. The learned predictor takes the form:*

$$\hat{\beta}(\alpha, \lambda, X) = (\hat{\Sigma} + \lambda I)^{-1}(\alpha\hat{\Sigma}_1\beta_1 + (1-\alpha)\hat{\Sigma}_2\beta_2).$$

Moreover, it holds that $L_1(\hat{\beta}(\alpha, \lambda, X))$ is equal to

$$\underbrace{\lambda^2 \operatorname{Tr}((\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}B^{\mathtt{sn}})}_{(T1)} + \underbrace{(1-\alpha)^2 \operatorname{Tr}(\hat{\Sigma}_2(\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\hat{\Sigma}_2 B^{\mathtt{df}})}_{(T2)}$$
$$+ \underbrace{2\lambda(1-\alpha) \cdot \operatorname{Tr}((\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\hat{\Sigma}_2 B^{\mathtt{mx}})}_{(T3)}.$$

Proof. For $1 \le i \le N$, let Y_i be the label for input X_i in the training dataset. For $i \in \{1, 2\}$ and $1 \le i \le N_i$, let $Y_{i,j} := \langle \beta_i, X_{i,j} \rangle$ be the label for the input $X_{i,j}$ according to β_i.

For the first part, it follows from standard analyses of ridge regression that the learned predictor takes the form:

$$\begin{aligned}
\hat{\beta}(\alpha, \lambda, X) &= (\hat{\Sigma} + \lambda I)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} X_i Y_i\right) \\
&= (\hat{\Sigma} + \lambda I)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} X_{i,1} Y_{i,1} + \frac{1}{N}\sum_{i=1}^{N} X_{i,2} Y_{i,2}\right) \\
&= (\hat{\Sigma} + \lambda I)^{-1}\left(\alpha\hat{\Sigma}_1\beta_1 + (1-\alpha)\hat{\Sigma}_2\beta_2\right)
\end{aligned}$$

as desired.

For the second part, we first observe that the difference $\beta_1 - \hat{\beta}(\alpha, \lambda, X)$ takes the form:

$$\begin{aligned}
\beta_1 - \hat{\beta}(\alpha, \lambda, X) &= \beta_1 - (\hat{\Sigma} + \lambda I)^{-1}\left(\alpha\hat{\Sigma}_1\beta_1 + (1-\alpha)\hat{\Sigma}_2\beta_2\right) \\
&= (\hat{\Sigma} + \lambda I)^{-1}\left(\lambda\beta_1 + (1-\alpha)\hat{\Sigma}_2(\beta_1 - \beta_2)\right).
\end{aligned}$$

This means that:

$$\begin{aligned}
&L_1(\hat{\beta}(\alpha, \lambda, X)) \\
&= (\beta_1 - \hat{\beta}(\alpha, \lambda, X))^T\Sigma(\beta_1 - \hat{\beta}(\alpha, \lambda, X)) \\
&= \left(\lambda\beta_1 + (1-\alpha)\hat{\Sigma}_2(\beta_1 - \beta_2)\right)^T (\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\left(\lambda\beta_1 + (1-\alpha)\hat{\Sigma}_2(\beta_1 - \beta_2)\right) \\
&= \lambda^2 \cdot \beta_1^T\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\beta_1 + (1-\alpha)^2 \cdot (\beta_1 - \beta_2)^T\hat{\Sigma}_2\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\hat{\Sigma}_2(\beta_1 - \beta_2) \\
&\quad + 2\lambda(1-\alpha) \cdot \beta_1^T\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\hat{\Sigma}_2(\beta_1 - \beta_2) \\
&= \lambda^2 \operatorname{Tr}((\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}B^{\mathtt{sn}}) + (1-\alpha)^2 \operatorname{Tr}(\hat{\Sigma}_2(\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\hat{\Sigma}_2 B^{\mathtt{df}}) \\
&\quad + 2\lambda(1-\alpha) \cdot \operatorname{Tr}((\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\hat{\Sigma}_2 B^{\mathtt{mx}}).
\end{aligned}$$

as desired. $\square$

Outline for the rest of this Appendix. The bulk of our analysis in this section boils down to analyzing Term 1 (T1), Term 2 (T2), and Term 3 (T3) in Claim 138. Our main technical tool is the random matrix machinery from Appendix B.4. In Appendix B.3.1, we provide useful sublemmas about intermediate deterministic equivalents that we apply to analyze Terms 2 and 3. We then analyze Term 1 (Appendix B.3.2), Term 2 (Appendix B.3.3), and Term 3 (Appendix B.3.4), and use this to prove Lemma 10 (Appendix B.3.5).

We apply the power scaling assumptions to derive a simpler expression for the deterministic equivalent (Lemma 150 in Appendix B.3.6). We then apply Lemma 150 to prove Theorem 11 (Appendix B.3.7), and we prove Corollary 12 (Appendix B.3.8). We also apply Lemma 150 to prove Theorem 13 (Appendix B.3.9), and we prove Corollary 14 (Appendix B.3.10). We defer auxiliary calculations to Appendix B.3.11.

B.3.1 Useful lemmas about intermediate deterministic equivalents

The results in this section consider $Z_1 := \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$, which we introduce when conditioning on the randomness of $\hat{\Sigma}_1$ when analyzing (T2) and (T3). We derive several properties of Z_1 and the effective regularizer $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$ below.

The first set of lemmas relate the trace of various matrices involving κ_1 and Z_1 to deterministic quantities. A subtlety is that κ_1 and Z_1 are correlated, so we cannot directly apply Marčenko-Pastur, and instead we must indirectly analyze this quantity.

Lemma 139. *Consider the setup of Lemma 10, and assume the notation above. Assume $\alpha < 1$. Let $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$, and let $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$. Suppose that B has bounded operator norm.*

$$\kappa_1 \operatorname{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1} B\right) \sim \frac{(1-\alpha)\kappa}{\lambda} \operatorname{Tr}\left((\Sigma + \kappa I)^{-1} B\right)$$

Proof. By Claim 143, we know that:

$$\begin{aligned}(1-\alpha)\operatorname{Tr}\left(\left(\hat{\Sigma} + \lambda I\right)^{-1} B\right) &= \operatorname{Tr}\left(\left(\hat{\Sigma}_2 + Z_1\right)^{-1} B\right) \\ &\sim_{(A)} \kappa_1 \operatorname{Tr}\left((\Sigma + \kappa_1 I)^{-1} B\right).\end{aligned}$$

where (A) applies Lemma 160 and Claim 144.

Furthermore, by Lemma 160, it holds that:

$$\lambda \operatorname{Tr}\left(\left(\hat{\Sigma} + \lambda I\right)^{-1} B\right) \sim \kappa \operatorname{Tr}\left((\Sigma + \kappa I)^{-1} B\right).$$

Putting this all together yields the desired result. $\square$

Lemma 140. *Consider the setup of Lemma 10, and assume the notation above. Assume $\alpha < 1$. Let $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$, and let $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$. Suppose that A and B have bounded operator norm. Then it holds that:*

$$\begin{aligned}&(\kappa_1)^2 \left(\mathrm{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1} A (\Sigma + \kappa_1 Z_1)^{-1} B\right) + E_1\right)\\ &\sim \frac{(1-\alpha)^2\kappa^2}{\lambda^2}\left(\mathrm{Tr}\left((\Sigma + \kappa I)^{-1} A (\Sigma + \lambda I)^{-1} B\right) + E_2\right)\end{aligned}$$

where

$$\begin{aligned}\kappa &= \kappa(\lambda, N, \Sigma)\\ E_1 &= \frac{\frac{1}{N(1-\alpha)}\mathrm{Tr}(A(\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1})}{1-\frac{1}{N(1-\alpha)}\mathrm{Tr}((\Sigma+\kappa_1 Z_1)^{-1})\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma}\cdot \mathrm{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}B\right)\\ E_2 &= \frac{\frac{1}{N}\mathrm{Tr}(A\Sigma(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\mathrm{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}\cdot \mathrm{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}B\right)\end{aligned}$$

Proof. By Claim 143, we know that:

$$\begin{aligned}(1-\alpha)^2\,\mathrm{Tr}\left(\left(\hat{\Sigma}+\lambda I\right)^{-1} A \left(\hat{\Sigma}+\lambda I\right)^{-1} B\right) &= \mathrm{Tr}\left(\left(\hat{\Sigma}_2+Z_1\right)^{-1} A \left(\hat{\Sigma}_2+Z_1\right)^{-1} B\right)\\ &\sim_{(A)} \kappa_1^2\left(\mathrm{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1} A (\Sigma+\kappa_1 Z_1)^{-1} B\right) + E_1\right).\end{aligned}$$

where (A) applies Lemma 160 and Claim 144.

Furthermore, by Lemma 160, it holds that:

$$\lambda^2\,\mathrm{Tr}\left(\left(\hat{\Sigma}+\lambda I\right)^{-1} A \left(\hat{\Sigma}+\lambda I\right)^{-1} B\right) \sim \kappa^2\left(\mathrm{Tr}\left((\Sigma+\kappa I)^{-1} A (\Sigma+\kappa I)^{-1} B\right) + E_2\right).$$

Putting this all together yields the desired result.

□

Lemma 141. *Consider the setup of Lemma 10, and assume the notation above. Assume $\alpha < 1$. Let $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$, and let $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$. Then it holds that:*

$$\kappa_1^2\frac{\mathrm{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\mathrm{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \sim \frac{(1-\alpha)^2\kappa^2}{\lambda^2}\frac{\mathrm{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\mathrm{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}$$

Proof. By Claim 143, we know that:

$$\begin{aligned}&(1-\alpha)^2\,\mathrm{Tr}\left(\left(\hat{\Sigma}+\lambda I\right)^{-1}\Sigma\left(\hat{\Sigma}+\lambda I\right)^{-1}\Sigma\right)\\ &= \mathrm{Tr}\left(\left(\hat{\Sigma}_2+Z_1\right)^{-1}\Sigma\left(\hat{\Sigma}_2+Z_1\right)^{-1}\Sigma\right)\\ &\sim_{(A)} \kappa_1^2\left(1+\frac{\frac{1}{N(1-\alpha)}\mathrm{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\mathrm{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}\right)\mathrm{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma\right)\\ &= \kappa_1^2\frac{\mathrm{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\mathrm{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}\end{aligned}$$

where (A) applies Lemma 160 and Claim 144.

Furthermore, by Lemma 160, it holds that:

$$\begin{aligned}
&\lambda^2 \operatorname{Tr}\left(\left(\hat{\Sigma}+\lambda I\right)^{-1} \Sigma \left(\hat{\Sigma}+\lambda I\right)^{-1} \Sigma\right) \\
&\sim_{(A)} \kappa^2 \left(1+\frac{\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2}}\right) \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma\,(\Sigma+\kappa I)^{-1}\Sigma\right) \\
&= \kappa^2 \left(\frac{\operatorname{Tr}\left(\Sigma^2\left(\Sigma+\kappa I\right)^{-2}\right)}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2}}\right).
\end{aligned}$$

where (A) applies Lemma 160.

Putting this all together yields the desired result. □

Next, we relate the random effective regularizer κ_1 to the deterministic effective regularizer $\kappa(\lambda, N, \Sigma)$.

Lemma 142. *Consider the setup of Lemma 10, and assume the notation above. Assume $\alpha < 1$. Let $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha} I$, and let $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$. Let $\kappa = \kappa(\lambda, N, \Sigma)$. Then, it holds that $\lambda\kappa_1 \sim \kappa$.*

Proof. Recall that $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$ is the unique value such that:

$$\frac{1}{\kappa_1} + \frac{1}{N(1-\alpha)} \operatorname{Tr}((Z_1^{-1/2}\Sigma Z_1^{-1/2} + \kappa_1 I)^{-1} Z_1^{-1/2}\Sigma Z_1^{-1/2}) = 1.$$

We can write this as:

$$1 + \frac{\kappa_1}{N(1-\alpha)} \operatorname{Tr}((\Sigma + \kappa_1 Z_1)^{-1}\Sigma) = \kappa_1.$$

Now we apply Lemma 139 to see that:

$$\kappa_1 = 1 + \frac{\kappa_1}{N(1-\alpha)} \operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma) \sim 1 + \frac{1}{N(1-\alpha)} \frac{(1-\alpha)\kappa}{\lambda} \operatorname{Tr}((\Sigma+\kappa I)^{-1}\Sigma).$$

We can write this to see that:

$$\kappa_1 \sim \frac{\kappa}{\lambda}\left(\frac{\lambda}{\kappa} + \frac{1}{N}\operatorname{Tr}((\Sigma+\kappa I)^{-1}\Sigma)\right) = \frac{\kappa}{\lambda}.$$

This implies that $\lambda\kappa_1 \sim \kappa$ as desired.

□

The proofs of these results relied on the following facts.

Claim 143. *Consider the setup of Lemma 10, and assume the notation above. Assume $\alpha < 1$. Let $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha} I$. Then it holds that:*

$$(\hat{\Sigma}+\lambda I)^{-1} = (1-\alpha)^{-1}(\hat{\Sigma}_2 + Z_1)^{-1}.$$

Proof. We observe that:

$$\begin{aligned}(1-\alpha)(\hat{\Sigma}+\lambda I)^{-1} &= (1-\alpha)(\alpha\hat{\Sigma}_1+(1-\alpha)\hat{\Sigma}_2+\lambda I)^{-1} \\ &= (1-\alpha)(1-\alpha)^{-1}\left(\hat{\Sigma}_2+\frac{\alpha}{1-\alpha}\hat{\Sigma}_1+\frac{\lambda}{1-\alpha}I\right)^{-1} \\ &= \left(\hat{\Sigma}_2+Z_1\right)^{-1},\end{aligned}$$

where $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$. □

Claim 144. *Consider the setup of Lemma 10, and assume the notation above. Assume $\alpha < 1$. Let $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$. Then it holds that Z_1 and Z_1^{-1} both have bounded operator norm.*

Proof. Since $\hat{\Sigma}_1$ is PSD, we observe that:

$$\|Z_1\|_{op} = \frac{\alpha}{1-\alpha}\|\hat{\Sigma}_1\|_{op} + \frac{\lambda}{1-\alpha}.$$

The fact that $\|\hat{\Sigma}_1\|_{op}$ is bounded follows from the boundedness requirements from Assumption 7. This proves that $\|Z_1\|_{op}$ is bounded.

To see that $\|Z_1^{-1}\|$ is also bounded, note that:

$$\|Z_1^{-1}\|_{op} \geq \frac{1-\alpha}{\lambda}$$

□

B.3.2 Analysis of Term 1 (T1)

We show the following deterministic equivalent for term 1. This analysis is identical to the analysis of the deterministic equivalent for single-objective linear regression (Bach, 2024; Wei et al., 2022), and we include it for completeness.

Lemma 145. *Consider the setup of Lemma 10, and assume the notation above. Then it holds that:*

$$\lambda^2 \operatorname{Tr}((\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}B^{sn}) \sim \frac{\kappa^2}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}\cdot \operatorname{Tr}(\Sigma(\Sigma+\kappa I)^{-2}B^{sn})$$

Proof. We apply Lemma 160 to see that:

$$\begin{aligned}&\lambda^2 \operatorname{Tr}((\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}B^{\mathtt{sn}}) \\ &\sim \kappa^2 \operatorname{Tr}((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}B^{\mathtt{sn}}) + \frac{\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})} \\ &= \frac{\kappa^2}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}\cdot \operatorname{Tr}(\Sigma(\Sigma+\kappa I)^{-2}B^{\mathtt{sn}}),\end{aligned}$$

as desired. □

B.3.3 Analysis of Term 2 (T2)

We show the following deterministic equivalent for term 2.

Lemma 146. *Consider the setup of Lemma 10, and assume the notation above. Then it holds that:*

$$\begin{aligned}
&(1-\alpha)^2 \operatorname{Tr}\left(\hat{\Sigma}_2(\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}\hat{\Sigma}_2 B^{df}\right) \\
&\sim \frac{(1-\alpha)^2}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}\left(\operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}\Sigma B^{df}\Sigma\right)\right) \\
&+ \frac{(1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}\cdot\left(\operatorname{Tr}\left(\Sigma B^{df}\right) - 2(1-\alpha)\operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma B^{df}\Sigma\right)\right)
\end{aligned}$$

The key idea of the proof is to unwrap the randomness in layers. First, we condition on $\hat{\Sigma}_1$ and replace the randomness $\hat{\Sigma}_2$ with a deterministic equivalent where the effective regularizer κ_1 depends on $\hat{\Sigma}_1$ (Lemma 147). At this stage, we unfortunately cannot directly deal with the randomness $\hat{\Sigma}_1$ with deterministic equivalence due to the presence of terms κ_1 which depend on $\hat{\Sigma}_1$, and we instead apply the sublemmas from the previous section.

The following lemma replaces the randomness $\hat{\Sigma}_2$ with a deterministic equivalent.

Lemma 147. *Consider the setup of Lemma 10, and assume the notation above. Assume that $\alpha < 1$. Let $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$, and let $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$. Then it holds that:*

$$\begin{aligned}
&(1-\alpha)^2 \operatorname{Tr}\left(\hat{\Sigma}_2(\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}\hat{\Sigma}_2 B^{df}\right) \\
&\sim \frac{\operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{df}\Sigma\right)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \\
&+ \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}\cdot\left(\operatorname{Tr}\left(\Sigma B^{df}\right) - 2\operatorname{Tr}\left(\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{df}\right)\right).
\end{aligned}$$

Proof. By Claim 143 we have that:

$$\begin{aligned}
&(1-\alpha)^2 \operatorname{Tr}\left(\hat{\Sigma}_2(\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}\hat{\Sigma}_2 B^{\mathrm{df}}\right) \\
&= \operatorname{Tr}\left(\hat{\Sigma}_2\left(\hat{\Sigma}_2+Z_1\right)^{-1}\Sigma\left(\hat{\Sigma}_2+Z_1\right)^{-1}\hat{\Sigma}_2 B^{\mathrm{df}}\right) \\
&\sim_{(A)} \operatorname{Tr}\left(\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathrm{df}}\right) + E \\
&= \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathrm{df}}\Sigma\right) + E
\end{aligned}$$

where (A) follows from Lemma 163 and Claim 144, and E is defined such that

$$E := \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}(\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}\cdot(\kappa_1)^2\operatorname{Tr}\left(Z_1(\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}Z_1 B^{\mathrm{df}}\right).$$

and $\kappa_1 = \kappa(\lambda, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$.

Note that:

$$\begin{aligned}
&(\kappa_1)^2 \operatorname{Tr}\left(Z_1 \left(\Sigma + \kappa_1 Z_1\right)^{-1} \Sigma \left(\Sigma + \kappa_1 Z_1\right)^{-1} Z_1 B^{\mathsf{df}}\right) \\
&= \operatorname{Tr}\left((\kappa_1 Z_1) \left(\Sigma + \kappa_1 Z_1\right)^{-1} \Sigma \left(\Sigma + \kappa_1 Z_1\right)^{-1} (\kappa_1 Z_1) B^{\mathsf{df}}\right) \\
&= \operatorname{Tr}\left(\left(I - \Sigma(\Sigma + \kappa_1 Z_1)^{-1}\right) \Sigma \left(I - \Sigma(\Sigma + \kappa_1 Z_1)^{-1}\right)^T B^{\mathsf{df}}\right) \\
&= \operatorname{Tr}\left(\Sigma B^{\mathsf{df}}\right) - 2\operatorname{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right) + \operatorname{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right).
\end{aligned}$$

Note that:

$$\begin{aligned}
&\operatorname{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right) \\
&+ \operatorname{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right) \cdot \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)}{1 - \frac{1}{N(1-\alpha)}\operatorname{Tr}(\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)} \\
&= \frac{\operatorname{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)}{1 - \frac{1}{N(1-\alpha)}\operatorname{Tr}(\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)}
\end{aligned}$$

□

Now we are ready to prove Lemma 146.

Proof of Lemma 146. The statement follows trivially if $\alpha = 1$. By Lemma 147, it holds that:

$$\begin{aligned}
&(1-\alpha)^2 \operatorname{Tr}\left(\hat{\Sigma}_2(\hat{\Sigma} + \lambda I)^{-1}\Sigma(\hat{\Sigma} + \lambda I)^{-1}\hat{\Sigma}_2 B^{\mathsf{df}}\right) \\
&\sim \frac{\operatorname{Tr}\left(\left(\Sigma + \kappa_1 Z_1\right)^{-1} \Sigma \left(\Sigma + \kappa_1 Z_1\right)^{-1} \Sigma B^{\mathsf{df}}\Sigma\right)}{1 - \frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)} \\
&+ \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)}{1 - \frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)} \cdot \left(\operatorname{Tr}\left(\Sigma B^{\mathsf{df}}\right) - 2\operatorname{Tr}\left((\Sigma + \kappa_1 Z_1)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)\right) \\
&\sim_{(A)} (1-\alpha)^2\left(\operatorname{Tr}\left(\left(\Sigma + \kappa I\right)^{-1}\Sigma\left(\Sigma + \kappa I\right)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)\right) \\
&+ \frac{\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa I)^{-2})}{1 - \frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa I)^{-2})} \cdot (1-\alpha)^2 \cdot \operatorname{Tr}\left((\Sigma + \kappa I)^{-1}\Sigma(\Sigma + \kappa I)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right) \\
&+ \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)}{1 - \frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma + \kappa_1 Z_1)^{-1}\Sigma(\Sigma + \kappa_1 Z_1)^{-1}\Sigma)} \cdot \left(\operatorname{Tr}\left(\Sigma B^{\mathsf{df}}\right) - 2(1-\alpha)\operatorname{Tr}\left((\Sigma + \kappa I)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)\right) \\
&= \frac{(1-\alpha)^2}{1 - \frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa I)^{-2})}\left(\operatorname{Tr}\left(\left(\Sigma + \kappa I\right)^{-1}\Sigma\left(\Sigma + \kappa I\right)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)\right) \\
&+ \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa_1 Z_1)^{-2})}{1 - \frac{1}{N(1-\alpha)}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa_1 Z_1)^{-2})} \cdot \left(\operatorname{Tr}\left(\Sigma B^{\mathsf{df}}\right) - 2(1-\alpha)\operatorname{Tr}\left((\Sigma + \kappa I)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)\right) \\
&\sim_{(B)} \frac{(1-\alpha)^2}{1 - \frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa I)^{-2})}\left(\operatorname{Tr}\left(\left(\Sigma + \kappa I\right)^{-1}\Sigma\left(\Sigma + \kappa I\right)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)\right) \\
&+ (1-\alpha)\frac{\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa I)^{-2})}{1 - \frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma + \kappa I)^{-2})} \cdot \left(\operatorname{Tr}\left(\Sigma B^{\mathsf{df}}\right) - 2(1-\alpha)\operatorname{Tr}\left((\Sigma + \kappa I)^{-1}\Sigma B^{\mathsf{df}}\Sigma\right)\right)
\end{aligned}$$

where (A) applies Lemma 140, Lemma 139, and (B) uses Lemma 141 and Lemma 142. □

B.3.4 Analysis of Term 3 (T3)

We show the following deterministic equivalent for term 3.

Lemma 148. *Consider the setup of Lemma 10 and assume the notation above. Let* $B^{mx} = (\beta_1 - \beta_2)\beta_1^T$*, and let* $\kappa = \kappa(\lambda, N, \Sigma)$*. Then it holds that:*

$$\begin{aligned}
&2\lambda(1-\alpha)\operatorname{Tr}\left((\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}\hat{\Sigma}_2 B^{mx}\right)\\
&\sim \frac{2(1-\alpha)\kappa}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}\operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}\Sigma B^{mx}\right)\\
&-2\frac{(1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}\cdot\kappa\operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma B^{mx}\right)
\end{aligned}$$

The analysis follows a similar structure to the analysis of (T2); we similarly unwrap the randomness in layers.

Lemma 149. *Consider the setup of Lemma 10 and assume the notation above. Assume* $\alpha < 1$*. Let* $Z_1 = \frac{\alpha}{1-\alpha}\hat{\Sigma}_1 + \frac{\lambda}{1-\alpha}I$*, and let* $\kappa_1 = \kappa(1, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$*. Then it holds that:*

$$\begin{aligned}
&2\lambda(1-\alpha)^2\operatorname{Tr}\left((\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}\hat{\Sigma}_2 B^{mx}\right)\\
&\sim 2\frac{\lambda\kappa_1}{(1-\alpha)}\frac{\operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{mx}\right)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}\\
&-2\frac{\lambda\kappa_1}{(1-\alpha)}\cdot\frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa_1 Z_1)^{-2})}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}\cdot\operatorname{Tr}\left(\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{mx}\right).
\end{aligned}$$

Proof. By Claim 143 we have that:

$$\begin{aligned}
&2\lambda(1-\alpha)\operatorname{Tr}\left((\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}\hat{\Sigma}_2 B^{\mathrm{mx}}\right)\\
&= 2\frac{\lambda}{(1-\alpha)}\operatorname{Tr}\left(\left(\hat{\Sigma}_2+Z_1\right)^{-1}\Sigma\left(\hat{\Sigma}_2+Z_1\right)^{-1}\hat{\Sigma}_2 B^{\mathrm{mx}}\right)\\
&\sim_{(A)} 2\frac{\lambda}{(1-\alpha)}\left(\kappa_1\operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathrm{mx}}\right)-E\right)
\end{aligned}$$

where (A) follows from Lemma 164 and Claim 144, and E is defined to be

$$\frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}\cdot(\kappa_1)^2\operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}Z_1 B^{\mathrm{mx}}\right).$$

and $\kappa_1 = \kappa(\lambda, N(1-\alpha), Z_1^{-1/2}\Sigma Z_1^{-1/2})$.

Note that:

$$\begin{aligned}
&(\kappa_1)^2 \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1} Z_1 B^{\mathtt{mx}}\right) \\
&= \kappa_1 \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}(\kappa_1 Z_1) B^{\mathtt{mx}}\right) \\
&= \kappa_1 \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma\left(I-(\Sigma+\kappa_1 Z_1)^{-1}\Sigma\right) B^{\mathtt{mx}}\right) \\
&= \kappa_1 \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathtt{mx}}\right) - \kappa_1 \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathtt{mx}}\right)
\end{aligned}$$

Moreover, note that:

$$\begin{aligned}
&2\frac{\lambda\kappa_1}{(1-\alpha)} \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&+ 2\frac{\lambda}{(1-\alpha)} \cdot \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \cdot \kappa_1 \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&= 2\frac{\lambda}{(1-\alpha)} \frac{\operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathtt{mx}}\right)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \cdot \kappa_1.
\end{aligned}$$

□

Now we are ready to prove Lemma 146.

Proof of Lemma 146. The statement follows trivially if $\alpha = 1$. By Lemma 147, it holds that:

$$\begin{aligned}
&2\lambda(1-\alpha)^2 \operatorname{Tr}\left((\hat{\Sigma}+\lambda I)^{-1}\Sigma(\hat{\Sigma}+\lambda I)^{-1}\hat{\Sigma}_2 B^{\mathtt{mx}}\right) \\
&\sim 2\frac{\lambda\kappa_1}{(1-\alpha)} \frac{\operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathtt{mx}}\right)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \\
&- \frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \cdot 2\frac{\lambda\kappa_1}{(1-\alpha)} \operatorname{Tr}\left((\Sigma+\kappa_1 Z_1)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&\sim_{(A)} 2(1-\alpha)\kappa \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&+ 2(1-\alpha)\kappa \frac{\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})} \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&- 2\frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \cdot \kappa \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&= 2\frac{(1-\alpha)\kappa}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})} \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&- 2\frac{\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)}{1-\frac{1}{N(1-\alpha)}\operatorname{Tr}((\Sigma+\kappa_1 Z_1)^{-1}\Sigma(\Sigma+\kappa_1 Z_1)^{-1}\Sigma)} \cdot \kappa \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&\sim_{(B)} 2\frac{(1-\alpha)\kappa}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})} \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma(\Sigma+\kappa I)^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&- 2(1-\alpha)\frac{\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})}{1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2})} \cdot \kappa \operatorname{Tr}\left((\Sigma+\kappa I)^{-1}\Sigma B^{\mathtt{mx}}\right)
\end{aligned}$$

where (A) applies Lemma 140, Lemma 139, and Lemma 142, and (B) uses Lemma 141 and Lemma 142.

□

B.3.5 Proof of Lemma 10

Lemma 10 follows from the sublemmas in this section.

Proof. We apply Claim 138 to decompose the error in terms (T1), (T2), and (T3). We replace these terms with deterministic equivalents using Lemma 145, Lemma 146, and Lemma 148. The statement follows from adding these terms. □

B.3.6 Reformulation of Lemma 10 using assumptions from Chapter 4.2.3

Under the assumptions from Chapter 4.2.3, we show the following:

Lemma 150. *Suppose that power scaling holds for the eigenvalues and alignment coefficients with scaling $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, and suppose that $P = \infty$. Suppose that $\lambda \in (0, 1)$, and $N \geq 1$. Let $L_1^{\boldsymbol{det}} := L_1^{\boldsymbol{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ be the deterministic equivalent from Lemma 10. Let $\kappa = \kappa(\lambda, N, \Sigma)$ from Definition 18. Let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)]$. Then it holds that:*

$$
\begin{aligned}
Q \cdot \mathbb{E}_{\mathcal{D}_W}[L_1^{\boldsymbol{det}}] &= \kappa^2 (1 - 2(1-\alpha)^2(1-\rho)) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma} + \kappa)^2} + (1-\alpha)^2 L^*(\rho) \\
&\quad + 2\kappa(1-\rho)(1-\alpha)(1 - 2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma} + \kappa)^2} \\
&\quad + 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \cdot (1 - 2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa},
\end{aligned}
$$

where $Q = 1 - \frac{1}{N}\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}$.

Before proving Lemma 150, we prove a number of sublemmas where we analyze each of the terms in Lemma 10 using the assumptions from Chapter 4.2.3. In the proofs in this section, we use the notation $F \approx F'$ to denote that $F = \Theta(F')$ where the Θ is allowed to hide dependence on the scaling exponents γ and δ. Moreover let $\Sigma = V \Lambda V^T$ be the eigendecomposition of Σ, where Λ is a diagonal matrix consisting of the eigenvalues.

Lemma 151. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, suppose that $P = \infty$. Assume the* *notation from Lemma 10. Let $\nu = \min(2(1+\gamma), \gamma+\delta)$. Then it holds that:*

$$\mathbb{E}_{\mathcal{D}_W}[T_1] := \kappa^2 \cdot \mathrm{Tr}(\Sigma\Sigma_\kappa^{-2}\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{sn}}]) = \kappa^2 \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2}.$$

Proof. Observe that:

$$\begin{aligned}\mathrm{Tr}(\Sigma\Sigma_\kappa^{-2}\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{sn}}]) &= \mathrm{Tr}(\Lambda(\Lambda+\kappa I)^{-2}\mathbb{E}_{\mathcal{D}_W}[V^T\beta_1\beta_1^T V]) \\ &= \sum_{i=1}^{P} \frac{i^{-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \cdot \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i\rangle^2] \\ &= \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2}\end{aligned}$$

□

Lemma 152. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$,, suppose that $P = \infty$. Assume the notation from Lemma 10. Then it holds that:*

$$\mathbb{E}_{\mathcal{D}_W}[T_2] := (1-\alpha)^2\left(\mathrm{Tr}\left(\Sigma_\kappa^{-2}\Sigma^3\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{df}}]\right)\right) = 2(1-\alpha)^2(1-\rho)\sum_{i=1}^{P}\frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}.$$

Proof. First, we observe that

$$\begin{aligned}\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1-\beta_2, v_i\rangle^2] &= \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i\rangle^2] + \mathbb{E}_{\mathcal{D}_W}[\langle \beta_2, v_i\rangle^2] - 2\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i\rangle\langle \beta_2, v_i\rangle] \\ &= i^{-\delta} + i^{-\delta} - 2\rho i^{-\delta} = 2(1-\rho)i^{-\delta}.\end{aligned}$$

It is easy to see that:

$$\begin{aligned}\mathrm{Tr}\left(\Sigma_\kappa^{-2}\Sigma^3\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{df}}]\right) &= \mathrm{Tr}(\Lambda^3(\Lambda+\kappa I)^{-2}\mathbb{E}_{\mathcal{D}_W}[V^T(\beta_1-\beta_2)(\beta_1-\beta_2)^T V]) \\ &= \sum_{i=1}^{P}\frac{i^{-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}\cdot\mathbb{E}_{\mathcal{D}_W}[\langle\beta_1-\beta_2, v_i\rangle^2] \\ &= 2(1-\rho)\sum_{i=1}^{P}\frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}.\end{aligned}$$

□

Lemma 153. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, suppose that $P = \infty$. Assume the notation from Lemma 10. Then it holds that:*

$$\mathbb{E}_{\mathcal{D}_W}[T_3] := 2(1-\alpha)\kappa \cdot \mathrm{Tr}\left(\Sigma_\kappa^{-2}\Sigma^2 B^{\mathtt{mx}}\right) = 2(1-\alpha)\kappa(1-\rho)\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}.$$

Proof. First, we observe that

$$\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1 - \beta_2, v_i \rangle \langle \beta_1, v_i \rangle] = \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle^2] - \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle \langle \beta_2, v_i \rangle] = i^{-\delta} - \rho i^{-\delta} = (1-\rho) i^{-\delta}.$$

Observe that:

$$\begin{aligned}
\operatorname{Tr}\left(\Sigma_\kappa^{-2} \Sigma^2 B^{\mathtt{mx}}\right) &= \operatorname{Tr}(\Lambda^2 (\Lambda + \kappa I)^{-2} \mathbb{E}_{\mathcal{D}_W}[V^T (\beta_1 - \beta_2) \beta_1^T V]) \\
&= \sum_{i=1}^{P} \frac{i^{-2(1+\gamma)}}{(i^{-1-\gamma} + \kappa)^2} \cdot \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1 - \beta_2, v_i \rangle \langle \beta_1, v_i \rangle] \\
&= (1-\rho) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma} + \kappa)^2}.
\end{aligned}$$

This means that:

$$\mathbb{E}_{\mathcal{D}_W}[T_3] = 2(1-\alpha)\kappa(1-\rho) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma} + \kappa)^2}.$$

□

Lemma 154. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, suppose that $P = \infty$. Assume the notation from Lemma 10. Then it holds that:*

$$\begin{aligned}
|\mathbb{E}_{\mathcal{D}_W}[T_4]| &:= 2\kappa(1-\alpha) \frac{1}{N} \operatorname{Tr}(\Sigma^2 \Sigma_\kappa^{-2}) \cdot \operatorname{Tr}\left(\Sigma_\kappa^{-1} \Sigma \mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{mx}}]\right) \\
&= 2\kappa(1-\alpha)(1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma} + \kappa)^2} \right) \left(\sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma} + \kappa} \right)
\end{aligned}$$

Proof. First, we observe that

$$\mathbb{E}_{\mathcal{D}_W}[\langle \beta_1 - \beta_2, v_i \rangle \langle \beta_1, v_i \rangle] = \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle^2] - \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1, v_i \rangle \langle \beta_2, v_i \rangle] = i^{-\delta} + -\rho i^{-\delta} = (1-\rho) i^{-\delta}.$$

Observe that:

$$\begin{aligned}
\operatorname{Tr}\left(\Sigma_\kappa^{-1} \Sigma \mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{mx}}]\right) &= \operatorname{Tr}(\Lambda (\Lambda + \kappa I)^{-1} \mathbb{E}_{\mathcal{D}_W}[V^T (\beta_1 - \beta_2) \beta_1^T V]) \\
&= \sum_{i=1}^{P} \frac{i^{-1-\gamma}}{i^{-1-\gamma} + \kappa} \cdot \mathbb{E}_{\mathcal{D}_W}[\langle \beta_1 - \beta_2, v_i \rangle \langle \beta_1, v_i \rangle] \\
&= (1-\rho) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma} + \kappa}.
\end{aligned}$$

Now, apply Lemma 156, we see that:

$$
\begin{aligned}
|\mathbb{E}_{\mathcal{D}_W}[T_4]| &:= 2\kappa(1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\cdot\operatorname{Tr}\left(\Sigma_\kappa^{-1}\Sigma B^{\mathtt{mx}}\right) \\
&=_{(A)} 2\kappa(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\left(\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa}\right)
\end{aligned}
$$

where (A) follows from Lemma 156. □

Lemma 155. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, suppose that $P = \infty$. Assume the notation from Lemma 10, and similarly let*

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}_W}[T_5] &:= (1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\cdot\left(\operatorname{Tr}\left(\Sigma\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{df}}]\right)-2(1-\alpha)\operatorname{Tr}\left(\Sigma_\kappa^{-1}\Sigma^2\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{df}}]\right)\right) \\
&= 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot\left(\sum_{i=1}^{P}i^{-\delta-1-\gamma}-2(1-\alpha)\cdot\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)}\right).
\end{aligned}
$$

Proof. First, we observe that

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}_W}[\langle\beta_1-\beta_2,v_i\rangle^2] &= \mathbb{E}_{\mathcal{D}_W}[\langle\beta_1,v_i\rangle^2]+\mathbb{E}_{\mathcal{D}_W}[\langle\beta_2,v_i\rangle^2]-2\mathbb{E}_{\mathcal{D}_W}[\langle\beta_1,v_i\rangle\langle\beta_2,v_i\rangle] \\
&= i^{-\delta}+i^{-\delta}-2\rho i^{-\delta} = 2(1-\rho)i^{-\delta}.
\end{aligned}
$$

Now, observe that $\mathbb{E}_{\mathcal{D}_W}[T_5]$ is equal to:

$$
\begin{aligned}
&:= (1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\cdot\left(\operatorname{Tr}\left(\Sigma\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{df}}]\right)-2(1-\alpha)\operatorname{Tr}\left(\Sigma_\kappa^{-1}\Sigma^2\mathbb{E}_{\mathcal{D}_W}[B^{\mathtt{df}}]\right)\right) \\
&= (1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\cdot\left(\operatorname{Tr}\left(\Lambda\mathbb{E}_{\mathcal{D}_W}[V^T(\beta_1-\beta_2)(\beta_1-\beta_2)^TV]\right)\right) \\
&- (1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\cdot\left(2(1-\alpha)\operatorname{Tr}\left((\Lambda+\kappa I)^{-1}\Lambda^2\mathbb{E}_{\mathcal{D}_W}[V^T(\beta_1-\beta_2)(\beta_1-\beta_2)^TV]\right)\right) \\
&= (1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\cdot\left(\sum_{i=1}^{P}i^{-1-\gamma}\langle\beta_1-\beta_2,v_i\rangle^2-2(1-\alpha)\cdot\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)}\langle\beta_1-\beta_2,v_i\rangle^2\right) \\
&= 2(1-\alpha)(1-\rho)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\cdot\left(\sum_{i=1}^{P}i^{-\delta-1-\gamma}-2(1-\alpha)\cdot\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)}\right) \\
&=_{(A)} 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot\left(\sum_{i=1}^{P}i^{-\delta-1-\gamma}-2(1-\alpha)\cdot\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)}\right).
\end{aligned}
$$

where (A) uses Lemma 156. □

The proofs of these sublemmas use the following fact.

Lemma 156. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, and suppose that $P = \infty$. Assume the notation from Lemma 10. Then it holds that:*

$$
\operatorname{Tr}\left(\Sigma^2(\Sigma+\kappa I)^{-2}\right) = \sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}.
$$

Proof. We see that:

$$\begin{aligned}(1-\alpha)\frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2}) &= (1-\alpha)\frac{1}{N}\operatorname{Tr}(V\Lambda^2(\Lambda+\kappa I)^{-2}V^T)\\ &= (1-\alpha)\frac{1}{N}\operatorname{Tr}(\Lambda^2(\Lambda+\kappa I)^{-2})\\ &= \sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}.\end{aligned}$$

□

Now, we are ready to prove Lemma 150.

Proof of Lemma 150. By Lemma 156, we know:

$$Q = 1-\frac{1}{N}\operatorname{Tr}(\Sigma^2(\Sigma+\kappa I)^{-2}) = 1-\frac{1}{N}\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}.$$

Moreover, we have that:

$$\begin{aligned}
&Q\cdot\mathbb{E}_{\mathcal{D}_W}[L_1^{\mathtt{det}}]\\
&=_{(A)}\mathbb{E}_{\mathcal{D}_W}[T_1+T_2+T_3+T_4+T_5]\\
&=_{(B)}\kappa^2\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2}+2(1-\alpha)^2(1-\rho)\sum_{i=1}^{P}\frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}\\
&+2\kappa(1-\rho)(1-\alpha)\sum_{i=1}^{P}\frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}\\
&-2\kappa(1-\rho)(1-\alpha)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\left(\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa}\right)\\
&+2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot\left(\sum_{i=1}^{P}i^{-\delta-1-\gamma}-2(1-\alpha)\cdot\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)}\right).
\end{aligned}$$

where (A) follows from Lemma 10, and (B) follows from Lemmas 151-155.

By Claim 137, we know that:

$$L^*(\rho) = 2(1-\rho)\sum_{i=1}^{P}i^{-\delta-1-\gamma} = 2(1-\rho)\sum_{i=1}^{P}\frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma})^2}.$$

This means that:

$$
\begin{aligned}
&L^*(\rho) - 2(1-\rho)\sum_{i=1}^{P} \frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&= 2(1-\rho)\sum_{i=1}^{P}\left(\frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma})^2} - \frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}\right) \\
&= 2(1-\rho)\sum_{i=1}^{P}\left(\frac{i^{-\delta-3(1+\gamma)}\cdot((i^{-1-\gamma}+\kappa)^2-(i^{-1-\gamma})^2)}{(i^{-1-\gamma})^2\cdot(i^{-1-\gamma}+\kappa)^2}\right) \\
&= 2\kappa^2(1-\rho)\sum_{i=1}^{P}\left(\frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma})^2\cdot(i^{-1-\gamma}+\kappa)^2}\right) + 4\kappa(1-\rho)\sum_{i=1}^{P}\left(\frac{i^{-\delta-3(1+\gamma)}\cdot i^{-1-\gamma}}{(i^{-1-\gamma})^2\cdot(i^{-1-\gamma}+\kappa)^2}\right) \\
&= 2\kappa^2(1-\rho)\sum_{i=1}^{P}\left(\frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) + 4\kappa(1-\rho)\sum_{i=1}^{P}\left(\frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}\right)
\end{aligned}
$$

Applying this and some other algebraic manipulations, we obtain that:

$$
\begin{aligned}
&Q\cdot L_1^{\texttt{det}} \\
&= \kappa^2(1-2(1-\alpha)^2(1-\rho))\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + (1-\alpha)^2 L^*(\rho) \\
&+ 2\kappa(1-\rho)(1-\alpha)(1-2(1-\alpha))\sum_{i=1}^{P}\frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&- 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\left(\sum_{i=1}^{P} i^{-\delta-1-\gamma} - \sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa}\right) \\
&+ 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot\left(\sum_{i=1}^{P} i^{-\delta-1-\gamma} - 2(1-\alpha)\cdot\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)}\right) \\
&= \kappa^2(1-2(1-\alpha)^2(1-\rho))\sum_{i=1}^{P}\frac{i^{-\delta-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + (1-\alpha)^2 L^*(\rho) \\
&+ 2\kappa(1-\rho)(1-\alpha)(1-2(1-\alpha))\sum_{i=1}^{P}\frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot(1-2(1-\alpha))\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa}.
\end{aligned}
$$

□

B.3.7 Proof of Theorem 11

We now prove Theorem 11. In the proof, we again use the notation $F \approx F'$ to denote $F = \Theta(F')$. The main ingredient is Lemma 150, coupled with the auxiliary calculations in Appendix B.3.11.

Proof. The proof boils down to three steps: (1) obtaining an exact expression, (2) obtaining an up-to-constants asymptotic expression in terms of κ and Q, and (3) substituting in κ and Q.

Step 1: Exact expression. We apply Lemma 150 to see that:

$$\begin{aligned} Q \cdot L_1^{\mathtt{det}} = {} & \kappa^2(1-2(1-\alpha)^2(1-\rho))\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + (1-\alpha)^2 L^*(\rho) \\ & + 2\kappa(1-\rho)(1-\alpha)(1-2(1-\alpha))\sum_{i=1}^{P}\frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\ & + 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot(1-2(1-\alpha))\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa}, \end{aligned}$$

where $Q = 1 - \frac{1}{N}\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}$, where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)]$, and where $\kappa = \kappa(\Sigma, N, \lambda)$ as defined in Definition 18.

Step 2: Asymptotic expression in terms of κ and Q. We show that

$$Q \cdot L_1^{\mathtt{det}} \approx \kappa^{\frac{\nu}{1+\gamma}} + (1-\alpha)^2(1-\rho) + (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}.$$

We analyze this expression term-by-term and repeatedly apply Lemma 157. We see that:

$$\kappa^2(1-2(1-\alpha)^2(1-\rho))\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \approx_{(A)} \kappa^{\frac{\nu}{1+\gamma}}(1-2(1-\alpha)^2(1-\rho)) \approx_{(B)} \kappa^{\frac{\nu}{1+\gamma}},$$

where (A) uses Lemma 157 and (B) uses that $\alpha \geq 0.5$. Moreover, we observe that:

$$(1-\alpha)^2 L^*(\rho) \approx_{(C)} (1-\alpha)^2(1-\rho),$$

where (C) uses Claim 137. Moreover, we see that:

$$\begin{aligned}
&2\kappa(1-\rho)(1-\alpha)(1-2(1-\alpha))\sum_{i=1}^{P}\frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2}\\
&\approx_{(D)} (1-\alpha)(1-\rho)(1-2(1-\alpha))\max\left(\kappa,\kappa^{\frac{\delta+\gamma}{1+\gamma}}\right)\\
&=_{(E)} O\left((1-\alpha)\sqrt{1-\rho}\max\left(\kappa,\kappa^{\frac{\delta+\gamma}{2(1+\gamma)}}\right)\right)\\
&= O\left(\sqrt{(1-\alpha)^2(1-\rho)\cdot\kappa^{\frac{\min(2(1+\gamma),\gamma+\delta)}{1+\gamma}}}\right)\\
&=_{(F)} O\left(\kappa^{\frac{\min(2(1+\gamma),\gamma+\delta)}{1+\gamma}}+(1-\alpha)^2(1-\rho)\right)\\
&= O\left(\kappa^{\frac{\nu}{1+\gamma}}+(1-\alpha)^2(1-\rho)\right)
\end{aligned}$$

where (D) uses Lemma 157, (E) uses that $1-\rho \le 1$ and that $\kappa = O(1)$ (which follows from Lemma 159 and the assumption that $\lambda \in (0,1)$) and (F) follows from AM-GM. Finally, observe that:

$$\begin{aligned}
&2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot(1-2(1-\alpha))\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa}\\
&\approx (1-2(1-\alpha))\cdot(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa}\\
&\approx_{(G)} (1-2(1-\alpha))\cdot(1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}
\end{aligned}$$

where (G) uses Lemma 157 twice.

Putting this all together, we see that:

$$Q\cdot L_1^{\texttt{det}} \approx \kappa^{\frac{\nu}{1+\gamma}}+(1-\alpha)^2(1-\rho)+(1-2(1-\alpha))\cdot(1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}.$$

We split into two cases based on α. When $\alpha \ge 0.75$, we observe that

$$(1-2(1-\alpha))\cdot(1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N} \approx (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N},$$

and when $\alpha \in [0.5, 0.75]$, we observe that

$$(1-2(1-\alpha))\cdot(1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N} = O\left((1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right)$$

and

$$(1-\alpha)^2(1-\rho) \approx_{(H)} (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}$$

where (H) follows from the fact that $\kappa = \Omega(N^{-1-\gamma})$ by Lemma 159. Altogether, this implies that:

$$Q \cdot L_1^{\texttt{det}} \approx \kappa^{\frac{\nu}{1+\gamma}} + (1-\alpha)^2(1-\rho) + (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N},$$

as desired.

Step 2: Substitute in κ and Q. Finally, we apply Lemma 158 to see that:

$$Q^{-1} = \left(1 - \frac{1}{N}\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)^{-1} = \Theta(1).$$

We apply Lemma 159 to see that

$$\kappa = \kappa(\Sigma, N, \Sigma) = \max(N^{-1-\gamma}, \lambda).$$

Plugging this into the expression derived in Step 2, we obtain the desired expression. □

B.3.8 Proof of Corollary 12

We prove Corollary 12 using Theorem 11.

Proof. We apply Theorem 11 to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^{\texttt{det}}] = \Theta\left(\underbrace{\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu})}_{\text{finite data error}} + \underbrace{(1-\alpha)^2 \cdot (1-\rho)}_{\text{mixture error}} + \underbrace{(1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho)}_{\text{overfitting error}}\right).$$

We split into three cases: $N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$, $(1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$, and $N \geq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$.

Case 1: $N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$. We observe that the finite data error dominates regardless of λ. This is because the condition implies that

$$\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) \geq (1-\alpha)(1-\rho),$$

which dominates both the mixture error and the overfitting error.

Case 2: $(1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$. We show that the finite error term and overfitting error dominate. Let $\tilde{N} = \min(\lambda^{-\frac{1}{1+\gamma}}, N)$. We can bound the sum of the finite data error and the overfitting error as:

$$\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) + (1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho) = \tilde{N}^{-\nu} + (1-\alpha)(1-\rho)\frac{\tilde{N}}{N}.$$

Taking a derivative (and verifying the second order condition), we see that this expression is minimized when:

$$\nu \cdot \tilde{N}^{-\nu-1} = \frac{(1-\alpha)(1-\rho)}{N}$$

which solves to:

$$\tilde{N} = \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{-\frac{1}{1+\nu}}\right).$$

The lower bound on N guarantees that:

$$\begin{aligned}
\tilde{N} &= \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{-\frac{1}{1+\nu}}\right) \\
&= O\left(\left((1-\alpha)^{1+\frac{1}{\nu}}(1-\rho)^{1+\frac{1}{\nu}}\right)^{-\frac{1}{1+\nu}}\right) = O\left((1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}}\right) \\
&= O(N),
\end{aligned}$$

which ensures that $\tilde{N}$ can be achieved by some choice of λ. In particular, we can take $\lambda = \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{1+\gamma}{\nu+1}}\right)$.

The resulting sum of the finite error and the overfitting error is:

$$\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) + (1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right) = \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu}{\nu+1}}\right).$$

The upper bound on N guarantees that this dominates the mixture error:

$$\Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu}{\nu+1}}\right) = \Omega\left(\left((1-\alpha)^{1+\frac{2+\nu}{\nu}}(1-\rho)^{1+\frac{1}{\nu}}\right)^{\frac{\nu}{\nu+1}}\right) = \Omega((1-\alpha)^2(1-\rho))$$

as desired.

Case 3: $N \geq (1-\alpha)^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$**.** We show that the mixture and the overfitting error terms dominate. First, we observe that the sum of the mixture error and the finite data error is:

$$(1-\alpha)^2(1-\rho) + (1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho) = \Theta\left((1-\alpha)(1-\rho)\left(1-\alpha+\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)\right).$$

This is minimized by taking $\lambda = \Theta((N(1-\alpha))^{-1-\gamma})$, which yields $\Theta((1-\alpha)^2(1-\rho))$.

The upper bound on N and the setting of λ guarantees that this term dominates the finite data error:

$$\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) = O((N(1-\alpha))^{-\nu}) \leq O\left((1-\alpha)^{-\nu}(1-\alpha)^{2+\nu}(1-\rho)\right) = O((1-\alpha)^2(1-\rho),$$

as desired.

□

B.3.9 Proof of Theorem 13

We prove Theorem 13.

Proof of Theorem 13. Like the proof of Theorem 11, the proof boils down to three steps: (1) obtaining an exact expression, (2) obtaining an up-to-constants asymptotic expression in terms of κ, and (3) substituting in κ.

Step 1: Exact expression. We first apply Lemma 150 to obtain the precise loss:

$$
\begin{aligned}
&Q \cdot \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] \\
&= \kappa^2(1 - 2(1-\alpha)^2(1-\rho)) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + (1-\alpha)^2 L^*(\rho) \\
&+ 2\kappa(1-\rho)(1-\alpha)(1-2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \cdot (1-2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa},
\end{aligned}
$$

where $Q = 1 - \frac{1}{N}\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}$ and where $\kappa = \kappa(\Sigma, N, \lambda)$ as defined in Definition 18. This can be written as:

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] - (1-\alpha)^2 L^*(\rho) \\
&= Q^{-1} \cdot \kappa^2(1 - 2(1-\alpha)^2(1-\rho)) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1} \cdot 2\kappa(1-\rho)(1-\alpha)(1-2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1} \cdot 2(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \cdot (1-2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&+ \frac{1-Q}{Q}(1-\alpha)^2 L^*(\rho).
\end{aligned}
$$

Step 2: Asymptotic expression in terms of κ. We use the notation $F \approx F'$ to denote that $F = \Theta(F')$. We obtain:

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] - (1-\alpha)^2 L^*(\rho) \\
&\approx_{(A)} \kappa^2(1 - 2(1-\alpha)^2(1-\rho)) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ \kappa(1-\rho)(1-\alpha)(1-2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ (1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \cdot (1-2(1-\alpha)) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&+ (1-Q)(1-\alpha)^2 L^*(\rho) \\
&\approx_{(B)} \kappa^2 \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + \kappa(1-\rho)(1-\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ (1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&+ (1-\alpha)^2 L^*(\rho) \cdot \frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right).
\end{aligned}
$$

where (A) uses that Q^{-1} is a constant by Lemma 158 and (B) uses that $\alpha \geq 0.75$ and the definition of Q. Now, using the bounds from Lemma 157, and the bound from Claim 137, we obtain:

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] - (1-\alpha)^2 L^*(\rho) \\
&\approx \kappa^{\frac{\min(2(1+\gamma), \gamma+\delta)}{1+\gamma}} + (1-\rho)(1-\alpha)\max\left(\kappa, \kappa^{\frac{\gamma+\delta}{1+\gamma}}\right) + (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N} + \frac{\kappa^{-\frac{1}{1+\gamma}}}{N}(1-\alpha)^2(1-\rho) \\
&\approx \kappa^{\frac{\nu}{1+\gamma}} + (1-\rho)(1-\alpha)\kappa^{\frac{\nu'}{1+\gamma}} + (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}.
\end{aligned}
$$

Step 3: Substituting in κ. Finally, we apply Lemma 159 to see that

$$\kappa = \kappa(\Sigma, N, \Sigma) = \max(N^{-1-\gamma}, \lambda).$$

Plugging this into the expression derived in Step 2, we obtain the desired expression. □

B.3.10 Proof of Corollary 14

We prove Corollary 14 using Theorem 13.

Proof. We apply Theorem 11 to see that $\mathbb{E}_{\mathcal{D}_W}[L_1^{\texttt{det}} - L_1(\beta(\alpha, 0))]$ is equal to:

$$\Theta\left(\underbrace{\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu})}_{\text{finite data error}} + \underbrace{(1-\rho)(1-\alpha)\max(\lambda^{\frac{\nu'}{1+\gamma}}, N^{-\nu'})}_{\text{mixture finite data error}} + \underbrace{(1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho)}_{\text{overfitting error}}\right).$$

We split into three cases: $N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$, $(1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}$, and $N \geq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}$.

Case 1: $N \leq (1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}}$. We observe that the finite data error dominates regardless of λ. This is because the condition implies that

$$\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) \geq (1-\alpha)(1-\rho),$$

which dominates both the mixture finite data error and the overfitting error.

Case 2: $(1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N \leq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}$. We show that the finite error term and overfitting error dominate. Let $\tilde{N} = \min(\lambda^{-\frac{1}{1+\gamma}}, N)$. We can bound the sum of the finite data error and the overfitting error as:

$$\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) + (1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho) = \tilde{N}^{-\nu} + (1-\alpha)(1-\rho)\frac{\tilde{N}}{N}.$$

Taking a derivative (and verifying the second order condition), we see that this expression is minimized when:

$$\nu \cdot \tilde{N}^{-\nu-1} = \frac{(1-\alpha)(1-\rho)}{N}$$

which solves to:

$$\tilde{N} = \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{-\frac{1}{1+\nu}}\right).$$

The lower bound on N guarantees that:

$$\begin{aligned}
\tilde{N} &= \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{-\frac{1}{1+\nu}}\right) \\
&= O\left(\left((1-\alpha)^{1+\frac{1}{\nu}}(1-\rho)^{1+\frac{1}{\nu}}\right)^{-\frac{1}{1+\nu}}\right) \\
&= O\left((1-\alpha)^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}}\right) \\
&= O(N)
\end{aligned}$$

which ensures that $\tilde{N}$ can be achieved by some choice of λ. In particular, we can take $\lambda = \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{1+\gamma}{\nu+1}}\right)$.

The resulting sum of the finite error and the overfitting error is:

$$\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) + (1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right) = \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu}{\nu+1}}\right).$$

The upper bound on N and the choice of λ guarantees that this dominates the mixture finite data error, as shown below:

$$\begin{aligned}
&(1-\rho)(1-\alpha)\max(\lambda^{\frac{\nu'}{1+\gamma}}, N^{-\nu'}) \\
&= \Theta\left((1-\rho)(1-\alpha)\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu'}{\nu+1}}\right) \\
&= \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu}{\nu+1}}(1-\alpha)(1-\rho)\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu'-\nu}{\nu+1}}\right) \\
&= \Theta\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu}{\nu+1}}(1-\alpha)^{\frac{\nu'+1}{\nu+1}}(1-\rho)^{\frac{\nu'+1}{\nu+1}}N^{\frac{\nu-\nu'}{\nu+1}}\right) \\
&= O\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu}{\nu+1}}(1-\alpha)^{\frac{\nu'+1}{\nu+1}}(1-\rho)^{\frac{\nu'+1}{\nu+1}}(1-\alpha)^{-\frac{\nu'+1}{\nu+1}}(1-\rho)^{-\frac{\nu'+1}{\nu+1}}\right) \\
&= O\left(\left(\frac{(1-\alpha)(1-\rho)}{N}\right)^{\frac{\nu}{\nu+1}}\right)
\end{aligned}$$

as desired.

Case 3: $N \geq (1-\alpha)^{-\frac{\nu'+1}{\nu-\nu'}}(1-\rho)^{-\frac{\nu'+1}{\nu-\nu'}}$**.** We show that the mixture finite data error and the overfitting error terms dominate. First, we observe that the sum of the mixture error and the finite data error is:

$$\begin{aligned}
&(1-\rho)(1-\alpha)\max(\lambda^{\frac{\nu'}{1+\gamma}}, N^{-\nu'}) + (1-\alpha)\left(\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)(1-\rho) \\
&= \Theta\left((1-\alpha)(1-\rho)\left(\lambda^{\frac{\nu'}{1+\gamma}} + \frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)\right)
\end{aligned}$$

This is minimized by taking $\lambda = \Theta(N^{-\frac{1+\gamma}{\nu'+1}})$, which yields $\Theta((1-\alpha)(1-\rho)N^{-\frac{\nu'}{\nu'+1}})$.

The upper bound on N and the setting of λ guarantees that this term dominates the finite data error:

$$\begin{aligned}
\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) &= \Theta(N^{-\frac{\nu}{\nu'+1}}) \\
&\leq \Theta\left((1-\alpha)(1-\rho)N^{-\frac{\nu'}{\nu'+1}}(1-\alpha)^{-1}(1-\rho)^{-1}N^{-\frac{\nu-\nu'}{\nu'+1}}\right) \\
&= O\left((1-\alpha)(1-\rho)N^{-\frac{\nu'}{\nu'+1}}(1-\alpha)^{-1}(1-\rho)^{-1}(1-\alpha)(1-\rho)\right) \\
&= O\left((1-\alpha)(1-\rho)N^{-\frac{\nu'}{\nu'+1}}\right)
\end{aligned}$$

as desired.

□

B.3.11 Auxiliary calculations under power scaling assumptions

We show the following auxiliary calculations which we use when analyzing the terms in Lemma 10 under the power scaling assumptions. Throughout this section, we again use the notation $F \approx F'$ to denote that $F = \Theta(F')$.

Lemma 157. *Suppose that power-law scaling holds for eigenvalues and alignment coefficients with scaling exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, and suppose that $P = \infty$.*

Let $\kappa = \kappa(\lambda, N, \Sigma)$ be defined according to Definition 18. Then the following holds:

$$\sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \approx \kappa^{-2}\kappa^{\frac{\min(2(1+\gamma),\gamma+\delta)}{1+\gamma}}$$

$$\sum_{i=1}^{P} \frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \approx 1$$

$$\sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \approx 1$$

$$\sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \approx \max(1, \kappa^{\frac{\delta-1}{1+\gamma}})$$

$$\sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa} \approx \max(1, \kappa^{\frac{\delta-1}{1+\gamma}})$$

$$\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \approx \kappa^{-\frac{1}{1+\gamma}}$$

$$\sum_{i=1}^{P} \frac{i^{-1-\gamma}}{i^{-1-\gamma}+\kappa} \approx \kappa^{-\frac{1}{1+\gamma}}$$

$$\sum_{i=1}^{P} \frac{i^{-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \approx \kappa^{-2}\kappa^{\frac{\gamma}{1+\gamma}}$$

Proof. To prove the first statement, observe that:

$$\begin{aligned}
\sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} &= \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&\approx \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} i^{1+\gamma-\delta} + \kappa^{-2} \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} i^{-\delta-1-\gamma} \\
&\approx \max(1, \kappa^{-\frac{2+\gamma-\delta}{1+\gamma}}) + \kappa^{-2}\kappa^{\frac{\delta+\gamma}{1+\gamma}} \\
&= \kappa^{-2}\max(\kappa^2, \kappa^{\frac{\gamma+\delta}{1+\gamma}}) + \kappa^{-2}\kappa^{\frac{\delta+\gamma}{1+\gamma}} \\
&\approx \kappa^{-2}\max(\kappa^2, \kappa^{\frac{\gamma+\delta}{1+\gamma}}) \\
&\approx \kappa^{-2}\kappa^{\frac{\min(2(1+\gamma),\gamma+\delta)}{1+\gamma}}.
\end{aligned}$$

To prove the second statement, we use Lemma 159 and the assumption that $\lambda \in (0, 1)$ to see $\kappa = \Theta(\max(\lambda, N^{-1-\gamma})) = O(1)$. This means that

$$\sum_{i=1}^{P} \frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} = \Omega\left(\sum_{i=1}^{P} i^{-\delta-3(1+\gamma)}\right) = \Omega(1).$$

Moreover, we see that:

$$\sum_{i=1}^{P} \frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} = O\left(\sum_{i=1}^{P} \frac{i^{-\delta-3(1+\gamma)}}{(i^{-1-\gamma})^2}\right) = O\left(\sum_{i=1}^{P} i^{-\delta-1-\gamma)}\right) = \Omega(1).$$

To prove the third statement, we use Lemma 159 and the assumption that $\lambda \in (0,1)$ to see $\kappa = \Theta(\max(\lambda, N^{-1-\gamma})) = O(1)$. This means that

$$\sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} = \Omega\left(\sum_{i=1}^{P} i^{-\delta-2(1+\gamma)}\right) = \Omega(1).$$

Moreover, we see that:

$$\sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{i^{-1-\gamma}+\kappa} = O\left(\sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{i^{-1-\gamma}}\right) = O\left(\sum_{i=1}^{P} i^{-\delta-1-\gamma}\right) = O(1).$$

To prove the fourth statement, observe that:

$$\begin{aligned}
\sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} &\approx \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} + \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-\delta-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&\approx \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} i^{-\delta} + \kappa^{-2} \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} i^{-\delta-2-2\gamma} \\
&\approx \max(1, \kappa^{-\frac{1-\delta}{1+\gamma}}) + \kappa^{-2}\kappa^{\frac{\delta+1+2\gamma}{1+\gamma}} \\
&\approx \max(1, \kappa^{\frac{\delta-1}{1+\gamma}}).
\end{aligned}$$

To prove the fifth statement, observe that:

$$\begin{aligned}
\sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa} &= \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa} + \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa} \\
&\approx \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} i^{-\delta} + \kappa^{-1} \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} i^{-\delta-1-\gamma} \\
&\approx \max(1, \kappa^{-\frac{1-\delta}{1+\gamma}}) + \kappa^{-1}\kappa^{\frac{\delta+\gamma}{1+\gamma}} \\
&\approx \max(1, \kappa^{\frac{\delta-1}{1+\gamma}}).
\end{aligned}$$

To prove the sixth statement, observe that:

$$\begin{aligned}
\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} &= \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} + \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&\approx \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} 1 + \kappa^{-2} \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} i^{-2-2\gamma} \\
&\approx \kappa^{-\frac{1}{1+\gamma}} + \kappa^{-2}\kappa^{\frac{1+2\gamma}{1+\gamma}} \\
&\approx \kappa^{-\frac{1}{1+\gamma}}.
\end{aligned}$$

To prove the seventh statement, observe that:

$$\begin{aligned}
\sum_{i=1}^{P} \frac{i^{-1-\gamma}}{i^{-1-\gamma}+\kappa} &= \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-1-\gamma}}{i^{-1-\gamma}+\kappa} + \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-1-\gamma}}{i^{-1-\gamma}+\kappa} \\
&\approx \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} 1 + \kappa^{-1} \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} i^{-1-\gamma} \\
&\approx \kappa^{-\frac{1}{1+\gamma}} + \kappa^{-1}\kappa^{\frac{\gamma}{1+\gamma}} \\
&\approx \kappa^{-\frac{1}{1+\gamma}}.
\end{aligned}$$

To prove the eighth statement, observe that:

$$\begin{aligned}
\sum_{i=1}^{P} \frac{i^{-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} &= \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&\approx \sum_{i \le \kappa^{-\frac{1}{1+\gamma}}} i^{1+\gamma} + \kappa^{-2} \sum_{i \ge \kappa^{-\frac{1}{1+\gamma}}} i^{-1-\gamma} \\
&\approx \max(1, \kappa^{-\frac{2+\gamma}{1+\gamma}}) + \kappa^{-2}\kappa^{\frac{\gamma}{1+\gamma}} \\
&= \kappa^{-2}\max(\kappa^2, \kappa^{\frac{\gamma}{1+\gamma}}) + \kappa^{-2}\kappa^{\frac{\gamma}{1+\gamma}} \\
&\approx \kappa^{-2}\max(\kappa^2, \kappa^{\frac{\gamma}{1+\gamma}}) \\
&\approx \kappa^{-2}\kappa^{\frac{\gamma}{1+\gamma}}
\end{aligned}$$

□

Lemma 158. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, and suppose that $P = \infty$. Assume* *the notation from Lemma 10, and similarly let*

$$Q := 1 - \frac{1}{N}\operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2}).$$

Then it holds that $Q^{-1} = \Theta(1)$.

Proof. Let $\Sigma = V\Lambda V^T$ be the eigendecomposition of Σ, where Λ is a diagonal matrix consisting of the eigenvalues. By Definition 18, we see that:

$$\frac{\lambda}{\kappa} + \frac{1}{N}\operatorname{Tr}(\Sigma\Sigma_\kappa^{-1}) = 1.$$

This implies that:

$$\begin{aligned} Q &= 1 - \frac{1}{N}\operatorname{Tr}(\Sigma\Sigma_\kappa^{-1}) + \frac{1}{N}\left(\operatorname{Tr}(\Sigma\Sigma_\kappa^{-1}) - \operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\right) \\ &= \frac{\lambda}{\kappa} + \frac{1}{N}\left(\operatorname{Tr}(\Sigma\Sigma_\kappa^{-1}) - \operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2})\right). \end{aligned}$$

Observe that:

$$\begin{aligned} \operatorname{Tr}(\Sigma\Sigma_\kappa^{-1}) - \operatorname{Tr}(\Sigma^2\Sigma_\kappa^{-2}) &= \operatorname{Tr}(\Lambda(\Lambda + \kappa I)^{-1}) - \operatorname{Tr}(\Lambda^2(\Lambda + \kappa I)^{-2}) \\ &= \sum_{i=1}^{P}\left(\frac{i^{-1-\gamma}}{i^{-1-\gamma} + \kappa} - \frac{i^{-2-\gamma}}{(i^{-1-\gamma} + \kappa)^2}\right) \\ &= \kappa\sum_{i=1}^{P}\frac{i^{-1-\gamma}}{(i^{-1-\gamma} + \kappa)^2}. \end{aligned}$$

This means that:

$$\begin{aligned} Q &= \frac{\lambda}{\kappa} + \frac{\kappa}{N}\sum_{i=1}^{P}\frac{i^{-1-\gamma}}{(i^{-1-\gamma} + \kappa)^2} \\ &\approx_{(A)} \frac{\lambda}{\kappa} + \Theta(\left(\frac{\kappa}{N}\kappa^{-2}\kappa^{\frac{\gamma}{1+\gamma}}\right)) \\ &= \frac{\lambda}{\kappa} + \Theta\left(\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right). \end{aligned}$$

where (A) uses Lemma 157.

Case 1: $\kappa = \Theta(\lambda)$. In this case, we see that

$$Q = \frac{\lambda}{\kappa} + \Theta\left(\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right) = \Theta(1).$$

This means that $Q^{-1} = \Theta(1)$.

Case 2: $\kappa = \Theta(N^{-1-\gamma})$. In this case, we see that

$$Q = \frac{\lambda}{\kappa} + \Theta\left(\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right) = \Omega\left(\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right) = \Omega(1).$$

This means that $Q^{-1} = \Theta(1)$.

□

Lemma 159. *Suppose that power-law scaling holds for the eigenvalues with scaling exponent γ, and suppose that $P = \infty$. Then it holds that $\kappa(\lambda, M, \Sigma) = \Theta(\max(\lambda, M^{-1-\gamma}))$.*

Proof. Let $\Sigma = V\Lambda V^T$ be the eigendecomposition of Σ, where Λ is a diagonal matrix consisting of the eigenvalues. Observe that:

$$\begin{aligned}
\mathrm{Tr}((\Sigma + \kappa I)^{-1}\Sigma) &= \mathrm{Tr}(\Lambda(\Lambda + \kappa I)^{-1}) \\
&= \sum_{i=1}^{P} \frac{i^{-1-\gamma}}{i^{-1-\gamma} + \kappa} \\
&\approx_{(A)} \kappa^{-\frac{1}{1+\gamma}}.
\end{aligned}$$

where (A) follows from Lemma 157. Using Definition 18, we see that for $\kappa = \kappa(\lambda, M, \Sigma)$, it holds that:

$$\frac{\lambda}{\kappa} + \frac{1}{M}\Theta(\kappa^{-1-\gamma}) = 1.$$

This implies that $\kappa = \Theta(\max(\lambda, M^{-1-\gamma}))$ as desired. □

B.4 Machinery from random matrix theory

In this section, we introduce machinery from random matrix theory that serves as the backbone for our analysis of multi-objective scaling laws in Appendix B.3. In Appendix B.4.1, we give a recap of known Marčenko-Pastur properties. In Appendix B.4.2, we use these known properties to derive random matrix theory results which are tailored to our analysis.

B.4.1 Recap of Marčenko-Pastur properties

We introduce Marčenko-Pastur properties, following the treatment in Bach (2024). Informally speaking, Marčenko-Pastur laws show that a random matrix $(\hat{\Sigma} + \lambda I)^{-1}$ (where $\hat{\Sigma}$ is a sample covariance) behaves similarly to a deterministic matrix of the form $(\hat{\Sigma} + \kappa I)^{-1}$, where $\kappa = \kappa(\lambda, M, \Sigma)$ is an *effective regularizer*.

Deriving this formally requires placing several structural assumptions on number of data points $N \geq 1$, the number of parameters $P \geq 1$, the distribution $\mathcal{D}_F$, and the vectors β_1 and

β_2. We adopt assumptions from Bach (2024) which guarantee that a Marčenko-Pastur law holds for Σ, and we further introduce a boundedness assumption for technical reasons.

Assumption 7. *We assume that: (1) $X \sim \mathcal{D}_F$ takes the form $X = Z\Sigma^{1/2}$ where Z has bounded subgaussian i.i.d components with mean zero and unit variance, (2) N and P approach ∞ with $\frac{P}{N}$ tending to $\gamma > 0$, (3) the spectral measure $\frac{1}{P}\sum_{i=1}^{P}\delta_{\lambda_i}$ of Σ converges to a probability measure with compact support, and Σ is invertible and bounded in operator norm, and (4) for $j \in \{1, 2\}$, the measure $\sum_{i=1}^{P}\langle v_i, \beta_j\rangle^2$ converges to a measure with bounded mass, and β_j has bounded ℓ_2 norm.*

The effective regularizer $\kappa(\lambda, M, \Sigma)$ is defined as follows.

Definition 18 (Effective regularizer)**.** *For $\lambda \geq 0$, $M \geq 1$, and a P-dimensional positive semidefinite matrix Σ with eigenvalues λ_i for $1 \leq i \leq P$, the value $\kappa(\lambda, M, \Sigma)$ is the unique value $\kappa \geq 0$ such that:*

$$\frac{\lambda}{\kappa} + \frac{1}{N}\sum_{i=1}^{P}\frac{\lambda_i}{\lambda_i + \kappa} = 1.$$

We are now ready to state the key random matrix theory results proven in Bach (2024). Following Bach (2024), the asymptotic equivalence notation $u \sim v$ means that u/v tends to 1 as N and P go to ∞.

Lemma 160 (Restatement of Proposition 1 in Bach (2024))**.** *Let $\hat{\Sigma} = \frac{1}{M}\sum_{i=1}^{M} X_i X_i^T$ be the sample covariance matrix from M i.i.d. samples from $X_1, \ldots, X_M \sim \mathcal{D}_F$. Let $\kappa = \kappa(\lambda, N, \Sigma)$. Suppose that A and B have bounded operator norm, and suppose that Assumption 7 holds. Then it holds that:*

$$\lambda \operatorname{Tr}\left((\hat{\Sigma} + \lambda I)^{-1} A\right) \sim \kappa \operatorname{Tr}\left((\Sigma + \kappa I)^{-1} A\right)$$

$$\lambda^2 \operatorname{Tr}\left((\hat{\Sigma} + \lambda I)^{-1} A (\hat{\Sigma} + \lambda I)^{-1} B\right) \sim \kappa^2 \operatorname{Tr}\left((\Sigma + \kappa I)^{-1} A (\Sigma + \kappa I)^{-1} B\right) + \kappa^2 \frac{\frac{1}{N}\operatorname{Tr}\left(A\Sigma(\Sigma + \kappa I)^{-2}\right)}{1 - \frac{1}{N}\operatorname{Tr}\left(\Sigma^2(\Sigma + \kappa I)^{-2}\right)} \operatorname{Tr}\left((\Sigma + \kappa I)^{-1}\Sigma(\Sigma + \kappa I)^{-1} B\right).$$

We note that the requirement that B has bounded operator norm in Lemma 160 is what forces us to require that $\|\beta_1\|$ and $\|\beta_2\|$ are bounded. However, Wei et al. (2022) showed that the norm can be unbounded in several real-world settings, and thus instead opt to assume a local Marčenko-Pastur law and derive scaling laws based on this assumption. We suspect it may be possible to derive our scaling law with an appropriate analogue of the local Marčenko-Pastur law, which would also have the added benefit of allowing one to relax other requirements in Assumption 7 such as gaussianity. We view such an extension as an interesting direction for future work.

B.4.2 Useful random matrix theory facts

We derive several corollaries of Lemma 160 tailored to random matrices that arise in our analysis of multi-objective scaling laws.

Lemma 161. *Assume that $\mathcal{D}_F$ satisfies the Marčenko-Pastur property (Assumption 7). Let Z be a positive definite matrix such that Z^{-1} has bounded operator norm, and let A be a matrix with bounded operator norm. Let $\hat{\Sigma} = \frac{1}{M}\sum_{i=1}^{M} X_i X_i^T$ be the sample covariance matrix from M i.i.d. samples from $X_1, \ldots, X_M \sim \mathcal{D}_F$. Then it holds that:*

$$\lambda \cdot \mathrm{Tr}((\hat{\Sigma} + \lambda Z)^{-1} A) \sim \kappa \cdot \mathrm{Tr}((\Sigma + \kappa Z)^{-1} A). \tag{B.1}$$

If A also has bounded trace and Z has bounded operator norm, then it holds that:

$$\mathrm{Tr}(\hat{\Sigma}(\hat{\Sigma} + \lambda Z)^{-1} A) \sim \mathrm{Tr}(\Sigma \cdot (\Sigma + \kappa Z)^{-1} A) \tag{B.2}$$

where $\kappa = \kappa(\lambda, M, Z^{-1/2}\Sigma Z^{-1/2})$.

Proof. For (B.1), observe that:

$$\begin{aligned}
\lambda \cdot \mathrm{Tr}((\hat{\Sigma} + \lambda Z)^{-1} A) &= \lambda \cdot \mathrm{Tr}(Z^{-1/2}(Z^{-1/2}\hat{\Sigma} Z^{-1/2} + \lambda I)^{-1} Z^{-1/2} A) \\
&= \lambda \cdot \mathrm{Tr}((Z^{-1/2}\hat{\Sigma} Z^{-1/2} + \lambda I)^{-1} Z^{-1/2} A Z^{-1/2}) \\
&\sim_{(A)} \kappa \cdot \mathrm{Tr}((Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2} A Z^{-1/2}) \\
&= \kappa \cdot \mathrm{Tr}(Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2} A) \\
&= \kappa \cdot \mathrm{Tr}((\Sigma + \kappa Z)^{-1} A).
\end{aligned}$$

where (A) applies Lemma 160 (using the fact that since A and Z^{-1} have bounded operator norm, it holds that $Z^{-1/2} A Z^{-1/2}$ has bounded operator norm).

For (B.2), observe that:

$$\begin{aligned}
\mathrm{Tr}(\hat{\Sigma}(\hat{\Sigma} + \lambda Z)^{-1} A) &=_{(A)} \mathrm{Tr}\left(\left(I - \lambda Z^{1/2}\left(Z^{-1/2}\hat{\Sigma} Z^{-1/2} + \lambda I\right)^{-1} Z^{-1/2}\right) A\right) \\
&=_{(B)} \mathrm{Tr}(A) - \lambda \cdot \mathrm{Tr}\left(\left(Z^{-1/2}\hat{\Sigma} Z^{-1/2} + \lambda I\right)^{-1} Z^{-1/2} A Z^{1/2}\right) \\
&\sim_{(C)} \mathrm{Tr}(A) - \kappa \cdot \mathrm{Tr}\left(\left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2} A Z^{1/2}\right) \\
&=_{(D)} \mathrm{Tr}\left(\left(I - \kappa Z^{1/2}\left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2}\right) A\right) \\
&=_{(E)} \mathrm{Tr}(\Sigma(\Sigma + \kappa Z)^{-1} A)
\end{aligned}$$

where (A) and (E) follows from Claim 165, (B) and (D) use the fact that $\mathrm{Tr}(A)$ is bounded, and (C) follows from Lemma 160 (using the fact that since A, Z, and Z^{-1} have bounded operator norm, it holds that $Z^{-1/2} A Z^{1/2}$ has bounded operator norm).

□

Lemma 162. *Assume that $\mathcal{D}_F$ satisfies the Marčenko-Pastur property (Assumption 7). Let Z be any positive definite matrix such that Z and Z^{-1} have bounded operator norm, and let A*

and B have bounded operator norm. Let $\hat{\Sigma} = \frac{1}{M}\sum_{i=1}^{M} X_i X_i^T$ be the sample covariance matrix from M i.i.d. samples from $X_1, \ldots, X_M \sim \mathcal{D}_F$. Then it holds that:

$$\begin{aligned}
&\lambda^2 \operatorname{Tr}((\hat{\Sigma} + \lambda Z)^{-1} A (\hat{\Sigma} + \lambda Z)^{-1} B) \\
&= \lambda^2 \operatorname{Tr}(Z^{-1/2}(Z^{-1/2}\hat{\Sigma}Z^{-1/2} + \lambda I)^{-1} Z^{-1/2} A Z^{-1/2} (Z^{-1/2}\hat{\Sigma}Z^{-1/2} + \lambda I)^{-1} B) \\
&\sim \kappa^2 \operatorname{Tr}((\Sigma + \kappa Z)^{-1} A (\Sigma + \kappa Z)^{-1} B) \\
&+ \kappa^2 \frac{\frac{1}{M}\operatorname{Tr}((\Sigma + \kappa Z)^{-1}\Sigma(\Sigma + \kappa Z)^{-1} A)}{1 - \frac{1}{M}\operatorname{Tr}((\Sigma + \kappa Z)^{-1}\Sigma(\Sigma + \kappa Z)^{-1}\Sigma)} \operatorname{Tr}((\Sigma + \kappa Z)^{-1}\Sigma(\Sigma + \kappa Z)^{-1} B)
\end{aligned}$$

where $\kappa = \kappa(\lambda, M, Z^{-1/2}\Sigma Z^{-1/2})$.

Proof. Let $q = \frac{\frac{1}{M}\operatorname{Tr}(Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2}+\kappa I)^{-2} Z^{-1/2} A Z^{-1/2})}{1-\frac{1}{M}\operatorname{Tr}(Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2}+\kappa I)^{-2} Z^{-1/2}\Sigma Z^{-1/2})}$.

Observe that:

$$\begin{aligned}
&\lambda^2 \operatorname{Tr}((\hat{\Sigma} + \lambda Z)^{-1} A (\hat{\Sigma} + \lambda Z)^{-1} B) \\
&\lambda^2 \operatorname{Tr}\left(Z^{-1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2} + \lambda I\right)^{-1} Z^{-1/2} A Z^{-1/2} \left(Z^{-1/2}\hat{\Sigma}Z^{-1/2} + \lambda I\right)^{-1} Z^{-1/2} B\right) \\
&= \lambda^2 \operatorname{Tr}\left(\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2} + \lambda I\right)^{-1} Z^{-1/2} A Z^{-1/2} \left(Z^{-1/2}\hat{\Sigma}Z^{-1/2} + \lambda I\right)^{-1} Z^{-1/2} B Z^{-1/2}\right) \\
&\sim_{(A)} \kappa^2 \operatorname{Tr}\left(\left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2} A Z^{-1/2} \left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2} B Z^{-1/2}\right) \\
&+ \kappa^2 q \operatorname{Tr}\left(\left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2}\Sigma Z^{-1/2} \left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2} B Z^{-1/2}\right) \\
&= \kappa^2 \operatorname{Tr}\left(Z^{-1/2}\left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2} A Z^{-1/2} \left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2} B\right) \\
&+ \kappa^2 q \operatorname{Tr}\left(Z^{-1/2}\left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2}\Sigma Z^{-1/2} \left(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I\right)^{-1} Z^{-1/2} B\right) \\
&= \kappa^2 \operatorname{Tr}\left((\Sigma + \kappa Z)^{-1} A (\Sigma + \kappa Z)^{-1} B\right) + q\kappa^2 \operatorname{Tr}\left((\Sigma + \kappa Z)^{-1} \Sigma (\Sigma + \kappa Z)^{-1} B\right),
\end{aligned}$$

where (A) follows from Lemma 160 (using the fact that since A, B, Z, and Z^{-1} have bounded operator norm, it holds that $Z^{-1/2}AZ^{1/2}$, Σ, and $Z^{-1/2}BZ^{1/2}$ have bounded operator norm).

We can simplify q as follows:

$$\begin{aligned}
q &= \frac{\frac{1}{M}\operatorname{Tr}(Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-2} Z^{-1/2} A Z^{-1/2})}{1 - \frac{1}{M}\operatorname{Tr}(Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-2} Z^{-1/2}\Sigma Z^{-1/2})} \\
&= \frac{\frac{1}{M}\operatorname{Tr}((Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2} A Z^{-1/2})}{1 - \frac{1}{M}\operatorname{Tr}((Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2}\Sigma Z^{-1/2})} \\
&= \frac{\frac{1}{M}\operatorname{Tr}(Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2} A)}{1 - \frac{1}{M}\operatorname{Tr}(Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2}\Sigma Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2} + \kappa I)^{-1} Z^{-1/2}\Sigma)} \\
&= \frac{\frac{1}{M}\operatorname{Tr}((\Sigma + \kappa Z)^{-1}\Sigma(\Sigma + \kappa Z)^{-1} A)}{1 - \frac{1}{M}\operatorname{Tr}((\Sigma + \kappa Z)^{-1}\Sigma(\Sigma + \kappa Z)^{-1}\Sigma)}.
\end{aligned}$$

□

Lemma 163. *Assume that $\mathcal{D}_F$ satisfies the Marčenko-Pastur property (Assumption 7). Let Z be any positive definite matrix such that Z and Z^{-1} have bounded operator norm. Let A and B have bounded operator norm, and suppose also that $\mathrm{Tr}(AB)$ is bounded. Let $\hat{\Sigma} = \frac{1}{M}\sum_{i=1}^{M} X_i X_i^T$ be the sample covariance matrix from M i.i.d. samples from $X_1, \ldots, X_M \sim \mathcal{D}_F$. Then it holds that:*

$$\mathrm{Tr}(\hat{\Sigma}(\hat{\Sigma}+\lambda Z)^{-1}A(\hat{\Sigma}+\lambda Z)^{-1}\hat{\Sigma}B) \sim \mathrm{Tr}(\Sigma(\Sigma+\kappa Z)^{-1}A(\Sigma+\kappa Z)^{-1}\Sigma B) + E, \tag{B.3}$$

where:

$$E := \frac{\frac{1}{M}\mathrm{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}A)}{1-\frac{1}{M}\mathrm{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}\Sigma)} \cdot \kappa^2 \,\mathrm{Tr}\left((\Sigma+\kappa Z)^{-1}\Sigma\,(\Sigma+\kappa Z)^{-1} ZBZ\right),$$

and $\kappa = \kappa(\lambda, M, Z^{-1/2}\Sigma Z^{-1/2})$.

Proof. Observe that:

$$\begin{aligned}
&\mathrm{Tr}(\hat{\Sigma}(\hat{\Sigma}+\lambda Z)^{-1}A(\hat{\Sigma}+\lambda Z)^{-1}\hat{\Sigma}B)\\
&= \mathrm{Tr}(\hat{\Sigma}(\hat{\Sigma}+\lambda Z)^{-1}A\left(\hat{\Sigma}(\hat{\Sigma}+\lambda Z)^{-1}\right)^T B)\\
&=_{(A)} \mathrm{Tr}\left(\left(I-\lambda Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}\right)A\left(I-\lambda Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}\right)^T B\right)\\
&=_{(B)} \mathrm{Tr}(AB) - \lambda\,\mathrm{Tr}\left(A\left(Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}\right)^T B\right)\\
&- \lambda\,\mathrm{Tr}\left(Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}AB\right)\\
&+ \lambda^2\,\mathrm{Tr}\left(Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}A\left(Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}\right)^T B\right)\\
&= \mathrm{Tr}(AB) - \lambda\,\mathrm{Tr}\left(AZ^{-1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{1/2}B\right)\\
&- \lambda\,\mathrm{Tr}\left(Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}AB\right)\\
&+ \lambda^2\,\mathrm{Tr}\left(Z^{1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{-1/2}AZ^{-1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1}Z^{1/2}B\right)\\
&= \mathrm{Tr}(AB) - \underbrace{\lambda\,\mathrm{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1}ZBA\right)}_{(1)} - \underbrace{\lambda\,\mathrm{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1}ABZ\right)}_{(2)}\\
&+ \underbrace{\lambda^2\,\mathrm{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1}A\left(\hat{\Sigma}+\lambda Z\right)^{-1}ZBZ\right)}_{(3)}
\end{aligned}$$

where (A) follows from Claim 165, (B) uses that $\mathrm{Tr}(AB)$ is bounded, For term (1) and term (2), we apply Lemma 161 to see that:

$$\lambda \operatorname{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1} ZBA\right) \sim \kappa\lambda \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} ZBA\right)$$
$$\lambda \operatorname{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1} ABZ\right) \sim \kappa \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} ABZ\right).$$

For term (3), we apply Lemma 162 to see that

$$\begin{aligned}
&\lambda^2 \operatorname{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1} A\left(\hat{\Sigma}+\lambda Z\right)^{-1} ZBZ\right)\\
&\sim \kappa^2 \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A(\Sigma+\kappa Z)^{-1} ZBZ\right)\\
&+\kappa^2 \frac{\frac{1}{M}\operatorname{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}A)}{1-\frac{1}{M}\operatorname{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}\Sigma)} \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}ZBZ\right)\\
&\sim \kappa^2 \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A(\Sigma+\kappa Z)^{-1} ZBZ\right)+E
\end{aligned}$$

This means that:

$$\begin{aligned}
\operatorname{Tr}\left(\hat{\Sigma}(\hat{\Sigma}+\lambda Z)^{-1}A(\hat{\Sigma}+\lambda Z)^{-1}\hat{\Sigma}\right) &\sim \operatorname{Tr}(AB)-\kappa\operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} ZBA\right)-\kappa\operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} ABZ\right)\\
&\quad+\kappa^2\operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A(\Sigma+\kappa Z)^{-1} ZBZ\right)+E\\
&=_{(C)} \operatorname{Tr}(\Sigma(\Sigma+\kappa Z)^{-1}A(\Sigma+\kappa Z)^{-1}\Sigma B)+E,
\end{aligned}$$

where (C) uses an analogous analysis to the beginning of the proof. □

Lemma 164. *Assume that $\mathcal{D}_F$ satisfies the Marčenko-Pastur property (Assumption 7). Let Z be any positive definite matrix such that Z and Z^{-1} have bounded operator norm, and let A and B have bounded operator norm. Let $\hat{\Sigma}=\frac{1}{M}\sum_{i=1}^{M} X_i X_i^T$ be the sample covariance matrix from M i.i.d. samples from $X_1,\ldots,X_M \sim \mathcal{D}_F$. Then it holds that:*

$$\lambda \operatorname{Tr}\left((\hat{\Sigma}+\lambda Z)^{-1}A(\hat{\Sigma}+\lambda Z)^{-1}\hat{\Sigma}B\right) \sim \kappa\operatorname{Tr}\left((\Sigma+\kappa Z)^{-1}A(\Sigma+\kappa Z)^{-1}\Sigma B\right)-E, \tag{B.4}$$

where:

$$E := \frac{\frac{1}{M}\operatorname{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}A)}{1-\frac{1}{M}\operatorname{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}\Sigma)}\cdot\kappa^2\operatorname{Tr}\left((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1} ZB\right)$$

and $\kappa=\kappa(\lambda, N, Z^{-1/2}\Sigma Z^{-1/2})$.

Proof. Observe that:

$$
\begin{aligned}
&\lambda \operatorname{Tr}\left((\hat{\Sigma}+\lambda Z)^{-1} A(\hat{\Sigma}+\lambda Z)^{-1} \hat{\Sigma} B\right) \\
&=_{(A)} \lambda \operatorname{Tr}\left(Z^{-1/2}(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I)^{-1} Z^{-1/2} A\left(I-\lambda Z^{-1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1} Z^{1/2}\right) B\right) \\
&= \lambda \operatorname{Tr}\left(Z^{-1/2}(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I)^{-1} Z^{-1/2} A B\right) \\
&- \lambda^2 \operatorname{Tr}\left(Z^{-1/2}\left(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I\right)^{-1} Z^{-1/2} A Z^{-1/2}(Z^{-1/2}\hat{\Sigma}Z^{-1/2}+\lambda I)^{-1} Z^{1/2} B\right) \\
&= \underbrace{\lambda \operatorname{Tr}\left((\hat{\Sigma}+\lambda Z)^{-1} A B\right)}_{(1)} - \underbrace{\lambda^2 \operatorname{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1} A(\hat{\Sigma}+\lambda Z)^{-1} Z B\right)}_{(2)}
\end{aligned}
$$

where (A) follows from Claim 165.

For term (1), we apply Lemma 161 see that:

$$\lambda \operatorname{Tr}\left((\hat{\Sigma}+\lambda Z)^{-1} A B\right) \sim \kappa \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A B\right).$$

For term (2), we apply Lemma 162 to see that

$$
\begin{aligned}
&\lambda^2 \operatorname{Tr}\left(\left(\hat{\Sigma}+\lambda Z\right)^{-1} A(\hat{\Sigma}+\lambda Z)^{-1} Z B\right) \\
&\sim \kappa^2 \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A\,(\Sigma+\kappa Z)^{-1} Z B\right) \\
&+ \kappa^2 \frac{\frac{1}{M}\operatorname{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1} A)}{1-\frac{1}{M}\operatorname{Tr}((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1}\Sigma)} \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1}\Sigma(\Sigma+\kappa Z)^{-1} Z B\right) \\
&\sim \kappa^2 \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A\,(\Sigma+\kappa Z)^{-1} Z B\right) + E.
\end{aligned}
$$

This means that:

$$
\begin{aligned}
&\lambda \operatorname{Tr}\left((\hat{\Sigma}+\lambda Z)^{-1} A(\hat{\Sigma}+\lambda Z)^{-1} \hat{\Sigma} B\right) \\
&\sim \kappa \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A B\right) + \kappa^2 \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A\,(\Sigma+\kappa Z)^{-1} Z B\right) - E \\
&= \kappa \operatorname{Tr}\left(Z^{-1/2}(Z^{-1/2}\Sigma Z^{-1/2}+\kappa I)^{-1} Z^{-1/2} A\left(I-\kappa Z^{1/2}\left(Z^{-1/2}\Sigma Z^{-1/2}+\kappa I\right)^{-1} Z^{-1/2}\right) B\right) - E \\
&=_{(A)} \kappa \operatorname{Tr}\left((\Sigma+\kappa Z)^{-1} A(\Sigma+\kappa Z)^{-1}\Sigma B\right) - E,
\end{aligned}
$$

where (A) uses an analogous analysis to the beginning of the proof. □

The proofs of these results relied on the following basic matrix fact.

Claim 165. *Let A be any matrix and let B be any symmetric positive definite matrix. Then it holds that:*

$$A(A+\lambda B)^{-1} = I - \lambda B^{1/2}\left(B^{-1/2} A B^{-1/2} + \lambda I\right)^{-1} B^{-1/2}.$$

Proof. Observe that:

$$
\begin{aligned}
& A(A+\lambda B)^{-1} \\
& = AB^{-1/2}\left(B^{-1/2}AB^{-1/2}+\lambda I\right)^{-1}B^{-1/2} \\
& = B^{1/2}\left(B^{-1/2}AB^{-1/2}\right)\left(B^{-1/2}AB^{-1/2}+\lambda I\right)^{-1}B^{-1/2} \\
& = B^{1/2}\left(B^{-1/2}AB^{-1/2}+\lambda I\right)\left(B^{-1/2}AB^{-1/2}+\lambda I\right)^{-1}B^{-1/2} - B^{1/2}\lambda\left(B^{-1/2}AB^{-1/2}+\lambda I\right)^{-1}B^{-1/2} \\
& = I - \lambda B^{1/2}\left(B^{-1/2}AB^{-1/2}+\lambda I\right)^{-1}B^{-1/2}.
\end{aligned}
$$

□

B.5 Extension: Market-entry threshold with richer form for L_2^*

In this section, we modify the safety requirement to take into account the impact of dataset size N and regularization parameter λ, and we extend our model and analysis of the market-entry threshold accordingly. We show that the characterization in Theorem 5 directly applies to this setting, and we also show relaxed versions of Theorem 8 and Theorem 9. Altogether, these extended results illustrate that our qualitative insights from Chapters 4.3-4.4 hold more generally.

We define a modified approximation of the safety violation $\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$. This modified approximation is defined analogously to $L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$. To formalize this, we define a deterministic equivalent $L_2^{\texttt{det}}$ for the safety violation to be

$$L_2^{\texttt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) := L_1^{\texttt{det}}(\beta_2, \beta_1, \mathcal{D}_F, \lambda, N, 1-\alpha). \tag{B.5}$$

It follows from Lemma 10 that $L_2(\hat{\beta}(\alpha, \lambda, X)) \sim L_2^{\texttt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$: here, we use the fact that $L_2(\hat{\beta}(\alpha, \lambda, X))$ is distributed identically to $L_1(\hat{\beta}(1-\alpha, \lambda, X))$. Now, using this deterministic equivalent, we define $\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) = L_2^{\texttt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$.

Using this formulation of $\tilde{L}_2$, we define a modified market entry threshold where we replace all instances of original approximation L_2^* with the modified approximation $\tilde{L}_2$. In particular, a company C faces reputational damage if:

$$\mathbb{E}_{(\beta_1,\beta_2)\sim\mathcal{D}_W}\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \alpha_C) \geq \tau_C.$$

The company selects $\alpha \in [0.5, 1]$ and $\lambda \in (0, 1)$ to maximize their performance subject to their safety constraint, as formalized by the following optimization program:[1]

$$(\tilde{\alpha}_C, \tilde{\lambda}_C) = \underset{\alpha\in[0.5,1],\lambda\in(0,1)}{\arg\min}\ \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N_C, \alpha)] \text{ s.t. } \mathbb{E}_{\mathcal{D}_W}[\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \alpha)] \leq \tau_C.$$

We define the modified market-entry threshold as follows.

[1] Unlike in Chapter 4.2, there might not exist $\alpha \in [0.5, 1]$ and $\lambda \in (0, 1)$ which satisfy the safety constraint, if N_C is too small.

Definition 19. *The modified market-entry threshold* $\tilde{N}_E^*(N_I, \tau_I, \tau_E, \mathcal{D}_W, \mathcal{D}_F)$ *is the minimum value of* $N_E \in \mathbb{Z}_{\geq 1}$ *such that* $\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_E, N_E, \tilde{\alpha}_E)] \leq \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)]$.

In this section, we analyze the modified market entry threshold $\tilde{N}_E^*(N_I, \tau_I, \tau_E, \mathcal{D}_W, \mathcal{D}_F)$. We show an extension of Theorem 5 (Appendix B.5.1). We then derive a simplified version of the deterministic equivalent $L_2^{\texttt{det}}$ (Appendix B.5.2). Finally, we show a weakened extension of Theorem 8 (Appendix B.5.3) and a weakened extension of Theorem 9 (Appendix B.5.4). These weakened extensions derive upper bounds (rather than tight bounds) on the modified market entry threshold, and also assume that $\delta \leq 1$.

B.5.1 Extension of Theorem 5

We study the market entry $\tilde{N}_E^*$ threshold in the environment of Theorem 5 where the incumbent has infinite data and the new company faces no safety constraint. We show that the modified market entry threshold takes the same form as the market entry threshold in Theorem 5.

Theorem 166 (Extension of Theorem 5)**.** *Suppose that power-law scaling holds for the eigenvalues and alignment coefficients, with scaling exponents* $\gamma, \delta > 0$ *and correlation coefficient* $\rho \in [0, 1)$*, and suppose that* $P = \infty$*. Suppose that the incumbent company has infinite data (i.e.,* $N_I = \infty$*), and that the entrant faces no constraint on their safety (i.e.,* $\tau_E = \infty$*). Suppose that the safety constraint* τ_I *satisfies* (4.1)*. Then, it holds that:*

$$\tilde{N}_E^*(\infty, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F) = \Theta\left(\left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^{-2/\nu}\right),$$

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)] = \Theta(1 - \rho)$*, and where* $\nu := \min(2(1 + \gamma), \delta + \gamma)$*.*

Theorem 166 shows that the qualitative insights from Theorem 5—including that the new company can enter with finite data—readily extend to this setting.

To prove Theorem 166, we build on the notation and analysis from Appendix B.1. It suffices to show that each company C will select $\alpha_C = \tilde{\alpha}_C$ and $\lambda_C = \tilde{\lambda}_C$. This follows trivially for the entrant $C = E$ since they face no safety constraint, and there is no different between the two settings. The key ingredient of the proof is to compute $\tilde{\alpha}_I$ and $\tilde{\lambda}_I$ for the incumbent (i.e., an analogue of Lemma 136 in Appendix B.1).

To do this, we first upper bound the following function of the safety loss and performance loss for general parameters λ and α.

Lemma 167. *For any* α *and* λ*, it holds that:*

$$\sqrt{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, \lambda))]} + \sqrt{\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, \lambda))]} \geq \sqrt{\mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma (\beta_1 - \beta_2)^T]}.$$

Proof. Note that:

$$\begin{aligned}
T &:= \sqrt{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha,\lambda))]} + \sqrt{\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha,\lambda))]} \\
&= \sqrt{(\beta_1 - \beta(\alpha,\lambda))^T \Sigma(\beta_1 - \beta(\alpha,\lambda))} + \sqrt{(\beta_2 - \beta(\alpha,\lambda))^T \Sigma(\beta_2 - \beta(\alpha,\lambda))} \\
&= \sqrt{(\lambda\beta_1 + (1-\alpha)\Sigma(\beta_1 - \beta_2))^T \Sigma(\Sigma + \lambda I)^{-2}(\lambda\beta_1 + (1-\alpha)\Sigma(\beta_1 - \beta_2))} \\
&+ \sqrt{(\lambda\beta_2 + \alpha\Sigma(\beta_2 - \beta_1))^T \Sigma(\Sigma + \lambda I)^{-2}(\lambda\beta_2 + \alpha\Sigma(\beta_2 - \beta_1))} \\
&= \sqrt{(\lambda\beta_1 + (1-\alpha)\Sigma(\beta_1 - \beta_2))^T \Sigma(\Sigma + \lambda I)^{-2}(\lambda\beta_1 + (1-\alpha)\Sigma(\beta_1 - \beta_2))} \\
&+ \sqrt{(-\lambda\beta_2 + \alpha\Sigma(\beta_1 - \beta_2))^T \Sigma(\Sigma + \lambda I)^{-2}(-\lambda\beta_2 + \alpha\Sigma(\beta_1 - \beta_2))}.
\end{aligned}$$

Now note that for any PSD matrix Σ' and any distribution, note that the following triangle inequality holds:

$$\sqrt{\mathbb{E}[(X_1 + X_2)^T \Sigma'(X_1 + X_2)]} \le \sqrt{\mathbb{E}[X_1^T \Sigma' X_1]} + \sqrt{\mathbb{E}[X_2^T \Sigma' X_2]}.$$

We apply this for $X_1 = \lambda\beta_1 + (1-\alpha)\Sigma(\beta_1 - \beta_2)$, $X_2 = -\lambda\beta_2 + \alpha\Sigma(\beta_1 - \beta_2)$, and distribution $\mathcal{D}_W$. This means that we can lower bound:

$$\begin{aligned}
T &\ge \sqrt{\mathbb{E}_{\mathcal{D}_W}[((\Sigma + \lambda I)(\beta_1 - \beta_2))^T \Sigma(\Sigma + \lambda I)^{-2}((\Sigma + \lambda I)(\beta_1 - \beta_2))]} \\
&= \sqrt{\mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2))^T \Sigma(\beta_1 - \beta_2))]}
\end{aligned}$$

as desired. □

Now, we are ready to compute $\tilde{\alpha}_I$ and $\tilde{\lambda}_I$ for the incumbent.

Lemma 168. *Let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta_2)^T]$. Suppose that $N_I = \infty$, and suppose that the safety constraint τ_I satisfies (4.1). Then it holds that $\alpha_I = \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}}$, and $\lambda_I = 0$ is optimal for the incumbent. Moreover, it holds that:*

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, \infty, \tilde{\alpha}_I)] = \left(\sqrt{L^*(\rho)} - \sqrt{\min(L^*(\rho), \tau_I)}\right)^2.$$

Proof. First, we apply Lemma 170 with $N = \infty$ to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, \infty, \alpha)] = \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha, \lambda))]$$

and

$$\mathbb{E}_{\mathcal{D}_W}[L_2^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, \infty, \alpha)] = \mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha, \lambda))].$$

Let $\alpha^* = \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}}$. By the assumption in the lemma statement, we know that:

$$\alpha^* \ge \sqrt{\frac{\mathbb{E}_{\mathcal{D}_W}[L_2^*(\beta_1, \beta_2, \mathcal{D}_F, 0.5)]}{L^*(\rho)}} = 0.5.$$

Observe that:

$$\begin{aligned}
&\sqrt{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha^*,0))]} + \sqrt{\min(\tau_I, L^*(\rho))} \\
&= \sqrt{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha^*,0))]} + \sqrt{\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha^*,0))]} \\
&= \sqrt{(1-\alpha^*)^2\mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)^T]} + \sqrt{(\alpha^*)^2\mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)^T]} \\
&= \sqrt{\mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)^T]}
\end{aligned}$$

We show that $(\tilde{\alpha}_I, \tilde{\lambda}_I) = (\alpha^*, 0)$. Assume for sake of contradiction that $(\alpha, \lambda) \neq (\alpha^*, 0)$ satisfies the safety constraint $\mathbb{E}_{\mathcal{D}_W}[\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \alpha)] \leq \min(\tau_I, L^*(\rho))$ and achieves strictly better performance loss:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, \infty, \alpha)] < \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, 0, \infty, \alpha^*)].$$

Then it would hold that:

$$\begin{aligned}
\sqrt{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha,\lambda))]} + \sqrt{\mathbb{E}_{\mathcal{D}_W}[L_2(\beta(\alpha,\lambda))]} &< \sqrt{\mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\alpha^*,0))]} + \sqrt{\min(\tau_I, L^*(\rho))} \\
&= \sqrt{\mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)^T]},
\end{aligned}$$

which contradicts Lemma 167.

To analyze the loss, note that:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, \infty, \tilde{\alpha}_I)] \\
&\qquad = \mathbb{E}_{\mathcal{D}_W}[L_1(\beta(\tilde{\alpha}_I, \tilde{\lambda}_I))] \\
&\qquad = (1-\tilde{\alpha}_I)^2 L^*(\rho) \\
&\qquad = (\sqrt{L^*(\rho)} - \sqrt{\min(L^*(\rho), \tau_I)})^2
\end{aligned}$$

□

We now prove Theorem 166.

Proof of Theorem 166. We analyze $(\tilde{\alpha}_C, \tilde{\lambda}_C)$ first for the incumbent $C = I$ and then for the entrant $C = E$.

Analysis of the incumbent $C = I$. By Lemma 168, we see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, \infty, \tilde{\alpha}_I)] = \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2.$$

Analysis of the entrant $C = E$. This analysis follows identically to the analogous case in the proof of Theorem 5, and we repeat the proof for completeness. Since the entrant faces no safety constraint, the entrant can choose any $\alpha \in [0.5, 1]$. We apply Corollary 12 to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda_E, N, \alpha_E)] = \inf_{\alpha\in[0.5,1]} \inf_{\lambda>0} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)] = \Theta\left(N^{-\nu}\right),$$

which means that:

$$N_E^*(\infty, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F) = \Theta\left(\left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^{-2/\nu}\right)$$

as desired. We can further apply Claim 137 to see that $L^*(\rho) = \Theta(1-\rho)$. □

B.5.2 Bounds on the excess loss for safety

We bound the excess loss $\alpha^2 L^*(\rho) - \mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}]$. We assume that $\alpha \geq 0.5$ and we further assume that $\delta \leq 1$.

Lemma 169. *Suppose that power scaling holds for the eigenvalues and alignment coefficients with scaling $\gamma > 0$ and $\delta \in (0,1]$, and correlation coefficient $\rho \in [0,1)$, and suppose that $P = \infty$. Suppose that $\alpha \geq 0.5$, $\lambda \in (0,1)$, and $N \geq 1$. Let $L_2^{\mathtt{det}} := L_2^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ be defined according to (B.5). Let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta_2)]$. Then it holds that:*

$$\alpha^2 L^*(\rho) - \mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}] = O\left(\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu})\right)$$

and

$$\mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}] - \alpha^2 L^*(\rho) = O\left(\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) + (1-\alpha)(1-\rho)\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right),$$

where $\nu = \min(2(1+\gamma), \delta + \gamma) = \delta + \gamma$.

To prove Lemma 169, we first simplify the deterministic equivalent $L_2^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ using the assumptions from Chapter 4.2.3.

Lemma 170. *Suppose that power scaling holds for the eigenvalues and alignment coefficients with scaling $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, and suppose that $P = \infty$. Suppose that $\lambda \in (0,1)$, and $N \geq 1$. Let $L_2^{\mathtt{det}} := L_2^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)$ be defined according to (B.5). Let $\kappa = \kappa(\lambda, N, \Sigma)$ from Definition 18. Let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta_2)]$. Then it holds that:*

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}] - L^*(\rho) &= Q^{-1} \cdot \kappa^2 \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + Q^{-1} 2\kappa\alpha(1-\alpha)(1-\rho) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&\quad + Q^{-1} 2\alpha(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \cdot \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&\quad - 2\alpha^2\kappa(1-\rho) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa},
\end{aligned}$$

where $Q = 1 - \frac{1}{N}\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}$.

Proof. First, we apply Lemma 150, coupled with the fact that $L_2^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha) := L_1^{\mathtt{det}}(\beta_2, \beta_1, \mathcal{D}_F, \lambda, N, 1-\alpha)$, to see that:

$$\begin{aligned}
Q \cdot \mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}] = \kappa^2(1 - 2\alpha^2(1-\rho)) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + \alpha^2 L^*(\rho) \\
+ 2\kappa(1-\rho)\alpha(1-2\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
+ 2\alpha(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \cdot (1-2\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa},
\end{aligned}$$

where $Q = 1 - \frac{1}{N}\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}$. Using that $(Q^{-1}-1)\alpha^2 L^*(\rho) = Q^{-1}\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) 2\alpha^2(1-\rho)\left(\sum_{i=1}^{P} i^{-\delta-1-\gamma}\right)$, this means that:

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}] - \alpha^2 L^*(\rho) = Q^{-1}\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) 2\alpha^2(1-\rho)\left(\sum_{i=1}^{P} i^{-\delta-1-\gamma}\right) \\
+ Q^{-1} \cdot \kappa^2(1-2\alpha^2(1-\rho)) \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
+ Q^{-1} 2\kappa(1-\rho)\alpha(1-2\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
+ Q^{-1} 2\alpha(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right) \cdot (1-2\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa}
\end{aligned}$$

By expanding some of these terms, we see that:

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}] - \alpha^2 L^*(\rho) \\
&= Q^{-1} 2\alpha^2 (1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \cdot \sum_{i=1}^{P} i^{-\delta-1-\gamma} \\
&+ Q^{-1} \cdot \kappa^2 \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} - Q^{-1} 2\alpha^2 (1-\rho) \cdot \kappa^2 \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1} 2\kappa (1-\rho) \alpha (1-\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} - Q^{-1} 2\kappa (1-\rho) \alpha^2 \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1} 2\alpha (1-\alpha)(1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \cdot \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&- Q^{-1} 2\alpha^2 (1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \cdot \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa}.
\end{aligned}
$$

When we collect terms, we obtain:

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}] - \alpha^2 L^*(\rho) \\
&= Q^{-1} \cdot \kappa^2 \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + Q^{-1} 2\kappa (1-\rho) \alpha (1-\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1} 2\alpha (1-\alpha)(1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \cdot \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&- Q^{-1} 2\kappa (1-\rho) \alpha^2 \left(\sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} + \sum_{i=1}^{P} \frac{\kappa \cdot i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \\
&+ Q^{-1} 2\alpha^2 (1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \cdot \left(\sum_{i=1}^{P} i^{-\delta-1-\gamma} - \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \right) \\
&= Q^{-1} \cdot \kappa^2 \sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} + Q^{-1} 2\kappa (1-\rho) \alpha (1-\alpha) \sum_{i=1}^{P} \frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1} 2\alpha (1-\alpha)(1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \cdot \sum_{i=1}^{P} \frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&- Q^{-1} 2\kappa (1-\rho) \alpha^2 \left(\sum_{i=1}^{P} \frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)} \right) \\
&+ Q^{-1} 2\kappa \alpha^2 (1-\rho) \frac{1}{N} \left(\sum_{i=1}^{P} \frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2} \right) \cdot \frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa}.
\end{aligned}
$$

Combining the last two terms gives us the desired statement. □

Now, we are ready to prove Lemma 169.

Proof. For the first bound, we observe that:

$$\begin{aligned}
&\alpha^2 L^*(\rho) - \mathbb{E}_{\mathcal{D}_W}[L_2^{\texttt{det}}] \\
&\leq_{(A)} 2\alpha^2\kappa(1-\rho)\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{i^{-1-\gamma}+\kappa} \\
&=_{(B)} O\left(\alpha^2(1-\rho)\kappa^{\frac{\min(1+\gamma,\delta+\gamma)}{1+\gamma}}\right) \\
&=_{(C)} O\left(\kappa^{\frac{\nu}{1+\gamma}}\right) \\
&=_{(D)} O\left(\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu})\right)
\end{aligned}$$

where (A) uses Lemma 170, (B) uses Lemma 157, (C) uses that $\delta \leq 1$ and $\rho \in [0, 1)$, and (D) uses Lemma 159.

For the second bound, we observe that:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_W}[L_2^{\texttt{det}}] - \alpha^2 L^*(\rho) \\
&\leq_{(A)} Q^{-1}\cdot\kappa^2\sum_{i=1}^{P}\frac{i^{-\delta-1-\gamma}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1}2\kappa\alpha(1-\alpha)(1-\rho)\sum_{i=1}^{P}\frac{i^{-\delta-2(1+\gamma)}}{(i^{-1-\gamma}+\kappa)^2} \\
&+ Q^{-1}2\alpha(1-\alpha)(1-\rho)\frac{1}{N}\left(\sum_{i=1}^{P}\frac{i^{-2-2\gamma}}{(i^{-1-\gamma}+\kappa)^2}\right)\cdot\sum_{i=1}^{P}\frac{i^{-\delta-2-2\gamma}}{i^{-1-\gamma}+\kappa} \\
&=_{(B)} O\left(\kappa^{\frac{\min(2(1+\gamma),\gamma+\delta)}{1+\gamma}} + \alpha(1-\alpha)(1-\rho)\kappa^{\frac{\min(1+\gamma,\gamma+\delta)}{1+\gamma}} + \alpha(1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right) \\
&=_{(C)} O\left(\kappa^{\frac{\gamma+\delta}{1+\gamma}} + (1-\alpha)(1-\rho)\kappa^{\frac{\gamma+\delta}{1+\gamma}} + (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right) \\
&= O\left(\kappa^{\frac{\gamma+\delta}{1+\gamma}} + (1-\alpha)(1-\rho)\frac{\kappa^{-\frac{1}{1+\gamma}}}{N}\right) \\
&=_{(D)} O\left(\max(\lambda^{\frac{\nu}{1+\gamma}}, N^{-\nu}) + (1-\alpha)(1-\rho)\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N)}{N}\right)
\end{aligned}$$

where (A) uses Lemma 170, (B) uses Lemma 157 and Lemma 158, (C) uses that $\delta \leq 1$ and $\alpha \geq 0.5$, and (D) uses Lemma 159. □

B.5.3 Extension of Theorem 8

We next study the market entry $\tilde{N}_E^*$ threshold in the environment of Theorem 8 where the incumbent has *finite data* and the new company faces no safety constraint. We place the further assumption that $\delta \leq 1$. We compute the following upper bound on the modified market entry threshold.

Theorem 171 (Extension of Theorem 8)**.** *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, and suppose that $P = \infty$. Assume that $\tau_E = \infty$. Suppose that the safety constraint τ_I satisfies (4.1). Then we have that $\tilde{N}_E^* = \tilde{N}_E^*(N_I, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F)$ satisfies:*

$$\tilde{N}_E^* := \begin{cases} O\left(N_I\right) & \text{if } N_I \leq \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ O\left(N_I^{\frac{1}{\nu+1}} \cdot \tilde{G}_I^{-\frac{1}{2(\nu+1)}}(1-\rho)^{-\frac{1}{2(\nu+1)}}\right) & \text{if } \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ O\left(\tilde{G}_I^{-\frac{1}{\nu}}\right) & \text{if } N_I \geq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}}, \end{cases}$$

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)] = \Theta(1-\rho)$, where $\alpha^* = \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}}$, where $\tilde{\alpha} := \sqrt{(1-\alpha^*)+(\alpha^*)^2}$, where $\tilde{G}_I = (1-\tilde{\alpha})^2(1-\rho)$, and where $\nu = \min(2(1+\gamma), \gamma+\delta) = \gamma+\delta$.

Theorem 171 shows that the key qualitative finding from Theorem 8—that the new company can enter with $N_E = o(N_I)$ data as long as the incumbent's dataset size is sufficiently large—readily extends to this setting. We note that the bound in Theorem 171 and the bound in Theorem 8 take slightly different forms: the term $G_I = (\sqrt{L^*(\rho)} - \sqrt{\min(L^*(\rho), \tau_I)})^2 = \Theta((1-\alpha^*)^2(1-\rho))$ is replaced by $\tilde{G}_I = (1-\tilde{\alpha})^2(1-\rho)$. We expect some of these differences arise because the bound in Theorem 171 is not tight, rather than fundamental distinctions between the two settings. Proving a tight bound on the modified market entry threshold is an interesting direction for future work.

To prove this, we compute a lower bound on the incumbent's loss.

Lemma 172. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0,1)$, and suppose that $P = \infty$. Assume that $\tau_E = \infty$. Suppose that the safety constraint τ_I satisfies (4.1). Then we have that:*

$$\begin{aligned} &\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \\ &= \begin{cases} \Omega\left(N_I^{-\nu}\right) & \text{if } N_I \leq \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ \Omega\left(N_I^{-\frac{\nu}{\nu+1}} \cdot \tilde{G}_I^{\frac{\nu}{2(\nu+1)}}(1-\rho)^{\frac{\nu}{2(\nu+1)}}\right) & \text{if } \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ \Omega\left(\tilde{G}_I\right) & \text{if } N_I \geq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}}. \end{cases} \end{aligned}$$

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1-\beta_2)^T\Sigma(\beta_1-\beta_2)] = \Theta(1-\rho)$, where $\alpha^* = \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}}$, where $\tilde{\alpha} := \sqrt{(1-\alpha^*)+(\alpha^*)^2}$, where $\tilde{G}_I = (1-\tilde{\alpha})^2(1-\rho)$ and where $\nu = \min(2(1+\gamma), \gamma+\delta) = \gamma+\delta$.

Proof. By Corollary 12 and Lemma 159, we know that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] = \Omega(\kappa^{\frac{\nu}{1+\gamma}}) = \Omega(\max(\lambda^{\frac{\nu}{1+\gamma}}, N_I^{-\nu})).$$

Let $C_{\delta,\gamma}$ be an implicit constant[2] such that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \geq C_{\delta,\gamma} \max(\lambda^{\frac{\nu}{1+\gamma}}, N_I^{-\nu}) \tag{B.6}$$

By Lemma 169, there also exists an implicit constant $C'_{\delta,\gamma}$ such that:

$$\alpha^2 L^*(\rho) - \mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N_I, \alpha)] \leq C'_{\delta,\gamma} \max(\lambda^{\frac{\nu}{1+\gamma}}, N_I^{-\nu}). \tag{B.7}$$

We now split into two cases: (1) $\frac{C'_{\delta,\gamma}}{C_{\delta,\gamma}} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \geq (1-\alpha^*)L^*(\rho)$, and (2) $\frac{C'_{\delta,\gamma}}{C_{\delta,\gamma}} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \leq (1-\alpha^*)L^*(\rho)$.

Case 1: $\frac{C'_{\delta,\gamma}}{C_{\delta,\gamma}} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \geq (1-\alpha^*)L^*(\rho)$**.** It follows from (B.6) that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \geq C_{\delta,\gamma} \max(\lambda^{\frac{\nu}{1+\gamma}}, N_I^{-\nu}) \geq C_{\delta,\gamma} N_I^{-\nu}.$$

Using the condition for this case, this implies that:

$$\begin{aligned}
N_I &\leq \left(\frac{1}{C_{\delta,\gamma}} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \right)^{-\frac{1}{\nu}} \\
&\leq \left(\frac{1}{C'_{\delta,\gamma}} (1-\alpha^*)L^*(\rho) \right)^{-\frac{1}{\nu}} \\
&= O\left(((1-\tilde{\alpha})(1-\rho))^{-\frac{1}{\nu}} \right) \\
&= O\left(\tilde{G}_I^{-\frac{1}{2\nu}} (1-\rho)^{-\frac{1}{2\nu}} \right).
\end{aligned}$$

This proves that N_I is up to constants within the first branch of the expression in the lemma statement. Since the bound in the lemma statement only changes by constants (that depend on δ and γ) between the first branch and second branch, this proves the desired expression for this case.

Case 2: $\frac{C'_{\delta,\gamma}}{C_{\delta,\gamma}} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \leq (1-\alpha^*)L^*(\rho)$**.** Note that $\alpha^* = \sqrt{\frac{\min(\tau_I, L^*(\rho))}{L^*(\rho)}}$ is the mixture parameter that achieves the safety constraint in the infinite-data ridgeless setting. The incumbent's safety constraint means that:

$$\mathbb{E}_{\mathcal{D}_W}[L_2^{\mathtt{det}}(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \leq (\alpha^*)^2 L^*(\rho).$$

[2]We need to introduce an implicit constant because of $O()$ is permitted to hide constants that depend on δ and γ.

By (B.7), this implies that

$$(\tilde{\alpha}_I)^2 L^*(\rho) \leq C'_{\delta,\gamma} \cdot \max(\lambda^{\frac{\delta+\gamma}{1+\gamma}}, N_I^{-\delta-\gamma}) + (\alpha^*)^2 L^*(\rho).$$

Now, applying (B.6) and the assumption for this case, we see that:

$$\begin{aligned}(\tilde{\alpha}_I)^2 L^*(\rho) &\leq \frac{C'_{\delta,\gamma}}{C_{\delta,\gamma}} \cdot \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] + (\alpha^*)^2 L^*(\rho) \\ &\leq (1-\alpha^*)L^*(\rho) + (\alpha^*)^2 L^*(\rho).\end{aligned}$$

This implies that:

$$\tilde{\alpha}_I \leq \sqrt{(1-\alpha^*) + (\alpha^*)^2}.$$

Let $\tilde{\alpha} := \sqrt{(1-\alpha^*) + (\alpha^*)^2}$. Plugging this into Corollary 12, we see that:

$$\begin{aligned}&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \\ &\geq \inf_{\alpha \in [0.5, \tilde{\alpha}]} \inf_{\lambda > 0} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N_I, \alpha)] \\ &= \Theta\left(\inf_{\lambda>0} \mathbb{E}_{\mathcal{D}_W}\left[L_1^*\left(\beta_1, \beta_2, \Sigma, \lambda, N_I, \tilde{\alpha}\right)\right]\right) \\ &= \begin{cases} \Theta\left(N_I^{-\nu}\right) & \text{if } N_I \leq (1-\tilde{\alpha})^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \Theta\left(\left(\frac{N_I}{(1-\tilde{\alpha})(1-\rho)}\right)^{-\frac{\nu}{\nu+1}}\right) & \text{if } (1-\tilde{\alpha})^{-\frac{1}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \leq N_I \leq (1-\tilde{\alpha})^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}} \\ \Theta((1-\tilde{\alpha})^2(1-\rho)) & \text{if } N_I \geq (1-\tilde{\alpha})^{-\frac{2+\nu}{\nu}}(1-\rho)^{-\frac{1}{\nu}}, \end{cases} \\ &= \begin{cases} \Theta\left(N_I^{-\nu}\right) & \text{if } N_I \leq \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ \Theta\left(N_I^{-\frac{\nu}{\nu+1}} \cdot \tilde{G}_I^{\frac{\nu}{2(\nu+1)}}(1-\rho)^{\frac{\nu}{2(\nu+1)}}\right) & \text{if } \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ \Theta\left(\tilde{G}_I\right) & \text{if } N_I \geq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}}. \end{cases}\end{aligned}$$

The statement follows in this case. □

We are now ready to prove Theorem 171.

Proof of Theorem 171. We analyze $(\tilde{\alpha}_C, \tilde{\lambda}_C)$ first for the incumbent $C = I$ and then for the entrant $C = E$. Like in the theorem statement, let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta_2)] = \Theta(1-\rho)$ (Claim 137) and $G_I := (\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))})^2$, and $\nu = \min(2(1+\gamma), \delta+\gamma)$.

Analysis of the incumbent $C = I$. We apply Lemma 172 to see that:

$$\begin{aligned}&\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, N_I, \tilde{\alpha}_I)] \\ &= \begin{cases} \Omega\left(N_I^{-\nu}\right) & \text{if } N_I \leq \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ \Omega\left(N_I^{-\frac{\nu}{\nu+1}} \cdot \tilde{G}_I^{\frac{\nu}{2(\nu+1)}}(1-\rho)^{\frac{\nu}{2(\nu+1)}}\right) & \text{if } \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ \Omega\left(\tilde{G}_I\right) & \text{if } N_I \geq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}}. \end{cases}\end{aligned}$$

Analysis of the entrant $C = E$. Since the entrant faces no safety constraint, the entrant can choose any $\alpha \in [0.5, 1]$. We apply Corollary 11 to see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_E, N, \tilde{\alpha}_E)] = \inf_{\alpha \in [0.5,1]} \inf_{\lambda > 0} \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \lambda, N, \alpha)] = \Theta\left(N^{-\nu}\right),$$

which means that $N_E^*(N_I, \tau_I, \infty, \mathcal{D}_W, \mathcal{D}_F)$ equals:

$$\begin{cases} O\left(N_I\right) & \text{if } N_I \leq \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \\ O\left(N_I^{\frac{1}{\nu+1}} \cdot \tilde{G}_I^{-\frac{1}{2(\nu+1)}}(1-\rho)^{-\frac{1}{2(\nu+1)}}\right) & \text{if } \tilde{G}_I^{-\frac{1}{2\nu}}(1-\rho)^{-\frac{1}{2\nu}} \leq N_I \leq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \\ O\left(\tilde{G}_I^{-\frac{1}{\nu}}\right) & \text{if } N_I \geq \tilde{G}_I^{-\frac{1}{2}-\frac{1}{\nu}}(1-\rho)^{\frac{1}{2}} \end{cases}$$

as desired.

□

B.5.4 Extension of Theorem 9

We next study the market entry $\tilde{N}_E^*$ threshold in the environment of Theorem 9 where the incumbent has infinite data and the new company faces a *nontrivial safety constraint*. We place the further assumption that $\delta \leq 1$. We compute the following upper bound on the modified market entry threshold.

Theorem 173 (Extension of Theorem 9)**.** *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, and suppose that $P = \infty$. Suppose that the safety constraints τ_I and τ_E satisfy (4.2). Then it holds that $\tilde{N}_E^* = \tilde{N}_E^*(\infty, \tau_I, \tau_E, \mathcal{D}_W, \mathcal{D}_F)$ satisfies:*

$$\tilde{N}_E^* := O\left(\max\left(\tilde{D}^{-\frac{1}{\nu}}, \tilde{D}^{-\frac{\nu+1}{\nu}}\left(G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} + \frac{1}{2}G_I - \frac{1}{2}G_E\right)\right)\right),$$

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta)] = \Theta(1-\rho)$, where $\nu = \min(2(1+\gamma), \delta + \gamma) = \delta + \gamma$, where $G_I := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2$ and $G_E := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_E, L^*(\rho))}\right)^2$, and where:

$$\tilde{D} := \alpha_E^* \cdot (G_I - G_E) - \frac{(G_I - G_E)^2}{4 \cdot L^*(\rho)}.$$

Theorem 173 shows that the key qualitative finding from Theorem 9—that the new company can enter with finite data, as long as they face a strictly weaker safety constraint than the incumbent company—readily extends to this setting. We note that the bound in Theorem 173 and the bound in Theorem 9 take slightly different forms. Some of these differences are superficial: while the bound in Theorem 173 contains two—rather than three—regimes, the third regime in Theorem 9 does not exist in the case where $\delta \leq 1$. Other differences are more substantial: for example, the bound in Theorem 173 scales with $\tilde{D}$ while the bound in Theorem 9 scales with D. However, we expect some of this difference arises because the bound in Theorem 173 is not tight, rather than fundamental distinctions between the two settings. Proving a tight bound on the modified market entry threshold is an interesting direction for future work.

We compute an upper bound on the number of data points N_E that the new company needs to achieve at most loss $\left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2$ on performance.

Lemma 174. *Suppose that the power-law scaling assumptions from Chapter 4.2.3 hold with exponents $\gamma, \delta > 0$ and correlation coefficient $\rho \in [0, 1)$, and suppose that $P = \infty$. Suppose that the safety constraints τ_I and τ_E satisfy (4.1). For sufficiently large constant $C_{\delta,\gamma}$, if*

$$N_E \geq C_{\delta,\gamma} \cdot \max\left(\tilde{D}^{-\frac{1}{\nu}}, \tilde{D}^{-\frac{\nu+1}{\nu}}\left(G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} + \frac{1}{2}G_I - \frac{1}{2}G_E\right)\right),$$

then it holds that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_E, N_E, \tilde{\alpha}_E)] \leq G_I,$$

where $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta)] = \Theta(1-\rho)$, where $\nu = \min(2(1+\gamma), \delta+\gamma) = \delta + \gamma$, where $G_I := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2$ and $G_E := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_E, L^*(\rho))}\right)^2$, and where:

$$\tilde{D} := \alpha_E^* \cdot (G_I - G_E) - \frac{(G_I - G_E)^2}{4 \cdot L^*(\rho)}.$$

Proof. It suffices to construct $\tilde{\alpha}$ and $\tilde{\lambda}$ such that

$$\mathbb{E}_{\mathcal{D}_W}[\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})] \leq \tau_E$$

and

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})] \leq G_I$$

for $N_E = \Omega\left(\max\left(\tilde{D}^{-\frac{1}{\nu}}, \tilde{D}^{-\frac{\nu+1}{\nu}}\left(G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} + \frac{1}{2}G_I - \frac{1}{2}G_E\right)\right)\right)$.

To define $\tilde{\alpha}$ and $\tilde{\lambda}$, it is inconvenient to work with the following intermediate quantities. Let $\alpha_E^* = \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_E, L^*(\rho))}\right)^2$ and let $\alpha_I^* = \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2$. We define an error function:

$$f(N_E, \alpha, \lambda) := \max(\lambda^{\frac{\nu}{1+\gamma}}, N_E^{-\nu}) + (1-\alpha)(1-\rho)\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N_E)}{N_E}$$

We define:

$$\tilde{\alpha} := \alpha_E^* + \frac{1}{2}(1-\alpha_E^*)^2 - \frac{1}{2}(1-\alpha_I^*)^2 = \alpha_I^* + \frac{\alpha_E^* - \alpha_I^*}{2}.$$

and

$$\tilde{\lambda} := \inf_{\lambda \in (0,1)} f(N_E, \tilde{\alpha}, \lambda).$$

At these values of $\tilde{\alpha}$ and $\tilde{\lambda}$ and under the condition on N_E, observe that:

$$\begin{aligned} f(N_E, \tilde{\alpha}, \tilde{\lambda}) &= \Theta\left(\max\left(N_E^{-\nu}, \left(\frac{N_E}{(1-\tilde{\alpha})(1-\rho)}\right)^{-\frac{\nu}{\nu+1}}\right)\right) \\ &= \Theta\left(\max\left(N_E^{-\nu}, \left(\frac{N_E}{G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} + \frac{1}{2}G_I + \frac{1}{2}G_E}\right)^{-\frac{\nu}{\nu+1}}\right)\right) \\ &= O\left(\tilde{D}\right), \end{aligned}$$

where the implicit constant can be reduced by increasing the implicit constant on N_E.

The remainder of the analysis boils down to showing that $\mathbb{E}_{\mathcal{D}_W}[\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})] \leq \tau_E$ and $\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})] \leq G_I$. To show this, we first derive an error function and bound these losses in terms of the error function.

Bounding $\mathbb{E}_{\mathcal{D}_W}[\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})] \leq \tau_E$. Observe that:

$$\begin{aligned} &\mathbb{E}_{\mathcal{D}_W}[\tilde{L}_2(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})] \\ &=_{(A)} \tilde{\alpha}^2 L^*(\rho) + O\left(\max(\lambda^{\frac{\nu}{1+\gamma}}, N_E^{-\nu}) + (1-\alpha)(1-\rho)\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N_E)}{N_E}\right) \\ &= (\alpha_E^* + \frac{1}{2}(1-\alpha_E^*)^2 - \frac{1}{2}(1-\alpha_I^*)^2)L^*(\rho) + O\left(f(N_E, \tilde{\alpha})\right) \\ &\leq \left((\alpha_E^*)^2 L^*(\rho) + \frac{((1-\alpha_I^*)^2 - (1-\alpha_E^*)^2)^2}{4} - \alpha_E^*((1-\alpha_I^*)^2 - (1-\alpha_E^*)^2)\right) L^*(\rho) + \tilde{D} \\ &= \tau_E + \frac{(G_I - G_E)^2}{4 \cdot L^*(\rho)} - \alpha_E^*(G_I - G_E)\alpha_E^* \cdot (G_I - G_E) - \frac{(G_I - G_E)^2}{4 \cdot L^*(\rho)} \\ &= \tau_E \end{aligned}$$

where (A) follows from Lemma 169. This gives us the desired bound.

Bounding $\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})]$. Observe that:

$$\begin{aligned} &\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}, N_E, \tilde{\alpha})] \\ &=_{(A)} (1-\tilde{\alpha})^2 L^*(\rho) + O\left(\max(\lambda^{\frac{\nu}{1+\gamma}}, N_E^{-\nu}) + (1-\alpha)(1-\rho)\frac{\min(\lambda^{-\frac{1}{1+\gamma}}, N_E)}{N_E}\right) \\ &\leq (1 - \alpha_E^* - \frac{1}{2}(1-\alpha_E^*)^2 + \frac{1}{2}(1-\alpha_I^*)^2)^2 L^*(\rho) + O\left(f(N_E, \tilde{\alpha})\right) \\ &\leq \left((1-\alpha_E^*)^2 + \frac{((1-\alpha_I^*)^2 - (1-\alpha_E^*)^2)^2}{4} - (1-\alpha_E^*)((1-\alpha_I^*)^2 - (1-\alpha_E^*)^2)\right) L^*(\rho) + \tilde{D} \\ &\leq G_E + (G_I - G_E)(1-\alpha_E^*) + \frac{(G_I - G_E)^2}{4L^*(\rho)} + \alpha_E^* \cdot (G_I - G_E) - \frac{(G_I - G_E)^2}{4 \cdot L^*(\rho)} \\ &= G_I. \end{aligned}$$

where (A) uses Theorem 13, coupled with the fact that $\delta \leq 1$ (which means that $\nu' = \nu$, so the mixture finite data error is subsumed by the finite data error) and coupled with Lemma 159. This gives us the desired bound.

□

We are now ready to prove Theorem 173.

Proof of Theorem 173. We analyze $(\tilde{\alpha}_C, \tilde{\lambda}_C)$ first for the incumbent $C = I$ and then for the entrant $C = E$. Like in the theorem statement, let $L^*(\rho) = \mathbb{E}_{\mathcal{D}_W}[(\beta_1 - \beta_2)^T \Sigma(\beta_1 - \beta)] = \Theta(1-\rho)$, let $\nu = \min(2(1+\gamma), \delta + \gamma) = \delta + \gamma$, let $G_I := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho))}\right)^2$ and $G_E := \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_E, L^*(\rho))}\right)^2$, and let:

$$\tilde{D} := \alpha_E^* \cdot (G_I - G_E) - \frac{(G_I - G_E)^2}{4 \cdot L^*(\rho)}.$$

Analysis of the incumbent $C = I$. To compute $\tilde{\alpha}_I$ and $\tilde{\lambda}_I$, we apply Lemma 168. The assumption $\tau_I \geq \mathbb{E}_{\mathcal{D}_W}[L_2(\beta_1, \beta_2, \Sigma, 0.5)]$ in the lemma statement can be rewritten as $\tau_I \geq 0.25 L^*(\rho)$, which guarantees the assumptions in Lemma 168 are satisfied. By Lemma 168, we see that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, \infty, \tilde{\alpha}_I)] = \left(\sqrt{L^*(\rho)} - \sqrt{\min(\tau_I, L^*(\rho)}\right)^2 = G_I.$$

Analysis of the entrant $C = E$. We apply Lemma 174 to see for sufficiently large constant $C_{\delta,\gamma}$, if

$$N_E \geq C_{\delta,\gamma} \cdot \max\left(\tilde{D}^{-\frac{1}{\nu}}, \tilde{D}^{-\frac{\nu+1}{\nu}}\left(G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} + \frac{1}{2}G_I - \frac{1}{2}G_E\right)\right),$$

then it holds that:

$$\mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_E, N_E, \tilde{\alpha}_E)] \leq G_I = \mathbb{E}_{\mathcal{D}_W}[L_1^*(\beta_1, \beta_2, \mathcal{D}_F, \tilde{\lambda}_I, \infty, \tilde{\alpha}_I)].$$

This means that:

$$\tilde{N}_E^* = O\left(\max\left(\tilde{D}^{-\frac{1}{\nu}}, \tilde{D}^{-\frac{\nu+1}{\nu}}\left(G_E^{\frac{1}{2}}(1-\rho)^{\frac{1}{2}} + \frac{1}{2}G_I - \frac{1}{2}G_E\right)\right)\right)$$

as desired. □

Appendix C

Appendix for Chapter 5

C.1 Example bandit setups

We consider the following risky-safe arm setups in our results. The first setup is a risky-safe arm bandit setup in continuous time, where user rewards are undiscounted.

Setup 1 (Undiscounted, continuous time risky-safe arm setup). *Consider a risky-safe arm bandit setup where the algorithm class $\mathcal{A}_{all}^{cont}$ equals*

$$\{A_f \mid f : [0,1] \to [0,1] \text{ is measurable}, f(0) = 0, f(1) = 1, f \text{ is continuous at } 0 \text{ and } 1\ \}.$$

The bandit setup is in continuous time: if a platform chooses algorithm $A \in \mathcal{A}_{all}^{cont}$, then at a given time step with information state $\mathcal{I}$, the user of that platform devotes a $\mathbb{P}[A(\mathcal{I}) = 1]$ fraction of the time step to the risky arm and the remainder of the time step to the safe arm. Let the prior be initialized so $p_0 := p(\mathcal{I}) = \mathbb{P}_{X \sim \mathcal{D}_1^{Prior}}[X = h] \in (0,1)$. Let the rewards be such that the full-information payoff $hp_0 + s(1 - p_0) = 0$. Let the background information quality be $\sigma_b < \infty$. Let the time horizon $T = \infty$ be infinite, and suppose the user utility is undiscounted.[1]

The next setup is again a risky-safe arm bandit setup in continuous time, but this time with discounted rewards.

Setup 2 (Discounted, continuous time risky-safe arm setup). *Consider a risky-safe arm bandit setup where the algorithm class is $\mathcal{A}_{all}^{cont}$. The bandit setup is in continuous time: if a platform chooses algorithm $A \in \mathcal{A}_{all}^{cont}$, then at a given time step with information state $\mathcal{I}$, the user of that platform devotes a $\mathbb{P}[A(\mathcal{I}) = 1]$ fraction of the time step to the risky arm and the remainder of the time step to the safe arm. Let the high reward h be 1, the low reward l be 0, and let the prior be initialized to some $p(\mathcal{I}) \geq \mathbb{P}_{X \sim \mathcal{D}_1^{Prior}}[X = 1] > s$ where*

[1] Formally, this means that the user utility is the limit $\lim_{T\to\infty} T \cdot \mathbb{E}\left[\frac{1}{T}\int d\pi(t)\right]$ as the time horizon goes to ∞, or alternatively the limit $\lim_{\beta\to 0} T \cdot \mathbb{E}\left[\int e^{-\beta t} d\pi(t)\right]$ as the discount factor vanishes. See Bolton and Harris (2000a) for a justification that these limits are well-defined.

s is the safe arm reward. Let the time horizon $T = \infty$ be infinite, suppose that there is no background information $\sigma_b = \infty$, and suppose the user utility is discounted with discount factor $\beta \in (0, \infty)$.

Finally, we consider another discounted risky-safe bandit setup, but this time with discrete time and finite time horizon.

Setup 3 (Discrete, risky-safe arm setup). *Consider a risky-safe arm bandit setup where the algorithm class is $\mathcal{A} = \left\{A_{f_\varepsilon^{TS}} \mid \varepsilon \in [0,1]\right\}$, where $A_{f_\varepsilon^{TS}}$ denotes the ε-Thompson sampling algorithm given by $f^T S_\varepsilon(p) = \varepsilon + (1-\varepsilon)p$. The bandit setup is in discrete time: if a platform chooses algorithm $A \in \mathcal{A}$, then at a given time step with information state $\mathcal{I}$, the user of that platform chooses the risky arm with probability $\mathbb{P}[A(\mathcal{I}) = 1]$ and the safe arm with probability $\mathbb{P}[A(\mathcal{I}) = 0]$. Let the time horizon $T < \infty$ be finite, suppose that the user utility is discounted with discount factor $\beta \in (0, 1]$, that there is no background information $\sigma_b = 0$, and let the prior be initialized to $p(\mathcal{I}) \in (0, 1)$*

C.2 Further details about the model choice

We examine two aspects our model—the choice of equilibrium set $\mathcal{E}$ and the action space of users—in greater detail.

C.2.1 What would change if users can play mixed strategies?

Suppose that $\mathcal{E}_{A_1,A_2}$ were defined to be the set of *all* equilibria for the users, rather than only pure strategy equilibria. The main difference is that all users might no longer choose the same platform at equilibrium, which would change the nature of the set $\mathcal{E}_{A_1,A_2}$. In particular, even when both platforms choose the same algorithm A, there is a symmetric mixed equilibrium where all users randomize equally between the two platforms. At this mixed equilibrium, the utility of the users is $\mathbb{E}_{X \sim Bin(N,1/2)}[R_A(X)]$, since the number of users at each platform would follows a binomial distribution. This quantity might be substantially lower than $R_A(N)$ depending on the nature of the bandit algorithms. As a result, the user quality level $Q(A, A)$, which is measured by the *worst* equilibrium for the users in $\mathcal{E}$, could be substantially lower than $R_A(N)$. Moreover, the condition for (A, A) to be an equilibrium for the platforms would still be that $R_A(N) \geq \max_{A'} R_{A'}(1)$, so there could exist a platform equilibria with user quality level much lower than $\max_{A'} R_{A'}(1)$. Intuitively, the introduction of mixtures corresponds to users no longer coordinating between their choices of platforms—-this leads to no single platform accumulating all of the data, thus lowering user utility.

C.2.2 What would change if users could change platforms at each round?

Our model assumes that users choose a platform at the beginning of the game which they participate on for the duration of the game. In this section, we examine this assumption in greater detail, informally exploring what would change if the users could switch platforms.

First, we provide intuition that in the shared data setting, there would be no change in the structure of the equilibrium as long as the equilibrium class $\mathcal{A}$ is closed under mixtures (i.e. if $A_1, A_2 \in \mathcal{A}$, then the algorithm that plays A_1 with probability p_1 and A_2 with probability p_2 must be in $\mathcal{A}$). A natural model for users switching platforms would be that users see the public information state at every round and choose a platform based on this information state (and algorithms for the platforms). A user's strategy is thus a mapping from an information state $\mathcal{I}$ to $\{1, 2\}$, and the platform would receive utility for a user depending on the fraction of time that they spend on that platform. Suppose that symmetric (mixed) equilibria for users are guaranteed to exist for any choice of platform algorithms, and we define the platform's utility by the minimal number of (fractional) users that they receive at any symmetric mixed equilibrium. In this model, we again see that (A, A) is a symmetric equilibrium for the platform if and only if A is an symmetric pure strategy equilibrium in the game G defined in Chapter 5.4. (To see this, note if A is not a symmetric pure strategy equilibrium, then the platform can achieve higher utility by choosing A' that is a deviation for a player in the game G. If (A, A) is a symmetric pure strategy equilibrium, then). Thus, the alignment results will remain the same.

In the separate data setting, even defining a model where users can switch platforms is more subtle since it is unclear how the information state of the users should be defined. One possibility would be that each user keeps track of their own information state based on the rewards that they observe. Studying the resulting equilibria would require reasoning about the evolution of user information states and furthermore may not capture practical settings where users see the information of other users. Given these challenges, we defer the analysis of users switching platforms in the case of separate data to future work.

C.3 Proof of Theorem 15

We prove Theorem 15.

Proof of Theorem 15. We split into two cases: (1) either $R_{A_1}(1) = \max_{A'} R_{A'}(1)$ or $R_{A_2}(1) = \max_{A'} R_{A'}(1)$, and (2) $R_{A_1}(1) < \max_{A'} R_{A'}(1)$ or $R_{A_2}(1) < \max_{A'} R_{A'}(1)$.

Case 1: $R_{A_1}(1) = \max_{A'} R_{A'}(1)$ **or** $R_{A_2}(1) = \max_{A'} R_{A'}(1)$**.** We show that (A_1, A_2) is an equilibrium.

Suppose first that $R_{A_1}(1) = \max_{A'} R_{A'}(1)$ and $R_{A_2}(1) = \max_{A'} R_{A'}(1)$. We see that the strategies $\mathbf{p} = [1]$ and $\mathbf{p} = [2]$, where the user chooses platform 1, is in the set of equilibria $\mathcal{E}_{A_1,A_2}$. This means that $v_1(A_1; A_2) = v_1(A_2; A_1) = 0$. Suppose that platform 1 chooses another algorithm A''. Since $R_{A''}(1) \leq \max_{A'} R_{A'}(1)$, we see that $\mathbf{p} = [2, \ldots, 2]$ is still an equilibrium. Thus, $v_1(A'; A_2) = 0$. This implies that A_1 is a best response for platform 1, and an analogous argument shows A_2 is a best response for platform 2. When the platforms choose (A_1, A_2), at either of the user equilibria $\mathbf{p} = [1]$ or $\mathbf{p} = [2]$, the user utility is $\max_{A'} R_{A'}(1)$. Thus $Q(A_1, A_2) = \max_{A'} R_{A'}(1)$.

Now, suppose that exactly one of $R_{A_1}(1) = \max_{A'} R_{A'}(1)$ and $R_{A_2}(1) = \max_{A'} R_{A'}(1)$ holds. WLOG, suppose $R_{A_1}(1) = \max_{A'} R_{A'}(1)$. Since $R_{A_2}(1) < \max_{A'} R_{A'}(1)$, we see that $[2] \notin \mathcal{E}_{A_1,A_2}$. On the other hand, $[1] \in \mathcal{E}_{A_1,A_2}$. This means that $v_1(A_1; A_2) = 1$ and $v_2(A_2; A_1) = 0$. Thus, A_1 is a best response for platform 1 trivially because $v_1(A'; A_2) \leq 1$ for all A' by definition. We next show that A_2 is a best response for platform 2. If the platform 2 plays another algorithm A', then [1] will still be in equilibrium for the users since platform 1 offers the maximum possible utility. Thus, $v_1(A'; A) = 0$, and A_2 is a best response for platform 2. When the platforms choose (A_1, A_2), the only user equilibria is $\mathbf{p} = [1]$ where the user utility is $\max_{A'} R_{A'}(1)$. Thus $Q(A_1, A_2) = \max_{A'} R_{A'}(1)$.

Case 2: $R_{A_1}(1) < \max_{A'} R_{A'}(1)$ or $R_{A_2}(1) < \max_{A'} R_{A'}(1)$. It suffices to show that (A_1, A_2) is not an equilibrium. WLOG, suppose that $R_{A_2}(1) \leq R_{A_1}(1)$. We see that $[1] \in \mathcal{E}_{A_1,A_2}$. Thus, $v_2(A_2; A_1) = 0$. However, if platform 2 switches to $A'' \in \arg\max_{A' \in \mathcal{A}} R_{A'}(1)$, then $\mathcal{E}_{A_1,A''}$ is equal to $\{[2]\}$ and so $v_2(A''; A_1) = 1$. This means that A_2 is not a best response for platform 2, and thus (A_1, A_2) is not an equilibrium. □

C.4 Proofs for Chapter 5.4

In the proofs of Theorems 16 and 17, the key technical ingredient is that pure strategy equilibria for users take a simple form. In particular, under strict information monotonicity, we show that in every pure strategy equilibrium $p^* \in \mathcal{E}^{A_1,A_2}$, all of the users choose the same platform.

Lemma 175. *Suppose that every algorithm $A \in \mathcal{A}$ is either strictly information monotonic or information constant (see Assumption 1). For any choice of platform algorithms $A_1, A_2 \in \mathcal{A}$ such that at least one of A_1 and A_2 is strictly information monotonic, it holds that:*

$$\mathcal{E}_{A_1,A_2} \subseteq \{[1, \ldots, 1], [2, \ldots, 2]\}.$$

Proof. WLOG, assume that A_1 is strictly information monotonic. Assume for sake of contradiction that the user strategy profile $[1, \ldots, 1, 2, \ldots, 2]$ (with $N_1 > 0$ users choosing platform 1 and $N_2 > 0$ users choosing platform 2) is in $\mathcal{E}_{A_1,A_2}$. Since $[1, \ldots, 1, 2, \ldots, 2]$ is an equilibrium, a user choosing platform 1 not want to switch to platform 2. The utility that they currently receive is $R_{A_1}(N_1)$ and the utility that they would receive from switching is $R_{A_2}(N_2 + 1)$, so this means:

$$R_{A_2}(N_2 + 1) \leq R_{A_1}(N_1).$$

Similarly, since $[1, \ldots, 1, 2, \ldots, 2]$ is an equilibrium, a user choosing platform 2 not want to switch to platform 1. The utility that they currently receive is $R_{A_2}(N_2)$ and the utility that they would receive from switching is $R_{A_1}(N_1+1)$, so this means:

$$R_{A_1}(N_1+1) \le R_{A_2}(N_2).$$

Putting this all together, we see that:

$$R_{A_2}(N_2+1) \le R_{A_1}(N_1) < R_{A_1}(N_1+1) \le R_{A_2}(N_2),$$

which is a contradiction since A_2 is either strictly information monotonic or information constant. □

C.4.1 Proof of Theorem 16

We prove Theorem 16.

Proof of Theorem 16. Since the algorithm class $\mathcal{A}$ is utility rich (Assumption 2), we know that for any $\alpha \in [\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N)]$, there exists an algorithm $A^* \in \mathcal{A}$ such that $R_{A^*}(N) = \alpha$. We claim that (A^*, A^*) is an equilibrium and we show that $Q(A^*, A^*) = \alpha$.

To show that (A^*, A^*) is an equilibrium, suppose that platform 1 chooses any algorithm $A \in \mathcal{A}$. We claim that $[2, 2, \ldots, 2] \in \mathcal{E}_{A_1, A_2}$. To see this, notice that the utility that a user receives from choosing platform 2 is $R_{A^*}(N)$, and the utility that they would receive if they deviate to platform is $R_{A^*}(1)$. By definition, we see that:

$$R_{A^*}(N) = \alpha \ge \max_{A' \in \mathcal{A}} R_{A'}(1) \ge R_{A^*}(1),$$

so choosing platform 2 is a best response for the user. This means that $v_1(A; A_2) = 0$ for any algorithm $A \in \mathcal{A}$. This in particular means that A^* is a best response for platform 1. By an analogous argument, we see that A^* is a best response for platform 2, and so (A^*, A^*) is an equilibrium.

We next show that the user quality level is α. It suffices to examine the set of pure strategy equilibria for users when the platforms play (A^*, A^*). By assumption, either A^* is information constant or A^* is strictly information monotonic. If A^* is information constant, then $R_{A^*}(n)$ is constant in n. This means that any pure strategy equilibrium $\mathbf{p}$ generates utility α for all users, so $Q(A^*, A^*) = \alpha$ as desired. If A^* is strictly information monotonic, then we apply Lemma 175 to see that $\mathcal{E}_{A^*, A^*} \subseteq \{[1, \ldots, 1], [2, \ldots, 2]\}$. In fact, since the platforms play the same algorithm, we see that

$$\mathcal{E}_{A^*, A^*} = \{[1, \ldots, 1], [2, \ldots, 2]\}.$$

The user utility at these equilibria is $R_{A^*}(N) = \alpha$, so $Q(A^*, A^*) = \alpha$ as desired. □

C.4.2 Proof of Theorem 17

We prove Theorem 17.

Proof of Theorem 17. First, we show the upper bound of $Q(A_1, A_2) \leq \max_{A' \in \mathcal{A}} R_{A'}(N)$. In fact, we show this upper bound for any selection of user actions $\mathbf{p}$, user $1 \leq i \leq N$, and platform algorithms (A_1, A_2). If a user chooses platform 1 and $0 \leq n \leq N-1$ other users also choose platform 1, then the user's utility is $R_{A_1}(n+1)$ which by Assumption 1 can be upper bounded by $R_{A_1}(N) \leq \max_{A' \in \mathcal{A}} R_{A'}(N)$. Similarly, if a user chooses platform 2, their utility can also be upper bounded by $R_{A_2}(N-n) \leq R_{A_2}(N) \leq \max_{A' \in \mathcal{A}} R_{A'}(N)$. This establishes the desired upper bound.

The remainder of the proof boils down to lower bounding Q_{A_1,A_2} at any equilibrium (A_1, A_2) by $\max_{A' \in \mathcal{A}} R_{A'}(1)$. We divide into two cases: (1) $R_{A_1}(N) < \max_{A' \in \mathcal{A}} R_{A'}(1)$ and $R_{A_2}(N)) < \max_{A' \in \mathcal{A}} R_{A'}(1)$, (2) at least one of $R_{A_1}(N) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$ and $R_{A_2}(N) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$ holds.

Case 1: $R_{A_1}(N) < \max_{A' \in \mathcal{A}} R_{A'}(1)$ and $R_{A_2}(N)) < \max_{A' \in \mathcal{A}} R_{A'}(1)$. We show that (A_1, A_2) is not an equilibrium. WLOG suppose that $R_{A_2}(N) \leq R_{A_1}(N)$. First, we claim that $v_2(A_2; A_1) = 0$. It suffices to show that $[1, \ldots, 1] \in \mathcal{E}_{A_1,A_2}$. To see this, notice that the utility that a user derives from choosing platform 1 is $R_{A_1}(N)$, while the utility that a user would derive from choosing platform 2 is $R_{A_2}(1)$. Moreover, we have that:

$$R_{A_1}(N) \geq R_{A_2}(N) \geq R_{A_2}(1),$$

since A_2 is either strictly information monotonic or information constant by assumption. This means that choosing platform 1 is a best response for the user, so $[1, \ldots, 1] \in \mathcal{E}_{A_1,A_2}$.

Next, we claim that A_2 is not a best response for platform 2. It suffices to show that for $A'' \in \arg\max_{A' \in \mathcal{A}} R_{A'}(1)$, platform 2 receives utility $v_2(A''; A_1) > v_2(A_2; A_1) = 0$. It suffices to show that $[1, \ldots, 1] \notin \mathcal{E}_{A_1,A''}$. To see this, notice that the utility that a user derives from choosing platform 1 is $R_{A_1}(N)$, while the utility that a user would derive from choosing platform 2 is $R_{A''}(1)$. Moreover, we have that:

$$R_{A_1}(N) < \max_{A' \in \mathcal{A}} R_{A'}(1) = R_{A''}(1),$$

so choosing platform 1 is not a best response for the user. Thus, $[1, \ldots, 1] \notin \mathcal{E}_{A_1,A''}$ and $v_2(A''; A_1) > 0$.

This means that (A_1, A_2) is not an equilibrium as desired.

Case 2: at least one of $R_{A_1}(N) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$ and $R_{A_2}(N) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$ holds. WLOG suppose that $R_{A_1}(N) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$. We break into 2 cases: (1) at least one of A_1 and A_2 is strictly information monotonic, and (2) both A_1 and A_2 are information constant.

Subcase 1: at least one of A_1 and A_2 is strictly information monotonic. It suffices to show that at every equilibrium in $\mathcal{E}_{A_1,A_2}$, the user utility is at least $\max_{A' \in \mathcal{A}} R_{A'}(1)$. We can apply Lemma 175, which tells us that $\mathcal{E}_{A_1,A_2} \subseteq \{[1, \ldots, 1], [2, \ldots, 2]\}$.

At $[1, \ldots, 1]$, the user utility is $R_{A_1}(N) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$ as desired.

Suppose now that $[2, \ldots, 2] \in \mathcal{E}_{A_1, A_2}$. The user utility at this equilibria is $R_{A_2}(N)$, so it suffices to show that $R_{A_2}(N) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$. Since $[2, \ldots, 2] \in \mathcal{E}_{A_1, A_2}$, we see that $v_1(A_1; A_2) = 0$. Thus, if platform 1 changed to $A \in \arg\max_{A' \in \mathcal{A}} R_{A'}(1)$, it must hold that $v_1(A; A_2) \leq v_1(A_1; A_2) = 0$ since (A_1, A_2) is a platform equilibrium. This means that $[2, \ldots, 2] \in \mathcal{E}_{A, A_2}$. The utility that a user would derive from choosing platform 1 is $R_A(1)$, while the utility that a user derives from choosing platform 2 is $R_{A_2}(N)$. Since $[2, \ldots, 2]$ is an equilibrium, it must hold that

$$R_{A_2}(N) \geq R_A(1) = \max_{A' \in \mathcal{A}} R_{A'}(1).$$

as desired.

Subcase 2: both A_1 and A_2 are information constant. Regardless of the actions of other users, if a user chooses platform 1 they receive utility $R_{A_1}(N)$ and if a user chooses platform 2 they receive utility $R_{A_2}(N)$. Thus, if any user chooses platform 1 in equilibrium, then $R_{A_1}(N) \geq R_{A_2}(N)$; and if any user chooses platform 2 in equilibrium, then $R_{A_2}(N) \geq R_{A_1}(N)$. The quality level $Q(A_1, A_2)$ is thus $\max(R_{A_1}(N), R_{A_2}(N)) \geq \max_{A' \in \mathcal{A}} R_{A'}(1)$ as desired. □

C.5 Proofs for Chapter 5.5

To analyze the equilibria in the shared data setting, we relate the equilibria of our game to the equilibria of an N-player game G that is closely related to strategic experimentation Bolton and Harris (1999; 2000a;b). In Appendix C.5.1, we formally establish the relationship between the equilibria in G and the equilibria in our game. In Appendix C.5.2, we provide a recap of the results from strategic experimentation literature for the risky-safe arm bandit problem (Bolton and Harris, 2000a; 1999). In Appendices C.5.3-C.5.5, we prove our main results using these tools.

C.5.1 Proof of Lemma 20

In Lemma 20 (restated below), we show that the symmetric equilibria of our game are equivalent to the symmetric pure strategy equilibria of the following game. Let G be an N player game where each player chooses an algorithm in $\mathcal{A}$ within the same bandit problem setup as in our game. The players share an information state $\mathcal{I}$ corresponding to the posterior distributions of the arms. At each time step, all of the N users arrive at the platform, player i pulls the arm drawn from $A_i(\mathcal{I})$, and the players all update $\mathcal{I}$. The utility received by a player is given by their expected discounted cumulative reward.

Lemma 20. *The solution (A, A) is in equilibrium if and only if A is a symmetric pure strategy equilibrium of the game G described above.*

Proof. We first claim that $\mathcal{E}_{A,A}$ is equal to the set of all possible strategy profiles $\mathbf{p}$. To see this, notice that the user utility is $R_A(N)$ regardless of their action or the actions of other users. Thus, *any* strategy profile is in equilibrium.

Suppose first that A is a symmetric pure strategy equilibrium of the game G. Consider the solution (A, A)—since $\mathcal{E}_{A,A}$ contains $[2, \ldots, 2]$, we see that $v_1(A; A) = 0$. Suppose that platform 1 instead chooses $A' \neq A$. We claim that $[2, \ldots, 2] \in \mathcal{E}_{A',A}$. To see this, when all of the other users choose A, then no user wishes to deviate to any other algorithm, including A', since A is a symmetric pure strategy equilibrium in G. Thus, $v_1(A'; A) = 0 = v_1(A; A$, and A is a best response for platform 1 as desired. An analogous argument applies to platform 1, thus showing that (A, A) is an equilibrium.

To prove the converse, suppose that (A, A) is in equilibrium. Since $\mathcal{E}_{A,A}$ contains $[2, \ldots, 2]$, we see that $v_1(A; A) = 0$. It suffices to show that for every $A' \in \mathcal{A}$, it holds that $v_1(A'; A) = 0 = v_1(A; A)$. Assume for sake of contradiction that A is not a symmetric pure strategy equilibrium in the game G. Then there exists a deviation A' for which a user achieves strictly higher utility than the algorithm A in the game G. Suppose that platform 1 chooses A'. Then we see that $\mathcal{E}_{A',A}$ does *not* contain $[2, \ldots, 2]$, since a user will wish to deviate to platform 1. This means that $v_1(A'; A) > 0$ which is a contradiction. □

C.5.2 Recap of strategic experimentation in the risky-safe arm bandit problem Bolton and Harris (1999; 2000a;b)

To analyze the game G, we leverage the literature on strategic experimentation for the risky-safe arm problem. We provide a high-level summary of results from Bolton and Harris (1999; 2000a;b), deferring the reader to Bolton and Harris (1999; 2000a;b) for a full treatment.

Bolton and Harris (1999; 2000a;b) study an infinite time-horizon, continuous-time game with N users updating a common information source, focusing on the risky-safe arm bandit problem. The observed rewards $\mathcal{D}^{\text{Noise}}$ are given by the mean reward with additive noise with variance σ. All N users arrive at the platform at every time step and choose a *fraction* of the time step to devote to the risky arm (see Bolton and Harris (1999) for a discussion of how this relates to mixed strategies). The players receive background information according to $N(\sqrt{\zeta}h, \sigma^2)$. (By rescaling, we can equivalently think of this as background noise from $N(h, \sigma_b^2)$ where $\sigma_b = \sigma/\sqrt{\zeta}$.) Each player's utility is defined by the (discounted) rewards of the arms that they pull.

Bolton and Harris (1999; 2000a;b) study the Markov equilibria of the resulting game, so user strategies correspond to mappings from the posterior probability that the risky arm has high reward to the fraction of the time step devoted to the risky arm. We denote the user strategies by measurable functions $f : [0, 1] \to [0, 1]$. The function f is a symmetric pure strategy equilibrium if for any choice of prior, f is optimal for each user in the game with that choice of prior.

Undiscounted setting. The cleanest setting is the case of undiscounted rewards. To make the undiscounted user utility over an infinite time-horizon well-defined, the full-information payoff is subtracted from the user utilities. For simplicity, let's focus on the setting where the full information payoff is 0. In this case, the undiscounted user utility is equal to $\mathbb{E}\left[\int d\pi(t)\right]$, where $d\pi(t)$ denotes the payoff received by the user (see Bolton and Harris (2000b) for a justification that this expectation is well-defined).

With undiscounted rewards, the user utility achieved by a set of strategies $f_1, \ldots, f_n$ permits a clean closed-form solution.

Lemma 176 (Informal restatement of results from Bolton and Harris (2000b)). *Suppose that N players choose the strategies $f_1, \ldots, f_N$ respectively. If the prior is p, then the utility $K(p; f_1, \ldots, f_n)$ of player 1 is equal to:*

$$K(p; f_1, \ldots, f_N) = \int G(p,q) \frac{(1 - f_1(q))s + (qh + (1-q)l)f_1(q) - (hq + s(1-q))}{\frac{\sigma^2}{\sigma_b^2} + \sum_{i=1}^N f_i(q)} dq$$

where:

$$G(p,q) = \begin{cases} \frac{2\sigma^2 p}{(h-l)^2 q^2 (1-q)} & \text{if } p \leq q \\ \frac{\sigma^2(1-p)}{(h-l)^2 q (1-q)^2} & \text{if } p \geq q. \end{cases}$$

Proof sketch. We provide a proof sketch, deferring the full proof to Bolton and Harris (2000a). For ease of notation, we let $K(p)$ denote $K(p; f_1, \ldots, f_N)$. The change in the posterior when the posterior is p is mean 0 and variance $\left(\frac{\sigma^2}{\sigma_b^2} + \sum_{i=1}^n f_i(p)\right)\Phi(p)$ where $\Phi(p) = \left(\frac{p(1-p)(h-l)}{\sigma}\right)^2$. The utility of player 1 is equal to the sum of the current payoff and the continuation payoff. It can be shown that this is equal to:

$$K(p) + \left[(1 - f_1(p))s + (ph + (1-p)l)f_1(p) - (hp + s(1-p)) + \left(\frac{\sigma^2}{\sigma_b^2} + \sum_{i=1}^n f_i(p)\right)\Phi(p)\frac{K''(p)}{2}\right] dt.$$

This means that:

$$0 = (1 - f_1(p))s + (ph + (1-p)l)f_1(p) - (hp + s(1-p)) + \left(\frac{\sigma^2}{\sigma_b^2} + \sum_{i=1}^n f_i(p)\right)\Phi(p)\frac{K''(p)}{2}.$$

We can directly solve this differential equation to obtain the desired expression. □

Bolton and Harris (2000a;b) characterize the symmetric equilibria, and we summarize below the relevant results for our analysis.

Theorem 177 (Informal restatement of results from Bolton and Harris (2000a;b)). *In the undiscounted game, there is a unique symmetric equilibrium f^*. When the prior is in $(0,1)$, the equilibrium utility is strictly smaller than the optimal utility in the N user team problem (where users coordinate) and strictly larger than the optimal utility in the 1 user team problem.*

Proof sketch. The existence of equilibrium boils down to characterizing the best response of a player given the experimentation levels of other players. This permits a full characterization of all equilibria (see Bolton and Harris (2000b)) from which we can deduce that there is a unique symmetric equilibrium f^*. Moreover, $f^*(p)$ is 0 if p is sufficiently low, 1 if p is sufficiently high, and interpolates between 0 and 1 for intermediate values of p.

To see that the utility when all users play f^* is strictly smaller than that of the optimal utility in the N user team problem, the calculations in Bolton and Harris (2000a) show that there is a unique strategy f_N that achieves the N-user team optimal utility and this is a cutoff strategy. On the other hand, f^* is not a cutoff strategy, so it cannot achieve the team optimal utility.

To see that the utility when all users play f^* than that that of the 1 user team problem, suppose that a user instead chooses the optimal 1 user strategy f_1. Using Lemma 176 and the fact that f^* involves nonzero experimentation at some posterior probabilities p, we see that the user would achieve strictly higher utility than in a single-user game where they play f'. This means that playing f_1 results in utility strictly higher than the single-user optimal utility. Since f^* is a best response, this means that playing f^* also results in strictly higher utility than the single-user optimal. □

Discounted setting. With discounted rewards, the analysis turns out to be much more challenging since the user utility does not permit a clean closed-form solution. Nonetheless, Bolton and Harris (1999) are able to prove properties about the equilibria.We summarize below the relevant results for our analysis.

Theorem 178 (Informal restatement of results from Bolton and Harris (1999))**.** *In the undiscounted game when there is no background information ($\sigma_b = \infty$), the following properties hold:*

1. *For any choice of discount factor, there is a unique symmetric equilibrium f^*.*
2. *The function f^* is monotonic and strictly increasing from 0 to 1 in an interval $[c_{low}, c_{high}]$ where $c_{low} < c_{high} < s$.*
3. *When the prior is initialized above c_{low}, the posterior never reaches c_{low} but converges to c_{low} in the limit.*
4. *When the prior is initialized above c_{low}, the equilibrium utility of f^* is at least as large as the optimal utility in the 1 user problem. The equilibrium utility of f^* is strictly smaller than the optimal utility in the N user team problem (where users coordinate).*

We note that we define the discounted utility as $\mathbb{E}\left[\int e^{-\beta t} d\pi(t)\right]$, while Bolton and Harris (1999) defines the discounted utility as $\mathbb{E}\left[\beta \int e^{-\beta t} d\pi(t)\right]$; however, this constant factor difference does not make a difference in any of the properties in Theorem 178.

C.5.3 Proof of Theorem 18

Proof of Theorem 18. The key ingredient in the proof of Theorem 18 is Lemma 20. This result shows that (A, A) is an equilibrium for the platforms if and only if A is a symmetric pure strategy equilibrium in the game G. Moreover, we see that at any equilibrium in $\mathcal{E}_{A,A}$, the users achieve utility $R_A(N)$ since the information state is shared between platforms. Notice that $R_A(N)$ is also equal to the utility that users achieve in the game G if they all choose A. It thus suffices to show that the equilibrium utility is a unique value $\alpha^* \in (\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N))$ at every symmetric pure strategy equilibrium in G.

To analyze the game G, we leverage the results in the literature on strategic experimentation (see Theorem 177). We know by Theorem 177 that there is a unique equilibrium f^* in the undiscounted strategic experimentation game. However, the equilibrium concepts are slightly different because f is a symmetric equilibrium in the undiscounted strategic experimentation game if for *any* choice of prior, f is optimal for each user in the game with that choice of prior; on the other hand, f is a symmetric equilibrium in G if for the specific choice of prior of the bandit setup, f is optimal for each user in the game with that choice of prior. We nonetheless show that f^* is the unique symmetric equilibrium in the game G.

We first claim that the algorithm A_{f^*} is a symmetric pure strategy equilibrium in G. To see this, notice that if a user in G were to deviate to $A_{f'} \in \mathcal{A}^{\text{cont}}_{\text{all}}$ for some f' and achieve higher utility, then it would also be better for a user in the game in Bolton and Harris (1999) to deviate to f' when the prior is $\mathbb{P}_{X \sim \mathcal{D}^{\text{Prior}}_1}[X = 1]$. This is not possible since f is an equilibrium in G, so A_{f^*} is a symmetric pure strategy equilibrium in G. This establishes that an equilibrium exists in G.

We next claim that if A_f is a symmetric pure strategy equilibrium in G, then $f(c) = f^*(c)$ for all $c \in (0, 1)$. Let $S = \{c \in (0, 1) \mid f(c) \neq f^*(c)\}$ be the set of values where f and f^* disagree. Assume for sake of contradiction that S has positive measure.

1. The first possibility is that only a measure zero set of $c \in S$ are realizable when all users play f. However, this is not possible: because of background information and because the prior is initialized in $(0, 1)$, the posterior eventually converges to 0 to 1 (but never reaches at any finite time). This means that every posterior in $(0, 1)$ is realizable.

2. The other possibility is that a positive measure of values in S are realizable all users play f. Let the solution f^{**} be defined by $f^{**}(c) = f(c)$ for all c that are realizable in f and $f^{**} = f^*(c)$ otherwise. It is easy to see that f^{**} would be an equilibrium in the strategic experimentation game, which is a contradiction because of uniqueness of equilibria in this game.

This implies that f and f^* agree on all but a measure 0 set of $[0, 1]$.

This proves that there is a unique symmetric equilibrium in the game G.

We next prove that the utility achieved by this symmetric equilibrium is strictly in between the single-user optimal and the global-optimal. This follows from the fact that the utility of

A_{f^*} in G is equal to the utility of f^* in the strategic experimentation game, coupled with Theorem 177.

□

C.5.4 Proof of Theorem 19

Proof of Theorem 19. The proof of Theorem 19 follows similarly to the proof of Theorem 18. Like in the proof of Theorem 18, it thus suffices to show that the equilibrium utility is a unique value $\alpha^* \in [\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N))$ at every symmetric pure strategy equilibrium in G.

To analyze the game G, we leverage the results in the literature on strategic experimentation (see Theorem 178). Again, the equilibrium concepts are slightly different because f is a symmetric equilibrium in the strategic experimentation game if for *any* choice of prior, f is optimal for each user in the game with that choice of prior; on the other hand, f is a symmetric equilibrium in G if for the specific choice of prior of the bandit setup, f is optimal for each user in the game with that choice of prior. Let f^* be the unique symmetric pure strategy equilibrium in the strategic experimentation game (see property (1) in Theorem 178).

We first claim that the algorithm A_{f^*} is a symmetric pure strategy equilibrium in G. To see this, notice that if a user in G were to deviate to $A_{f'} \in \mathcal{A}_{\text{all}}$ for some f' and achieve higher utility, then it would also be better for a user in the strategic experimentation game to deviate to f' when the prior is $\mathbb{P}_{X \sim \mathcal{D}_1^{\text{Prior}}}[X = 1]$. This is not possible since f is an equilibrium in G, so A_{f^*} is a symmetric pure strategy equilibrium in G. This establishes that an equilibrium exists in G.

We next claim that if A_f is a symmetric pure strategy equilibrium in G, then $f(c) = f^*(c)$ for all $c > c_{low}$, where c_{low} is defined according to Theorem 178. Note that by the assumption in Setup 1, the prior is initialized above s, which means that it is initialized above c_{low}. Let $S = \{c > c_{low} \mid f(c) \neq f^*(c)\}$ be the set of values where f and f^* disagree. Assume for sake of contradiction that S has positive measure.

1. The first possibility is that only a measure 0 set of $c \in S$ are realizable when all users play f and the prior is initialized to $\mathbb{P}_{X \sim \mathcal{D}_1^{\text{Prior}}}[X = 1]$. However, this is not possible because then the trajectories of f and f^* would be identical so the same values would be realized for f and f^*, but we already know that all of $c > c_{low}$ is realizable for f^*.

2. The other possibility is that a positive measure of values in S are realizable when all users play f and the prior is initialized to $\mathbb{P}_{X \sim \mathcal{D}_1^{\text{Prior}}}[X = 1]$. Let the solution f^{**} be defined by $f^{**}(c) = f(c)$ for all c that are realizable in f and $f^{**} = f^*(c)$ otherwise. It is easy to see that f^* would be an equilibrium in the game in Bolton and Harris (1999), which is a contradiction because of uniqueness of equilibria in this game.

This implies that f and f^* agree on all but a measure 0 set of $[c_{low}, 1]$.

Now, let us relate the utility achieved by a symmetric equilibrium in G to the utility achieved in the strategic experimentation game. Since the prior is initialized above c_{high}, property (3) in Theorem 178 tells us that the posterior never reaches c_{low} but can come arbitrarily close to c_{low}. This in particular means if $f(c) = f^*(c)$ for all $c > c_{low}$, then the utility achieved when all users choose a strategy f where is equivalent to the utility achieved when all users choose f^*.

Thus, it follows from property (4) in Theorem 178 that when the prior is initialized sufficiently high, the equilibrium utility is strictly smaller than the utility that users achieve in the team problem (which is equal to the global optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(N)$). Moreover, it also follows from property (4) in Theorem 178 that the equilibrium utility is always at least as large as the utility $\max_{A' \in \mathcal{A}} R_{A'}(1)$ that users achieve in the single user game.

□

C.5.5 Proof of Theorem 21

To prove Theorem 21, the key ingredient is the following fact about the game G.

Lemma 179. *Suppose that every algorithm $A \in \mathcal{A}$ is side information monotonic. If A is a symmetric pure strategy equilibrium of G, then the equilibrium utility at G is at least $\max_{A'} R_{A'}(1)$.*

Proof of Lemma 179. Let A be a symmetric pure strategy equilibrium in G. To see this, notice that an user can always instead play $A^* = \arg\max_{A'} R_{A'}(1)$. Since A is a best response for this user, it suffices to show that playing A^* results in utility at least $\max_{A'} R_{A'}(1)$. By definition, the utility that the user would receive from playing A^* is $\mathcal{U}^{\text{shared}}(1; \mathbf{2}_{N-1}, A^*, A)$. By side information monotonicity, we know that the presence of the background posterior updates by other users cannot reduce this user's utility, and in particular

$$\mathcal{U}^{\text{shared}}(1; \mathbf{2}_{N-1}, A^*, A) \geq R_{A^*}(1) = \max_{A'} R_{A'}(1)$$

as desired. □

Now we can prove Theorem 21 from Lemma 20 and Lemma 179.

Proof of Theorem 21. By Lemma 20, if the solution (A, A) is an equilibrium for the platforms, then it is a symmetric pure strategy equilibrium of the game G. To lower bound $Q(A, A)$, notice that the quality level $Q(A, A)$ is equal to the utility of A in the game G. By Lemma 179, this utility is at least $\max_{A'} R_{A'}(1)$, so $Q(A, A) \geq \max_{A'} R_{A'}(1)$ as desired. To upper bound $Q(A, A)$, notice that

$$Q(A, A) = R_A(N) \leq \max_{A' \in \mathcal{A}} R_{A'}(N),$$

as desired. □

C.6 Proofs for Chapter 5.6

We prove Lemmas 22-25.

C.6.1 Proof of Lemma 22

Proof of Lemma 22. First, we show strict information monotonicity. Applying Lemma 176, we see that

$$R_A(n) = K(p; A, \dots, A) = \int G(p,q) \frac{(1 - A(q))s + (qh + (1-q)l)A(q)}{\frac{\sigma^2}{\sigma_b^2} + nA(q)} dq.$$

Note that the value $(1 - A(q))s + (qh + (1-q)l)A(q)$ is always *negative* based on how we set up the rewards. We see that as N increases, the denominator $NA(q)$ weakly increases at every value of q. This means that $R_A(n)$ is weakly increasing in n. To see that $R_A(n)$ *strictly* increases in n, we note that there exists an open neighborhood around $q = 1$ where $A(q) > 0$. In this open neighborhood, we see that $nA(q)$ strictly increases in n, which means that $R_A(n)$ strictly increases.

Next, we show side information monotonicity. Applying Lemma 176, we see that

$$\mathcal{U}^{\text{shared}}(1; \mathbf{2}_n, A, A') = K(p; A, A', \dots, A') = \int G(p,q) \frac{(1 - A(q))s + (qh + (1-q)l)A(q)}{\frac{\sigma^2}{\sigma_b^2} + A(q) + nA'(q)} dq.$$

This expression is weakly larger for $n > 0$ than $n = 0$.

□

C.6.2 Proof of Lemma 23

To prove Lemma 23, the key technical ingredient is that if the posterior becomes more informative, then the reward increases.

Lemma 180 (Impact of Increased Informativeness)**.** *Consider the bandit setup in Setup 3. Let $0 < p_{prior} < 1$ be the prior probability, and let p' be the random variable for the posterior if another sample from the risky arm is observed. Let $K(p)$ denote the expected cumulative discounted reward that a user receives when they are the only user participating on a platform offering the ε-Thompson sampling algorithm $A_{f_\varepsilon^{TS}}$. Then the following holds:*

$$\mathbb{E}_{p'}[K(p')] > K(p_{prior}).$$

Proof. Notice that $f_\varepsilon^{TS}(p) = \varepsilon + (1-\varepsilon)p$. For notational simplicity, let $f = f_\varepsilon^{TS}$.

Let K_t denote the cumulative discounted reward for time steps t to T received by the user. We proceed by backwards induction on t that for any $0 < p < 1$ it holds that:

$$\mathbb{E}[K_t(p')] > K_t(p)$$

where p is the posterior at the beginning of time step t and where p' be the random variable for the posterior if another sample from the risky arm is observed before the start of time step t.

The base case is $t = T$, where we see that the reward is

$$s(1 - f(q)) + (qh + (1 - q)l)(f(q)) = s(1 - f(q)) + (q(h - l) + l)f(q)$$

if the posterior is q. We see that this is a strictly convex function in q for our choice of algorithm A_f, which means that $\mathbb{E}[K_T(p'; n)] > K_T(p; n)$ as desired.

We now assume that the statement holds for $t + 1$ and we prove it for t. For the purposes of our analysis, we generate a series of 2 samples s_1, s_2 from the risky arm. We generate these samples recursively. Let $p_0 = p$, and let p_1 denote the posterior given by p conditioned on s_1, and let p_2 denote the posterior conditioned on s_1 and s_2. The sample s_{i+1} is drawn from a noisy observed reward for h with probability p_i and a noisy observed reward for l with probability p_i. We assume that the algorithms use these samples (in order).

Our goal is to compare two instances: Instance 1 is initialized at p at time step t and Instance 2 is given a sample s_1 before time step t. The reward of an algorithm can be decomposed into two terms: the *current payoff* and the *continuation payoff* for remaining time steps.

The current payoff for Instance 1 is:

$$s(1 - f(p)) + (p(h - l) + l)f(p)$$

and the current payoff for Instance 2 is:

$$\mathbb{E}[s(1 - f(p_1)) + (p_1(h - l) + l)f(p_1)].$$

Since this is a strictly convex function of the posterior, and p_1 is a posterior update of p with the risky arm, we see that the expected current payoff for Instance 1 is strictly larger than the xpected current payoff for Instance 2.

The expected continuation payoff for Instance 1 is equal to the β-discounted version of:

$$(1 - f(p)) \cdot K_{t+1}(p) + f(p)\mathbb{E}_{s_1}[K_{t+1}(p_1)]$$

and the expected continuation payoff for Instance 2 is equal to the β-discounted version of:

$$\begin{aligned}
&\mathbb{E}_{s_1}[f(p_1) \cdot \mathbb{E}_{s_2}[K_{t+1}(p_2)]] + (1 - (f(p_1))[K_{t+1}(p_1)]] \\
_{(1)} &\geq \mathbb{E}_{s_1}[f(p_1) \cdot K_{t+1}(p_1) + (1 - (A(p_1))[K_{t+1}(p_1)]] \\
&= \mathbb{E}_{s_1}[K_{t+1}(p_1)] \\
&\geq_{(2)} (1 - f(p)) \cdot K_{t+1}(p) + f(p)\mathbb{E}_{s_1}[K_{t+1}(p_1)]
\end{aligned}$$

where (1) and (2) use the induction hypothesis for $t + 1$.

This proves the desired statement.

□

We prove Lemma 23.

Proof of Lemma 23. We first show strict information monotonicity and then we show side information monotonicity. Our key technical ingredient is Lemma 180.

Proof of strict information monotonicity. We show that for any $n \geq 1$, it holds that $R_A(n+1) > R_A(n)$. To show this, we construct a sequence of $T+1$ "hybrids". In particular, for $0 \leq t \leq T$, let the tth hybrid correspond to the bandit instance where 2 users participate in the algorithm in the first t time steps and 1 users participate in the algorithm in the remaining time steps. These hybrids enable us to isolate and analyze the gain of one additional observed sample.

For each $2 \leq t \leq T$, it suffices to show that the $t-1$th hybrid incurs larger cumulative discounted reward than the $t-2$th hybrid. Notice a user participating in both of these hybrids at all time steps achieves the same expected reward in the first $t-1$ time steps for these two hybrids. The first time step where the two hybrids deviate in expectation is the tth time step (after the additional information from the $t-1$th hybrid has propagated into the information state). Thus, it suffices to show that $t-1$th hybrid incurs larger cumulative discounted reward than the $t-2$th hybrid between time steps t and T. Let $\mathcal{H}$ be the history of actions and observed rewards from the first $t-2$ time steps, the arm and the reward for the 1st user at the $t-1$th time step. We condition on $\mathcal{H}$ for this analysis. Let a denote the arm chosen by the 2nd user at the $t-1$th time step in the $t-1$th hybrid. We split into cases based on a.

The first case is that a is the safe arm. In this case, the hybrids have the same expected reward for time steps t to T.

The second case is a is the risky arm. This happens with nonzero probability given $\mathcal{H}$ based on the structure of ε-Thompson sampling and based on the structure of the risky-safe arm problem. We condition on $\mathcal{H}' = \mathcal{H} \cup \{a = \text{risky}\}$ for the remainder of the analysis and show that the $t-1$th hybrid achieves strictly higher reward than the tth hybrid.

To show this, let $K_t(q)$ denote the cumulative discounted reward incurred from time step t to time step T if the posterior at the start of time step t is q. Let p be the posterior given by conditioning on $\mathcal{H}$ (which is the same as conditioning on $\mathcal{H}'$). Let p' denote the distribution over posteriors given by updating with an additional sample from the risky arm. Showing that he $t-1$th hybrid achieves strictly higher reward than the tth hybrid reduces to showing that $\mathbb{E}[K_t(p')] > K_t(p)$. Since the discounting is geometric, this is equivalent to showing that $\mathbb{E}[K_1(p')] > K_1(p)$, which follows from Lemma 180 as desired.

Proof of side information monotonicity. We apply the same high-level approach as in the proof of strict information monotonicity, but construct a different set of hybrids. For $0 \leq t \leq T$, let the tth hybrid correspond to the bandit instance where both the user who plays algorithm A and the other user also updates the shared information state with the algorithm A' in the first t time steps and only the single user playing A updates the information state in the remaining time steps.

For each $2 \leq t \leq T$, it suffices to show that the $t-1$th hybrid incurs larger cumulative discounted reward than the $t-2$th hybrid. Notice a user participating in both of these hybrids at all time steps achieves the same expected reward in the first $t-1$ time steps for these two hybrids. The first time step where the two hybrids deviate in expectation is the tth time step (after the additional information from the $t-1$th hybrid has propagated into the information state). Thus, it suffices to show that $t-1$th hybrid incurs larger cumulative discounted reward than the $t-2$th hybrid between time steps t and T. Let $\mathcal{H}$ be the history of actions and observed rewards from the first $t-2$ time steps, the arm and the reward for the 1st user at the $t-1$th time step. We condition on $\mathcal{H}$ for this analysis. Let a denote the arm chosen by the 2nd user at the $t-1$th time step in the $t-1$th hybrid. We split into cases based on a.

The first case is that a is the safe arm. In this case, the hybrids have the same expected reward for time steps t to T.

The second case is a is the risky arm. If this happens with zero probability conditioned on $\mathcal{H}$, we are done. Otherwise, we condition on $\mathcal{H}' = \mathcal{H} \cup \{a = \text{risky}\}$ for the remainder of the analysis and show that the $t-1$th hybrid achieves higher reward than the tth hybrid.

To show this, let $K_t(q)$ denote the cumulative discounted reward incurred from time step t to time step T if the posterior at the start of time step t is q. Let p be the posterior given by conditioning on $\mathcal{H}$ (which is the same as conditioning on $\mathcal{H}'$). Let p' denote the distribution over posteriors given by updating with an additional sample from the risky arm. Showing that he $t-1$th hybrid achieves strictly higher reward than the tth hybrid reduces to showing that $\mathbb{E}[K_t(p')] > K_t(p)$. Since the discounting is geometric, this is equivalent to showing that $\mathbb{E}[K_1(p')] > K_1(p)$, which follows from Lemma 180 as desired.

□

C.6.3 Proof of Lemma 24

We prove Lemma 24.

Proof of Lemma 24. Applying Lemma 176, we see that:

$$R_A(n) = K(p; A, \ldots, A) = \int G(p,q) \frac{(1-A(q))s + (qh + (1-q)l)A(q)}{\frac{\sigma^2}{\sigma_b^2} + nA(q)} dq.$$

It follows immediately that the set $\{R_A(N) \mid A \in \mathcal{A}_{\text{all}}^{\text{cont}}\}$ is connected. To see that the supremum is achieved, we use the characterization in Bolton and Harris (2000b) of the team optimal as a cutoff strategy within $\mathcal{A}_{\text{all}}^{\text{cont}}$. It thus suffices to show that there exists A' such that $R_{A'}(N) \leq \max_{A \in \mathcal{A}} R_A(1)$. To see this, let $f_{1-\varepsilon}$ be the cutoff strategy where $f_{1-\varepsilon}(p) = 1$ for $p \geq 1-\varepsilon$ and $f_{1-\varepsilon}(p) = 0$ for $p < 1-\varepsilon$. We see that there exists $\varepsilon > 0$ such that $R_{A_{f_{1-\varepsilon}}}(N) \leq \max_{A \in \mathcal{A}} R_A(1)$.

□

C.6.4 Proof of Lemma 25

We prove Lemma 25.

Proof of Lemma 25. First, we show that there exists $A' \in \mathcal{A}$ such that $R_{A'}(N) \leq \max_{A \in \mathcal{A}} R_A(1)$. To see this, notice that:

$$R_{A_1}(N) = R_{A_1}(1) \leq \max_{A \in \mathcal{A}_{closure}} R_A(1)$$

since the reward of uniform exploration is independent of the number of other users.

It now suffices to show that the set $\{R_{A_\varepsilon}(N) \mid \varepsilon \in [0,1]\}$ is closed for every $A \in \mathcal{A}$.

To analyze the expected β-discounted utility of an algorithm, it is convenient to formulate it in terms of the sequences of actions and rewards observed by the algorithm. Let the realized history denote the sequence $\mathcal{H} = (a_1^1, o_1^1), \ldots, (a_1^N, o_1^N), \ldots, (a_T^1, o_T^1), \ldots, (a_T^N, o_T^N)$ of (pulled arm, observed reward) pairs observed at each time step. An algorithm A' induces a distribution $\mathcal{D}_{A'}$ over realized histories (that depends on the prior distributions $\mathcal{D}_i^{\text{Prior}}$). If we let

$$f((a_1^1, o_1^1), \ldots, (a_1^N, o_1^N), \ldots, (a_T^1, o_T^1), \ldots, (a_T^N, o_T^N)) := \sum_{t=1}^{T} r_{a_t^1} \beta^t,$$

then the expected β-discounted cumulative reward of an algorithm A' is:

$$\begin{aligned} R_A(N) &= \mathbb{E}_{(a_1^1, o_1^1), \ldots, (a_1^N, o_1^N), \ldots, (a_T^1, o_T^1), \ldots, (a_T^N, o_T^N) \sim \mathcal{D}_{A'}} \left[\sum_{t=1}^{T} r_{a_t^1} \beta^t \right] \\ &= \mathbb{E}_{\mathcal{H} \sim \mathcal{D}_{A'}} \left[f(\mathcal{H}) \right]. \end{aligned}$$

Since the mean rewards are bounded, we see that:

$$f((a_1^1, o_1^1), \ldots, (a_1^N, o_1^N), \ldots, (a_T^1, o_T^1), \ldots, (a_T^N, o_T^N)) \in \left[\min_{1 \leq i \leq k} r_i, \max_{1 \leq i \leq k} r_i \right].$$

Now, notice that the total-variation distance

$$TV(\mathcal{D}_{A_{\varepsilon_1}}, \mathcal{D}_{A_{\varepsilon_2}}) \leq 1 - (\varepsilon_1 - \varepsilon_2)^{NT}$$

because with $(\varepsilon_1 - \varepsilon_2)^{NT}$ probability, A_{ε_1} behaves identically to A_{ε_2}. We can thus conclude that:

$$\begin{aligned} |R_{A_{\varepsilon_1}}(N) - R_{A_{\varepsilon_2}}(N)| &\leq \left| \max_{1 \leq i \leq k} r_i - \min_{1 \leq i \leq k} r_i \right| TV(\mathcal{D}_{A_{\varepsilon_1}}, \mathcal{D}_{A_{\varepsilon_2}}) \\ &\leq \left| \max_{1 \leq i \leq k} r_i - \min_{1 \leq i \leq k} r_i \right| \left(1 - (\varepsilon_1 - \varepsilon_2)^{NT}\right). \end{aligned}$$

This proves that R_{A_ε} changes continuously in ε which proves the desired statement. $\square$

Appendix D

Appendix for Chapter 6

D.1 Additional discussion

D.1.1 Comparison with measures of market power in economics

Traditional measures of market power in economic theory are based on classical markets of homogeneous goods, where a firm's primary action is choosing a price to sell the good or the quantity of the good to sell. The scalar nature of these quantities enables them to be easily compared across different market contexts and firms. In addition, the utility of the firm and the utility of participants are inversely related: a higher price yields greater utility for the firm and lower utility for all participants. This simple relationship enables directly reasoning about participant welfare and profit of firms. However, a digital economy is much more complex (Stigler Committee, 2019; Crémer et al., 2019) and classical measures can struggle to accurately characterize these economies. As an example, consider the following two textbook definitions of market power.

- **Lerner index.** The Lerner index (Lerner, 1934) quantifies the pricing power of a firm, measuring by how much the firm can raise the price above marginal costs. Marginal costs reflect the price that would arise in a perfectly competitive market. A major issue of applying this standard definition of market power to digital economies is that it is not clear what the competitive reference state should look like, "We have lost the competitive benchmark,"[1] as Jacques Crémer said. Thus, measures based on profit margin cannot directly be adopted as a proxy for market power in digital economies.
- **Market share.** Measures such as the Herfindahl–Hirschman index (HHI), which is used by the US federal trade commission[2] to measure market competitiveness, are based on *market share*: the fraction of participants who participate in a given firm.

[1]Opening statement at the 2019 Antitrust and competition conference – digital platforms, markets, and democracy

[2]See `https://www.justice.gov/atr/herfindahl-hirschman-index` (retrieved January, 2022).

> However, the validity of market share as a proxy for power relies on a specific model of competition where the elasticity of demand is low. This model is challenging to justify[3] in the context of digital economies where opening an account on a platform is very simple and usually free of charge. In addition, not all participants with accounts on a digital platform are equally active and inactive participants should not factor into the market power of a firm in the same way as active participants. Market share is not sufficiently expressive to make this distinction.

In contrast, performative power is a causal notion of influence that does not require a precise specification of the market but is still sensitive to the nuances of the market. The definition therefore could serve as a useful tool in markets that resist a clean mathematical specification.

Behavioral aspects. Complex consumer behavior that plays a critical role in digital marketplaces. As outlined by the Stigler Committee (2019), "the findings from behavioral economics demonstrate an under-recognized market power held by incumbent digital platforms." In particular, behavioral aspects of consumers—such as tendencies for single-homing, vulnerability to addiction, and the impact of framing and nudging on participant behavior (e.g. Thaler and Sunstein, 2008; Fogg, 2002)—can be exploited by firms in digital economies, but do not factor into traditional measures of market power. By focusing on changes in participant features, performative power has the potential to capture the effects of these behavioral patterns while again sidestepping the challenges of explicitly modeling them.

D.1.2 From performative power to consumer harm

Performative power focuses on measuring power rather than harm. The relationship to harm depends on the choice and interpretation of the outcome variable and requires additional substantive arguments. In general, this connection can be achieved if the attributes $z(u)$ consists of the sensitive features that are impacted by the firm, the distance function is aligned with the utility function of participants, and the set $\mathcal{F}$ reflects actions that are taken by the firm. We implement this strategy to establish an exact correspondence between performative power and harm in the strategic classification setup.

Relating performative power and user burden in strategic classification. The fact that a monopoly firm has nonzero performative power has consequences for the optimization strategies that it would use, as we discussed in Chapter 6.3. To make this explicit, let's contrast the solutions of ex-ante and ex-post optimization in a simple one-dimensional setting.

Example 13 (1-dimensional setting)**.** *Consider a 1-dimensional feature vector $x \in \mathrm{R}$ and suppose that the posterior $p(x) = \mathbb{P}[Y = 1 \mid X = x]$ is strictly increasing in x with*

[3]This critique is similar to the disconnect between the Cournot model and the Bertrand model in classical economics (de Bornier, 1992). E.g., "concentration is worse than just a noisy barometer of market power" (Syverson, 2019).

$\lim_{x\to-\infty} p(x) = 0$, and $\lim_{x\to\infty} p(x) = 1$. Now consider a set of actions $\mathcal{F}$ that corresponds to the set of all threshold functions and set $\mathrm{dist}(x, x') = c(x, x') = |x - x'|$. Let θ_{SL} be the supervised learning threshold from ex-ante optimization, which is the unique value where $p(\theta_{\mathrm{SL}}) = 0.5$. Then, the ex-post threshold lies at $\theta_{\mathrm{PO}} = \theta_{SL} + \Delta\gamma$.

In Example 13 ex-post optimization leads to a higher acceptance threshold than ex-ante optimization. Thus, for any setting where the participants utility is decreasing in the threshold (e.g., the class of utility functions that Milli et al. (2019) call *outcome monotonic*), this implies that ex-post optimization creates stronger negative externalities for participants than ex-ante optimization. Furthermore, the effect grows with the performative power of the firm. In the extreme case of the monopoly setting with no outside options, ex-post optimization can leave certain participants with a net utility of 0 and thus can transfer the entire utility from these participants to the firm.

D.1.3 Monopoly power in heterogeneous setting

Different participants are typically impacted differently by a classifier, depending on their relative position to the decision boundary, as visualized in Figure 6.1a. As a result of this heterogeneity, the upper bound in (6.5) is not necessarily tight, because the firm can not extract the full utility from all participants simultaneously.

We investigate the effect of heterogeneity in a concrete 1-dimensional setting where $\mathrm{dist}_X(x, x') = c(x, x') = |x - x'|$. Consider a set of actions $\mathcal{F}$ that corresponds to the set of all threshold functions. Suppose that the posterior $p(x) = \mathbb{P}[Y = 1 \mid X = x]$ satisfies the following regularity assumptions: $p(x)$ is strictly increasing in x with $\lim_{x\to-\infty} p(x) = 0$, and $\lim_{x\to\infty} p(x) = 1$. Now, let θ_{SL} be the supervised learning threshold, which is the unique value where $p(\theta_{\mathrm{SL}}) = 0.5$. We can then obtain the following bound on the performative power P with respect to any $\mathcal{F}$ assuming the firm's classifier is θ_{SL} in the current economy (see Proposition 181):

$$0.5\Delta\gamma \underset{\mathcal{D}_{\mathrm{orig}}}{\mathbb{P}} \left[x \in [\theta_{\mathrm{SL}}, \theta_{\mathrm{SL}} + 0.5\Delta\gamma]\right] \leq \mathrm{P} \leq \Delta\gamma\,. \tag{D.1}$$

This bound illustrates how performative power in strategic classification depends on the fraction of participants that fall in between the old and the new threshold. As long as the density in this region is non-zero, a platform that offers $\Delta\gamma > 0$ utility will also have strictly positive performative power, providing a lower bound on P.

Proposition 181. *Suppose that $\mathrm{dist}(x, x') = c(x, x') = |x - x'|$. Consider a set of actions $\mathcal{F}$ that corresponds to the set of all threshold functions. Suppose that the posterior $p(x) = \mathbb{P}[Y = 1 \mid X = x]$ satisfies the following regularity assumptions: $p(x)$ is strictly increasing in x with $\lim_{x\to-\infty} p(x) = 0$, and $\lim_{x\to\infty} p(x) = 1$. Now, let θ_{SL} be the supervised learning threshold, which is the unique value where $p(\theta_{\mathrm{SL}}) = 0.5$. If the firm's classifier is θ_{SL} in the current economy, then performative power P with $\mathcal{F}$ can be bounded as:*

$$0.5\gamma \underset{\mathcal{D}_{\mathrm{orig}}}{\mathbb{P}} \left[x \in [\theta_{\mathrm{SL}}, \theta_{\mathrm{SL}} + 0.5\Delta\gamma]\right] \leq \mathrm{P} \leq 2\Delta\gamma\,. \tag{D.2}$$

D.1.4 Background on the search advertisement study

We provide additional context on the search advertisement study by Narayanan and Kalyanam (2015) on which we build the establish a lower-bound on performative power. They examine position effects in search advertising, where advertisements are displayed across a number of ordered slots whenever a keyword is searched. They show that the position effect of display slot 1 versus display slot 2 is 0.0048 clicks per impression (see Table 2 in their manuscript).

To arrive at this number, the authors implemented a regression discontinuity approach to estimate the position effect. The input is a sample of data (k, p, z, y) where k is a keyword, $p \in \{1, 2\}$ is the position of the advertisement in the list of displayed content, z is the AdRank score, and y is the click-through-rate (CTR). The following local linear regression estimator

$$y = \alpha + \xi \mathrm{I}[p = 1] + \gamma_1 z + \gamma_2 z \mathrm{I}[p = 1] + g(k) \tag{D.3}$$

is applied to a subset of the data within an appropriate window size $\lambda > 0$ around the threshold for fitting $\alpha, \xi, \gamma_1, \gamma_2, g$.

We are interested in the value ξ which is an estimate of the *position effect* of the display slot. To connect the causal effect estimate ξ to the causal effect β as in Definition 4 we treat each incoming keyword query as a distinct "viewer". Following the query, the viewer u either clicks on the advertisement in one of the display slots or does not click on any advertisement. The value $z_s(u)[i]$ corresponds to the probability that item i is consumed by viewer u under the scoring rule s. For $i = 0$, the value $z_s(u)[0]$ corresponds to the probability that the viewer does not click on any advertisement. If item i is displayed, $z_s(u)[i]$ corresponds to the click-through-rate. Hence $\beta = \gamma$ and $P \geq \gamma$.

D.2 Proofs

D.2.1 Auxiliary results

The proofs for Chapter 6.3 leverage the following lemma, which bounds the diameter of Θ with respect to Wasserstein distance in distribution map.

Lemma 182. *Let* P *be the performative power with respect to* Θ. *For any* $\theta, \theta' \in \Theta$, *it holds that* $\mathcal{W}(\mathcal{D}(\theta), \mathcal{D}(\theta')) \leq 2\mathrm{P}$.

Proof. Let θ_{curr} be the current classifier weights. We use the fact that for any weights $\theta'' \in \Theta$, it holds that $\mathcal{W}(\mathcal{D}(\theta_{\text{curr}}), \mathcal{D}(\theta'')) \leq \frac{1}{|U|} \sum_{u \in \mathcal{U}} \mathbb{E}[\text{dist}(z(u), z_{\theta''}(u))]$ where the expectation is over randomness in the potential outcomes. This follows from the definition of Wasserstein distance—in particular that we can instantiate the mass-moving function by mapping each participant to themselves. Thus, we see that:

$$\begin{aligned}
\mathcal{W}(\mathcal{D}(\theta), \mathcal{D}(\theta')) &\leq \mathcal{W}(\mathcal{D}(\theta), \mathcal{D}(\theta_{\text{curr}})) + \mathcal{W}(\mathcal{D}(\theta_{\text{curr}}), \mathcal{D}(\theta')) \\
&\leq \frac{1}{|U|}\sum_{u\in\mathcal{U}} \mathbb{E}[\text{dist}(z(u), z_\theta(u))] + \frac{1}{|U|}\sum_{u\in\mathcal{U}} \mathbb{E}[\text{dist}(z(u), z_{\theta'}(u))] \\
&\leq 2 \sup_{\theta''\in\Theta} \frac{1}{|U|}\sum_{u\in\mathcal{U}} \mathbb{E}[\text{dist}(z(u), z_{\theta''}(u))] \\
&\leq 2\text{P},
\end{aligned}$$

where the last line uses the definition of performative power that bounds the effect of any θ in the action set Θ on the participant data z. $\square$

D.2.2 Proof of Proposition 26

Let ϕ be the previous deployment inducing the distribution on which the supervised learning threshold θ_{SL} is computed. Let θ^* be an optimizer of $\min_{\theta\in\Theta} \text{R}(\theta_{\text{PO}}, \theta)$, where we recall the definition of the decoupled performative risk as $\text{R}(\phi, \theta) := \mathbb{E}_{z\sim\mathcal{D}(\phi)}\, \ell(\theta;\, z)$. Then, we see that for any ϕ:

$$\begin{aligned}
&\text{PR}(\theta_{\text{SL}}) - \text{PR}(\theta_{\text{PO}}) \\
&= (\text{R}(\theta_{\text{SL}}, \theta_{\text{SL}}) - \text{R}(\phi, \theta_{\text{SL}})) + \text{R}(\phi, \theta_{\text{SL}}) - \text{R}(\theta_{\text{PO}}, \theta_{\text{PO}}) \\
&\leq (\text{R}(\theta_{\text{SL}}, \theta_{\text{SL}}) - \text{R}(\phi, \theta_{\text{SL}})) + \text{R}(\phi, \theta^*) - \text{R}(\theta_{\text{PO}}, \theta^*) \\
&\leq L_z \mathcal{W}(\mathcal{D}(\theta_{\text{SL}}), \mathcal{D}(\phi)) + L_z \mathcal{W}(\mathcal{D}(\phi), \mathcal{D}(\theta_{\text{PO}})) \\
&\leq 4 L_z P\,.
\end{aligned}$$

The first inequality follows because θ^* minimizes risk on the distribution $\mathcal{D}(\theta_{\text{PO}})$, while θ_{SL} minimizes risk on $\mathcal{D}(\phi)$. The second inequality follows from the dual of the Wasserstein distance where L_z is the Lipschitz constant of the loss function in the data argument z. The last inequality follows from Lemma 182.

Now, suppose that ℓ is γ-strongly convex. Then we have that:

$$\text{R}(\theta, \theta_{\text{PO}}) - \text{R}(\theta, \theta_{\text{SL}}) \geq \frac{\gamma}{2}\|\theta_{\text{PO}} - \theta_{\text{SL}}\|^2$$

Again applying Lemma 182,

$$\begin{aligned}
\text{PR}(\theta_{\text{SL}}) &= \text{R}(\theta_{\text{SL}}, \theta_{\text{SL}}) \\
&\leq \text{R}(\theta, \theta_{\text{SL}}) + L_z \mathcal{W}(\mathcal{D}(\phi), \mathcal{D}(\theta_{\text{SL}})) \\
&\leq \text{R}(\phi, \theta_{\text{SL}}) + 2L_z P \\
&\leq \text{R}(\phi, \theta_{\text{PO}}) - \frac{\gamma}{2}\|\theta_{\text{PO}} - \theta_{\text{SL}}\|^2 + 2L_z P \\
&\leq \text{R}(\theta_{\text{PO}}, \theta_{\text{PO}}) + L_z \cdot \mathcal{W}(\mathcal{D}(\phi), \mathcal{D}(\theta_{\text{PO}})) - \frac{\gamma}{2}\|\theta_{\text{PO}} - \theta_{\text{SL}}\|^2 + 2L_z P \\
&\leq \text{PR}(\theta_{\text{PO}}) + 4L_z P - \frac{\gamma}{2}\|\theta_{\text{PO}} - \theta_{\text{SL}}\|^2.
\end{aligned}$$

Using that $\mathrm{PR}(\theta_{\mathrm{PO}}) \leq \mathrm{PR}(\theta_{\mathrm{SL}})$, we find that

$$\frac{\gamma}{2}\|\theta_{\mathrm{PO}} - \theta_{\mathrm{SL}}\|^2 \leq 4L_z P\,.$$

Rearranging gives

$$\|\theta_{\mathrm{PO}} - \theta_{\mathrm{SL}}\| \leq \sqrt{\frac{8L_z P}{\gamma}}\,.$$

D.2.3 Proof of Proposition 27

Let's focus on firm i, fixing classifiers selected by the other firms. Let's take PR and R to be defined with respect to $\mathcal{D}(\cdot) = \mathcal{D}(\theta_1, \ldots, \theta_{i-1}, \cdot, \theta_{i+1}, \ldots, \theta_C)$. Let $\theta^* = \arg\min_\theta \mathrm{R}(\theta_i, \theta)$. We see that:

$$\begin{aligned}
\mathrm{PR}(\theta^i) &\leq \mathrm{PR}(\theta^*) \\
&\leq \mathrm{R}(\theta_i, \theta^*) + L_z \mathcal{W}(\mathcal{D}(\theta_i), \mathcal{D}(\theta^*)) \\
&\leq \min_\theta \mathrm{R}(\theta_i, \theta) + L_z \left(\frac{1}{|\mathcal{U}|}\sum_{u \in \mathcal{U}} \mathbb{E}[\mathrm{dist}(z(u), z_{\theta^*}(u))]\right) \\
&\leq \min_\theta \mathrm{R}(\theta_i, \theta) + L_z P_i.
\end{aligned}$$

Rewriting this, we see that:

$$\mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell_i(\theta^i;\, z)] \leq \min_\theta \mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell_i(\theta;\, z)] + L_z P_i.$$

If, in addition, ℓ_i is γ-strongly convex, then we know that:

$$L_z P_i \geq \mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell_i(\theta^i;\, z)] - \min_\theta \mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell(\theta;\, z)] \geq \frac{\gamma}{2}\|\theta^i - \min_\theta \mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell_i(\theta;\, z)]\|^2.$$

Rearranging, we obtain that

$$\left\|\theta^i - \min_\theta \mathop{\mathbb{E}}_{z\sim\mathcal{D}}[\ell(\theta;\, z)]\right\|_2 \leq \sqrt{\frac{2L_z P_i}{\gamma}}\,.$$

D.2.4 Proof of Corollary 28

Let P be the performative power associated with the variables $z_\theta^{C=1}$. We first claim that the performative power of any firm in the mixture model is at most P/C. This follows from the fact that for a given firm the potential outcome $z_\theta(u)$ is equal to $z(u)$ with probability $1 - 1/C$ and equal to $z_\theta^{C=1}(u)$ with probability $1/C$.

Let's focus on platform i, fixing classifiers selected by the other platforms. Let's take PR and R to be with respect to $\mathcal{D}(\cdot) = \mathcal{D}(\theta^*, \cdots, \theta^*, \cdot, \theta^*, \ldots, \theta^*)$. Now, we can apply Proposition 27 to see that

$$\mathrm{PR}(\theta^*) = \mathop{\mathbb{E}}_{z\sim\mathcal{D}^{C=1}(\theta^*)}[\ell(\theta^*; z)] \leq \min_{\theta} \mathop{\mathbb{E}}_{z\sim\mathcal{D}^{C=1}(\theta^*)}[\ell(\theta; z)] + \frac{L_z P}{C}.$$

Thus, in the limit as $C \to \infty$, it holds that

$$\mathop{\mathbb{E}}_{z\sim\mathcal{D}^{C=1}(\theta^*)}[\ell(\theta^*; z)] \to \min_{\theta} \mathop{\mathbb{E}}_{z\sim\mathcal{D}^{C=1}(\theta^*)}[\ell(\theta; z)]$$

as desired.

D.2.5 Proof of Lemma 29

Fix a classifier f and a unit u. By Assumption 3, we know that $x(u)$ and $x_f(u)$ are both in $\mathcal{X}_{\Delta\gamma}(u)$. The claim follows from

$$\mathrm{dist}(x_{\mathrm{orig}}(u), x_f(u)) \leq \sup_{x' \in \mathcal{X}_{\Delta\gamma}(u)} \mathrm{dist}(x_{\mathrm{orig}}(u), x').$$

D.2.6 Proof of Proposition 30

The proof is by construction of a classifier $f^* : \mathrm{R}^m \to \{0, 1\}$. For each individual u we define the set

$$\tilde{\mathcal{X}}(u) := \mathop{\arg\sup}_{x' \in \mathcal{X}_{\Delta\gamma}(u)} \mathrm{dist}(x(u), x').$$

Now let f^* be such that

$$f^*(x) = \begin{cases} 1 & x \in \tilde{\mathcal{X}}(u) \text{ with } u = x[1] \\ 0 & x \notin \tilde{\mathcal{X}}(u) \text{ with } u = x[1] \end{cases}$$

where we used that the first coordinate of the feature vector x uniquely identifies the individual. The effect of f^* on a population $\mathcal{U}$ corresponds to

$$\frac{1}{|\mathcal{U}|} \sum_{u\in\mathcal{U}} \mathrm{dist}(x(u), x_{f^*}(u)) = \frac{1}{|\mathcal{U}|} \sum_{u\in\mathcal{U}} \sup_{x' \in \mathcal{X}_{\Delta\gamma}(u)} \mathrm{dist}(x(u), x') \tag{D.4}$$

Thus for any $\mathcal{F}$ that contains f^* the performative power is maximized.

D.2.7 Proof of Corollary 31

Applying Lemma 29, it suffices to show that the diameter of the set $\mathcal{X}_{\Delta\gamma}(u)$ can be upper bounded by $2L\Delta\gamma$ for any $u \in \mathcal{U}$. We see that for any $x, x' \in \mathcal{X}_{\Delta\gamma}(u)$, it holds that:

$$\begin{aligned}\operatorname{dist}(x, x') &\leq L \cdot c(x, x') \\ &\leq L \cdot (c(x_{\text{orig}}(u), x) + c(x_{\text{orig}}(u), x')) \\ &\leq 2L\Delta\gamma,\end{aligned}$$

using that c is a metric.

D.2.8 Proof of Proposition 32

To prove this proposition, we show the following two intermediate results which are proven in the next two sections:

Proposition 183. *Consider the 1-dimensional setup specified in Chapter 6.4.3, and suppose that the economy is at a symmetric state where both firms choose classifier θ. For any $\mathcal{F}$, consider one of the firms, let $\mathcal{F}$ denote their action set and let $\theta_{\min}$ be the minimum threshold classifier in $\mathcal{F}$. Then, the performative power of the firm is upper bounded by:*

$$\mathrm{P} \leq L \min(c(\theta_{\min}, \theta), \gamma) + \gamma L p_{\text{reach}}([\theta_{\min}, \theta]).$$

where $p_{\text{reach}}([\theta_{\min}, \theta]) := \mathbb{P}_{\mathcal{D}_{\text{orig}}}[x \in [\xi(\theta_{\min}), \xi(\theta)]]$ with $\xi(\theta')$ being the unique value such that $\xi(\theta') < \theta'$ and $c(\xi(\theta'), \theta') = 1$.

Proposition 184. *Consider the 1-dimensional setup described in Chapter 6.4.3. Then, a symmetric solution $[\theta^*, \theta^*]$ is an equilibrium if and only if θ^* satisfies*

$$\mathbb{E}_{(x,y)\sim\mathcal{D}_{\text{orig}}}[y = 1 \mid x \geq \xi(\theta^*)] = \tfrac{1}{2}, \tag{D.5}$$

where $\xi(\theta^*)$ is the unique value such that $c(\xi(\theta^*), \theta^*) = \gamma$ and $\xi(\theta^*) < \theta^*$. Both firms earn zero utility at this equilibrium. Moreover, the set $\mathcal{F}^+(\theta^*)$ of actions that a firm can take at equilibrium that achieve nonnegative utility is exactly equal to $[\theta^*, \infty)$, assuming the other firm chooses the classifier θ^*.

We now prove Proposition 32 from these intermediate results. We apply Proposition 183 to see that the performative power is upper bounded by

$$B := L \min(c(\theta_{\min}, \theta^*), \gamma) + L\gamma p_{\text{reach}}([\theta_{\min}, \theta^*])$$

where (θ^*, θ^*) is a symmetric state. Using Proposition 27, we see that $\mathcal{F}(\theta^*) = [\theta^*, \infty)$. This means that $\theta_{\min} = \theta^*$, and so $\xi(\theta_{\min}) = \xi(\theta^*)$. Thus, $B = 0$ which demonstrates that the performative power is upper bounded by 0, and is thus equal to 0.

D.2.9 Proof of Proposition 183

Consider a classifier $f \in \mathcal{F}(\theta)$ with threshold θ', and suppose that a firm changes their classifier to f. It suffices to show that:

$$\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\mathbb{E}[\mathrm{dist}(x(u), x_f(u))] \leq L\min(c(\theta_{\min},\theta),\gamma) + L\gamma p_{\mathrm{reach}}([\theta_{\min},\theta]).$$

For technical convenience, we reformulate this in terms of the cost function c. Based on the definition of L, it suffices to show that:

$$\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\mathbb{E}[c(x(u), x_f(u))] \leq \min(c(\theta_{\min},\theta),\gamma) + \gamma p_{\mathrm{reach}}([\theta_{\min},\theta]).$$

Case 1: $\theta' > \theta$**.** Participants either are indifferent between θ and θ' or prefer θ to θ'. Due to the tie breaking rule, the firm will thus lose all of its participants. Thus, all participants will switch to the other firm and adapt their features to that firm which has threshold θ. This is the same behavior as these participants had in the current state, so $x_f(u) = x(u)$ for all participants u. This means that

$$\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\mathbb{E}[c(x(u), x_f(u))] = 0$$

as desired.

Case 2: $\theta' < \theta$**.** Participants either are indifferent between θ and θ' or prefer θ' to θ. Due to the tie breaking rule, the firm will thus gain all of the participants. We break into several cases:

$$\begin{cases} x_f(u) = x(u) = x_{\mathrm{orig}}(u) & \text{if } x_{\mathrm{orig}}(u) < \xi(\theta') \\ x_f(u) = \theta', x(u) = x_{\mathrm{orig}}(u) & \text{if } x_{\mathrm{orig}}(u) \in [\xi(\theta'), \min(\theta', \xi(\theta)))] \\ x_f(u) = x(u) = x_{\mathrm{orig}}(u) & \text{if } x_{\mathrm{orig}}(u) \in (\theta', \xi(\theta)) \\ x_f(u) = \theta', x(u) = \theta & \text{if } x_{\mathrm{orig}}(u) \in (\xi(\theta), \theta') \\ x_f(u) = x_{\mathrm{orig}}(u), x(u) = \theta & \text{if } x_{\mathrm{orig}}(u) \in [\max(\theta', \xi(\theta)), \theta] \\ x_f(u) = x(u) = x_{\mathrm{orig}}(u) & \text{if } x_{\mathrm{orig}}(u) \geq \theta. \end{cases}$$

The only cases that contribute to $\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\mathbb{E}[c(x(u), x_f(u))]$ are the second, fourth, and fifth cases. Thus, we can upper bound $\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\mathbb{E}[c(x(u), x_f(u))]$ by:

$$\underbrace{\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}\mid x_{\mathrm{orig}}(u)\in[\xi(\theta'),\min(\theta',\xi(\theta))]}\mathbb{E}[c(x(u), x_f(u))]}_{(A)} + \underbrace{\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}\mid x_{\mathrm{orig}}(u)\in(\xi(\theta),\theta')}\mathbb{E}[c(x(u), x_f(u))]}_{(B)}$$

$$+\underbrace{\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}|x_{\text{orig}}(u)\in[\max(\theta',\xi(\theta)),\theta]}\mathbb{E}[c(x(u),x_f(u))]}_{(C)}$$

For (A), we see that

$$\begin{aligned}(A) &= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}|x_{\text{orig}}(u)\in[\xi(\theta'),\min(\theta',\xi(\theta)))}\mathbb{E}[c(x_{\text{orig}}(u),\theta')]\\ &\leq \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}|x_{\text{orig}}(u)\in[\xi(\theta'),\min(\theta',\xi(\theta)))}\mathbb{E}[c(\xi(\theta'),\theta')]\\ &= \gamma\cdot\mathbb{P}_{\mathcal{D}_{\text{orig}}}[x\in[\xi(\theta'),\min(\theta',\xi(\theta))))]\\ &\leq \gamma\cdot\mathbb{P}_{\mathcal{D}_{\text{orig}}}[x\in[\xi(\theta'),\xi(\theta))].\end{aligned}$$

For (B), we see that:

$$\begin{aligned}(B) &= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}|x_{\text{orig}}(u)\in(\xi(\theta),\theta')}\mathbb{E}[c(\theta,\theta')]\\ &= c(\theta,\theta')\cdot\mathbb{P}_{\mathcal{D}_{\text{orig}}}[x\in(\xi(\theta),\theta')]\\ &= \min(c(\theta,\theta'),\gamma)\cdot\mathbb{P}_{\mathcal{D}_{\text{orig}}}[x\in(\xi(\theta),\theta')].\end{aligned}$$

For (C), we see that:

$$\begin{aligned}(C) &= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}|x_{\text{orig}}(u)\in[\max(\theta',\xi(\theta)),\theta]}\mathbb{E}[c(x_{\text{orig}}(u),\theta)]\\ &\leq \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}|x_{\text{orig}}(u)\in[\max(\theta',\xi(\theta)),\theta]}\mathbb{E}[\min\left(c(\theta',\theta),c(\xi(\theta),\theta)\right)]\\ &= \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}|x_{\text{orig}}(u)\in[\max(\theta',\xi(\theta)),\theta]}\mathbb{E}[\min\left(c(\theta',\theta),\gamma\right)]\\ &= \min(c(\theta',\theta),\gamma)\cdot\mathbb{P}_{\mathcal{D}_{\text{orig}}}[x\in[\max(\theta',\xi(\theta)),\theta]]\end{aligned}$$

Putting this all together, we obtain that:

$$\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\mathbb{E}[c(x(u),x_f(u))]\leq\gamma p_{\text{reach}}([\theta',\theta])+\min(c(\theta',\theta),\gamma)$$

for $p_{\text{reach}}([\theta',\theta]) := \mathbb{P}_{\mathcal{D}_{\text{orig}}}\left[x\in[\xi(\theta'),\xi(\theta)]\right]$ as desired. Since $\gamma p_{\text{reach}}([\theta',\theta])+\min(c(\theta',\theta),\gamma)$ is decreasing in θ', this expression is maximized when $\theta'=\theta_{\min}$. Thus we obtain an upper bound of

$$\gamma\cdot p_{\text{reach}}([\theta_{\min},\theta])+\min(c(\theta_{\min},\theta),\gamma)\,.$$

D.2.10 Proof of Proposition 184

The proof proceeds in two steps. First, we establish that $[\theta^*, \theta^*]$ is an equilibrium; next, we show that $[\theta, \theta]$ is not in equilibrium for $\theta \neq \theta^*$.

Establishing that $[\theta^*, \theta^*]$ is an equilibrium and $\mathcal{F}^+(\theta^*) = [\theta^*, \infty)$. First, we claim that $[\theta^*, \theta^*]$ is an equilibrium. At $[\theta^*, \theta^*]$, each participant chooses the first firm with 1/2 probability. The expected utility earned by a firm is:

$$\begin{aligned}
\frac{1}{2}\int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)(p(x) - (1 - p(x))\mathrm{d}x &= \int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)(p(x) - 0.5)\mathrm{d}x \\
&= \int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)p(x)\mathrm{d}x - 0.5\int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)\mathrm{d}x \\
&= \int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)\mathrm{d}x \left(\frac{\int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)p(x)\mathrm{d}x}{\int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)\mathrm{d}x} - \frac{1}{2}\right) \\
&= \left(\int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)\mathrm{d}x\right)\left(\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{\text{orig}}}[y = 1 \mid x \geq \xi(\theta)] - \frac{1}{2}\right) \\
&= 0\,.
\end{aligned}$$

If the firm chooses $\theta > \theta^*$, then since the cost function is strictly monotonic in its second argument, participants either are indifferent between θ and θ^* or prefer θ to θ^*. Due to the tie breaking rule, the firm will thus lose all of its participants and incur 0 utility. Thus the firm has no incentive to switch to θ.

If the firm chooses $\theta < \theta^*$, then it will gain all of the participants. The firm's utility will be:

$$\begin{aligned}
&\int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)(p(x) - (1 - p(x))\mathrm{d}x \\
&\qquad = \int_{\xi(\theta)}^{\xi(\theta^*)} p_{\text{orig}}(x)(p(x) - (1 - p(x))\mathrm{d}x + \int_{\xi(\theta^*)}^{\infty} p_{\text{orig}}(x)(p(x) - (1 - p(x))\mathrm{d}x \\
&\qquad = 2\int_{\xi(\theta)}^{\xi(\theta^*)} p_{\text{orig}}(x)(p(x) - 0.5)\mathrm{d}x.
\end{aligned}$$

It is not difficult to see that at θ^*, it must hold that $p(\xi(\theta^*)) \leq 0.5$. Since the posterior is strictly increasing, this means that $p(\xi(\theta)) < p(\xi(\theta^*)) = 0.5$, so the above expression is negative. This means that the firm will not switch to $\xi(\theta)$.

Moreover, this establishes that $\mathcal{F}(\theta*) = [\theta^*, \infty)$.

$[\theta, \theta]$ is not in equilibrium if $\xi(\theta^*)$ does not satisfy (D.5). If $\theta < \theta^*$, then the firm earns utility

$$\frac{1}{2}\left(\int_{\xi\theta}^{\infty} p_{\text{orig}}(x)(p(x) - (1 - p(x))\right)\mathrm{d}x\,,$$

which we already showed above was negative. Thus, the firm has incentive to change their threshold to above θ so that it loses the full participant base and gets 0 utility.

If $\theta > \theta^*$, then the firm earns utility

$$U = \frac{1}{2}\left(\int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)(p(x) - (1-p(x))\right)\mathrm{d}x\,,$$

which is strictly positive. Fix $\varepsilon > 0$, and suppose that the firm changes to a threshold θ' such that $c(\theta', \theta) = \varepsilon$. Then it would gain all of the participants and earn utility:

$$\begin{aligned}\int_{\xi(\theta')}^{\infty} p_{\text{orig}}(x)(p(x) - (1-p(x))\mathrm{d}x &= \int_{\xi(\theta')}^{\xi\theta} p_{\text{orig}}(x)(p(x) - (1-p(x))\mathrm{d}x \\ &\quad + \int_{\xi(\theta)}^{\infty} p_{\text{orig}}(x)(p(x) - (1-p(x))\mathrm{d}x \\ &= \int_{\xi(\theta')}^{\xi(\theta)} p_{\text{orig}}(x)(p(x) - (1-p(x))\mathrm{d}x + 2U.\end{aligned}$$

We claim that this expression approaches $2U$ as $\varepsilon \to 0$. To see this, note that $c(\xi(\theta'), \theta) \to \gamma$ and so $\xi(\theta') \to \xi(\theta)$ as $\varepsilon \to 0$. This implies that $\int_{\xi(\theta')}^{\xi(\theta)} p_{\text{orig}}(x)(p(x) - (1-p(x))\mathrm{d}x \to 0$ as desired. Thus, the expression approaches $2U > U$ as desired. This means that there exists ε such that the firm changing to θ' results in a strict improvement in utility.

D.2.11 Proof of Theorem 33

Recall the definition of the action set $\mathcal{S}$. We prove Theorem 33 by constructing a $s_{\text{swap}} \in \mathcal{S}$ and relating the effect of a change in the score function from s_{curr} to s_{swap} to the causal effect of position.

For $u \in \mathcal{U}$ let $i_1(u)$ and $i_2(u)$ denote the index of the content item shown to user u under s_{curr} in the first and second display slot, respectively. Now, let the score function s_{swap} be such that the content items displayed in the first two display slots are swapped relative to s_{curr}, simultaneously for all users $u \in \mathcal{U}$:

$$s_{\text{swap}}(u)[i] = \begin{cases} s_{\text{curr}}(u)[i_2(u)] & i = i_1(u) \\ s_{\text{curr}}(u)[i_1(u)] & i = i_2(u) \\ s_{\text{curr}}(u)[i] & \text{otherwise.} \end{cases} \tag{D.6}$$

It holds that $s_{\text{swap}} \in \mathcal{S}$, since $|s_{\text{curr}}(u)[i_1(u)] - s_{\text{curr}}(u)[i_2(u)]| \leq \delta$ for all $u \in \mathcal{U}$. We lower bound performative power as

$$\mathrm{P} = \sup_{s\in\mathcal{S}} \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \mathbb{E}\left[\|z(u) - z_s(u)\|_1\right] \geq \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} \mathbb{E}\left[\|z(u) - z_{s_{\text{swap}}}(u)\|_1\right] \tag{D.7}$$

To bound the difference between the counterfactual variable $z(u)$ and $z_{s_{\text{swap}}}(u)$, we decompose s_{swap} into a series of unilateral *swapped* score functions, one for each viewer. The score function s^u_{swap} associated with viewer u swaps the scores of content that currently appears in the first two display slots for viewer u and keeps the scores of the other viewers unchanged.

Assumption 4 implies that $z_{s_{\text{swap}}}(u) = z_{s^u_{\text{swap}}}(u)$, since there are no peer effects; $z_{s_{\text{swap}}}(u)$ is independent of $s_{\text{swap}}(u')$ for $u' \neq u$. Thus, we can aggregate the unilateral effects across all viewers $u \in \mathcal{U}$ to obtain the effect of s_{swap} as:

$$\mathrm{P} \geq \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}\left[\|z(u) - z_{s_{\text{swap}}}(u)\|_1\right] = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}\left[\|z(u) - z_{s^u_{\text{swap}}}(u)\|_1\right]. \tag{D.8}$$

Reasoning about unilateral effects allows us to relate the summands in (D.8) to the causal effect of position. In particular, focus on coordinate $i_1(u)$ in the norm, and let $Y_0(u) = z(u)[i_1(i)]$ and $Y_1(u) = z_{s^u_{\text{swap}}}(u)[i_1(u)]$. Then, we have

$$\mathrm{P} \geq \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}\left|z(u)[i_1] - z_{s^u_{\text{swap}}}(u)[i_1]\right| = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}\left|Y_0(u) - Y_1(u)\right| = \beta.$$

where the causal effect of position β is defined as in Definition 4.

D.2.12 Proof of Proposition 181

The upper bound follows from Corollary 32. For the lower bound, we take f to be the threshold classifier given by $\theta_{\text{SL}} + \Delta\gamma$. We see that for $x_{\text{orig}}(u) \in [\theta_{\text{SL}}, \theta_{\text{SL}} + \Delta\gamma]$, it holds that $x_f(u) = \theta_{\text{SL}} + \Delta\gamma$ and $x(u) = x_{\text{orig}}(u)$. This means that the performative power is at least:

$$\begin{aligned}
\mathrm{P} &= \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}[\mathrm{dist}(x(u), x_f(u))] \\
&= \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \mathbb{E}[|x(u) - x_f(u)|] \\
&\geq \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} I[x_{\text{orig}}(u) \in [\theta_{\text{SL}}, \theta_{\text{SL}} + \Delta\gamma]]\, \mathbb{E}[|\theta_{\text{SL}} + \Delta\gamma - x_{\text{orig}}(u)|] \\
&\geq \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} I[x_{\text{orig}}(u) \in [\theta_{\text{SL}}, \theta_{\text{SL}} + \frac{1}{2}\Delta\gamma]] \cdot \frac{1}{2}\Delta\gamma \\
&\geq \frac{1}{2}\Delta\gamma \mathop{\mathbb{P}}_{\mathcal{D}_{\text{orig}}}[x \in [\theta_{\text{SL}}, \theta_{\text{SL}} + \frac{1}{2}\Delta\gamma]],
\end{aligned}$$

as desired.

Appendix E

Appendix for Chapter 8

E.1 Details of the empirical setup in Chapter 8.3.4

The code can be found at `https://github.com/mjagadeesan/supply-side-equilibria`.

Dataset information. We use the MovieLens-100K dataset which consists of $N = 943$ users, 1682 movies, and 100,000 ratings (Harper and Konstan, 2015). We imported the dataset using the `scikit-surprise` library.

Calculation of user embeddings. For $D \in \{2, 3, 5, 10, 50\}$, we obtain D-dimensional user embeddings by running NMF (with D factors). In particular, we ran NMF using the `scikit-surprise` library on the full MovieLens-100K dataset with the default hyperparameters.

Calculation of single-genre equilibrium p^*. We calculate the single-genre equilibrium genre $p^* = \arg\max_{\|p\|=1 | p \in \mathbb{R}^D_{\geq 0}} \sum_{i=1}^N \log(\langle p, u_i \rangle)$. We write p^* as

$$p^* = \underset{\|p\| \leq 1 | p \in \mathbb{R}^D_{\geq 0}}{\arg\max} \sum_{i=1}^N \log(\langle p, u_i \rangle)$$

and solve the resulting optimization program. For $q = 2$, we directly use the `cvxpy` library with the default hyperparameters. For $q \neq 2$, we run projected gradient descent with learning rate 1.0 for 100 iterations where p is initialized as a standard normal clamped so all the coordinates are at least 1. The projection step onto $\|p\| \leq 1 \mid p \in \mathbb{R}^D_{\geq 0}$ uses the `cvxpy` library with the default hyperparameters.

Calculation of β_u. We directly calculate β_u according to the following formula:

$$\frac{\log(N)}{\log(N) - \log\left(\| \sum_{n=1}^N \frac{u_n}{\|u_n\|_*} \|_*\right)}.$$

Calculation of β_e. For this part, we first compute a restricted dataset with N randomly chosen users for computational tractability. We estimate β_e by binary searching with the lower bound initialized to 1 and the upper bound initialized to β_u, under the gap between the lower and upper bounds is $\leq \varepsilon = 0.05$. For each value β, we estimate whether the condition in (8.4) holds as follows. To compare the left-hand side $\max_{y \in \mathcal{S}^\beta} \prod_{i=1}^N y_i$ to the the right-hand side $\max_{y \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^N y_i$, we first compute $\tilde{p}^* = \mathrm{argmax}_{p \in \mathbb{R}^D_{\geq 0}, \|p\|=1} \prod_{i=1}^N \langle u_i, p \rangle$, we directly use the `cvxpy` library with the default hyperparameters. We then repeat the following procedure $T = 50$ times, which will correspond to estimating the convex hull $\bar{\mathcal{S}}^\beta$ with $T = 50$ different draws of randomly chosen vectors. In each trial $1 \leq t \leq 50$, we draw $m - 1 = 75$ unit q-norm random vectors in $\tilde{p}_1, \ldots, \tilde{p}_{m-1} \in \mathbb{R}^D_{\geq 0}$ by randomly sampling multivariate gaussians, taking the absolute value of the coordinates, and normalizing to have q-norm equal to 1. We let $\tilde{p}_m = \tilde{p}^*$ be the argmax computed previously. Then, for $1 \leq j \leq m$, we compute the vectors

$$\tilde{y}_j = \left[\frac{\langle u_1, \tilde{p}_j \rangle}{\langle u_1, \tilde{p}^*_j \rangle}, \ldots, \frac{\langle u_N, \tilde{p}_j \rangle}{\langle u_N, \tilde{p}^*_j \rangle} \right].$$

We then evaluate whether there exists $w \in \mathbb{R}^m_{\geq 0}, \sum_{j=1}^m w_j = 1$ such that $\sum_{j=1}^m \log(w_j \tilde{y}_j) \geq \tau$ using `cvxpy` library with the default hyperparameters except for the tolerance which is set to 10^{-7}. (The hyperparameter τ is equal to $0.85 \cdot 10^{-6}$ if $N = 20$, $1.5 \cdot 10^{-6}$ if $N = 30$, and $1.9 \cdot 10^{-6}$ if $N = 50$.) If the optimization program is feasible, we say that the trial passed; otherwise the trial failed. If the trial passes for *any* of the $T = 50$ trials, we interpret the condition in (8.4) as holding for that value of β.

E.2 Proofs for Chapter 8.2

E.2.1 Proof of Proposition 34

We restate and prove Proposition 34.

Proposition 34. *For any set of users and any $\beta \geq 1$, a pure strategy equilibrium does not exist.*

Proof of Proposition 34. Assume for sake of contradiction that the solution $p_1, \ldots, p_P$ is a pure strategy equilibrium. We divide into two cases based on whether there are ties. The cases are: (1) there exist $1 \leq j' \neq j \leq P$ and i such that $\langle p_j, u_i \rangle = \langle p_{j'}, u_i \rangle$, (2) there does not exist j, j' and i such that $\langle p_j, u_i \rangle = \langle p_{j'}, u_i \rangle$.

Case 1: there exist $1 \leq j' \neq j \leq P$ and i such that $\langle p_j, u_i \rangle = \langle p_{j'}, u_i \rangle$. Let producer j and producer j' be such that $\langle p_j, u_i \rangle = \langle p_{j'}, u_i \rangle$. The idea is that the producer j can leverage the discontinuity in their profit function (8.1) at p_j. In particular, consider the vector $p_j + \varepsilon u_i$. The number of users that they receive as $\varepsilon \to_+ 0$ is *strictly greater* than at p_j. The cost, on the other hand, is continuous in ε. This demonstrates that there exists $\varepsilon > 0$ such that:

$$\mathcal{P}(p_j + \varepsilon u_i; p_{-j}) > \mathcal{P}(p_j; p_{-j})$$

as desired. This is a contradiction.

Case 2: there does not exist j, j' and i such that $\langle p_j, u_i \rangle = \langle p_{j'}, u_i \rangle$. Since the sum of the expected number of users won by all of the producers is N, there exists a producer who wins a nonzero number of users in expectation. Let j be such a producer. Using the assumption that there are no ties (i.e. there does not exist j' and i such that $\langle p_j, u_i \rangle = \langle p_{j'}, u_i \rangle$), we know that producer j wins the following set of users:

$$\mathcal{N}_j := \{1 \le i \le N \mid \langle p_j, u_i \rangle > \langle p_{j'}, u_i \rangle \forall j' \neq j\}.$$

We see that $\mathcal{N}_j$ is nonempty by the assumption that producer j wins a nonzero number of users in expectation. We now leverage that the profit function of producer j is continuous at p_j. There exists $\varepsilon > 0$ such that $\langle p_j(1-\varepsilon), u_i \rangle > \langle p_{j'}, u_i \rangle$ for all $j' \neq j$ and all $i \in \mathcal{N}_j$, so that:

$$\mathcal{P}(p_j(1-\varepsilon); p_{-j}) > \mathcal{P}(p_j; p_{-j})$$

as desired. This is a contradiction.

□

E.2.2 Proof of Proposition 35

We restate and prove Proposition 35.

Proposition 35. *For any set of users and any $\beta \ge 1$, a symmetric mixed equilibrium exists.*

Proof of Proposition 35. We apply a standard existence result of symmetric, mixed strategy equilibria in discontinuous games (see Corollary 5.3 of (Reny, 1999)). We adopt the terminology of that paper and refer the reader to (Reny, 1999) for a formal definition of the conditions. Note that the game is symmetric by assumption, since the producers have symmetric utility functions. It suffices to show that: (1) the producer action space is convex and compact and (2) the game is diagonally better-reply secure.

Producer action space is convex and compact. In the current game, the producer action space is not compact. However, we show that we can define a slightly modified game, where the producer action space is convex and compact, without changing the equilibrium of the game. For the remainder of the proof, we analyze this modified game.

In particular, each producer must receive at least 0 profit at equilibrium since $\mathcal{P}(\vec{0}; p_{-1}) \ge 0$ regardless of the actions p_{-1} taken by other producers. If a producer chooses p such that $\|p\| > N^{1/\beta}$, then their utility will be strictly negative. Thus, we can restrict to $\left\{p \in \mathbb{R}^D_{\ge 0} \mid \|p\| \le 2N^{1/\beta}\right\}$ which is a convex compact set. We add a factor of 2 slack to guarantee that any best-response by a producer will be in the *interior* of the action space and not on the boundary.

Establishing diagonal better reply security. First, we show the payoff function $\mathcal{P}(\mu; [\mu, \ldots, \mu])$ (where μ is a distribution over the producer action space) is continuous in μ. Here we slightly abuse notation since $\mathcal{P}$ is technically defined over pure strategies in (8.1). We implicitly extend the definition to mixed strategies by considering expected profit. Using the fact that each producer receives a $1/P$ fraction of users in expectation at a symmetric solution, we see that:

$$\mathcal{P}(\mu; [\mu, \ldots, \mu]) = \frac{N}{P} - \int \|p\|^{\beta} d\mu.$$

Since the underlying topology on the set of distributions μ is the weak* topology, this implies continuity of the payoff.

Now, we construct, for each relevant payoff in the closure of the graph of the game's diagonal payoff function, an action that diagonal payoff secures that payoff. More formally, let (μ^*, α^*) be in the closure of the graph of the game's diagonal payoff function, and suppose that $(\mu^*, \ldots, \mu^*)$ is not an equilibrium. It suffices to show that a producer can secure a payoff of $\alpha > \alpha^*$ along the diagonal at $(\mu^*, \ldots, \mu^*)$. We construct μ^{sec} that secures a payoff of $\alpha > \alpha^*$ along the diagonal at $(\mu^*, \ldots, \mu^*)$.

Recall that $\alpha^* = \mathcal{P}(\mu^*, \ldots, \mu^*)$ by the continuity of the payoff function shown above. Since $(\mu^*, \ldots, \mu^*)$ is not an equilibrium, there exists $p \in \left\{p' \in \mathbb{R}^D_{\geq 0} \mid \|p'\| \leq N^{1/\beta}\right\}$ such that

$$\mathcal{P}(p; [\mu^*, \ldots, \mu^*]) > \mathcal{P}(\mu^*; [\mu^*, \ldots, \mu^*]) = \alpha^*.$$

Since we ultimately want to show that p achieves high profit in an open neighborhood of μ^*, we need to strengthen the above statement. We can achieve by this by appropriately perturbing p. First, we can perturb p to $\tilde{p}$ such that for each $1 \leq i \leq N$, the distribution $\langle p', u_i \rangle$ where $p' \sim \mu^*$ does not have a point mass at $\langle \tilde{p}, u_i \rangle$, and such that:

$$\mathcal{P}(\tilde{p}; [\mu^*, \ldots, \mu^*]) = \sum_{i=1}^{n} \left(\mathbb{P}_{p' \sim \mu^*}[\langle \tilde{p}, u_i \rangle > \langle p', u_i \rangle]\right)^{P-1} - c(\tilde{p}) > \alpha^*.$$

Now, we construct p^{sec} as a perturbation of $\tilde{p}$ to add ε slack to the constraint $\langle \tilde{p}, u_i \rangle > \langle p', u_i \rangle$. In particular, we observe that there exists $\varepsilon^* > 0$ and $p^{\text{sec}} \in \mathbb{R}^D_{\geq 0}$ such that

$$\mathcal{P}(p^{\text{sec}}; [\mu^*, \ldots, \mu^*]) \geq \sum_{i=1}^{n} \left(\mathbb{P}_{p' \sim \mu^*}[\langle p^{\text{sec}}, u_i \rangle > \langle p', u_i \rangle + \varepsilon^* \|u_i\|_2]\right)^{P-1} - c(p^{\text{sec}}) > \alpha^*. \tag{E.1}$$

We claim that μ^{sec} taken to be the point mass at p^{sec} will secure a payoff of

$$\alpha = \frac{\sum_{i=1}^{n} \left(\mathbb{P}_{p' \sim \mu^*}[\langle p^{\text{sec}}, u_i \rangle > \langle p', u_i \rangle + \varepsilon^* \|u_i\|_2]\right)^{P-1} - c(p^{\text{sec}}) + \alpha^*}{2} > \alpha^*$$

along the diagonal at $(\mu^*, \ldots, \mu^*)$. For each $1 \leq i \leq N$, we define the event A_i to be:

$$A_i = \{p' \mid \langle p^{\text{sec}}, u_i \rangle > \langle p', u_i \rangle\}$$

and define the event A_i^{ε} as:

$$A_i^{\varepsilon} = \{p' \mid \langle p^{\text{sec}}, u_i \rangle > \langle p', u_i \rangle + \varepsilon \|u_i\|_2\}.$$

In this notation, we can rewrite equation (E.1) as:

$$\mathcal{P}(p^{\text{sec}}; [\mu^*, \ldots, \mu^*]) \geq \sum_{i=1}^{n} \left(\mu^*(A_i^{\varepsilon^*})\right)^{P-1} - c(p^{\text{sec}}) > \alpha^*$$

and α as:

$$\alpha = \frac{\sum_{i=1}^{n} \left(\mu^*(A_i^{\varepsilon^*})\right)^{P-1} - c(p^{\text{sec}}) + \alpha^*}{2} > \alpha^*$$

Consider the metric on $\mathbb{R}^D_{\geq 0}$ given by the ℓ_2 norm. For $\varepsilon > 0$ let $B_\varepsilon(\mu^*)$ denote the ε-ball with respect to the Prohorov metric; using the definition of the weak* topology, we see that $B_\varepsilon(\mu^*)$ is an open set with respect to the weak* topology. For every $p' \in A_i^\varepsilon$, we see that A_i contains the open neighborhood $B_\varepsilon(p')$ with respect to the ℓ_2 norm. By the definition of the Prohorov metric, we know that for all $\mu' \in B_\varepsilon(\mu^*)$, it holds that

$$\mu'(A_i) \geq \mu^*(A_i^\varepsilon) - \varepsilon$$

This implies that

$$\mathcal{P}(p^{\text{sec}}; [\mu', \ldots, \mu']) \geq \sum_{i=1}^{n} \left(\mu'(A_i)\right)^{P-1} - c(p^{\text{sec}}) \geq \sum_{i=1}^{n} \left(\mu^*(A_i^\varepsilon) - \varepsilon\right)^{P-1} - c(p^{\text{sec}}).$$

Putting this all together, we see that:

$$\mathcal{P}(p^{\text{sec}}; [\mu', \ldots, \mu']) \geq \left(\sum_{i=1}^{n} \left(\mu^*(A_i^\varepsilon)\right)^{P-1} - c(p^{\text{sec}})\right) - \sum_{i=1}^{n} \left(\underbrace{\left(\mu^*(A_i^\varepsilon)\right)^{P-1} - \left(\mu^*(A_i^\varepsilon) - \varepsilon\right)^{P-1}}_{(A)}\right).$$

Using that (A) goes to 0 as ε goes to 0, we see that for sufficiently small ε, it holds that:

$$\sum_{i=1}^{n} \left(\left(\mu^*(A_i^\varepsilon)\right)^{P-1} - \left(\mu^*(A_i^\varepsilon) - \varepsilon\right)^{P-1}\right) \leq \frac{\mathcal{P}(p^{\text{sec}}; [\mu^*, \ldots, \mu^*]) - \alpha^*}{3}.$$

As long as ε is also less than ε^*, this means that:

$$\begin{aligned}
\mathcal{P}(p^{\text{sec}}; [\mu', \ldots, \mu']) &\geq \left(\sum_{i=1}^{n} \left(\mu^*(A_i^\varepsilon)\right)^{P-1} - c(p^{\text{sec}})\right) - \frac{\sum_{i=1}^{n} \left(\mu^*(A_i^\varepsilon)\right)^{P-1} - c(p^{\text{sec}}) - \alpha^*}{3} \\
&= \frac{2 \cdot \left(\sum_{i=1}^{n} \left(\mu^*(A_i^\varepsilon)\right)^{P-1} - c(p^{\text{sec}})\right) + \alpha^*}{3} \\
&\geq \frac{2 \cdot \left(\sum_{i=1}^{n} \left(\mu^*(A_i^{\varepsilon^*})\right)^{P-1} - c(p^{\text{sec}})\right) + \alpha^*}{3} \\
&> \alpha
\end{aligned}$$

for all $\mu' \in B_\varepsilon(\mu^*)$, as desired. $\square$

E.2.3 Proof of Proposition 36

In this proof, we consider the payoff function $\mathcal{P}(\mu_1; [\mu_2, \ldots, \mu_P])$ (where μ is a distribution over the producer action space) defined to be the expected profit attained if a producer plays μ_1 when other producers play $\mu_2, \ldots, \mu_P$. Strictly speaking, this is an abuse of notation since $\mathcal{P}$ is technically defined over pure strategies in (8.1). We implicitly extend the definition to mixed strategies by considering *expected* profit.

Proof of Proposition 36. Let μ be a symmetric equilibrium, and assume for sake of contradiction that there is an atom at $p \in \mathbb{R}^d$ with probability mass $\alpha > 0$. It suffices to construct a vector p' that achieves profit

$$\mathcal{P}(p'; [\mu, \ldots, \mu]) > \mathcal{P}(\vec{0}; [\mu, \ldots, \mu]) = \mathcal{P}(\mu; [\mu, \ldots, \mu]).$$

Consider the vector $p' = p + \varepsilon u_1$ for some $\varepsilon > 0$. For any given realization of actions by other producers, and for any given user, the vector p' never wins the user with lower probability than the vector p. We construct an event and a user where the vector p' wins the user with strictly higher probability than the vector p. Let E be the event that all of the other producers choose the p vector. This event happens with probability α^{P-1}. Conditioned on E, the vector p' wins user u_1; on the other hand, the vector p wins user u_1 with probability $1/P$. Since the cost function is continuous in ε, there exists ε such that $\mathcal{P}(p; [\mu, \ldots, \mu]) > \mathcal{P}(\vec{0}; [\mu, \ldots, \mu]) = \mathcal{P}(\mu; [\mu, \ldots, \mu])$. This is a contradiction.

□

E.2.4 Derivation of Example 3

We first apply Proposition 35 to see that a symmetric mixed equilibrium exists. Next, we use the fact that every symmetric mixed equilibrium equilibrium is by definition a single-genre equilibrium in 1-dimension. Finally, we apply Lemma 185 to see that the cumulative distribution function is $F(p) = \min(1, p^{\beta/P-1})$.

E.3 Proofs for Chapter 8.3

In Chapter E.3.1, we prove Theorem 38, and in Chapter E.3.2, we prove the corollaries of Theorem 38 in Chapter 8.3 (with the exception of Corollary 42, whose proof we defer to Chapter E.4.3).

E.3.1 Proof of Theorem 38

To prove Theorem 38, our main technical ingredient is Lemma 39, which shows that the existence of a single-genre equilibrium boils down to a minimax theorem and is restated below.

Lemma 39 (Informal)**.** *There exists a symmetric equilibrium* μ *with* $|\mathrm{Genre}(\mu)| = 1$ *if and only if:*

$$\inf_{y \in \mathcal{S}^\beta} \left(\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^{N} \frac{y'_i}{y_i} \right) = \sup_{y' \in \mathcal{S}^\beta} \left(\inf_{y \in \mathcal{S}^\beta} \sum_{i=1}^{N} \frac{y'_i}{y_i} \right). \tag{8.5}$$

Before diving into the proof of Lemma 39 and Theorem 38, we describe an intermediate result will be useful in the proof of Lemma 39. Suppose that there exists an equilibrium μ such that $\mathrm{Genre}(\mu) = \{p^*\}$ contains a single direction. Then μ is fully determined by the distribution over quality $\|p\|$ where $p \sim \mu$; therefore, let F denote the cdf of $\|p\|$ for $p \sim \mu$. We can derive a closed-form expression for F; in fact, we show that it is identical to the cdf of the 1-dimensional setup in Example 3.

Lemma 185. *Suppose that* μ *is a symmetric equilibrium such that* $\mathrm{Genre}(\mu)$ *contains a single vector. Let* F *be the cdf of the distribution over* $\|p\|$ *where* $p \sim \mu$*. Then, it holds that:*

$$F(r) = \min\left(1, \left(\frac{r^\beta}{N}\right)^{1/(P-1)}\right). \tag{E.2}$$

The intuition for Lemma 185 is that a single-genre equilibrium essentially reduces the producer's decision to a 1-dimensional space, and so inherits the structure of the 1-dimensional equilibrium.

To formalize the lemmas in this proof sketch, we will define a set $\mathcal{S}_{>0}$ which deletes all points with a zero coordinate from $\mathcal{S}$. More formally:

$$\mathcal{S}_{>0} := \left\{ \mathbf{U}p \mid \|p\| \leq 1, p \in \mathbb{R}^D_{\geq 0} \right\} \cap \mathbb{R}^N_{>0}.$$

For notational convenience, we also define:

$$\mathcal{B} := \left\{ p \in \mathbb{R}^D_{\geq 0} \mid \|p\| \leq 1 \right\},$$

$$\mathcal{B}_{>0} := \left\{ p \in \mathbb{R}^D_{\geq 0} \mid \|p\| \leq 1, \langle p, u_i \rangle > 0 \forall i \right\},$$

which are both convex sets. We further define:

$$\mathcal{D} := \left\{ p \in \mathbb{R}^D_{\geq 0} \mid \|p\| = 1 \right\}$$

and

$$\mathcal{D}_{>0} := \left\{ p \in \mathbb{R}^D_{\geq 0} \mid \|p\| = 1, \langle p, u_i \rangle > 0 \forall i \right\}.$$

Note that it follows from definition that:

$$\mathcal{S} = \{ \mathbf{U}p \mid p \in \mathcal{B} \}$$

$$\mathcal{S}_{>0} = \{ \mathbf{U}p \mid p \in \mathcal{B}_{>0} \}$$

The proof will proceed by proving Lemma 185 and Lemma 39, and then proving Theorem 38 from these lemmas. In Chapter E.3.1, we prove a useful auxiliary lemma about single-genre equilibria; in Chapter E.3.1, we prove Lemma 185; in Appendix E.3.1, we formalize and prove Lemma 39; and in Chapter E.3.1, we prove Theorem 38.

Auxiliary lemma

We show that at a single-genre equilibrium, it must hold that the direction vector has nonzero inner product with every user.

Lemma 186. *Suppose that μ is a symmetric equilibrium such that* $\text{Genre}(\mu)$ *contains a single vector p^*. Then $p^* \in span(u_1, \ldots, u_N)$ (which also means that $\langle p^*, u_i \rangle > 0$ for all i.)*

Proof. Assume for sake of contradiction that $\langle p^*, u_i \rangle = 0$ for some i. Suppose that $p' \in \text{supp}(\mu)$, and consider the vector $p' + \varepsilon \frac{u_i}{||u_i||}$. We see that $p' + \varepsilon \frac{u_i}{||u_i||}$ wins user u_i with probability 1 whereas p' wins user u_i with probability $1/P$. The probability that $p + \varepsilon u_i$ wins any other user is also at least the probability that p' wins u_i. By leveraging this discontinuity, we see there exists ε such that $\mathcal{P}(p' + \varepsilon \frac{u_i}{||u_i||}; [\mu, \ldots, \mu]) > \mathcal{P}(p'; [\mu, \ldots, \mu]) + (1 - \frac{1}{P})$ which is a contradiction. □

Proof of Lemma 185

We restate and prove Lemma 185.

Lemma 185. *Suppose that μ is a symmetric equilibrium such that* $\text{Genre}(\mu)$ *contains a single vector. Let F be the cdf of the distribution over $\|p\|$ where $p \sim \mu$. Then, it holds that:*

$$F(r) = \min\left(1, \left(\frac{r^\beta}{N}\right)^{1/(P-1)}\right). \tag{E.2}$$

Proof. Next, we show that $F(r) = 0$ only if $r = 0$. Since the distribution μ is atomless (by Proposition 36), we can view the support as a closed set. Let $r_{\min}$ be the minimum magnitude element in the support of μ. Since μ is atomless, this means that with probability 1, every producer will have magnitude greater than $r_{\min}$. This, coupled with Lemma 186, means that the producer the expected number of users achieved at $r_{\min} p$ is 0, and $\mathcal{P}(r_{\min} p; [\mu, \ldots \mu]) = -r_{\min}^\beta$. However, since $r_{\min} p \in \text{supp}(\mu)$, it must hold that:

$$-r_{\min}^\beta = \mathcal{P}(r_{\min} p; [\mu, \ldots, \mu]) \geq \mathcal{P}(\vec{0}; [\mu, \ldots, \mu]) \geq 0.$$

This means that $r_{\min} = 0$.

Next, we show that the equilibrium profit at $(\mu, \ldots, \mu)$ is equal to 0. To see this, suppose that if the producer chooses $\vec{0}$. Since μ is atomless and since $\langle p^*, u_i \rangle > 0$ for all i (by Lemma 186), we see that if a producer chooses $\vec{0} \in \text{supp}(\mu)$, they receive 0 users in expectation. This means that $\mathcal{P}(\vec{0}; [\mu, \ldots, \mu]) = 0$ as desired.

Next, we show that $F(r) = \left(\frac{r^\beta}{N}\right)^{1/(P-1)}$ for any $rp^* \in \text{supp}(\mu)$. To show this, notice that the producer must earn the same profit—here, zero profit—for any $p \in \text{supp}(\mu)$. This means that for any $rp^* \in \text{supp}(\mu)$, it must hold that $NF(r)^{P-1} - r^\beta = 0$. Solving, we see that $F(r) = \left(\frac{r^\beta}{N}\right)^{1/(P-1)}$.

Finally, we show that the support of F is exactly $[0, N^{1/\beta}]$. First, we already showed that $r_{\min} = 0$ which means that 0 is the minimum magnitude element in the support. Moreover, $r = N^{1/\beta}$ must be the maximum magnitude element in the support since it is the unique value for which $F(r) = 1$. Now, we show that $\text{supp}(F)$ is equal to $[0, N^{1/\beta}]$. Note that the set $\text{supp}(F) \cup [N^{1/\beta}, \infty) \cup (-\infty, 0]$ is a finite union of closed sets and is thus closed. Let $S' := \mathbb{R} \setminus (\text{supp}(F) \cup [N^{1/\beta}, \infty) \cup (-\infty, 0])$; it suffices to prove that $S' = \emptyset$. Assume for sake of contradiction that $S' \neq \emptyset$. Since S' is open, there exists $x \in (0, N^{1/\beta})$ and $\varepsilon > 0$ such that $(x, x+\varepsilon) \subseteq S'$. Let $r_1 = \inf_{y \in \text{supp}(F), y \leq x} y$ and let $r_2 = \sup_{y \in \text{supp}(F), y \geq x+\varepsilon} y$. Note that both r_1 and r_2 are in $\text{supp}(F)$ (since it is closed), and $(r_1, r_2) \cap \text{supp}(F) = \emptyset$. By the structure of F, since $F(r_2) > F(r_1)$, this means that the cdf jumps from $F(x)$ to $F(x+\varepsilon)$ anyway so there would be atoms (but there are no atoms by Proposition 36). This proves that the support is $[0, N^{1/\beta}]$.

In conclusion, we have shown that $F(r) = \left(\frac{r^\beta}{N}\right)^{1/(P-1)}$ for any $r \in [0, N^{1/\beta}]$. The min with 1 comes from the fact that $F(r) = 1$ for $r \geq N^{1/\beta}$. □

Formal Statement and Proof of Lemma 39

We begin with a proof sketch of Lemma 39. For μ to be an equilibrium, no alternative q should do better than $p \sim \mu$, which yields the following necessary and sufficient condition after plugging into the profit function (8.1):

$$\sup_q \left(\sum_{i=1}^N \frac{1}{N}\left(\frac{\langle q, u_i\rangle}{\langle p^*, u_i\rangle}\right)^\beta - \|q\|^\beta\right) = \mathbb{E}_{p' \sim \mu}\left[\sum_{i=1}^N \frac{1}{N}\left(\frac{\langle p', u_i\rangle}{\langle p^*, u_i\rangle}\right)^\beta - \|p'\|^\beta\right] \tag{E.3}$$

The term $\frac{1}{N}(\cdot)^\beta$ is the probability $(F(\cdot))^{P-1}$ that q outperforms the max of $P-1$ samples from μ.

We next change variables according to $y_i = \langle p^*, u_i\rangle^\beta$ and $y'_i = \langle \frac{q}{||q||}, u_i\rangle^\beta$ and simplify to see that μ is an equilibrium if and only if $\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^n \frac{y'_i}{y_i} = N$. Thus, there exists a single-genre equilibrium if and only if

$$\inf_{y \in \mathcal{S}^\beta} \sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{y_i} = N. \tag{E.4}$$

While the left-hand side of equation (E.4) is challenging to reason about directly, we show that the dual $\sup_{y' \in \mathcal{S}^\beta} \inf_{y \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{y_i}$ is in fact equal to N.

With this proof sketch in mind, we are ready to formalize and prove Lemma 39.

Lemma 187 (Formalization of Lemma 39)**.** *There exists a symmetric equilibrium μ with $|\text{Genre}(\mu)| = 1$ if and only if:*

$$\inf_{p^* \in \mathcal{B}_{>0}} \sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{(\langle p^*, u_i\rangle)^\beta} = \sup_{y' \in \mathcal{S}^\beta} \inf_{p^* \in \mathcal{B}_{>0}} \sum_{i=1}^N \frac{y'_i}{(\langle p^*, u_i\rangle)^\beta}. \tag{E.5}$$

It turns out to be more convenient to use a (slightly less intuitive) variant of Lemma 187 to prove Theorem 38. We state and prove Lemma 188 below.

Lemma 188. *There exists a symmetric equilibrium μ with $|\mathrm{Genre}(\mu)| = 1$ if and only if:*

$$\inf_{p^* \in \mathcal{B}_{>0}} \sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} \leq N. \tag{E.6}$$

The main ingredient in the proof of Lemma 188 is the following characterization of a single-genre equilibrium in a given direction.

Lemma 189. *There is a symmetric equilibrium μ with $\mathrm{Genre}(\mu) = \{p^*\}$ if and only if:*

$$\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} \leq N. \tag{E.7}$$

Proof. First, by Lemma 186, we see that the denominator is nonzero for every term in the sum, so equation (E.7) is well-defined.

If μ is a single-genre equilibrium, then the cdf of the magnitudes follows the form in Lemma 185. Thus, it suffices to identify necessary and sufficient conditions for that solution (that we call μ_{p^*}) to be a symmetric equilibrium.

The solution μ_{p^*} is an equilibrium if and only if no alternative q should do better than $p \sim \mu$. The profit level at μ_{p^*} is 0 by the structure of the cdf. Putting this all together, we see a necessary and sufficient for μ_{p^*} to be an equilibrium is:

$$\sup_{q \in \mathbb{R}^D_{\geq 0}} \left(\sum_{i=1}^N F\left(\frac{\langle q, u_i \rangle}{\langle p^*, u_i \rangle}\right)^{P-1} - \|q\|^\beta \right) \leq 0,$$

where the term $\frac{1}{N}\left(\cdot\right)^\beta$ is the probability $\left(F(\cdot)\right)^{P-1}$ that q outperforms the max of $P-1$ samples from μ. Using the structure of the cdf, we can write this as:

$$\sup_{q \in \mathbb{R}^D_{\geq 0}} \left(\sum_{i=1}^N \min\left(1, \frac{1}{N}\left(\frac{\langle q, u_i \rangle}{\langle p^*, u_i \rangle}\right)^\beta\right) - \|q\|^\beta \right) \leq 0.$$

We can equivalently write this as:

$$\sup_{q \in \mathbb{R}^D_{\geq 0}} \left(\frac{1}{||q||^\beta} \sum_{i=1}^N \min\left(1, \frac{1}{N}\left(\frac{\langle q, u_i \rangle}{\langle p^*, u_i \rangle}\right)^\beta\right) - 1 \right) \leq 0,$$

which we can equivalently write as

$$\sup_{q \in \mathcal{D}} \sup_{r > 0} \left(\frac{1}{r^\beta} \sum_{i=1}^N \min\left(1, \frac{r^\beta}{N}\left(\frac{\langle q, u_i \rangle}{\langle p^*, u_i \rangle}\right)^\beta\right) - 1 \right) \leq 0.$$

For any direction q, if we disregard the first min with 1, the expression would be constant in r. With the minimum, the objective $\left(\frac{1}{r^\beta}\sum_{i=1}^N \min\left(1, \frac{1}{N}\left(\frac{\langle q,u_i\rangle}{\langle p^*,u_i\rangle}\right)^\beta\right) - 1\right)$ is weakly decreasing in r. Thus, $\sup_{r>0}\left(\frac{1}{r^\beta}\sum_{i=1}^N \min\left(1, \frac{1}{N}\left(\frac{\langle q,u_i\rangle}{\langle p^*,u_i\rangle}\right)^\beta\right) - 1\right)$ is attained as $r \to 0$. In fact, the maximum is attained at a value r if $r\langle q, u_i\rangle < N^{1/\beta}\langle p^*, u_i\rangle$ for all i. This holds for *some* $r > 0$ since $\langle p^*, u_i\rangle > 0$ for all i by Lemma 186. Thus we can equivalently formulate the condition as:

$$\sup_{q\in\mathcal{D}}\left(\left(\sum_{i=1}^N \frac{1}{N}\left(\frac{\langle q,u_i\rangle}{\langle p^*,u_i\rangle}\right)^\beta\right) - 1\right) \leq 0,$$

which we can write as:

$$\sup_{q\in\mathcal{D}}\sum_{i=1}^N \left(\frac{\langle q,u_i\rangle}{(\langle p^*,u_i\rangle)}\right)^\beta \leq N.$$

This is equivalent to:

$$\sup_{q\in\mathcal{B}}\sum_{i=1}^N \left(\frac{\langle q,u_i\rangle}{(\langle p^*,u_i\rangle)}\right)^\beta \leq N.$$

A change of variables gives us the desired formulation. □

Now, we can deduce Lemma 188.

Proof of Lemma 188. First, suppose that equation (E.6) does not hold. Then it must be true that:

$$\sup_{y'\in\mathcal{S}^\beta}\sum_{i=1}^N \frac{y'_i}{(\langle p^*,u_i\rangle)^\beta} > N$$

for every direction $p^* \in \mathcal{D}_{>0}$. This means that no direction in $\mathcal{D}_{>0}$ can be a single-genre equilibrium. We can further rule out directions in $\mathcal{D} \setminus \mathcal{D}_{>0}$ by applying Lemma 186.

Now, suppose that equation (E.6) does hold. It is not difficult to see that the optimum

$$\inf_{p^*\in\mathcal{B}_{>0}}\sup_{y'\in\mathcal{S}^\beta}\sum_{i=1}^N \frac{y'_i}{(\langle p^*,u_i\rangle)^\beta}$$

is attained at some direction $p^* \in \mathcal{D}_{>0}$. Applying Lemma 189, we see that there exists a single-genre equilibrium in the direction p^*. □

Proof of Lemma 187

To prove Lemma 187 from Lemma 188, we require the following additional lemma that helps us analyze the right-hand side of equation (E.5).

Lemma 190. *For any set $\mathcal{R} \subseteq \mathbb{R}^N_{>0}$, it holds that:*

$$\sup_{y' \in \mathcal{R}} \inf_{y \in \mathcal{R}} \sum_{i=1}^{N} \frac{y'_i}{y_i} = N.$$

Proof. By taking $y' = y$, we see that:

$$\sup_{y' \in \mathcal{R}} \inf_{y \in \mathcal{R}} \sum_{i=1}^{N} \frac{y'_i}{y_i} \leq N.$$

To show equality, notice by AM-GM that:

$$\sum_{i=1}^{N} \frac{y'_i}{y_i} \geq N \left(\prod_{i=1}^{n} \frac{y'_i}{y_i} \right)^{1/N} = N \left(\frac{\prod_{i=1}^{n} y'_i}{\prod_{i=1}^{N} y_i} \right)^{1/N}.$$

We can take $y' = \arg\max_{y'' \in \mathcal{R}} \prod_{i=1}^{n} y''_i$, and obtain a lower bound of N as desired. (If the $\arg\max$ does not exist, then note that if we take y' where $\prod_{i=1}^{n} y'_i$ is sufficiently close to the optimum $\sup_{y'' \in \mathcal{R}} \prod_{i=1}^{n} y''_i$, we have that $\inf_{y \in \mathcal{R}} \left(\frac{\prod_{i=1}^{n} y'_i}{\prod_{i=1}^{N} y_i} \right)^{1/N}$ is sufficiently close to 1 as desired.) □

Now we are ready to prove Lemma 187.

Proof of Lemma 187. First, we see that:

$$\begin{aligned} N &= \sup_{y' \in \mathcal{S}^\beta_{>0}} \inf_{y \in \mathcal{S}^\beta_{>0}} \sum_{i=1}^{N} \frac{y'_i}{y_i} \\ &= \sup_{y' \in \mathcal{S}^\beta} \inf_{y \in \mathcal{S}^\beta_{>0}} \sum_{i=1}^{N} \frac{y'_i}{y_i} \\ &= \sup_{y' \in \mathcal{S}^\beta} \inf_{p^* \in \mathcal{B}_{>0}} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta}, \end{aligned}$$

where the first equality follows from Lemma 190.

Now, let's combine this with Lemma 188 to see that a necessary and sufficient condition for the existence of a single-genre equilibrium is:

$$\inf_{p^* \in \mathcal{B}_{>0}} \sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} \leq \sup_{y' \in \mathcal{S}^\beta} \inf_{p^* \in \mathcal{B}_{>0}} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} \tag{E.8}$$

Weak duality tells us that $\inf_{p^* \in \mathcal{B}_{>0}} \sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} \geq \sup_{y' \in \mathcal{S}^\beta} \inf_{p^* \in \mathcal{B}_{>0}} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta}$, so equation (E.8) is equivalent to:

$$\inf_{p^* \in \mathcal{B}_{>0}} \sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} = \sup_{y' \in \mathcal{S}^\beta} \inf_{p^* \in \mathcal{B}_{>0}} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta}.$$

□

Finishing the proof of Theorem 38

Proof of Theorem 38. Recall that by Lemma 188, a single genre equilibrium exists if and only if equation (E.6) is satisfied.

We can rewrite the left-hand side of equation (E.6) as follows:

$$\inf_{p^* \in \mathcal{B}_{>0}} \left(\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{\langle p^*, u_i \rangle^\beta} \right) = \inf_{p^* \in \mathcal{B}_{>0}} \left(\sup_{y' \in \bar{\mathcal{S}}^\beta} \sum_{i=1}^N \frac{y'_i}{\langle p^*, u_i \rangle^\beta} \right),$$

since the objective is linear in y'. Now, observing that the objective is convex in p and concave in y', we can apply Sion's min-max theorem[1] to see that:

$$\inf_{p^* \in \mathcal{B}_{>0}} \left(\sup_{y' \in \bar{\mathcal{S}}^\beta} \sum_{i=1}^N \frac{y'_i}{\langle p^*, u_i \rangle^\beta} \right) = \sup_{y' \in \bar{\mathcal{S}}^\beta} \left(\inf_{p^* \in \mathcal{B}_{>0}} \sum_{i=1}^N \frac{y'_i}{\langle p^*, u_i \rangle^\beta} \right) = \sup_{y' \in \bar{\mathcal{S}}^\beta} \left(\inf_{y \in \mathcal{S}^\beta_{>0}} \sum_{i=1}^N \frac{y'_i}{y_i} \right).$$

Thus, we have the following necessary and sufficient condition for a single-genre equilibrium to exist:

$$\sup_{y' \in \bar{\mathcal{S}}^\beta} \left(\inf_{y \in \mathcal{S}^\beta_{>0}} \sum_{i=1}^N \frac{y'_i}{y_i} \right) \leq N. \tag{E.9}$$

First, we show that if (8.4) does not hold, then there does not exist a single-genre equilibrium. Let $y' = \arg\max_{y'' \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^n y''_i$. (The maximum exists because $\prod_{i=1}^n y''_i$ is a continuous function and $\bar{\mathcal{S}}^\beta$ is a compact set.) We see that:

$$\sum_{i=1}^N \frac{y'_i}{y_i} \geq N \left(\frac{\prod_{i=1}^n y'_i}{\prod_{i=1}^n y_i} \right)^{1/N} \geq N \left(\frac{\max_{y'' \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^n y''_i}{\max_{y'' \in \mathcal{S}^\beta_{>0}} \prod_{i=1}^n y''_i} \right)^{1/N} = N \left(\frac{\max_{y'' \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^n y''_i}{\max_{y'' \in \mathcal{S}^\beta} \prod_{i=1}^n y''_i} \right)^{1/N} > N,$$

which proves that:

$$\inf_{p^* \in \mathcal{B}_{>0}} \left(\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^N \frac{y'_i}{\langle p^*, u_i \rangle^\beta} \right) = \sup_{y' \in \bar{\mathcal{S}}^\beta} \left(\inf_{y \in \mathcal{S}^\beta_{>0}} \sum_{i=1}^N \frac{y'_i}{y_i} \right) > N.$$

Thus equation (E.9) is not satisfied and a single-genre equilibrium does not exist as desired.

Next, we show that if (8.4) holds, then there exists a single-genre equilibrium. Let $y^* = \arg\max_{y'' \in \mathcal{S}^\beta} \prod_{i=1}^n y''_i = \arg\max_{y'' \in \mathcal{S}^\beta} \sum_{i=1}^n \log(y''_i)$. (The maximum exists because $\prod_{i=1}^n y''_i$ is a continuous function and $\mathcal{S}^\beta$ is a compact set.) By assumption, we see that y^* is also the maximizer over $\bar{S}^\beta$. We further see that $y^* \in \mathcal{S}^\beta_{>0}$. Using convexity of $\bar{\mathcal{S}}^\beta$, this means that for any $y' \in \bar{\mathcal{S}}^\beta$, it must hold that $\langle y' - y^*, \nabla \left(\sum_{i=1}^n \log(y^*_i) \right) \rangle \leq 0$. We can write this as:

$$\langle y' - y^*, \nabla \sum_{i=1}^n \frac{1}{y^*_i} \rangle \leq 0.$$

[1] Note that $\bar{\mathcal{S}}^\beta$ is compact and convex and $\mathcal{B}_{>0}$ is convex (but not compact). We apply the non-compact formulation of Sion's min-max theorem in (Ha, 1981).

This can be written as:

$$\sum_{i=1}^{n} \frac{y_i' - y_i^*}{y_i^*} \leq 0,$$

which implies that:

$$\sum_{i=1}^{n} \frac{y_i'}{y_i^*} \leq N.$$

Thus, we have that

$$\sup_{y' \in \bar{\mathcal{S}}^{\beta}} \left(\inf_{y \in \mathcal{S}^{\beta}_{>0}} \sum_{i=1}^{N} \frac{y_i'}{y_i^*} \right) \leq N,$$

and thus equation (E.9) is satisfied so a single-genre equilibrium does not exist as desired.

Next, we show that if all equilibria have multiple genres for some β, then all equilibria have multiple genres for all $\beta' \geq \beta$. Notice that equation 8.4 can equivalently be restated as:

$$\max_{y \in \mathcal{S}} \prod_{i=1}^{N} y_i = \max_{y \in \bar{\mathcal{S}}^{\beta}} \left(\prod_{i=1}^{N} y_i \right)^{1/\beta}. \tag{E.10}$$

It thus suffices to show that:

$$\max_{y \in \bar{\mathcal{S}}^{\beta}} \left(\prod_{i=1}^{N} y_i \right)^{1/\beta} \leq \max_{y \in \bar{\mathcal{S}}^{\beta'}} \left(\prod_{i=1}^{N} y_i \right)^{1/\beta'}$$

for all $\beta' \geq \beta$. To see this, let y denote the maximizer of $\max_{y \in \bar{\mathcal{S}}^{\beta}} \left(\prod_{i=1}^{N} y_i \right)^{1/\beta}$ (this is achieved since we are taking a maximum of a continuous function over a compact set). By definition, we see that y can be written as a convex combination $\sum_{j=1}^{P} \lambda_j (x_i^j)^{\beta}$ where $x^1, \ldots, x^P$ denote vectors in $\mathcal{S}$ and where $\sum_{j=1}^{P} \lambda_j = 1$. In this notation, we see that:

$$\max_{y \in \bar{\mathcal{S}}^{\beta}} \left(\prod_{i=1}^{N} y_i \right)^{1/\beta} = \left(\prod_{i=1}^{N} \left(\sum_{j=1}^{P} \lambda_j (x_i^j)^{\beta} \right)^{1/\beta} \right)$$

By taking y to be $\sum_{j=1}^{P} \lambda_j (x_i^j)^{\beta'}$, we see that:

$$\max_{y \in \bar{\mathcal{S}}^{\beta'}} \left(\prod_{i=1}^{N} y_i \right)^{1/\beta'} \geq \left(\prod_{i=1}^{N} \left(\sum_{j=1}^{P} \lambda_j (x_i^j)^{\beta'} \right)^{1/\beta'} \right).$$

Notice that for any $1 \leq i \leq N$, it holds that:

$$\left(\sum_{j=1}^{P} \lambda_j (x_i^j)^{\beta'} \right) = \left(\sum_{j=1}^{P} \lambda_j ((x_i^j)^{\beta})^{\beta'/\beta} \right) \geq \left(\sum_{j=1}^{P} \lambda_j ((x_i^j)^{\beta}) \right)^{\beta'/\beta},$$

where the last inequality follows from convexity of $f(c) = c^{\beta'/\beta}$ for $\beta' \geq \beta$. Putting this all together, we see that:

$$\max_{y \in \bar{\mathcal{S}}^{\beta'}} \left(\prod_{i=1}^{N} y_i\right)^{1/\beta'} \geq \left(\prod_{i=1}^{N} \left(\sum_{j=1}^{P} \lambda_j (x_i^j)^{\beta'}\right)^{1/\beta'}\right) \geq \left(\prod_{i=1}^{N} \left(\sum_{j=1}^{P} \lambda_j (x_i^j)^{\beta}\right)^{1/\beta}\right) = \max_{y \in \bar{\mathcal{S}}^{\beta}} \left(\prod_{i=1}^{N} y_i\right)^{1/\beta}$$

as desired. □

E.3.2 Proofs of corollaries of Theorem 38

We prove all of the corollaries of Theorem 38 in Chapter 8.3.2, except for Corollary 42 (proof deferred to Appendix E.4.2).

First, we prove Corollary 40, restated below.

Corollary 40. *The threshold β^* is always at least 1. That is, if $\beta = 1$, there exists a single-genre equilibrium.*

Proof. When $\beta = 1$, we see that $\mathcal{S}^\beta = \mathcal{S}^1$ is a linear transformation of a convex set (the unit ball restricted to $\mathbb{R}^D_{\geq 0}$), so it is convex. This means that $\bar{\mathcal{S}}^\beta = \mathcal{S}^\beta$, and so (8.4) is trivially satisfied. By Theorem 38, there exists a single-genre equilibrium. □

Next, we prove Corollary 41, restated below.

Corollary 41. *Let the cost function be $c(p) = \|p\|_q^\beta$. For any set of user vectors, it holds that $\beta^* \geq q$. If the user vectors are equal to the standard basis vectors $\{e_1, \ldots, e_D\}$, then β^* is equal to q.*

Proof. We split the proof into two steps: (1) showing that $\beta^* \geq q$ for any set of user vectors and (2) showing that $\beta^* \leq q$ for the standard basis vectors.

Showing that $\beta^* \geq q$ for any set of users. To show that $\beta^* \geq q$, by Theorem 38, it suffices to show that equation (8.4) is satisfied at $\beta = q$. Suppose that the right-hand side of (8.4):

$$\max_{y \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^{N} y_i$$

is maximized at some $y^* \in \bar{S}^\beta$. It suffices to construct $\tilde{y} \in \mathcal{S}^\beta$ such that

$$\prod_{i=1}^{N} \tilde{y}_i \geq \prod_{i=1}^{N} y_i^* \tag{E.11}$$

To construct $\tilde{y}$, we introduce some notation. By the definition of a convex hull, we can write y^* as

$$y^* = \sum_{k=1}^{m} \lambda_k y^k,$$

where $y^1, \ldots, y^m \in \mathcal{S}^\beta$ and where $\lambda_1, \ldots, \lambda_m \in [0,1]$ are such that $\sum_{k=1}^m \lambda_k = 1$. Let $p^1, \ldots, p^m \in \mathbb{R}^D_{\geq 0}$ be such that $\|p^k\|_q \leq 1$ for all $1 \leq k \leq m$ and y^k is given by the β-coordinate-wise powers of $\mathbf{U}p_k$. Now, we let $y = \mathbf{U}\tilde{p}$ where the dth coordinate of $\tilde{p}$ is given by:

$$\tilde{p}_d := \left(\sum_{k=1}^m \lambda_k ((p^k)_d)^q\right)^{1/q}.$$

It follows from definition that:

$$\|\tilde{p}\|_q = \left(\sum_{d=1}^D \sum_{k=1}^m \lambda_k ((p^k)_d)^q\right)^{1/q} = \left(\sum_{k=1}^m \lambda_k \sum_{d=1}^D ((p^k)_d)^q\right)^{1/q} \leq \left(\sum_{k=1}^m \lambda_k \|p^k\|_q^q\right)^{1/q} \leq 1,$$

which means that $\tilde{y} \in \mathcal{S}^\beta$.

The remainder of the proof boils down to showing (E.11). It suffices to show that for every $1 \leq i \leq N$, it holds that $\tilde{y}_i \geq y_i^*$. Notice that:

$$y_i^* = \sum_{k=1}^m \lambda_k (y^k)_i = \sum_{k=1}^m \lambda_k \langle u_i, p^k \rangle^q = \sum_{k=1}^m \lambda_k \left(\sum_{d=1}^D (u_i)_d (p^k)_d\right)^q,$$

and

$$\tilde{y}_i = \langle u_i, \tilde{p} \rangle^q = \left(\sum_{d=1}^D (u_i)_d \tilde{p}_d\right)^q = \left(\sum_{d=1}^D (u_i)_d \left(\sum_{k=1}^m \lambda_k ((p^k)_d)^q\right)^{1/q}\right)^q.$$

Thus, it suffices to show the following inequality:

$$\sum_{d=1}^D (u_i)_d \left(\sum_{k=1}^m \lambda_k ((p^k)_d)^q\right)^{1/q} \geq \left(\sum_{k=1}^m \lambda_k \left(\sum_{d=1}^D (u_i)_d (p^k)_d\right)^q\right)^{1/q}. \tag{E.12}$$

The high-level idea is that the proof boils down to the triangle inequality for an appropriately chosen norm over $\mathbb{R}^m$. For $z \in \mathbb{R}^m$, we let:

$$\|z\|_\lambda := \left(\sum_{k=1}^m \lambda_k z^q\right)^{1/q}.$$

To see that this is a norm, note that $\left(\sum_{k=1}^m \lambda_k z^q\right)^{1/q} = \left(\sum_{k=1}^m (\lambda_k^{1/q} z)^q\right)^{1/q}$. The norm properties of this function are implied by the norm properties of $\|\cdot\|_q$. By the triangle inequality, we see that:

$$\begin{aligned}
\sum_{d=1}^{D}(u_i)_d\left(\sum_{k=1}^{m}\lambda_k((p^k)_d)^q\right)^{1/q} &= \sum_{d=1}^{D}(u_i)_d\|[p_d^1,\dots,p_d^m]\|_\lambda \\
&\geq \|\sum_{d=1}^{D}(u_i)_d[p_d^1,\dots,p_d^m]\|_\lambda \\
&= \left(\sum_{k=1}^{m}\lambda_k\left(\sum_{d=1}^{D}(u_i)_d(p^k)_d\right)^q\right)^{1/q}
\end{aligned}$$

which implies equation (E.12).

Showing that $\beta^* \leq q$ for the standard basis vectors. By Theorem 38, it suffices to show, for any $\beta > q$, that equation (8.4) is not satisfied. First, we compute the left-hand side of equation (8.4):

$$\max_{y\in\mathcal{S}^\beta}\prod_{i=1}^{N}y_i = \left(\max_{x\in\mathbb{R}^D_{\geq 0},\|x\|_q=1}\prod_{i=1}^{D}x_i\right)^\beta = \left(\frac{1}{D}\right)^{\beta/q} < \left(\frac{1}{D}\right).$$

where the last line follows from AM-GM. Now, we compute the right-hand side:

$$\max_{y\in\bar{\mathcal{S}}^\beta}\prod_{i=1}^{N}y_i.$$

Consider $y^* = \left[\frac{1}{D},\dots,\frac{1}{D}\right]$. Notice that y is a convex combination of the standard basis vectors—all of which are in $\mathcal{S}$ and actually in $\mathcal{S}^\beta$ too—so $y \in \bar{\mathcal{S}}^\beta$. This means that

$$\max_{y\in\bar{\mathcal{S}}^\beta}\prod_{i=1}^{N}y_i \geq \prod_{i=1}^{N}y_i^* = \left(\frac{1}{D}\right).$$

This proves that:

$$\max_{y\in\mathcal{S}^\beta}\prod_{i=1}^{N}y_i < \max_{y\in\bar{\mathcal{S}}^\beta}\prod_{i=1}^{N}y_i,$$

so equation (8.4) is not satisfied as desired.

□

We prove Corollary 43, restated below.

Corollary 43. *Let $\|\cdot\|_*$ denote the dual norm of $\|\cdot\|$, defined to be $\|p\|_* = \max_{\|p\|=1,p\in\mathbb{R}^D_{\geq 0}}\langle q,p\rangle$. Let $Z := \|\sum_{n=1}^{N}\frac{u_n}{\|u_n\|_*}\|_*$. Then,*

$$\beta^* \leq \frac{\log(N)}{\log(N)-\log(Z)}. \tag{8.6}$$

Proof. WLOG assume that the users to have unit dual norm. By Theorem 38, it suffices to show that:

$$\max_{y \in \mathcal{S}^\beta} \prod_{i=1}^N y_i < \max_{y \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^N y_i.$$

First, let's lower bound the right-hand side. Consider the point $y = \frac{1}{N} \sum_{i'=1}^N z^{i'}$ where $z^{i'}$ is defined to be the β-coordinate-wise power of $\mathbf{U}\left(\arg\max_{||p||=1} \langle p, u_i \rangle\right)$. This means that

$$y_i \geq \frac{1}{N} z_i^i = \frac{1}{N} \left(\max_{||p||=1} \langle p, u_i \rangle \right)^\beta = \frac{||u_i||_*^\beta}{N} = \frac{1}{N}.$$

This means that:

$$\max_{y \in \bar{\mathcal{S}}^\beta} \prod_{i=1}^N y_i \geq \frac{1}{N^N}.$$

Next, let's upper bound the left-hand side. By AM-GM, we see that:

$$\begin{aligned}
\max_{y \in \mathcal{S}^\beta} \prod_{i=1}^N y_i &= \max_{||p||=1, p \in \mathbb{R}^D_{\geq 0}} \left(\prod_{i=1}^N \langle p, u_i \rangle \right)^\beta \\
&\leq \left(\frac{\sum_{i=1}^N \langle p, u_i \rangle}{N} \right)^{N\beta} \leq \left(\frac{\langle p, \sum_{i=1}^N u_i \rangle}{N} \right)^{N\beta} \\
&\leq \frac{\left(\| \sum_{i=1}^N u_i \|_* \right)^{N\beta}}{N^{N\beta}}.
\end{aligned}$$

Putting this all together, we see that it suffices for:

$$\frac{1}{N^N} > \frac{\left(\| \sum_{i=1}^N u_i \|_* \right)^{N\beta}}{N^{N\beta}},$$

which we can rewrite as:

$$N^{\beta - 1} > \left(\| \sum_{i=1}^N u_i \|_* \right)^\beta$$

which we can rewrite as:

$$N^{1 - 1/\beta} > \| \sum_{i=1}^N u_i \|_*.$$

□

We prove Corollary 44, restated below.

Corollary 44. *If there exists μ with $|\text{Genre}(\mu)| = 1$, then the corresponding producer direction maximizes Nash social welfare of the users:*

$$\text{Genre}(\mu) = \underset{\|p\|=1|p\in\mathbb{R}^D_{\geq 0}}{\arg\max} \sum_{i=1}^{N} \log(\langle p, u_i \rangle). \tag{8.7}$$

Proof. Corollary 44 follows as a consequence of the proof of Theorem 38. We apply Lemma 189 to see that if μ is a single-genre equilibrium with $\text{Genre}(\mu) = \{p^*\}$, then:

$$\sup_{y' \in \mathcal{S}^\beta} \sum_{i=1}^{N} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} \leq N.$$

We see that:

$$N \geq \sup_{y' \in \mathcal{S}^\beta} \frac{y'_i}{(\langle p^*, u_i \rangle)^\beta} \geq N \sup_{y' \in \mathcal{S}^\beta} \left(\frac{\prod_{i=1}^{N} y'_i}{\prod_{i=1}^{N} (\langle p^*, u_i \rangle)^\beta} \right)^{1/N} \geq N \left(\frac{\sup_{y' \in \mathcal{S}^\beta} \prod_{i=1}^{N} y'_i}{\prod_{i=1}^{N} (\langle p^*, u_i \rangle)^\beta} \right)^{1/N}.$$

This implies that:

$$\prod_{i=1}^{N} y_i = \prod_{i=1}^{N} (\langle p^*, u_i \rangle)^\beta \geq \sup_{y' \in \mathcal{S}^\beta} \prod_{i=1}^{N} y'_i,$$

where $y \in \mathcal{S}^\beta$ is defined so that $y_i = \langle p^*, u_i \rangle^\beta$. This implies that:

$$p^* \in \underset{||p||\leq 1, p\in\mathbb{R}^D_{\geq 0}}{\arg\max} \sum_{i=1}^{N} \log(\langle p, u_i \rangle) = \underset{||p||=1, p\in\mathbb{R}^D_{\geq 0}}{\arg\max} \sum_{i=1}^{N} \log(\langle p, u_i \rangle)$$

as desired. ∎

E.4 Proofs and Details for Chapter 8.4

In Appendix E.4.1, we provide an overview of how we leverage Lemma 50 to analyze equilibria in the setting of two populations of users. In Appendix E.4.2, we prove Corollary 42. In Appendix E.4.3, we prove the results from Chapter 8.4.1, and in Chapter E.4.4, we prove the results from Chapter 8.4.2. In Appendix E.4.5, we formalize the infinite-producer limit, which we study in Chapter 8.4.3, and in Appendix E.4.6, we prove results from Chapter 8.4.3. In Appendix E.4.7, we prove several auxiliary lemmas that we used along the way.

E.4.1 Overview of proof techniques

Before diving into proof techniques, we observe that it suffices to study a simpler setting with *two normalized users* and a rescaled cost function.

Claim 191. *A distribution μ is an equilibria for a marketplace with 2 populations of users of size $N/2$ located at vectors u_1 and u_2 and with producer cost function $c(p) = \|p\|_2^\beta$ if and only if μ is an equilibria for a marketplace with 2 users located at vectors $\frac{u_1}{\|u_1\|}$ and $\frac{u_2}{\|u_2\|}$ and with producer cost function $c(p) = \frac{2}{N}\|p\|_2^\beta$.*

Thus, we focus on marketplaces with 2 users located at vectors u_1 and u_2 such that $\|u_1\| = \|u_2\| = 1$ and with producer cost function $c(p) = \alpha\|p\|_2^\beta$ for $\alpha > 0$.

The proofs in this section boil down to leveraging conditions (C1)-(C3) in Lemma 50, restated below.

Lemma 50. *Let $\mathbf{U} = [u_1; u_2; \ldots; u_N]$ be the $N \times D$ matrix of users vectors. Given a set $S \subseteq \mathbb{R}_{\geq 0}^N$ and distributions $H_1, \ldots, H_N$ over $\mathbb{R}_{\geq 0}$, suppose that the following conditions hold:*

(C1) Every $z^* \in S$ is a maximizer of the equation:

$$\max_{z \in \mathbb{R}_{\geq 0}^D} \sum_{i=1}^{N} H_i(z_i) - c_{\mathbf{U}}(z), \tag{8.9}$$

where $c_{\mathbf{U}}(z) := \min\left\{c(p) \mid p \in \mathbb{R}_{\geq 0}^D, \mathbf{U}p = z\right\}$.

(C2) There exists a random variable Z with support S, such that the marginal distribution Z_i has cdf equal to $H_i(z)^{1/(P-1)}$.

(C3) Z is distributed as $\mathbf{U}Y$ with $Y \sim \mu$, for some distribution μ over $\mathbb{R}_{\geq 0}^D$.

Then, the distribution μ from (C3) is a symmetric mixed Nash equilibrium. Moreover, every symmetric mixed Nash equilibrium μ is associated with some $(H_1, \ldots, H_N, S)$ that satisfy (C1)-(C3).

Proof of Lemma 50. The intuition is that the the set S captures the support of the realized inferred user values $[\langle u_1, p\rangle, \ldots, \langle u_N, p\rangle]$ for $p \sim \mu$ and the distribution H_i captures the distribution of the maximum inferred user value$\max_{1 \leq j \leq P-1}\langle u_i, p_j\rangle$ for user u_i.

To formalize this, we reparameterize from content vectors in $\mathbb{R}_{\geq 0}^D$ to realized inferred user values in $\mathbb{R}_{\geq 0}^N$. That is, we transform the content vector $p \in \mathbb{R}_{\geq 0}^D$ into the vector of realized inferred user values given by $z = \mathbf{U}p$. This reparameterization allows us to cleanly reason about the number of users that a producer wins: a producer wins a user u_i if and only if they have the highest value in the ith coordinate of z. In this parametrization, the cost of production can be computed through an induced function $c_{\mathbf{U}}$ given by $c_{\mathbf{U}}(z) := \min\left\{c(p) \mid p \in \mathbb{R}_{\geq 0}^D, z = \mathbf{U}p\right\}$ if $z \in \left\{\mathbf{U}p \mid p \in \mathbb{R}_{\geq 0}^D\right\}$.

In this reparameterization, the producer profit takes a clean form. If producer 1 chooses $z \in \mathbb{R}^N$, and other producers follow a distribution μ_Z over $\mathbb{R}^N$, then the expected profit of producer 1 is:

$$\sum_{i=1}^{N} H_i(z_i) - c_{\mathbf{U}}(z),$$

where $H_i(\cdot)$ is the cumulative distribution function of the maximum realized inferred user value over the other $P-1$ producers, i.e. of the random variable $\max_{2 \le j \le P}(z_j)_i$ where $z_2, \ldots, z_P \sim \mu_Z$.

Recall that a distribution μ corresponds to a symmetric mixed Nash equilibrium if and only if every z in the support $S := \mathrm{supp}(\mu_Z)$ is a maximizer of equation (8.9) (where μ_Z is the distribution over $\mathbf{U}p$ for $p \sim \mu$).

□

Leveraging (C1)

To leverage (C1), we use the first-order and second-order conditions for z to be a maximizer of equation (8.9). In order to obtain useful closed-form expressions, we explicitly compute the induced cost function in terms of the angle θ^* between the user vectors.

Lemma 192. *Let there be 2 users located at $u_1, u_2 \in \mathbb{R}^D_{\ge 0}$ such that $\|u_1\| = \|u_2\| = 1$, and let $\theta^* := \cos^{-1}(\langle u_1, u_2 \rangle) > 0$ be the angle between the user vectors. Let the cost function be $c(p) = \alpha \|p\|_2^\beta$ for $\alpha > 0$. For any $z \in \left\{\mathbf{U}p \mid p \in \mathbb{R}^D_{\ge 0}\right\}$, the induced cost function is given by:*

$$c_{\mathbf{U}}(z) = \alpha \sin^{-\beta}(\theta^*) \left(z_1^2 + z_2^2 - 2 z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}}.$$

First-order condition. The first order condition implies that we can compute the densities h_1 and h_2 of H_1 and H_2 in terms of the $c_{\mathbf{U}}$. The densities $h_1(z_1)$ and $h_2(z_2)$ depend on the gradient $\nabla_z c_{\mathbf{U}}$ and *both* coordinates z_1 and z_2.

Lemma 193. *Let there be 2 users located at $u_1, u_2 \in \mathbb{R}^D_{\ge 0}$ such that $\|u_1\| = \|u_2\| = 1$, and let $\theta^* := \cos^{-1}(\langle u_1, u_2 \rangle) > 0$ be the angle between the user vectors. Let the cost function be $c(p) = \alpha \|p\|_2^\beta$ for $\alpha > 0$. For any $z \in \left\{\mathbf{U}p \mid p \in \mathbb{R}^D_{\ge 0}\right\}$, the first-order condition of equation (8.9) can be written as:*

$$\begin{bmatrix} h_1(z_1) \\ h_2(z_2) \end{bmatrix} = \nabla_z (c_{\mathbf{U}}(z)).$$

More specifically, it holds that:

$$\begin{bmatrix} h_1(z_1) \\ h_2(z_2) \end{bmatrix} = \beta \alpha \sin^{-\beta}(\theta^*) \left(z_1^2 + z_2^2 - 2 z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-1} \begin{bmatrix} z_1 - z_2 \cos(\theta^*) \\ z_2 - z_1 \cos(\theta^*) \end{bmatrix},$$

and if we represent $z = \mathbf{U}[r\cos(\theta), r\sin(\theta)]$, then it also holds that:

$$\begin{bmatrix} h_1(z_1) \\ h_2(z_2) \end{bmatrix} = \beta \alpha r^{\beta-1} \begin{bmatrix} \frac{\sin(\theta^* - \theta)}{\sin(\theta^*)} \\ \frac{\sin(\theta)}{\sin(\theta^*)} \end{bmatrix}.$$

Second-order condition. When we also take advantage of the second-order condition, we can identify the "direction" that the support must point at $z \in S$ terms of the location of z, the cost function parameter β, and the angle θ^* between the two populations of users.

Lemma 194. *Let there be 2 users located at $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$ such that $\|u_1\| = \|u_2\| = 1$, and let $\theta^* := \cos^{-1}(\langle u_1, u_2 \rangle) > 0$ be the angle between the user vectors. Let the cost function be $c(p) = \alpha \|p\|_2^\beta$ for $\alpha > 0$. If z is of the form $[r\cos(\theta), r\cos(\theta^* - \theta)]$ for $\theta \in [0, \theta^*]$, then the sign of $\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2}$ is equal to the sign of:*

$$\frac{\beta - 2}{\beta} \cos(\theta^* - 2\theta) - \cos(\theta^*).$$

Lemma 195. *Let there be 2 users located at $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$ such that $\|u_1\| = \|u_2\| = 1$, and let $\theta^* := \cos^{-1}(\langle u_1, u_2 \rangle) > 0$ be the angle between the user vectors. Let the cost function be $c(p) = \alpha \|p\|_2^\beta$ for $\alpha > 0$. Suppose that condition (C1) is satisfied for (H_1, H_2, S). If S contains a curve of the form $\{(z_1, g(z_1)) \mid x \in I\}$ for any open interval I and any differentiable function g, then for any $z_1 \in I$, it holds that:*

$$g'(z_1) \cdot \left(\frac{\beta - 2}{\beta} \cos(\theta^* - 2\theta) - \cos(\theta^*) \right) \leq 0.$$

Lemmas 194 and 195 demonstrate that if $\left(\frac{\beta-2}{\beta}\cos(\theta^* - 2\theta) - \cos(\theta^*)\right) > 0$, then the curve g must be decreasing, and if $\left(\frac{\beta-2}{\beta}\cos(\theta^* - 2\theta) - \cos(\theta^*)\right) < 0$, then the curve g must be increasing. This characterizes the "direction" of the curve in terms of the location z_1.

Leveraging (C3)

For the case of 2 users with cost function $c(p) = \|p\|_2^\beta$, the condition (C3) always holds, as long as condition (C1) holds. Since the two vectors u_1 and u_2 are linearly independent, the matrix $\mathbf{U}$ is invertible, so we can define μ to be the distribution given by $\mathbf{U}^{-1}Z$. The only remaining condition comes p being restricted to $\mathbb{R}^D_{\geq 0}$ rather than $\mathbb{R}^D$. This means that S must be contained in the convex cone generated by $[1, \cos(\theta^*)]$ and $[\cos(\theta^*), 1]$. This restriction on S is already implicitly implied by (C1): it is not difficult to see that all maximizers of (8.9) will be contained in this convex cone.

Leveraging (C2)

To leverage (C2), we obtain a functional equation that restricts the relationship between H_1, H_2, and S for a given value of P, and we instantiate this in two ways. First, when the support is a curve $(z_1, g(z_1))$, the marginal distributions Z_1 and Z_2 are related by a change of variables formula given by $Z_2 \sim g(Z_1)$. This translates into a condition on H_1 and H_2 that depends on the derivative g' and the number of producers P. Second, if the equilibrium were to contain finitely many genres, there would be a pair of functional equations relating the cdfs H_1 and H_2, the distribution over quality within each genre, and the number of producers P. We describe each of these settings in more detail below.

Case 1: support is a single curve. The first case where we instantiate (C2) is when S is equal to $\{(z_1, g(z_1)) \mid x \in M\}$ where M is a (well-behaved) subset of $\mathbb{R}_{\geq 0}$. Let h_1^* and h_2^* be the densities of the marginal distributions Z_1 and Z_2 respectively. Since $Z_2 \sim g(Z_1)$, the change of variables formula implies that the densities h_1^* and h_2^* are related as follows:

$$h_1^*(z_1) = h_2^*(g(z_1))|g'(z_1)|, \tag{E.13}$$

In order to use equation (E.13), we need to translate it into a condition on the distributions H_1 and H_2. Let h_1 and h_2 be the densities of H_1 and H_2 respectively. Then equation (E.13) can reformulated as:

$$\frac{h_1(x)}{(H_1(x))^{\frac{P-2}{P-1}}} = \frac{h_2(g(x))}{(H_2(g(x)))^{\frac{P-2}{P-1}}}|g'(x)|. \tag{E.14}$$

Equation (E.14) reveals that the constraint induced by the number of producers P can be messy in general, since it involves both the densities h_1 and h_2 and the cdfs H_1 and H_2. Intuitively, these complexities arise because H_i^* and H_i are related by a $(P-1)$th degree polynomial (put differently, the maximum of $P-1$ i.i.d. draws of a random variable does not generally have a clean structure). Nonetheless, equation (E.14) does simplify into a tractable form in special cases. For example, if $P = 2$, then the dependence on H_1 and H_2 vanishes. As another example, if g is *increasing*, then $H_1(x) = H_2(g(x))$ for any $P \geq 2$, so the dependence on H_1 and H_2 again vanishes.

Case 2: two-genre equilibria. The second case where we instantiate (C2) is when S is a subset of the union of two lines: that is,

$$S \subseteq \{(z_1, c_1 \cdot z_1) \mid z_1 \in \mathbb{R}_{\geq 0}\} \cup \{(z_1, c_2 \cdot z_1) \mid z_1 \in \mathbb{R}_{\geq 0}\},$$

where $\cos(\theta^*) \leq c_1, c_2 \leq \frac{1}{\cos(\theta^*)}$. Since linear transformations preserve lines through the origin, this means that the support of the distribution μ of $U^{-1}Z$ is also contained in the union of two lines through the origin: thus $|\text{Genre}(\mu)| \leq 2$.

A distribution Z can be entirely specified by the probabilities $\alpha_1 + \alpha_2$ that it places on each of the two lines and the conditional distribution of Z_1 along each of the lines (this in particular determines the conditional distribution of Z_2 along the lines). More specifically, the probabilities $\alpha_1 + \alpha_2$ will correspond to

$$\alpha_1 := \mathbb{P}_Z[Z \in \{(z_1, c_1 \cdot z_1) \mid z_1 \in \mathbb{R}_{\geq 0}\}]$$
$$\alpha_2 := \mathbb{P}_Z[Z \in \{(z_1, c_2 \cdot z_1) \mid z_1 \in \mathbb{R}_{\geq 0}\}],$$

and F_1 and F_2 will correspond to the cdfs of the conditional distributions

$$F_1 \sim Z_1 \mid Z \in \{(z_1, c_1 \cdot z_1\}$$
$$F_2 \sim Z_1 \mid Z \in \{(z_1, c_2 \cdot z_1\}$$

respectively. The (unique) distribution Z associated with $\alpha_1, \alpha_2, F_1, F_2$ satisfies (C2) if and only if the following pairs of functional equations are satisfied:

$$(\alpha_1 F_1(z_1) + \alpha_2 F_2(z_1)) = (H_1(z_1))^{\frac{1}{P-1}} \text{ and } \left(\alpha_1 F_1(c_1^{-1} z_2) + \alpha_2 F_2(c_2^{-1} z_2)\right) = (H_2(z_2))^{\frac{1}{P-1}}. \tag{E.15}$$

The functional equations can be solved to determine if there is a valid solution.

E.4.2 Proof of Corollary 42

We prove Corollary 42, restated below:

Corollary 42. *Suppose that there are N users split equally between two linearly independently vectors $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$, and let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2 \rangle}{\|u_1\|_2 \|u_2\|}\right)$. Let the cost function be $c(p) = \|p\|_2^\beta$. Then it holds that:*

$$\beta^* = \frac{2}{1 - \cos(\theta^*)}.$$

Proof. By Claim 191, we can assume that there are 2 normalized users $\|u_1\| = \|u_2\|$. We further assume WLOG that $u_1 = e_1$.

We claim that if there is a single-genre equilibrium, it must be in the direction of $[\cos(\theta^*/2), \sin(\theta^*/2)]$. By Corollary 44, if there is a single-genre equilibrium in a direction p, then it must maximize $\log(\langle p, u_1 \rangle) + \log(\langle p, u_1 \rangle)$. Let's let $p = [\cos(\theta), \sin(\theta)]$. Then, we see that:

$$\log(\langle p, u_1 \rangle) + \log(\langle p, u_2 \rangle) = \log(\cos(\theta)) + \log(\cos(\theta^* - \theta)) = \log\left(\frac{\cos(\theta^*) + \cos(\theta^* - 2\theta)}{2}\right),$$

which is uniquely maximized at $\theta = \theta^*/2$ as desired.

We first show that $\beta^* \leq \frac{2}{1-\cos(\theta^*)}$. Assume for sake of contradiction that there is a single-genre equilibrium. The above argument shows that it must be in the direction of $[\cos(\theta^*/2), \sin(\theta^*/2)]$. By Lemma 185, we know that the support of the equilibrium distribution is a line segment. If $\beta > \frac{2}{1-\cos(\theta^*)}$, we see that

$$\frac{\beta - 2}{\beta} \cos(\theta^* - 2\theta) - \cos(\theta^*) = 1 - \frac{2}{\beta} - \cos(\theta^*) < 0.$$

By Lemma 50 and Lemma 195, we see that the single-genre line $(z, g(z))$ must have $g'(z_1) \leq 0$ in its support, which is a contradiction.

We next show that $\beta^* \leq \frac{2}{1-\cos(\theta^*)}$. It suffices to show that the single-genre distribution in the direction of $[\cos(\theta^*/2), \sin(\theta^*/2)]$ with cdf given by $F(q) = \left(\frac{q^\beta}{2}\right)^{1/(P-1)}$. We apply Claim 50; it suffices to verify condition (C1). Notice that

$$H_1(w) = H_2(w) = \left(\frac{w^\beta}{2\cos^\beta(\theta^*/2)}\right).$$

Thus, equation (8.9) can be written as:

$$\max_z \left(\min(1, \frac{z_1^\beta}{2\cos^\beta(\theta^*/2)}) + \min(1, \frac{z_2^\beta}{2\cos^\beta(\theta^*/2)}) - c_{\mathbf{U}}(z) \right).$$

It suffices to show that that for all z, it holds that:

$$z_1^\beta + z_2^\beta - 2\cos^\beta(\theta^*/2)\left(\frac{z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)}{\sin^2(\theta^*)}\right)^\beta \leq 0.$$

Let $z = [r\cos(\theta), r\cos(\theta^* - \theta)]$. Then this reduces to:

$$\cos^\beta(\theta) + \cos^\beta(\theta^* - \theta) \leq 2\cos^\beta(\theta^*/2) \leq 0.$$

We observe that $\cos^\beta(\theta) + \cos^\beta(\theta^* - \theta)$ is maximized at $\theta = \theta^*/2$, which proves the desired statement.

□

E.4.3 Proofs for Chapter 8.4.1

We prove Proposition 45, restated below:

Proposition 45. *Suppose that there are N users split equally between two linearly independently vectors $u_1, u_2 \in \mathbb{R}^2_{\geq 0}$, and let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2\rangle}{\|u_1\|_2 \|u_2\|}\right)$ be the angle between the user vectors. Let the cost function be $c(p) = \|p\|_2^\beta$, and let $P \geq 2$. Let μ be a symmetric Nash equilibrium such that the distributions $\langle u_1, p\rangle$ and $\langle u_2, p\rangle$ over $\mathbb{R}_{\geq 0}$ are absolutely continuous. As long as $\beta \neq 2$ or $\theta^* \neq \pi/2$, the support of μ does not contain an ℓ_2-ball of radius ε for any $\varepsilon > 0$.*[2]

Proof of Proposition 45. Assume for sake of contradiction that the support of μ contains an ℓ_2-ball of radius $\varepsilon_1 > 0$. We apply Lemma 50 and show that condition (C1) is violated. Since μ contains a ball of ε_1-radius ball, we know that the distribution Z over $\mathbf{U}p$ over $p \sim \mu$ contains an ℓ_2 ball of radius $\varepsilon_2 > 0$. Let this ball be B. Notice that Z_1 and Z_2 are absolutely continuous by assumption, Z_1 and Z_2 have bounded support, and the function $m \mapsto m^{P-1}$ is Lipschitz on any bounded interval: this means that H_1 and H_2 are also absolutely continuous. This means that densities exist a.e. For $(z_1, z_2) \in B$, we can apply the first-order condition in Lemma 193 to obtain that:

$$h_1(z_1) = \frac{\partial c_{\mathbf{U}}(z)}{\partial z_1}$$

We see that this needs to be satisfied for $z = [z_1, m]$ where $m \in (z_2 - \varepsilon', z_2 + \varepsilon')$. This means that the mapping $m \mapsto \frac{\partial c_{\mathbf{U}}([z_1, m])}{\partial z_1}$ needs to be a constant on $m \in (z_2 - \varepsilon', z_2 + \varepsilon')$. This means

[2] The case of $\beta = 2$ and $\theta^* = \pi/2$ is degenerate and permits a range of possible equilibria.

that the derivative of this mapping with respect to z_2 needs to be 0, so:

$$\frac{\partial^2 c_{\mathbf{U}}([z_1, z_2])}{\partial z_1 \partial z_2} = 0 \tag{E.16}$$

for all $z \in B$.

We apply Lemma 194 to show that equation (E.16) cannot be zero on all of B. For all z that satisfy equation (E.16), Lemma 194 implies if we represent z as $\mathbf{U}[r\cos(\theta), r\sin(\theta)]$, then

$$\frac{\beta - 2}{\beta} \cos(\theta^* - 2\theta) = \cos(\theta^*).$$

If equation (E.16) holds for all $z \in B$, then it must hold at all θ within some nonempty interval. This is a contradiction as long as $\beta \neq 2$ or $\theta^* \neq \pi/2$.

For the special case where $\beta = 2$ and $\theta^* = \pi/2$, □

We next prove Theorem 46, restated below:

Theorem 46. *Suppose that there are N users split equally between two linearly independently vectors $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$, and let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2 \rangle}{\|u_1\|_2 \|u_2\|}\right)$ be the angle between the user vectors. Let the cost function be $c(p) = \|p\|_2^\beta$. Let μ be a a distribution on $\mathbb{R}^d$ such that the distributions $\langle u_1, p\rangle$ and $\langle u_2, p\rangle$ over $\mathbb{R}_{\geq 0}$ over $\mathbb{R}_{\geq 0}$ for $p \sim \mu$ are absolutely continuous and twice continuously differentiable within their supports. There are two regimes based on β and θ^*:*

1. *If $\beta < \beta^* = \frac{2}{1-\cos(\theta^*)}$ and if μ is a symmetric mixed equilibrium, then μ satisfies $|\text{Genre}(\mu)| = 1$.*
2. *If $\beta > \beta^* = \frac{2}{1-\cos(\theta^*)}$, if $|\text{Genre}(\mu)| < \infty$, and if the conditional distribution of $\|p\|$ along each genre is continuously differentiable, then μ is not an equilibrium.*

We split into two propositions: together, these propositions directly imply Theorem 46.

Proposition 196. *Consider the setup in Theorem 46. If $\beta < \beta^* = \frac{2}{1-\cos(\theta^*)}$ and μ is a symmetric mixed equilibrium, then μ satisfies $|\text{Genre}(\mu)| = 1$.*

Proposition 197. *Consider the setup in Theorem 46. If $\beta > \beta^* = \frac{2}{1-\cos(\theta^*)}$, if $|\text{Genre}(\mu)| < \infty$, and if the conditional distribution of $\|p\|$ along each genre is continuous differentiable, then μ is not an equilibrium.*

To prove Proposition 196, we leverage the machinery given by Lemma 50 as follows. Condition (C1) helps us show that the support S can be specified by $(w, g(w))$ for an increasing function w: in particular, Lemma 193 enables us to show that S must be one-to-one, and Lemma 195 enables us to pin down the sign of g'. Using condition (C2), which simplifies since g is increasing, we show a functional equation in terms of g that has a unique solution at the single-genre equilibrium. We formalize this below.

Proof of Proposition 196. By Claim 191, it suffices to focus on the case of 2 normalized users. By Lemma 50, it suffices to study (H_1, H_2, S) that satisfy (C1), (C2), and (C3).

Let $\text{supp}(H_1) = I_1$ and let $\text{supp}(H_2) = I_2$. Note that since the distributions are twice continuously differentiable, we know that the densities h_1 and h_2 exist and are continuously differentiable a.e on I_1 and I_2 respectively. We break the proof into several steps.

Step 1: there exists a one-to-one function g such that $S = \{(w, g(w)) \mid w \in I_1\}$ and where g is continuously differentiable and strictly increasing. We first show that $\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} < 0$ everywhere. By Lemma 194, it suffices to show that $\frac{\beta-2}{\beta}\cos(\theta^* - 2\theta) - \cos(\theta^*) < 0$. To see this, notice that

$$\frac{\beta-2}{\beta}\cos(\theta^* - 2\theta) - \cos(\theta^*) < 0 \leq \frac{\beta-2}{\beta} - \cos(\theta^*) = 1 - \cos(\theta^*) - \frac{2}{\beta} < 0$$

because $\beta < \frac{2}{1-\cos(\theta^*)}$.

We now show that the support S is equal to $\{(w, g(w)) \mid w \in I_1\}$ for some one-to-one function $g : I_1 \to I_2$. To show this, it suffices to show that the support does contain both (z_1, z_2) and (z_1, z_2') for $z_2 \neq z_2'$ (and, analogously, the support does not contain both (z_1', z_2) and (z_1, z_2) for $z_1 \neq z_1'$). Notice that for any fixed value of z_1, the function $z_2 \mapsto \frac{\partial c_{\mathbf{U}}([z_1, z_2])}{\partial z_1}$ is strictly decreasing. If (z_1, z_2) and (z_1, z_2') are both in the support, then by Lemma 193, it must be true that:

$$h_1(z_1) = \frac{\partial c_{\mathbf{U}}([z_1, z_2])}{\partial z_1} = \frac{\partial c_{\mathbf{U}}([z_1, z_2'])}{\partial z_1}.$$

However, since $z_2 \mapsto \frac{\partial c_{\mathbf{U}}([z_1, z_2])}{\partial z_1}$ is strictly decreasing, this means that $z_2 = z_2'$ as desired.

We can thus implicitly define the function g by the (unique) value such that:

$$Q(w, g(w)) - h_1(w) = 0$$

where

$$Q(z_1, z_2) := \frac{\partial c_{\mathbf{U}}([z_1, z_2])}{\partial z_1}.$$

Uniqueness follows from the fact that Q is a strictly decreasing function in its second argument, since $\frac{\partial Q(w, g(w))}{\partial z_2} = \frac{\partial^2 c_{\mathbf{U}}([w, g(w)])}{\partial z_1 \partial z_2} < 0$ as we showed previously. Since $h_1(w)$ is continuously differentiable and since:

$$\frac{\partial Q(w, g(w))}{\partial z_2} \neq 0$$

for $w \in I_1$, we can apply the implicit function theorem to see that $g(w)$ is continuously differentiable for $w \in I_1$.

We next show that g is increasing on I_1. Within the interior of I_1, by Lemma 195 along with the fact that $\frac{\beta-2}{\beta}\cos(\theta^* - 2\theta) - \cos(\theta^*) < 0$ everywhere, we see that g is a strictly increasing function on each contiguous portion of I_1. It thus suffices to show I_1 is an interval and that there are no gaps. If there is a gap, there must be a gap for both z_1 *and* z_2 at the same point z since the support is on-to-one and closed. However, if z is right above the gap, the producer would obtain higher utility by choosing $(1-\varepsilon)z$ for sufficiently small ε to ensure that $(1-\varepsilon)z$ is within the gap on both coordinates. This means that I_1 is an interval, which proves g is an increasing function.

Step 2: differential equation. We show that

$$g'(w)g(w) - g'(w)w\cos(\theta^*) = w - g(w)\cos(\theta^*), \tag{E.17}$$

for all $w \in \text{supp}(H_1)$.

First, we derive the the condition that we described in equation (E.14) and further simplify it using that g is increasing. Let $H_1^*(w) = H_1(w)^{\frac{1}{P-1}}$ and $H_2^*(w) = H_2(w)^{\frac{1}{P-1}}$. The densities h_1^* and h_2^* take the following form:

$$h_1^*(w) = (H_1^*)'(w) = \frac{1}{P-1}h_1(w)H_1(w)^{-\frac{P-2}{P-1}}$$
$$h_2^*(w) = (H_2^*)'(w) = \frac{1}{P-1}h_2(w)H_2(w)^{-\frac{P-2}{P-1}}.$$

In order for there to exist a distribution μ that satisfies condition (C2), it must hold that $H_1^*(w) = H_2^*(g(w))$ because g is increasing. (This also means that $H_1(w) = H_2(g(w))$.) This means that $h_1^*(w) = h_2^*(g(w))g'(w)$ and $H_1(w) = H_2(g(w))$. Plugging this into the above expressions for h_1^* and h_2^*, this means that:

$$h_1(w) = (P-1)h_1^*(w)H_1(w)^{\frac{P-2}{P-1}} = (P-1)g'(w)h_2^*(g(w))H_2(g(w))^{\frac{P-2}{P-1}} = h_2(w)g'(w).$$

This means that

$$g'(w) = \frac{h_1(w)}{h_2(w)} = \frac{w - g(w)\cos(\theta^*)}{g(w) - w\cos(\theta^*)},$$

where the last line follows from Lemma 193. This gives us the desired differential equation.

Step 3: solving the differential equation. We claim that the only valid solution to the differential equation (E.17) is $g(w) = w$. To see this, let $f(w) = \frac{g(w)}{w}$. This means that $wf(w) = g(w)$ and thus $f(w) + wf'(w) = g'(w)$. Plugging this into equation (E.17) and simplifying we obtain a separable differential equation. The solutions to this differential equation are $f(w) = 1$ and the following:

$$f_K^*(w) = K - \log(w) = \frac{1}{2}\left((1+\cos(\theta^*))\log(1+f(w)) - (1+\cos(\theta^*))\log(1-f(w))\right)$$

for some constant K. Notice that for f_K^* to even be well-defined, we know that $f_K^*(w) < 1$ everywhere.

Assume for sake of contradiction that there exists an equilibrium with support given by $\{(w, g(w)) \mid w \in I\}$ for $g(w) \neq w$. Then we know that $g(w) = f_K^*(w) \cdot x$ for some K. In order for this solution to even be well-defined, it would imply that $f_K^*(w) < 1$ everywhere. This implies that $g(w) < w$, for all $w \in I_1$. However, we know that the function g^{-1} must satisfy the differential equation too (and $g^{-1}(w) \neq w$), so by an analogous argument, we know that $g^{-1}(w) < w$ for all $w \in I_2$, which means that $w < g(w)$. This is a contradiction.

We can thus conclude that since $g(x) = x$, we have that $|\text{Genre}(\mu)| = 1$ as desired. □

To prove Proposition 197, we also leverage the machinery in Lemma 50. We use Lemma 193 to rule out all finite-genre equilibria except for two-genre equilibria. We can show that $H_1(w)$ and $H_2(w)$ grow proportionally to w^β. Then, we can implement this knowledge of H_1 and H_2 into the finite genre formulation of condition (C2) in equation (E.15) and show that no solutions to the functional equation exist for finite P. We formalize this below.

Proof of Proposition 197. By Claim 191, it suffices to focus on the case of 2 normalized users. We further assume WLOG that $u_1 = e_1$ and $u_2 = [\cos(\theta^*), \sin(\theta^*)]$. Since $\beta > \frac{2}{1-\cos(\theta^*)}$, we know by Corollary 42 that there is no single-genre equilibrium. Assume for sake of contradiction that there exists a *finite*-genre equilibrium μ with $|\text{Genre}(\mu)| \geq 2$. By Lemma 50, we know that there exists H_1, H_2 and S associated with μ that satisfy (C1)-(C3). Our proof boils down to two steps:

- *Step 1:* We show that $\text{Genre}(\mu) = \{\theta_1, \theta_2\}$ for some $\theta_1 < \theta^*/2 < \theta_2$.
- *Step 2:* We show that no two-genre distribution μ exists.

Step 1. Let us first translate the concept of genres to the reparameterized space. First, we consider the following set:

$$\text{Genre}_Z(S) := \left\{ \frac{1}{c_{\mathbf{U}}(z)}[z_1, z_2] \mid z \in S \right\}.$$

Since vectors in $\text{Genre}_Z(S)$ are of the form $[\cos(\theta), \cos(\theta^* - \theta)]$ by the normaalization by $c_{\mathbf{U}}(z)$, we can actually define a set of *angles*:

$$\text{Genre}_\Theta(S) := \left\{ \cos^{-1}(z_1) \mid [z_1, z_2] \in \text{Genre}_Z(S) \right\}.$$

We see that $\theta \in \text{Genre}_\Theta(S)$ if and only if $[\cos(\theta), \cos(\theta^* - \theta)] \in \text{Genre}_Z(S)$ if and only if $[\cos(\theta), \sin(\theta)] \in \text{Genre}(\mu)$. Elements of $\text{Genre}_\Theta(S)$ thus exactly corresponds to genres of $\text{Genre}(\mu)$.

We first observe that every $\theta \in \text{Genre}_\Theta(S)$ is in $(0, \theta^*)$. By (C1) of Lemma 50, the set S must be contained in the convex cone of $[1, \cos(\theta^*)]$ and $[\cos(\theta^*, 1]$, which implies that $\theta \in [0, \theta^*]$. It thus suffices to show that $\theta \neq 0$ and $\theta \neq \theta^*$. We show that $\theta \neq 0$ (the case of $\theta \neq \theta*$ follows from an analogous argument). In this case, we see that there must be some set of the form $\{[r, r\cos(\theta^*)] \mid r \in \mathbb{R}_{\geq 0}\}$ that is subset of S. If $\theta^* = \pi/2$, then this would mean the distribution given by H_2 would have a point mass at 0, which is clearly not possible at equilibrium. Otherwise, if $\theta^* < \pi/2$, we apply (C1) and Lemma 193, and we see that $h_2(r\cos(\theta^*)) = 0$. However, this is a contradiction, since there is positive probability mass on some line segment on this genre by assumption.

Now, we observe that the support of the cdfs H_1 and H_2 must be bounded *intervals* of the form $[0, z_1^{\max}]$ and $[0, z_2^{\max}]$. First, we show that $\max(\text{supp}(H_1)), \max(\text{supp}(H_1)) < \infty$. By (C1), we see that a producer must achieve nonzero profit (since they always so $c_{\mathbf{U}}(z) \leq 2$, which means that $z_1, z_2 \leq \frac{2}{\alpha}$ as desired. This means that we can set $z_1^{\max} = \max(\text{supp}(H_1))$ and $z_2^{\max} = \max(\text{supp}(H_2))$. Next, we show that the supports of H_1 and H_2 contain the full intervals $[0, z_1^{\max}]$ and $[0, z_2^{\max}]$, respectively. Assume for sake of contradiction that the support of H_1 does not contain some interval $(x, x+\varepsilon)$ for $\varepsilon > 0$ within $[0, z_1^{\max}]$. Let ε be defined so that $z_1 = x + \varepsilon \in \text{supp}(H_1)$. However, this means that there exists z_2 such that $[z_1, z_2] \in S$ and, moreover, $[z_1, z_2]$ must be located on a genre $\theta \in (0, \theta^*)$. We can thus reduce z_1 and hold z_2 fixed, while keeping $H_1(z_1) + H_2(z_2)$ fixed, and reducing the cost $c_{\mathbf{U}}(z)$, which violates the fact that $[z_1, z_2]$ is a maximizer of (8.9). An analogous argument shows that the support of H_2 is the full interval $[0, z_2^{\max}]$.

Next, we show that for $\theta, \theta' \in \text{Genre}_\Theta(S)$, it must hold that

$$\frac{\sin(\theta^* - \theta)}{\cos^{\beta-1}(\theta)} = \frac{\sin(\theta^* - \theta')}{\cos^{\beta-1}(\theta')} \text{ and } \frac{\sin(\theta)}{\cos^{\beta-1}(\theta^* - \theta)} = \frac{\sin(\theta')}{\cos^{\beta-1}(\theta^* - \theta')} \tag{E.18}$$

To prove this, suppose that $|\text{Genre}_Z(S)| = G$ and label the genres by the indices $1, \ldots, G$ arbitrarily. For $z_1 \in \text{supp}(H_1)$ let $T(z_1) \subseteq \{1, \ldots, G\}$ be the set of genres j where there exists z_2 such that $(z_1, z_2) \in S$ and $[z_1, z_2]$ points in the direction of $[\cos(\theta_j), \cos(\theta^* - \theta_j)]$. By Lemma 193, for all $i \in T(z_1)$, it must hold that:

$$h_1(z_1) = \beta z_1^{\beta-1} \alpha \cdot \frac{\sin(\theta^* - \theta_i)}{\sin(\theta^*)} \cdot \frac{1}{\cos(\theta_i)^{\beta-1}}.$$

This means that for $i, i' \in T(z_1)$, it holds that

$$\frac{\sin(\theta^* - \theta_i)}{\cos^{\beta-1}(\theta_i)} = \frac{\sin(\theta^* - \theta_{i'})}{\cos^{\beta-1}(\theta_{i'})}.$$

We now generalize this argument to arbitrary genres $\theta, \theta' \in \text{Genre}_\Theta(S)$. Consider $1 \leq i, i' \leq G$. Even though θ_i and $\theta_{i'}$ may not be in the same set $T(z_1)$, we show that there must be some "path" connecting θ_i and $\theta_{i'}$. To formalize this, for each genre $1 \leq i \leq G$, let $S_i = \{z_1 \in \text{supp}(H_1) \mid i \in T(z_i)\}$. Let's define an undirected graph vertices $[G]$ and an edge (i_1, i_2) if and only if $S_{i_1} \cap S_{i_2} \neq \emptyset$. The argument from the previous paragraph showed that if there an edge between i and i', then $\frac{\sin(\theta^* - \theta_i)}{\cos^{\beta-1}(\theta_i)} = \frac{\sin(\theta^* - \theta_{i'})}{\cos^{\beta-1}(\theta_{i'})}$. Moreover, if there exists a path from i to i' in this graph, then we can chain together equalities along each edge in the path to prove $\frac{\sin(\theta^* - \theta_i)}{\cos^{\beta-1}(\theta_i)} = \frac{\sin(\theta^* - \theta_{i'})}{\cos^{\beta-1}(\theta_{i'})}$. The only remaining case is that there is no path from i to i'. However, this would mean that the vertices $[G]$ can be divided into a partition $P_1, \ldots, P_n$ for $n > 1$ such that there is no edge across partitions. Note that $\cup_{1 \leq i \leq G} S_i = \text{supp}(H_1)$, which we already proved is equal to $[0, z_1^{\max}]$. Thus, this would mean that the disjoint, closed sets $\cup_{i \in P_1} S_i, \ldots, \cup_{i \in P_n} S_i$ have union equal to $[0, z_1^{\max}]$, which is not possible Sierpinski (1918). Thus we have shown that $\frac{\sin(\theta^* - \theta_i)}{\cos^{\beta-1}(\theta_i)} = \frac{\sin(\theta^* - \theta_{i'})}{\cos^{\beta-1}(\theta_{i'})}$ for any $1 \leq i, i' \leq G$ and an analogous argument shows that $\frac{\sin(\theta_i)}{\cos^{\beta-1}(\theta^* - \theta_i)} = \frac{\sin(\theta_{i'})}{\cos^{\beta-1}(\theta^* - \theta_{i'})}$. This proves equation (E.18).

We next show that there exist exactly 2 genres given by $\theta_1 < \theta_2$. Using Lemma 194, we see that for any θ, there are at most two values of $\theta' \neq \theta_1$ such that equation (E.18) can hold. Moreover, by Lemma 195, one of these values lies within the region where g' would have to be negative (which is not possible). Thus, there are at most two genres, and Lemma 194 further tells us that they lie on opposite sides of $\theta^*/2$.

Step 2. Condition (C2) gives us functional equations that the distribution μ must satisfy for $P < \infty$. More specifically, let F_1 be the cdf of the magnitude of the genre given by θ_1, and let F_2 be the cdf of the magnitude of the genre given by θ_2. Then we obtain the following functional equations:

$$\left(\alpha_1 F_1\left(\frac{z_1}{\cos(\theta_1)}\right) + \alpha_2 F_2\left(\frac{z_1}{\cos(\theta_2)}\right)\right)^{P-1} = H_1(z_1)$$
$$\left(\alpha_1 F_1\left(\frac{z_2}{\cos(\theta^* - \theta_1)}\right) + \alpha_2 F_2\left(\frac{z_2}{\cos(\theta^* - \theta_2)}\right)\right)^{P-1} = H_2(z_1).$$

For these functional equations to be useful, we need to compute the cdfs H_1 and H_2. This will involve some notation: as in the previous step, let the genres be $\{\theta_1, \theta_2\}$ where $\theta_1 < \theta^*/2 < \theta_2$. Let $r_1^{\max} := \max(\text{supp}(F_1))$ be the maximum value in the support of F_1 and let $r_2^{\max} := \max(\text{supp}(F_2))$ be the maximum value in the support of F_2. We define:

$$i_1 := \arg\max_{i \in \{1,2\}} r_i \cos(\theta_i) \qquad i_2 := \arg\max_{i \in \{1,2\}} r_i \cos(\theta^* - \theta_i)$$

which correspond to which genre produces the highest value of z_1 and z_2 respectively.

We apply Lemma 193 to see that for all z_1 and z_2 in the support of H_1 and H_2, it holds that:

$$h_1(z_1) = \beta z_1^{\beta-1} \alpha \cdot \frac{\sin(\theta^* - \theta_{i_1})}{\sin(\theta^*)} \cdot \frac{1}{\cos(\theta_{i_1})^{\beta-1}}$$
$$h_1(z_2) = \beta z_2^{\beta-1} \alpha \cdot \frac{\sin(\theta_{i_2})}{\sin(\theta^*)} \cdot \frac{1}{\cos(\theta^* - \theta_{i_2})^{\beta-1}}.$$

We can integrate with respect to z_1 and z_2 to obtain that $H_1(z_1) = c_1 z_1^{\beta}$ and $H_1(z_2) = c_2 z_2^{\beta}$, such that:

$$c_1 = \alpha \cdot \frac{\sin(\theta^* - \theta_{i_1})}{\sin(\theta^*)} \cdot \frac{1}{\cos(\theta_{i_1})^{\beta-1}} \tag{E.19}$$
$$c_2 = \alpha \cdot \frac{\sin(\theta_{i_2})}{\sin(\theta^*)} \cdot \frac{1}{\cos(\theta^* - \theta_{i_2})^{\beta-1}}. \tag{E.20}$$

WLOG assume that $c_1 \geq c_2$ for the remainder of the analysis.

Using this specification of H_1 and H_2, we can write the functional equations as

$$\alpha_1 F_1\left(\frac{z_1}{\cos(\theta_1)}\right) + \alpha_2 F_2\left(\frac{z_1}{\cos(\theta_2)}\right) = c_1^{\frac{1}{P-1}} z_1^{\frac{\beta}{P-1}}$$
$$\alpha_1 F_1\left(\frac{z_2}{\cos(\theta^* - \theta_1)}\right) + \alpha_2 F_2\left(\frac{z_2}{\cos(\theta^* - \theta_2)}\right) = c_2^{\frac{1}{P-1}} z_2^{\frac{\beta}{P-1}}.$$

By taking a derivative with respect to z_1 and z_2, we see that for any z_1 within the support of H_1 and z_2 within the support of H_2, it holds that:

$$\frac{\alpha_1}{\cos(\theta_1)} f_1\left(\frac{z_1}{\cos(\theta_1)}\right) + \frac{\alpha_2}{\cos(\theta_2)} f_2\left(\frac{z_1}{\cos(\theta_2)}\right) = c_1^{\frac{1}{P-1}} \frac{\beta}{P-1} z_1^{\frac{\beta}{P-1}-1}. \tag{E.21}$$

$$\frac{\alpha_1}{\cos(\theta^*-\theta_1)} f_1\left(\frac{z_2}{\cos(\theta^*-\theta_1)}\right) + \frac{\alpha_2}{\cos(\theta^*-\theta_2)} f_2\left(\frac{z_2}{\cos(\theta^*-\theta_2)}\right) = c_2^{\frac{1}{P-1}} \frac{\beta}{P-1} z_2^{\frac{\beta}{P-1}-1}. \tag{E.22}$$

We prove that these functional equations have no valid solution. To show this, we prove that any solution to equations (E.21) and (E.22) would have negative density somewhere. Where the negative density occurs depends on i_1 and i_2.

We thus do casework on i_1 and i_2. In this analysis, we will use the notation $z_1^{\max}$ to denote $\max(\mathrm{supp}(H_1))$ and $z_2^{\max}$ to denote $\max(\mathrm{supp}(H_2))$. Note that by definition, $z_1^{\max} = r_{i_1}\cos(\theta_{i_1})$ and $z_2^{\max} = r_{i_2}\cos(\theta^* - \theta_{i_2})$.

First, we reduce the number of cases needed by using the fact that $c_1 \geq c_2$ (which we assumed earlier WLOG). In particular, this turns out to imply that $i_2 \neq 1$. More precisely, we show:

$$r_1^{\max}\cos(\theta^*-\theta_1) < r_2^{\max}\cos(\theta^*-\theta_2) \tag{E.23}$$

To show this, assume for sake of contradiction that $r_1^{\max}\cos(\theta^*-\theta_1) \geq r_2^{\max}\cos(\theta^*-\theta_2)$. Then we'd have that

$$z_2^{\max} = r_1^{\max}\cos(\theta^*-\theta_1) < r_1^{\max}\cos(\theta_1) \leq z_1^{\max}$$

which would imply that $c_1 < c_2$, which is a contradiction.

We thus split into 2 cases based on i_1.

- **Case 1**: $r_2^{\max}\cos(\theta_2) < r_1^{\max}\cos(\theta_1)$
- **Case 2**: $r_1^{\max}\cos(\theta_1) \leq r_2^{\max}\cos(\theta_2)$

Let's first handle **Case 1**. Since $z_1^{\max} = r_1^{\max}\cos(\theta_1) > r_2^{\max}\cos(\theta_2)$, we see that

$$\frac{z_1^{\max}}{\cos(\theta_2)} > r_2^{\max}$$

is not in the support of F_2. This means that the density f_2 of F_2 at $\frac{z_1^{\max}}{\cos(\theta_2)}$ is equal to 0 and, moreover, there exists $z_1^* < z_1^{\max}$ sufficiently close to $z_1^{\max}$ such that z_1^* is in the support of H_1 and $\frac{z_1^*}{\cos(\theta_2)}$ is not in the support of F_2. At z_1^*, by equation (E.21), we see that:

$$\frac{\alpha_1}{\cos(\theta_1)} f_1\left(\frac{z_1^*}{\cos(\theta_1)}\right) = \frac{\alpha_1}{\cos(\theta_1)} f_1\left(\frac{z_1^*}{\cos(\theta_1)}\right) + \frac{\alpha_2}{\cos(\theta_2)} f_2\left(\frac{z_1^*}{\cos(\theta_2)}\right) = c_1^{\frac{1}{P-1}} \frac{\beta}{P-1} (z_1^*)^{\frac{\beta}{P-1}-1}.$$

Now, let's let z_2^* be such that:

$$z_2^* := z_1^* \frac{\cos(\theta^*-\theta_1)}{\cos(\theta_1)}.$$

At z_2^*, we see that the left-hand side of equation (E.22) satisfies

$$\begin{aligned}
&\frac{\alpha_1}{\cos(\theta^*-\theta_1)} f_1\left(\frac{z_2^*}{\cos(\theta^*-\theta_1)}\right) + \frac{\alpha_2}{\cos(\theta^*-\theta_2)} f_2\left(\frac{z_2^*}{\cos(\theta^*-\theta_2)}\right)\\
&\geq \frac{\alpha_1}{\cos(\theta^*-\theta_1)} f_1\left(\frac{z_2^*}{\cos(\theta^*-\theta_1)}\right)\\
&= \frac{\cos(\theta_1)}{\cos(\theta^*-\theta_1)}\left(\frac{\alpha_1}{\cos(\theta_1)} f_1\left(\frac{z_1^*}{\cos(\theta_1)}\right)\right)\\
&= \frac{\cos(\theta_1)}{\cos(\theta^*-\theta_1)}\left(c_1^{\frac{1}{P-1}}\frac{\beta}{P-1}(z_1^*)^{\frac{\beta}{P-1}-1}\right)\\
&= \frac{\cos(\theta_1)}{\cos(\theta^*-\theta_1)}\left(c_1^{\frac{1}{P-1}}\frac{\beta}{P-1}\left(z_2^*\frac{\cos(\theta_1)}{\cos(\theta^*-\theta_1)}\right)^{\frac{\beta}{P-1}-1}\right)\\
&= c_1^{\frac{1}{P-1}}(z_2^*)^{\frac{\beta}{P-1}-1}\frac{\beta}{P-1}\left(\frac{\cos(\theta_1)}{\cos(\theta^*-\theta_1)}\right)^{\frac{\beta}{P-1}}\\
&> c_2^{\frac{1}{P-1}}\frac{\beta}{P-1}(z_2^*)^{\frac{\beta}{P-1}-1},
\end{aligned}$$

where the last inequality uses that $c_1 \geq c_2$ (which we assumed WLOG earlier) and $\theta_1 < \theta^*/2$. However, this is a contradiction since (E.22) must hold.

Let's next handle **Case 2**. By equation (E.23), we know that $z_2^{\max} = r_2^{\max}\cos(\theta^*-\theta_2) > r_1^{\max}\cos(\theta^*-\theta_1)$, so there exists z_2 such that $z_2 \in (r_1^{\max}\cos(\theta^*-\theta_1), z_2^{\max})$. At this value of z_2, we see by equation (E.22) that

$$\frac{\alpha_2}{\cos(\theta^*-\theta_2)} f_2\left(\frac{z_2}{\cos(\theta^*-\theta_2)}\right) = c_2^{\frac{1}{P-1}}\frac{\beta}{P-1}z_2^{\frac{\beta}{P-1}-1}.$$

Since $r_2^{\max}\cos(\theta_2) = \frac{z_2^{\max}\cos(\theta_2)}{\cos(\theta^*-\theta_2)}$, we know that $z_1 = \frac{z_2\cos(\theta_2)}{\cos(\theta^*-\theta_2)}$ is in the support of H_1. By equation (E.21), for $z_1 = \frac{z_2\cos(\theta_2)}{\cos(\theta^*-\theta_2)}$:

$$\frac{\alpha_2}{\cos(\theta_2)} f_2\left(\frac{z_1}{\cos(\theta_2)}\right) \leq \frac{\alpha_1}{\cos(\theta_1)} f_1\left(\frac{z_1}{\cos(\theta_1)}\right) + \frac{\alpha_2}{\cos(\theta_2)} f_2\left(\frac{z_1}{\cos(\theta_2)}\right) = c_1^{\frac{1}{P-1}}\frac{\beta}{P-1}z_1^{\frac{\beta}{P-1}-1}.$$

Putting this all together, we see that:

$$\begin{aligned}
c_1^{\frac{1}{P-1}}\frac{\beta}{P-1}z_1^{\frac{\beta}{P-1}-1} \geq \frac{\alpha_2}{\cos(\theta_2)} f_2\left(\frac{z_1}{\cos(\theta_2)}\right) &= \frac{\alpha_2\cos(\theta^*-\theta_2)}{\cos(\theta_2)} f_2\left(\frac{z_2}{\cos(\theta^*-\theta_2)}\right)\\
&= c_2^{\frac{1}{P-1}}\frac{\cos(\theta^*-\theta_2)}{\cos(\theta_2)}\frac{\beta}{P-1}z_2^{\frac{\beta}{P-1}-1}\\
&= c_2^{\frac{1}{P-1}}\frac{\cos(\theta^*-\theta_2)}{\cos(\theta_2)}\frac{\beta}{P-1}\left(z_1\frac{\cos(\theta^*-\theta_2)}{\cos(\theta_2)}\right)^{\frac{\beta}{P-1}-1}\\
&= c_2^{\frac{1}{P-1}}\frac{\beta}{P-1}z_1^{\frac{\beta}{P-1}-1}\left(\frac{\cos(\theta^*-\theta_2)}{\cos(\theta_2)}\right)^{\frac{\beta}{P-1}}
\end{aligned}$$

This implies that:

$$\frac{c_1}{c_2} \geq \left(\frac{\cos(\theta^* - \theta_2)}{\cos(\theta_2)}\right)^\beta.$$

However, by equations (E.19) and (E.20), we also see that:

$$\frac{c_1}{c_2} = \frac{\sin(\theta^* - \theta_2)}{\sin(\theta_2)} \frac{\cos(\theta^* - \theta_2)^{\beta-1}}{\cos(\theta_2)^{\beta-1}} = \frac{\tan(\theta^* - \theta_2)}{\tan(\theta_2)} \frac{\cos(\theta^* - \theta_2)^{\beta}}{\cos(\theta_2)^{\beta}} < \frac{\cos(\theta^* - \theta_2)^{\beta}}{\cos(\theta_2)^{\beta}}, \tag{E.24}$$

where the last step uses that $\theta^* - \theta_2 < \theta_2$. This is a contradiction.

□

E.4.4 Proofs for Chapter 8.4.2

We prove Proposition 47, restated below.

Proposition 47. *Suppose that there are 2 users located at the standard basis vectors $e_1, e_2 \in \mathbb{R}^2$, and the cost function is $c(p) = \|p\|_2^\beta$. For $P = 2$ and $\beta \geq \beta^* = 2$, there is an equilibrium μ supported on the quarter-circle of radius $(2\beta^{-1})^{1/\beta}$, where the angle $\theta \in [0, \pi/2]$ has density $f(\theta) = 2\cos(\theta)\sin(\theta)$.*

Conceptually speaking, the machinery given by Lemma 50 enables us to systematically identify the equilibrium in the concrete market instance of Proposition 47. Condition (C1) is simple along the quarter circle: by Lemma 193, the densities $h_1(u)$ and $h_2(v)$ are *proportional* to u and v. Since the support of a single curve and $P = 2$, condition (C2) can be simplified to a clean condition on the densities h_1 and h_2 given by (E.13).

To actually prove Proposition 47, we only need to *verify* that the equilibrium μ in Proposition 47 which is easier.

Proof. By Lemma 50, it suffices to prove that (C1)-(C3) hold for H_1, H_2, and S associated with the distribution μ in the statement of the proposition. Conditions (C2) and (C3) follow by construction of μ, so it suffices to prove (C1).

First, we claim that

$$H_1(z_1) = \left(\frac{2}{\beta}\right)^{-2/\beta} z_1^2, \text{ and } H_2(z_2) = \left(\frac{2}{\beta}\right)^{-2/\beta} z_2^2.$$

We show that $H_2(z_2) = \left(\frac{2}{\beta}\right)^{-2/\beta} z_2^2$ (an analogous argument applies to H_1). We see that H_2 is supported on $\left[0, \left(\frac{2}{\beta}\right)^{1/\beta}\right]$ by construction, so it suffices to show that

$$h_2(z_2) = 2\left(\frac{2}{\beta}\right)^{-2/\beta} z_2$$

on this interval. Since $z_2 = \left(\frac{2}{\beta}\right)^{1/\beta} \sin(\theta)$, by the change of variables formula for $P = 2$, we see that

$$h_2(z_2) \left(\frac{2}{\beta}\right)^{1/\beta} \cos(\theta) = f(\theta) = 2\sin(\theta)\cos(\theta).$$

We can solve and obtain:

$$h_2(z_2) = 2\left(\frac{2}{\beta}\right)^{-1/\beta} \sin(\theta) = 2\left(\frac{2}{\beta}\right)^{-2/\beta} z_2,$$

as desired.

Now, we prove (C1). Applying Lemma 192, we see that:

$$H_1(z_1) + H_2(z_2) - c_{\mathbf{U}}(z) = \left(\min\left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_1^2, 1\right) + \min\left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_2^2, 1\right)\right) - (z_1^2 + z_2^2)^{\beta/2}.$$

Thus, equation (8.9) can be written as:

$$\max_{z_1, z_2 \geq 0} \left(\left(\min\left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_1^2, 1\right) + \min\left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_2^2, 1\right)\right) - (z_1^2 + z_2^2)^{\beta/2}\right) \tag{E.25}$$

We wish to show equation (E.25) is maximized whenever $z \in S$. Since $z_1^2 + z_2^2 = \left(\frac{2}{\beta}\right)^{2/\beta}$ for any $z \in S$, this follows from Lemma 199. $\square$

We prove Proposition 48, restated below.

Proposition 48. *Suppose that there are 2 users located at the standard basis vectors $e_1, e_2 \in \mathbb{R}^2$, with cost function $c(p) = \|p\|_2^\beta$. For $\beta = 2$, there is a multi-genre equilibrium μ with support equal to*

$$\left\{\left(x, (1 - x^{\frac{2}{P-1}})^{\frac{P-1}{2}}\right) \mid x \in [0, 1]\right\}, \tag{8.8}$$

and where the distribution of x has cdf equal to $\min(1, x^{2/(P-1)})$.

Again, the machinery given by Lemma 50 enables us to systematically identify the equilibrium in the concrete market instance of Proposition 48. Since we need to consider $P \neq 2$, the condition (C2) does not take as clean of a form: as shown by (E.14), it depends on both the densities h_1 and h_2 along with the cdfs H_1 and H_2. Nonetheless, in the special case of $\beta = 2$, we can compute the cdf in closed-form: Lemma 193 implies that the density $h_1(z_1)$ is entirely specified by z_1 and does not depend on z_2, so we can integrate over the density to explicitly compute the cdf. We can obtain the equilibria in Proposition 48 as a solution to a differential equation.

To prove Proposition 47, we again only need to *verify* that the equilibrium μ in Proposition 48 which is easier.

Proof of Proposition 48. By Lemma 50, it suffices to prove that (C1)-(C3) hold for H_1, H_2, and S for the distribution μ given in the statement of the proposition. Conditions (C2) and (C3) follow by construction of μ, so it suffices to prove (C1).

First, we claim that $H_1(z_1) = z_1^2$ and $H_1(z_1) = z_2^2$. We see that since the cdf of p_1 for $p \sim \mu$ is z_1, we know that $H_1(z_1) = z_1^2$ by construction. For z_2, first we note that the cdf of p_2 for $p_2 \sim \mu$ is given by:

$$\mathbb{P}_{p_2 \sim \mu}[p_2 \leq p_2'] = \mathbb{P}_{p_1 \sim \mu}\left[p_1 \geq (1 - (p_2')^{\frac{2}{P-1}})^{\frac{P-1}{2}}\right] = 1 - (1 - (p_2')^{\frac{2}{P-1}}) = (p_2')^{\frac{2}{P-1}}.$$

By definition, this means that $H_2(z_2) = z_2^2$ as desired.

Now, we prove (C1). Applying Lemma 192, we see that:

$$H_1(z_1) + H_2(z_2) - c_{\mathbf{U}}(z) = \left(\min\left(z_1^2, 1\right) + \min\left(z_2^2, 1\right)\right) - (z_1^2 + z_2^2).$$

Thus, equation (8.9) can be written as:

$$\max_{z_1, z_2 \geq 0} \left(\min\left(z_1^2, 1\right) + \min\left(\left(z_2^2, 1\right)\right) - (z_1^2 + z_2^2)^{\beta/2}\right) \tag{E.26}$$

We wish to show equation (E.26) is maximized whenever $z \in S$. Since $z_1^2 + z_2^2 = 1$ for any $z \in S$, this follows from Lemma 199 applied to $\beta = 2$. □

E.4.5 Formalization of the infinite-producer limit

Since our characterization result (Theorem 49) focuses on finite-genre equilibria, we restrict our formal definition of the infinite-producer limit to case of finite genres for technical convenience.

We arrive at a formalism by taking a limit of the conditions in Lemma 50 as $P \to \infty$. Let μ be a finite-genre distribution over $\mathbb{R}_{\geq 0}^D$. We can specify μ by the three attributes: the genres $d_1, \ldots, d_G$, the distributions F_g over $\mathbb{R}_{\geq 0}$ corresponding to the distribution of $\|p\|$ for p drawn from μ conditioned on p pointing in the direction of d_g, and the weights α_g corresponding to the probability that $p \sim \mu$ points in the direction of d_g. In particular, μ can be described as follows: with probability α_g, choose the vector $q_g d_g$ where q_g is drawn from a distribution with cdf F_g. We see that the corresponding function H_i from Lemma 50 will be equal to:

$$H_i(z_i) = \left(\sum_{g=1}^{G} \alpha_g F_g\left(\frac{\langle u_i, p\rangle}{\langle u_i, d_g\rangle}\right)\right)^{P-1}.$$

Note that the conditions (C2) and (C3) are essentially satisfied by construction; condition (C1) requires that

$$\max_{p \in \mathbb{R}_{\geq 0}^D} \left(\sum_{i=1}^{N} H_i(\langle u_i, p\rangle) - c(p)\right)$$

is maximized for any $p \in \operatorname{supp}(\mu)$. This can be rewritten as requiring that any $p^* \in \operatorname{supp}(\mu)$ satisfies:

$$p^* \in \operatorname*{arg\,max}_{p \in \mathbb{R}^D_{\geq 0}} \left(\sum_{i=1}^{N} \left(\sum_{g=1}^{G} \alpha_g F_g \left(\frac{\langle u_i, p \rangle}{\langle u_i, d_g \rangle} \right) \right)^{P-1} - c(p) \right).$$

Let's rewrite this in terms of winning producers: more formally, let $F_g^{\max}(\cdot) = (F_g(\cdot))^{P-1}$ denote the cumulative distribution function of the *maximum* quality in a genre, conditioned on all producers choosing that genre. We call the distributions $F_1^{\max}, \ldots, F_G^{\max}$ the *conditional quality distributions*. Then we obtain the following:

$$p^* \in \operatorname*{arg\,max}_{p \in \mathbb{R}^D_{\geq 0}} \left(\sum_{i=1}^{N} \left(\sum_{g=1}^{G} \alpha_g \left(F_g^{\max} \left(\frac{\langle u_i, p \rangle}{\langle u_i, d_g \rangle} \right) \right)^{1/(P-1)} \right)^{P-1} - c(p) \right). \tag{E.27}$$

Taking a limit as $P \to \infty$, we see that

$$\left(\sum_{g=1}^{G} \alpha_g F_g \left(\frac{\langle u_i, p \rangle}{\langle u_i, d_g \rangle} \right)^{1/(P-1)} \right)^{P-1} \to \prod_{g=1}^{G} \left(F_g \left(\frac{\langle u_i, p \rangle}{\langle u_i, d_g \rangle} \right) \right)^{\alpha_i}.$$

Thus, equation E.27 (informally speaking) approaches the following condition in the limit:

$$p^* \in \operatorname*{arg\,max}_{p \in \mathbb{R}^D_{\geq 0}} \left(\sum_{i=1}^{N} \left(\sum_{g=1}^{G} \alpha_g F_g \left(\frac{\langle u_i, p \rangle}{\langle u_i, d_g \rangle} \right)^{1/(P-1)} \right)^{P-1} - c(p) \right). \tag{E.28}$$

Motivated by equation E.28, we specify μ by three attributes—the *genres* $d_1, \ldots, d_G$, the *conditional quality* distributions $F_g^{\max}$ over $\mathbb{R}_{\geq 0}$, and the *weights* α_g corresponding to the probability that $p \sim \mu$ points in the direction of a given genre—as follows.

Definition 20 (Finite-genre equilibria for $P = \infty$)**.** *Let $u_1, \ldots, u_N \in \mathbb{R}^D_{\geq 0}$ be a set of users and let $c(p) = \|p\|_2^\beta$ be the cost function. A set of genres $d_1, \ldots, d_G \in \mathbb{R}^D_{\geq 0}$ such that $\|d_i\|_2 = 1$ for all $1 \leq g \leq G$, a set of conditional quality distributions $F_1, F_2, \ldots, F_G$ over $\mathbb{R}_{\geq 0}$, and a set of weights $\alpha_1, \ldots, \alpha_G \geq 0$ such that $\sum_{g=1}^{G} \alpha_g = 1$ forms a finite-genre equilibrium if the following condition holds for*

$$p^* \in \operatorname*{arg\,max}_{p \in \mathbb{R}^D_{\geq 0}} \left(\sum_{i=1}^{N} \left(\prod_{g=1}^{G} \left(F_g^{max} \left(\frac{\langle u_i, p \rangle}{\langle u_i, d_g \rangle} \right) \right)^{\alpha_i} \right) - c(p) \right) \tag{E.29}$$

for any $p^* = q_i d_i$ such that $1 \leq i \leq G$ and $q_i \in supp(F_i)$.

Using the formalization in Definition 20 of equilibria for $P = \infty$, we investigate the case of two homogeneous populations of users, and we characterize two-genre equilibria.

Theorem 198. *[Formal version of Theorem 49] Suppose that there are 2 users located at two linearly independently vectors $u_1, u_2 \in \mathbb{R}^D_{\geq 0}$, let $\theta^* := \cos^{-1}\left(\frac{\langle u_1, u_2 \rangle}{\|u_1\|_2 \|u_2\|}\right) < 0$ be the angle between them. Suppose we have cost function $c(p) = \|p\|_2^\beta$, $\beta > \beta^* = \frac{2}{1-\cos(\theta^*)}$, and $P = \infty$ producers. Then, the genres d_1, d_2, conditional quality distributions $F_1^{max} = F^{max}$ and $F_2^{max} = F^{max}$, and weights $\alpha_1 = \alpha_2 = 2$ form an equilibrium (as per Definition 20), where*

$$\{d_1, d_2\} := \left\{[\cos(\theta^G + \theta_{min}), \sin(\theta^G + \theta_{min})], [\cos(\theta^* - \theta^G + \theta_{min}), \sin(\theta^* - \theta^G + \theta_{min})]\right\}$$

such that $\theta^G := \arg\max_{\theta \leq \theta^*/2} \left(\cos^\beta(\theta) + \cos^\beta(\theta^* - \theta)\right)$ *and* $\theta_{min} := \min\left(\cos^{-1}\left(\frac{\langle u_1, e_1 \rangle}{\|u_1\|}\right), \cos^{-1}\left(\frac{\langle u_2, e_1 \rangle}{\|u_2\|}\right)\right)$*, and where*

$$F^{max}(q) := \begin{cases} C_2^{(2n+2)\beta} & \text{if } q \in C_1^{1/\beta} C_2^{2n+1} [C_2, 1] \text{ for } n \geq 0 \\ C_1^{-2} C_2^{-2n\beta} q^{2\beta} & \text{if } q \in C_1^{1/\beta} C_2^{2n} [C_2, 1] \text{ for } n \geq 0 \\ 1 & \text{if } q \geq C_1^{1/\beta}, \end{cases},$$

such that the constants are defined by $C_1 := \frac{\sin(\theta^*)\cos(\theta^G)}{\sin(\theta^* - \theta_G)}$ *and* $C_2 := \frac{\cos(\theta^* - \theta^G)}{\cos(\theta^G)}$.

E.4.6 Proofs for Chapter 8.4.3

To recover the equilibrium in the infinite-producer limit, we need to show that there exists a two-genre equilibrium and find this equilibrium. We can apply machinery that is conceptually similar to Lemma 50 enables us to systematically identify the particular equilibrium within the family of two-genre equilibrium. The first-order condition (Lemma 193) given by condition (C1) helps identify the location of the genre directions, and this further enables us to compute the cdfs H_1 and H_2. At this stage, the proof boils down to solving for the conditional quality distributions F_1 and F_2. We obtain an infinite-producer limit of the functional equations in (E.15) which can be solved directly.

To actually prove Theorem 198, we again only need to *verify* that the equilibrium μ in Theorem 198 which is easier.

Proof of Theorem 198. WLOG, we assume that $\|u_1\| = \|u_2\| = 1$. It suffices to verify that the genres, conditional quality distributions, and weights satisfy (E.29). Motivated by Lemma 50, we define:

$$H_1(z_1) = \sqrt{F_1^{\max}\left(\frac{z_1}{\langle u_1, d_1 \rangle}\right) F_2^{\max}\left(\frac{z_1}{\langle u_1, d_2 \rangle}\right)}$$

$$H_2(z_2) = \sqrt{F_1^{\max}\left(\frac{z_2}{\langle u_1, d_1 \rangle}\right) F_2^{\max}\left(\frac{z_2}{\langle u_1, d_2 \rangle}\right)}.$$

We define the support S to be

$$S := \{[\langle u_1, q d_1 \rangle, \langle u_2, q d_1 \rangle] \mid q_1 \in \text{supp}(F_1^{\max})\} \cup \{[\langle u_1, q d_1 \rangle, \langle u_2, q d_1 \rangle] \mid q_2 \in \text{supp}(F_2^{\max})\}.$$

Using this notation, we can rewrite (E.29) as requiring that:

$$\max_z \left(H_1(z_1) + H_2(z_2) - c_{\mathbf{U}}(z)\right) \tag{E.30}$$

is maximized for every $z \in S$.

First, we show that

$$\sin(\theta^G)\cos^{\beta-1}(\theta^G) = \sin(\theta^* - \theta^G)\cos^{\beta-1}(\theta^* - \theta^G) \tag{E.31}$$

This immediately follows from using that $\theta^G \in \arg\max_\theta \left(\cos^\beta(\theta) + \cos^\beta(\theta^* - \theta)\right)$ and applying the first-order condition.

For the remainder of the proof, we define:

$$c := \frac{\sin(\theta^* - \theta^G)}{\sin(\theta^*)\cos^{\beta-1}(\theta^G)} = \frac{\sin(\theta^G)}{\sin(\theta^*)\cos^{\beta-1}(\theta^* - \theta^G)},$$

Computing H_1 and H_2. We show that:

$$H_1(z_1) = \min\left(cz_1^\beta, 1\right) \text{ and } H_2(z_2) = \min\left(1, cz_2^\beta\right).$$

We show that

$$H_1(z_1) = \min\left(cz_1^\beta, 1\right), \tag{E.32}$$

and observe that the expression for H_2 follows from an analogous argument. By definition, we see that:

$$\begin{aligned} H_1(z_1) &= \sqrt{F_1\left(\frac{z_1}{\langle u_1, d_1\rangle}\right) F_2\left(\frac{z_1}{\langle u_1, d_2\rangle}\right)} \\ &= \sqrt{F\left(\frac{z_1}{\langle u_1, d_1\rangle}\right) F\left(\frac{z_1}{\langle u_1, d_2\rangle}\right)}. \end{aligned}$$

We know that either (1) $\langle u_1, d_1\rangle = \langle u_2, d_2\rangle = \cos(\theta^G)$ and $\langle u_1, d_2\rangle = \langle u_2, d_1\rangle = \cos(\theta^* - \theta^G)$, or (2) $\langle u_1, d_2\rangle = \langle u_2, d_1\rangle = \cos(\theta^G)$ and $\langle u_1, d_1\rangle = \langle u_2, d_2\rangle = \cos(\theta^* - \theta^G)$. WLOG, we assume that (1) holds. This means that:

$$\begin{aligned} H_1(z_1) &= \sqrt{F^{\max}\left(\frac{z_1}{\langle u_1, d_1\rangle}\right) F^{\max}\left(\frac{z_1}{\langle u_1, d_2\rangle}\right)} \\ &= \sqrt{F^{\max}\left(\frac{z_1}{\cos(\theta^G)}\right) F^{\max}\left(\frac{z_1}{\cos(\theta^* - \theta^G)}\right)} \end{aligned}$$

Let's reparameterize and let:

$$q_1 = \frac{z_1}{\cos(\theta^* - \theta^G)}.$$

This means that:

$$H_1(q_1 \cos(\theta^* - \theta^G)) = \sqrt{F^{\max}(q_1) F^{\max}\left(q_1 \frac{\cos(\theta^* - \theta^G)}{\cos(\theta^G)}\right)}.$$

Equation (E.32) reduces to

$$\sqrt{F^{\max}(q_1) F^{\max}\left(q_1 \frac{\cos(\theta^* - \theta^G)}{\cos(\theta^G)}\right)} = \min\left(1, c\cos(\theta^* - \theta^G)^\beta q_1^\beta\right).$$

which simplifies to

$$\sqrt{F^{\max}(q_1) F^{\max}\left(q_1 \frac{\cos(\theta^* - \theta^G)}{\cos(\theta^G)}\right)} = \min\left(1, \frac{\sin(\theta^G)\cos(\theta^* - \theta^G)}{\sin(\theta^*)} q_1^\beta\right)$$

which simplifies to

$$\sqrt{F^{\max}(q_1) F^{\max}(q_1 C_2)} = \min\left(1, C_3^{-1} q_1^\beta\right) \tag{E.33}$$

We verify equation (E.33) by doing casework on q_1. Note that $C_1^{1/\beta} = C_3^{1/\beta} C_2$. If $q_1 \geq C_3^{1/\beta} C_2^{-1/\beta}$, then we see that $F^{\max}(q_1) = F^{\max}(q_1 C_2) = 1$ and the equation holds. In fact, if $q_1 \geq C_3^{1/\beta} C_2^{1-1/\beta}$, then we see that $F^{\max}(q_1) = 1$ and

$$F^{\max}(q_1 C_2) = C_3^{-2} C_2^{2-2\beta} (q_1 C_2)^{2\beta} = C_3^{-2} C_2^2 q_1^{2\beta}$$

, so equation (E.33) is satisfied. Otherwise, if $q_1 = C_3^{1/\beta} C_2 C_2^{2n} \gamma$ for $n \geq 0$ and $\gamma \in [C_2, 1]$, then

$$F^{\max}(q_1) = C_3^{-2} C_2^{-2\beta - 2n\beta} q^{2\beta}$$

and

$$F^{\max}(q_1 C_2) = C_2^{(2n+2)\beta},$$

so:

$$\sqrt{F^{\max}(q_1) F^{\max}(q_1 C_2)} = \sqrt{C_3^{-2} C_2^{-2\beta - 2n\beta} C_2^{(2n+2)\beta}} = \sqrt{C_3^{-2} C_2^2 q^{2\beta}} = C_3^{-1} C_2 q_1^\beta$$

as desired. Finally, if $q_1 = C_1^{1/\beta} C_2^{1-1/\beta} C_2^{2n+1} \gamma$ for $n \geq 0$ and $\gamma \in [C_2, 1]$, then

$$F^{\max}(q_1) = C_2^{(2n+2)\beta}$$

and

$$F^{\max}(q_1 C_2) = C_3^{-2} C_2^{-(2n+4)\beta} q^{2\beta},$$

so:

$$\sqrt{F^{\max}(q_1) F^{\max}(q_1 C_2)} = \sqrt{C_2^{(2n+2)\beta} C_3^{-2} C_2^{(2n+4)\beta} q^{2\beta} C_2^{2\beta}} = C_3^{-1} q^\beta.$$

This proves the desired formulas for H_1 and an analogous argument applies to H_2.

Showing equation (E.30) is maximized at every $z \in S$. We need to show that for every $z \in S$, it holds that:

$$H_1(z_1) + H_2(z_2) - c_{\mathbf{U}}(z) = \max_{z'}(H_1(z'_1) + H_2(z'_2) - c_{\mathbf{U}}(z')).$$

Plugging in our expressions above, our goal is to show:

$$\min(1, cz_1^\beta) + \min(1, cz_2^\beta) - c_{\mathbf{U}}(z) = \max_{z'}(H_1(z'_1) + H_2(z'_2) - c_{\mathbf{U}}(z'))$$

for every $z \in S$.

We split into two steps: first, we show that

$$\min(1, cz_1^\beta) + \min(1, cz_2^\beta) - c_{\mathbf{U}}(z) = 0 \tag{E.34}$$

for every $z \in S$, and next we show that:

$$\max_{z'}(H_1(z'_1) + H_2(z'_2) - c_{\mathbf{U}}(z')) \leq 0. \tag{E.35}$$

To show (E.34), let's first consider $[z_1, z_2] = [r\cos(\theta^G), r\cos(\theta^G - \theta^*)] \in S$. Then we see that:

$$\begin{aligned}\min(1, cz_1^\beta) + \min(1, cz_2^\beta) - c_{\mathbf{U}}(z) &= cz_1^\beta + cz_2^\beta - c_{\mathbf{U}}(z) \\ &= r^\beta \left(c\cos^\beta(\theta^G) + c\cos^\beta(\theta^* - \theta^G) - 1\right)\end{aligned}$$

Thus, it suffices to show that:

$$\cos^\beta(\theta^G) + \cos^\beta(\theta^* - \theta^G) = \frac{1}{c}. \tag{E.36}$$

We now show equation (E.36):

$$\begin{aligned}\cos^\beta(\theta^G) + \cos^\beta(\theta^* - \theta^G) &=_{(A)} \frac{\cos(\theta^G)\cos^{\beta-1}(\theta^* - \theta^G)\sin(\theta^* - \theta^G)}{\sin(\theta^G)} + \cos^\beta(\theta^* - \theta^G) \\ &= \frac{\cos^{\beta-1}(\theta^* - \theta^G)}{\sin(\theta^G)}\left(\cos(\theta^G)\sin(\theta^* - \theta^G) + \cos(\theta^* - \theta^G)\sin(\theta^G)\right) \\ &= \frac{\cos^{\beta-1}(\theta^* - \theta^G)}{\sin(\theta^G)}\sin(\theta^*) \\ &= \frac{1}{c}.\end{aligned}$$

where (A) follows from applying equation (E.31). Let's now consider let's first consider $[z_1, z_2] = [r\cos(\theta^G - \theta^*), r\cos(\theta^*)] \in S$. Then, we see that

$$\min(1, cz_1^\beta) + \min(1, cz_2^\beta) - c_{\mathbf{U}}(z) = cz_1^\beta + cz_2^\beta - c_{\mathbf{U}}(z) = r^\beta \left(c\cos^\beta(\theta^G) + c\cos^\beta(\theta^* - \theta^G) - 1\right) = 0,$$

where the last equality follows from equation (E.36). This establishes equation (E.35).

Now, we show equation (E.35). Let's represent z' as $\mathbf{U}[r'\cos(\theta), r'\sin(\theta)]$. Then this becomes:

$$c(r')^\beta \cos^\beta(\theta) + c(r')^\beta \cos^\beta(\theta^* - \theta) \leq (r')^\beta.$$

Dividing by r'^β, we obtain:

$$\cos^\beta(\theta) + \cos^\beta(\theta^* - \theta) \leq \frac{1}{c}.$$

To show this, observe that:

$$\cos^\beta(\theta) + \cos^\beta(\theta^* - \theta) \leq \cos^\beta(\theta^G) + \cos^\beta(\theta^* - \theta^G) = \frac{1}{c}.$$

where the first inequality follows from the fact that θ^G is a maximizer of $\cos^\beta(\theta) + \cos^\beta(\theta^* - \theta)$ by definition, and the second equality follows from equation (E.36). This establishes equation (E.35).

This proves that equation (E.36) is maximized at every $z \in S$, and thus the conditions of Definition 20 are satisfied. □

E.4.7 Proofs of auxiliary lemmas

We state and prove Lemma 199, a lemma which we used in the proofs of Proposition 48 and Proposition 47.

Lemma 199. *For any $\beta \geq 2$, the expression*

$$\max_{z_1, z_2 \geq 0} \left(\left(\min \left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_1^2, 1 \right) + \min \left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_2^2, 1 \right) \right) - (z_1^2 + z_2^2)^{\beta/2} \right)$$

is maximized for any (z_1, z_2) such that $z_1^2 + z_2^2 = \left(\frac{2}{\beta}\right)^{2/\beta}$.

Proof. First, for z_1, z_2 such that $z_1^2 + z_2^2 = \left(\frac{2}{\beta}\right)^{2/\beta}$, we have that

$$\left(\frac{2}{\beta}\right)^{-2/\beta} (z_1^2 + z_2^2) - (z_1^2 + z_2^2)^{\beta/2} = 1 - \frac{2}{\beta}.$$

It thus suffices to prove that:

$$\left(\min \left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_1^2, 1 \right) + \min \left(\left(\frac{2}{\beta}\right)^{-2/\beta} z_2^2, 1 \right) \right) - (z_1^2 + z_2^2)^{\beta/2} \leq 1 - \frac{2}{\beta}$$

for any $z_1, z_2 \geq 0$. It suffices to prove the stronger statement that:

$$\left(\frac{2}{\beta}\right)^{-2/\beta} (z_1^2 + z_2^2) - (z_1^2 + z_2^2)^{\beta/2} \leq 1 - \frac{2}{\beta}$$

Let $c = z_1^2 + z_2^2$; then we can rewrite the desired condition as:

$$\max_{c \geq 0} \left(\left(\frac{2}{\beta}\right)^{-2/\beta} c^2 - c^\beta \right) \leq 1 - \frac{2}{\beta}.$$

A first-order condition tells us for $\beta \geq 2$, that $\left(\frac{2}{\beta}\right)^{-2/\beta} c^2 - c^\beta$ is maximized at $c = \left(\frac{2}{\beta}\right)^{1/\beta}$, which proves the desired statement. $\square$

We prove Lemma 192.

Proof of Lemma 192. It suffices to show that if $z_1 = \langle u_1, p \rangle$ and $z_2 = \langle u_2, p \rangle$, then:

$$\|p\|^2 = \frac{z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)}{\sin^2(\theta^*)} \tag{E.37}$$

WLOG, let $u_1 = e_1$ and let $u_2 = [\cos(\theta^*), \sin(\theta^*)]$. We see that:

$$\begin{aligned}
\frac{z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)}{\sin^2(\theta^*)} &= \frac{p_1^2 + (p_1 \cos(\theta^*) + p_2 \sin(\theta^*))^2 - 2p_1(p_1 \cos(\theta^*) + p_2 \sin(\theta^*)) \cos(\theta^*)}{\sin^2(\theta^*)} \\
&= \frac{p_1^2 \sin^2(\theta^*) + p_2^2 \sin^2(\theta^*)}{\sin^2(\theta^*)} \\
&= p_1^2 + p_2^2 \\
&= \|p\|_2^2,
\end{aligned}$$

which proves equation (E.37). $\square$

We prove Lemma 193.

Proof of Lemma 193. Since μ is a symmetric mixed equilibrium, z must be a maximizer of equation (8.9). The equation

$$\begin{bmatrix} h_1(z_1) \\ h_2(z_2) \end{bmatrix} = \nabla_z (c_{\mathbf{U}}(z))$$

is the first-order condition and thus holds for every z is in the support of μ.

Next, we show that:

$$\nabla_z (c_{\mathbf{U}}(z)) = \beta \alpha^\beta \sin^{-\beta}(\theta^*) \left(\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2} - 1} \right) \begin{bmatrix} z_1 - z_2 \cos(\theta^*) \\ z_2 - z_1 \cos(\theta^*) \end{bmatrix}.$$

By applying Lemma 192, we see that:

$$\begin{aligned}
\nabla_z (c_{\mathbf{U}}(z)) &= \nabla_z \left(\alpha^\beta \sin^{-2\beta}(\theta^*) \left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}} \right) \\
&= \alpha^\beta \sin^{-\beta}(\theta^*) \cdot \nabla_z \left(\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}} \right) \\
&= \beta \alpha^\beta \sin^{-\beta}(\theta^*) \left(\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2} - 1} \right) \begin{bmatrix} z_1 - z_2 \cos(\theta^*) \\ z_2 - z_1 \cos(\theta^*) \end{bmatrix},
\end{aligned}$$

as desired.

Finally, we show that

$$\nabla_z(c_{\mathbf{U}}(z)) = \beta\alpha^\beta \sin^{-\beta}(\theta^*) \left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-1} \begin{bmatrix} z_1 - z_2\cos(\theta^*) \\ z_2 - z_1 \cos(\theta^*) \end{bmatrix}.$$

We see that:

$$\begin{aligned}
\nabla_z(c_{\mathbf{U}}(z)) &= \beta\alpha^\beta \sin^{-\beta}(\theta^*) \left(\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-1}\right) \begin{bmatrix} z_1 - z_2\cos(\theta^*) \\ z_2 - z_1 \cos(\theta^*) \end{bmatrix} \\
&= \beta\alpha^\beta r^{\beta-2} \begin{bmatrix} \frac{z_1 - z_2\cos(\theta^*)}{\sin^2(\theta^*)} \\ \frac{z_2 - z_1\cos(\theta^*)}{\sin^2(\theta^*)} \end{bmatrix} \\
&= \beta\alpha^\beta r^{\beta-1} \begin{bmatrix} \frac{\cos(\theta) - \cos(\theta^* - \theta)\cos(\theta^*)}{\sin^2(\theta^*)} \\ \frac{\cos(\theta^* - \theta) - \cos(\theta)\cos(\theta^*)}{\sin^2(\theta^*)} \end{bmatrix} \\
&= \beta\alpha^\beta r^{\beta-1} \begin{bmatrix} \frac{\cos(\theta * - (\theta^* - \theta)) - \cos(\theta^* - \theta)\cos(\theta^*)}{\sin^2(\theta^*)} \\ \frac{\sin(\theta^*)\sin(\theta)}{\sin^2(\theta^*)} \end{bmatrix} \\
&= \beta\alpha^\beta r^{\beta-1} \begin{bmatrix} \frac{\sin(\theta^*)\sin(\theta^* - \theta)}{\sin^2(\theta^*)} \\ \frac{\sin(\theta)}{\sin(\theta^*)} \end{bmatrix} \\
&= \beta\alpha^\beta r^{\beta-1} \begin{bmatrix} \frac{\sin(\theta^* - \theta)}{\sin(\theta^*)} \\ \frac{\sin(\theta)}{\sin(\theta^*)} \end{bmatrix},
\end{aligned}$$

as desired.

□

We prove Lemma 194.

Proof of Lemma 194. By construction, we see that $z \in \left\{\mathbf{U}p \mid p \in \mathbb{R}^D_{\geq 0}\right\}$. We can apply Lemma 193 to see that

$$\begin{aligned}
\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} &= \frac{\partial^2}{\partial z_1 \partial z_2}\left(\sin^{-2\beta}(\theta^*) \left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}}\right) \\
&= \frac{\partial}{\partial z_2}\left(\beta\alpha^\beta \sin^{-\beta}(\theta^*) \left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-1} (z_1 - z_2\cos(\theta^*))\right) \\
&= \beta\alpha^\beta \sin^{-\beta}(\theta^*) \frac{\partial}{\partial z_2}\left(\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-1} (z_1 - z_2\cos(\theta^*))\right).
\end{aligned}$$

This is the same sign as:

$$\begin{aligned}
&\frac{\partial}{\partial z_2}\left(\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-1} (z_1 - z_2\cos(\theta^*))\right) \\
&= (\beta - 2)\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-2} (z_1 - z_2\cos(\theta^*))(z_2 - z_1\cos(\theta^*)) - \left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-1}\cos(\theta^*) \\
&= \left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)^{\frac{\beta}{2}-2}\left((\beta - 2)(z_1 - z_2\cos(\theta^*))(z_2 - z_1\cos(\theta^*)) - \cos(\theta^*)\left(z_1^2 + z_2^2 - 2z_1 z_2 \cos(\theta^*)\right)\right)
\end{aligned}$$

This is the same sign as:

$$(\beta-2)(z_1 - z_2\cos(\theta^*))(z_2 - z_1\cos(\theta^*)) - \cos(\theta^*)\left(z_1^2 + z_2^2 - 2z_1z_2\cos(\theta^*)\right).$$

Let's represent z as $[r\cos(\theta), r\cos(\theta^*-\theta)]$. The above expression is the same sign as:

$$\begin{aligned}&(\beta-2)(\cos(\theta)-\cos(\theta^*-\theta)\cos(\theta^*))(\cos(\theta^*-\theta)-\cos(\theta)\cos(\theta^*)) - \cos(\theta^*)\sin^2(\theta^*)\\ &= (\beta-2)(\sin(\theta^*)\sin(\theta^*-\theta))(\sin(\theta)\sin(\theta^*)) - \cos(\theta^*)\sin^2(\theta^*)\\ &= \sin^2(\theta^*)\left((\beta-2)\sin(\theta^*-\theta)\sin(\theta)-\cos(\theta^*)\right).\end{aligned}$$

This is the same sign as:

$$\begin{aligned}(\beta-2)\sin(\theta^*-\theta)\sin(\theta)-\cos(\theta^*) &= (\frac{\beta}{2}-1)(\cos(\theta^*-2\theta)-\cos(\theta^*))-\cos(\theta^*)\\ = \left(\frac{\beta}{2}-1\right)(\cos(\theta^*-2\theta) - \frac{\beta}{2}\cos(\theta^*).\end{aligned}$$

This is the same sign as:

$$\frac{\beta-2}{\beta}\cos(\theta^*-2\theta)-\cos(\theta^*).$$

□

We prove Lemma 195.

Proof of Lemma 195. By Lemma 194, we see that $\left(\frac{\beta-2}{\beta}\cos(\theta^*-2\theta)-\cos(\theta^*)\right)$ has the same sign as $\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1\partial z_2}$. Thus it suffices to show that $g'(z_1)\cdot\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1\partial z_2}\le 0$. When $\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1\partial z_2}=0$, the condition in the proposition statement is trivially satisfied. We thus assume for the remainder of the proof that $\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1\partial z_2}\neq 0$.

The second-order condition for z to be a maximizer of equation (8.9) is the following:

$$\begin{bmatrix} h_1'(z_1) & 0 & \\ 0 & b & h_2'(z_2)\end{bmatrix} - \nabla^2 c_{\mathbf{U}}(z) \preceq 0. \tag{E.38}$$

Let's apply Lemma 193, to see that:

$$h_1(x) = \frac{\partial c_{\mathbf{U}}([x, g(x)])}{\partial z_1}.$$

Since this holds in a neighborhood of z_1, we see that:

$$h_1'(z_1) = \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1^2} + g'(z_1)\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1\partial z_2}.$$

An analogous argument, coupled with the inverse function theorem, shows that:

$$h_2'(z_2) = \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_2^2} + \frac{1}{g'(z_1)} \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2}.$$

Plugging this into equation (E.38), we obtain:

$$\begin{aligned}
0 &\succeq \begin{bmatrix} h_1'(z_1) & 0 & \\ 0 & b & h_2'(z_2) \end{bmatrix} - \nabla^2 c_{\mathbf{U}}(z) \\
&= \begin{bmatrix} \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1^2} + g'(z_1) \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} & 0 \\ 0 & \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_2^2} + \frac{1}{g'(z_1)} \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} \end{bmatrix} - \nabla^2 c_{\mathbf{U}}(z) \\
&= \begin{bmatrix} g'(z_1) \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} & -\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} \\ -\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} & \frac{1}{g'(z_1)} \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} \end{bmatrix} \\
&= \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} \begin{bmatrix} g'(z_1) & -1 \\ -1 & \frac{1}{g'(z_1)}. \end{bmatrix}
\end{aligned}$$

When $\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} = 0$, the condition in the proposition statement is trivially satisfied. Since we've assumed that $\frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} \neq 0$, the eigenvectors are $[1, g'(u)]$ which has eigenvalue 0 and $[-g'(u), 1]$ which has eigenvalue

$$\frac{(g'(z_1))^2 + 1}{g'(z_1)} \cdot \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2}.$$

The sign of that eigenvalue is equal to the sign of $g'(z_1) \cdot \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2}$. Since the matrix must be negative semidefinite, we see that $g'(z_1) \cdot \frac{\partial^2 c_{\mathbf{U}}(z)}{\partial z_1 \partial z_2} \leq 0$.

□

E.5 Proofs for Chapter 8.5

We prove Proposition 52, restated below.

Proposition 52. *Suppose that*

$$\max_{||p|| \leq 1} \min_{1 \leq i \leq N} \left\langle p, \frac{u_i}{||u_i||} \right\rangle < N^{-P/\beta}. \tag{8.11}$$

Then for any symmetric equilibrium μ, the profit $\mathcal{P}^{eq}(\mu)$ is strictly positive.

Proof. Without loss of generality, we assume user vectors have unit norm $\|u_i\|$. Given an equilibrium μ, we will construct an explicit vector p that generates positive profit. This proves that the equilibrium profit is positive because no vector can achieve higher than the equilibrium profit. The vector p is of the form $(Q\left(\max_{p' \in \text{supp}(\mu)} ||p'||\right) + \varepsilon) \cdot u_{i^*}$ for some $i^* \in [1, N]$.

Cluster the set of unit vectors p into N groups $G_1, \ldots, G_N$, based on the user for whom they generate the lowest value. That is, each vector p belongs to the group G_i where $u_i = \arg\min_{1 \le i' \le N} \langle p, u_{i'} \rangle$. This means that if all producers choose (unit vector) directions in G_i, then the maximum inferred user value for u_i is

$$\max_{1 \le j \le P} \langle p_j, u_i \rangle \le \max_{||p|| \le 1} \min_{1 \le i \le N} \langle p, u_i \rangle = Q. \tag{E.39}$$

Let G_{i^*} be the group with highest probability of appearing in μ. That is, let $i^* \in \arg\max_i \mathbb{P}_{v \sim \mu}\left[\frac{v}{||v||} \in G_i\right]$.

Let E be the event that all of the other $P-1$ producers choose directions in G_{i^*}. The event E happens with probability at least $\mathbb{P}_{v \sim \mu}\left[\frac{v}{||v||} \in G_{i^*}\right] \ge (1/N)^{P-1}$. Since the inferred user value is linear in the magnitude of the producer action, we see that the maximum possible inferred user value for user u_i from the other producers is $Q\left(\max_{p' \in \text{supp}(\mu)} ||p'||\right)$. On the other hand, the action p results in inferred user value $(Q\left(\max_{p' \in \text{supp}(\mu)} ||p'||\right) + \varepsilon)$ for u_{i^*}, so it wins u_{i^*} with probability 1 on the event E. This means that the expected profit obtained by p is at most

$$\left(\frac{1}{N}\right)^{P-1} - \left(Q\left(\max_{p' \in \text{supp}(\mu)} ||p'||\right) + \varepsilon\right)^{\beta}.$$

Taking a limit as $\varepsilon \to_+ 0$, we obtain the profit can be set arbitrarily close to:

$$\left(\frac{1}{N}\right)^{P-1} - \left(Q\left(\max_{p' \in \text{supp}(\mu)} ||p'||\right)\right)^{\beta}. \tag{E.40}$$

It suffices to bound $\max_{p' \in \text{supp}(\mu)} ||p'||$. The action $p'' \in \arg\max_{p' \in \text{supp}(\mu)} ||p'||$ produces a profit of at most $N - \left(\max_{p \in \text{supp}(\mu)} ||p||\right)^{\beta}$. Thus, $\left(\max_{p \in \text{supp}(\mu)} ||p||\right)^{\beta} \le N$, so $\left(\max_{p \in \text{supp}(\mu)} ||p||\right) \le N^{1/\beta}$.

Plugging this into (E.40), we see that there exist actions that produces profit arbitrarily close to

$$\left(\frac{1}{N}\right)^{P-1} - NQ^{\beta}.$$

Thus, a strictly positive profit will be obtained if:

$$Q < \left(\frac{1}{N}\right)^{P/\beta},$$

as desired. □

We prove Proposition 53, restated below.

Proposition 53. *If μ is a single-genre equilibrium, then the profit $\mathcal{P}^{eq}(\mu)$ is equal to 0.*

Proof. Since μ is an equilibrium, all choices p in the support of μ achieve profit equal to the equilibrium profit. We apply Lemma 185 to see that the cdf of μ is $F(p) = \min\left(1, \left(\frac{p^\beta}{N}\right)^{1/(P-1)}\right)$, which shows that $p = 0$ is in $\text{supp}(\mu)$. For this choice of p, the cost is 0, but the producer also never wins any users, so the profit is also zero, as claimed. $\square$

Appendix F

Appendix for Chapter 9

F.1 Auxiliary definitions and lemmas

In our analysis of equilibria, it will be helpful to work with several quantities. We first define $\mathcal{C}_t$ to be the set of content that achieves 0 utility for that type. That is:

$$\mathcal{C}_t := \{[w_{\text{costly}}, w_{\text{cheap}}] \mid u([w_{\text{costly}}, w_{\text{cheap}}], t) = 0\} . \tag{F.1}$$

We also define an augmented version of these sets that also includes content with $w_{\text{costly}} = 0$ that achieving positive utility. That is, we define $\mathcal{C}_t^{\text{aug}}$ to be

$$\{[w_{\text{costly}}, w_{\text{cheap}}] \mid u([w_{\text{costly}}, w_{\text{cheap}}], t) = 0\} \cup \left\{[0, w_{\text{cheap}}] \mid w_{\text{cheap}} \in [0, \min_{w' \mid u(w',t)=0} w'_{\text{cheap}})]\right\} , \tag{F.2}$$

The set $\mathcal{C}_t^{\text{aug}}$ turns out to be closely related to the function f_t defined in (9.5).

Lemma 200. *The set $\mathcal{C}_t^{aug}$ can be written as:*

$$\mathcal{C}_t^{aug} = \{[f_t(w_{cheap}), w_{cheap}] \mid w_{cheap} \geq 0\}$$

where f_t is defined by (9.5).

Proof. First, we show that $\mathcal{C}_t^{\text{aug}} \subseteq \{(f_t(w_{\text{cheap}}), w_{\text{cheap}}) \mid w_{\text{cheap}} \geq 0\}$. If $w \in \mathcal{C}_t^{\text{aug}}$, then either $u(w,t) = 0$ or $w_{\text{costly}} = 0$ and $w_{\text{cheap}} \in [0, \min_{w' \mid u(w',t)=0} w'_{\text{cheap}})]$. If $u(w,t) = 0$, since investing in quality is costly, it must hold that $w_{\text{costly}} = f_t(w_{\text{cheap}})$. Next, suppose that $w_{\text{cheap}} \in [0, \min_{w' \mid u(w',t)=0} w'_{\text{cheap}})]$. We observe that $\min_{w' \mid u(w',t)=0} w'_{\text{cheap}}$ is the unique value of w'_{cheap} such that $u([0, w'_{\text{cheap}}], t) = 0$. This implies that $u([0, w_{\text{cheap}}], t) \geq 0$, so $f_t(w_{\text{cheap}}) = 0$ as desired.

Next, we show that $\{[f_t(w_{\text{cheap}}), w_{\text{cheap}}] \mid w_{\text{cheap}} \geq 0\} \subseteq \mathcal{C}_t^{\text{aug}}$. Let $w = [f_t(w_{\text{cheap}}), w_{\text{cheap}}]$ for some $w_{\text{cheap}} \geq 0$. If $u(w,t) = 0$, then $w \in \mathcal{C}_t^{\text{aug}}$ as desired. If $u(w,t) > 0$, then it must hold that $w_{\text{costly}} = 0$ (otherwise, it would be possible to lower w_{costly} while keeping utility nonnegative, which contradicts the fact that $f_t(w_{\text{cheap}}) = w_{\text{costly}}$), so $w \in \mathcal{C}_t^{\text{aug}}$. $\square$

We prove that the function f_t is weakly increasing.

Lemma 201. *The function f_t as defined in (9.5) is weakly increasing. Moreover, the function $M^E([f_t(w_{cheap}), w_{cheap}])$ is strictly increasing in w_{cheap}.*

Proof. Suppose that $w^1_{\text{cheap}} \geq w^2_{\text{cheap}}$. We claim that $f_t(w^1_{\text{cheap}}) \geq f_t(w^2_{\text{cheap}})$. To see this, note that

$$u([f_t(w^1_{\text{cheap}}), w^2_{\text{cheap}}], t) > u([f_t(w^1_{\text{cheap}}), w^1_{\text{cheap}}], t) \geq 0,$$

which proves the first statement.

To see that $M^{\text{E}}([f_t(w_{\text{cheap}}), w_{\text{cheap}}])$ is increasing, note that f_t is a weakly increasing function (see Lemma 201) and that M^{E} is strictly increasing in both of its arguments. $\square$

We next show the following properties of the optima of (9.7).

Lemma 202. *The optimization program $\inf_{w \in \mathbb{R}^2_{\geq 0}} c(w)$ s.t. $u(w,t) \geq 0, M^E(w) \geq m$ satisfies the following properties:*

1. *For any $m \in \left\{M^E(w) \mid w \in \mathcal{C}^{aug}_t\right\}$, the optimization program is feasible and any optimum w^* satisfies $w^* \in \mathcal{C}^{aug}_t$.*

2. *If $m \in \left\{M^E(w) \mid w \in \mathcal{C}^{aug}_t\right\}$ and $C^E_t(m) > 0$, the optimization program has a unique optimum w^* and moreover $M^E(w^*) = m$.*

Proof. Suppose that $m \in \left\{M^{\text{E}}(w) \mid w \in \mathcal{C}^{\text{aug}}_t\right\}$.

First, we show that the optimization program is feasible. Suppose that w is such that $M^{\text{E}}(w) = m$. Using the fact that $u([w'_{\text{costly}}, w_{\text{cheap}}], t)$ approaches ∞ as $w'_{\text{costly}} \to \infty$, we see that there exists $w'_{\text{costly}} \geq w_{\text{costly}}$ such that $M^{\text{E}}([w'_{\text{costly}}, w_{\text{cheap}}]) \geq M^{\text{E}}(w) = m$ and $u([w'_{\text{costly}}, w_{\text{cheap}}], t) \geq 0$, as desired.

Next, we show that there exists $w \in \mathbb{R}^2_{\geq 0}$ such that $u(w,t) \geq 0$, $M^{\text{E}}(w) \geq m$, and $c(w) = C^E_t(m)$. To make the domain compact, observe that there exists $w' \in \mathbb{R}^2_{\geq 0}$ such that $M^{\text{E}}(w') = m$ by assumption, which means that $C^E_t(m) \leq c(w')$. The set

$$\begin{aligned}
&\left\{w \in \mathbb{R}^2_{\geq 0} \mid c(w) \leq c(w'), u(w,t) \geq 0, M^{\text{E}}(w) \geq m\right\} \\
&= \left\{w \in \mathbb{R}^2_{\geq 0} \mid M^{\text{E}}(w) \geq m\right\} \cap \left\{w \in \mathbb{R}^2_{\geq 0} \mid u(w,t) \geq 0\right\} \cap c^{-1}\left([0, c(w')]\right).
\end{aligned}$$

The first two terms are closed, and the last term is compact (because the preimage of a continuous function of a compact set is compact). This means that the intersection is compact. Now, we use the fact that the inf of a continuous function over compact set is achievable.

Let w^* be an optima. We show the following two properties:

(P1) If $w^*_{\text{costly}}, w^*_{\text{cheap}} > 0$, then $M^{\text{E}}(w^*) = m$.

(P2) If $w^*_{\text{costly}} > 0$, then $u(w^*, t) = 0$.

First, we show (P1). Assume for sake of contradiction that $M^{\mathrm{E}}(w) > m$. Let d be the direction normal to $\nabla u(w)$ where the costly coordinate is negative and the cheap coordinate is negative. We see that

$$\begin{aligned}\langle d, \nabla M^{\mathrm{E}}(w^*)\rangle &= -|d_1|(\nabla M^{\mathrm{E}}(w^*))_1 - |d_2|(\nabla M^{\mathrm{E}}(w^*))_2 < 0\\ \langle d, \nabla c(w^*)\rangle &= -|d_1|(\nabla c(w^*))_1 - |d_2|(\nabla c(w^*))_2 < 0\\ \langle d, \nabla u(w^*)\rangle &= 0.\end{aligned}$$

This proves there exists $\varepsilon > 0$ such that $w' = w + \varepsilon d$ satisfies $M^{\mathrm{E}}(w') \geq m$, $u(w', t) \geq 0$, and $c(w') < c(w^*)$, which is a contradiction.

Next, we show (P2). Assume for sake of contradiction that $u(w^*, t) > 0$. Let d be the normal direction to $\nabla M^{\mathrm{E}}(w^*)$ where the costly coordinate is negative and the cheap coordinate is positive. We see that

$$\begin{aligned}\langle d, \nabla u(w^*, t)\rangle &= -|d_1|(\nabla u(w))_1 + d_2(\nabla u(w))_2 < 0\\ \langle d, \nabla M^{\mathrm{E}}(w^*)\rangle &= 0,\end{aligned}$$

Moreover, we can see that $\langle d, \nabla c(w^*)\rangle = -|d_1|(\nabla c(w))_1 + d_2(\nabla c(w))_2 < 0$, since this can be written as:

$$\frac{(\nabla c(w))_1}{(\nabla c(w))_2} > \frac{|d_2|}{|d_1|} = \frac{(\nabla M^{\mathrm{E}}(w))_1}{(\nabla M^{\mathrm{E}}(w))_2},$$

which holds by assumption. This proves there exists $\varepsilon > 0$ such that $w' = w + \varepsilon d$ satisfies $M^{\mathrm{E}}(w') \geq m$, $u(w', t) \geq 0$, and $c(w') < c(w^*)$, which is a contradiction.

We now show that $w^* \in \mathcal{C}_t^{\mathrm{aug}}$. First, suppose that $w^*_{\mathrm{costly}} = 0$. Then, using the fact that $u(w^*, t) \geq 0$, we see that $f_t(w^*_{\mathrm{cheap}}) = 0 = w_{\mathrm{costly}}*$, so by Lemma 200, $w^* \in \mathcal{C}_t^{\mathrm{aug}}$. Next, suppose that $w^*_{\mathrm{costly}} > 0$. Then we see that $u(w^*, t) = 0$ by (P2), so $w^* \in \mathcal{C}_t^{\mathrm{aug}}$.

For the remainder of the analysis, we assume that $c(w^*) = C_t^E(m) > 0$.

If gaming is costless ($(\nabla(c(w)))_2 = 0$ for all w) and $c(w^*) > 0$, then it must hold that $w^*_{\mathrm{costly}} > 0$. This implies that $u(w^*, t) = 0$. This means that there is a unique value $w \in \mathcal{C}_t^{\mathrm{aug}}$ such that $c(w) = C_t^E(m)$, so this implies that w^* is the unique optima. If $w^*_{\mathrm{cheap}} > 0$, then we can apply (P1) to see that $M^{\mathrm{E}}(w^*) = m$. If $w^*_{\mathrm{cheap}} = 0$, the fact that $[0, w^*_{\mathrm{costly}}] \in \mathcal{C}_t^{\mathrm{aug}}$ implies that $M^{\mathrm{E}}(w^*) = \inf_{w \in \mathcal{C}_t^{\mathrm{aug}}} M^{\mathrm{E}}(w)$. By the assumption that $m \in \left\{M^{\mathrm{E}}(w) \mid w \in \mathcal{C}_t^{\mathrm{aug}}\right\}$, this means that $m = M^{\mathrm{E}}(w^*)$ as desired.

If gaming is costly ($(\nabla(c(w)))_2 = 0$ for all w) and $C_t^E(m) > 0$, then there is a unique value $w \in \mathcal{C}_t^{\mathrm{aug}}$ such that $c(w) = c(w^*)$, which shows there is a unique optima. If $w^*_{\mathrm{cheap}} > 0$ and $w^*_{\mathrm{costly}} > 0$, then (P1) implies that $m = M^{\mathrm{E}}(w^*)$. If $w^*_{\mathrm{cheap}} = 0$, then the fact that $[0, w^*_{\mathrm{costly}}] \in \mathcal{C}_t^{\mathrm{aug}}$ implies that $M^{\mathrm{E}}(w^*) = \inf_{w \in \mathcal{C}_t^{\mathrm{aug}}} M^{\mathrm{E}}(w)$. Finally, suppose that $w^*_{\mathrm{costly}} = 0$. Assume for sake of contradiction that $M^{\mathrm{E}}([0, w^*_{\mathrm{cheap}}]) > m$. Then there exists $w_{\mathrm{cheap}} < w^*_{\mathrm{cheap}}$ such that $M^{\mathrm{E}}([0, w_{\mathrm{cheap}}]) \geq m$, $c([0, w'_{\mathrm{cheap}}]) < c([0, w^*_{\mathrm{cheap}}])$, and $u([0, w'_{\mathrm{cheap}}], t) \geq u([0, w^*_{\mathrm{cheap}}], t)$, which would mean that w^* is not an optima, which is a contradiction. $\square$

We next prove the following properties of the equilibrium characterizations for Example 4 in the case of homogeneous users. First, we analyze the marginal distribution of quality of the symmetric mixed equilibrium for engagement-based optimization.

Proposition 203. *Consider Chapter 4 with sufficiently high baseline utility $\alpha > -1$, bounded gaming costs $\gamma \in [0, 1)$, and homogeneous users ($\mathcal{T} = \{t\}$). LThe distribution $(W_{costly}, W_{cheap}) \sim \mu^e(P, c, u, \mathcal{T})$ (where $\mu^e(P, c, u, \mathcal{T})$ is specified as in Theorem 65) satisfies:*

$$\mathbb{P}[W_{costly} \leq w_{costly}] = \begin{cases} (-\alpha)^{1/(P-1)} & \text{if } 0 \leq w_{costly} \leq -\alpha \\ (\min(1, w_{costly} + \gamma \cdot t \cdot (w_{costly} + \alpha)))^{1/(P-1)} & \text{if } w_{costly} \geq \max(0, -\alpha). \end{cases}$$

Proof. Let $\beta_t = \min\{w_{\text{costly}} \mid u([w_{\text{costly}}, 0]) \geq 0\}$ be the minimum investment level. We apply the equilibrium characterization in Chapter 67. We split into two cases: (1) $\beta_t > 0$ and (2) $\beta_t = 0$.

Case 1: $\beta_t > 0$. The minimum-investment function f_t is strictly increasing so f_t^{-1} is well-defined. Using the the equilibrium characterization in Chapter 67, we observe that:

$$\mathbb{P}[W_{\text{costly}} \leq w_{\text{costly}}] = \begin{cases} (\min(1, c([\beta_t, 0])))^{1/(P-1)} & \text{if } 0 \leq w_{\text{costly}} \leq \beta_t \\ \left(\min(1, c([w_{\text{costly}}, f_t^{-1}(w_{\text{costly}})]))\right)^{1/(P-1)} & \text{if } w_{\text{costly}} \geq \beta_t. \end{cases}$$

Now, using the specification in Chapter 4, where $c([w_{\text{costly}}, w_{\text{cheap}}]) = w_{\text{costly}} + \gamma \cdot w_{\text{cheap}}$, $u(w, t) = w_{\text{costly}} - (w_{\text{cheap}}/t) + \alpha$, and $f_t(w_{\text{cheap}}) = \max(0, (w_{\text{cheap}}/t) - \alpha)$, we can simplify this expression. In particular, we see that $\beta_t = \max(0, -\alpha) = -\alpha \leq 1$ (since $\alpha > -1$ by assumption). Moreover, $f_t^{-1}(w_{\text{costly}}) = t \cdot (w_{\text{costly}} + \alpha)$ and $c([w_{\text{costly}}, f_t^{-1}(w_{\text{costly}})]) = w_{\text{costly}} + \gamma \cdot t \cdot (w_{\text{costly}} + \alpha)$. Together, this yields the desired expression for this case.

Case 1: $\beta_t = 0$. Even though the minimum-investment function f_t is no longer strictly increasing in general, it is strictly increasing on a restricted interval. Let

$$\delta_t = \inf\{w_{\text{cheap}} \mid w_{\text{cheap}} \geq 0, f_t(w_{\text{cheap}}) > 0\}$$

be the minimum value such that strictly positive quality is required to maintain nonnegative utility. We see that $f_t(w_{\text{cheap}})$ is strictly increasing for $w_{\text{cheap}} > \delta_t$. This means that for $w_{\text{costly}} > 0$, the inverse f_t^{-1} exists. Using the the equilibrium characterization in Chapter 67, we observe that:

$$\mathbb{P}[W_{\text{costly}} \leq w_{\text{costly}}] = \begin{cases} (\min(1, C_t(\delta_t)))^{1/(P-1)} & \text{if } w_{\text{costly}} = 0 \\ \left(\min(1, c([w_{\text{costly}}, f_t^{-1}(w_{\text{costly}})]))\right)^{1/(P-1)} & \text{if } w_{\text{costly}} \geq \delta_t. \end{cases}$$

Now, using the specification in Chapter 4, where $c([w_{\text{costly}}, w_{\text{cheap}}]) = w_{\text{costly}} + \gamma \cdot w_{\text{cheap}}$, $u(w, t) = w_{\text{costly}} - (w_{\text{cheap}}/t) + \alpha$, and $f_t(w_{\text{cheap}}) = \max(0, (w_{\text{cheap}}/t) - \alpha)$, we can simplify this expression. In particular, we see that $\delta_t = t \cdot \alpha$ and $C_t(\delta_t) = t \cdot \alpha \cdot \gamma$. Moreover, as above, we see that $f_t^{-1}(w_{\text{costly}}) = t \cdot (w_{\text{costly}} + \alpha)$ and $c([w_{\text{costly}}, f_t^{-1}(w_{\text{costly}})]) = w_{\text{costly}} + \gamma \cdot t \cdot (w_{\text{costly}} + \alpha)$. Together, this yields the desired expression. $\square$

Next, we analyze the marginal distribution of quality of the symmetric mixed equilibrium for investment-based optimization.

Proposition 204. *Consider Chapter 4 with bounded gaming costs $\gamma \in [0,1)$ and sufficiently high baseline utility $\alpha > -1$. Furthermore, suppose that either (a) users are homogeneous ($\mathcal{T} = \{t\}$), or (b) the baseline utility satisfies $\alpha \geq 0$. The distribution $(W_{costly}, W_{cheap}) \sim \mu^i(P, c, u, \mathcal{T})$ (where $\mu^i(P, c, u, \mathcal{T})$ is specified as in Theorem 65) satisfies:*

$$\mathbb{P}[W_{costly} \leq w_{costly}] = \begin{cases} (-\alpha)^{1/(P-1)} & \text{if } 0 \leq w_{costly} \leq -\alpha \\ (\min(1, w_{costly}))^{1/(P-1)} & \text{if } w_{costly} \geq \max(0, -\alpha). \end{cases}$$

Proof. Let $\beta_t = \min\{w_{\text{costly}} \mid u([w_{\text{costly}}, 0]) \geq 0\}$ be the minimum investment level. We apply the equilibrium characterization in Chapter 65. We observe that:

$$\mathbb{P}[W_{\text{costly}} \leq w_{\text{costly}}] = \begin{cases} (\min(1, c([\beta_t, 0])))^{1/(P-1)} & \text{if } 0 \leq w_{\text{costly}} \leq \beta_t \\ (\min(1, c([w_{\text{costly}}, 0])))^{1/(P-1)} & \text{if } w_{\text{costly}} \geq \beta_t. \end{cases}$$

Now, using the specification in Chapter 4, where $c([w_{\text{costly}}, 0]) = w_{\text{costly}}$ and $u(w,t) = w_{\text{costly}} - (w_{\text{cheap}}/t) + \alpha$, we can simplify these expressions. In particular, we see that $\beta_t = \max(0, -\alpha) \leq 1$ (since $\alpha > -1$ by assumption). Together, this yields the desired expression. □

Finally, we analyze the marginal distribution over T of (V, T) in Chapter 8 for Cases 2-3.

Lemma 205. *Consider the setup of Chapter 8. If $1 \leq a_{t_1}/a_{t_2} \leq 1.5$, then it holds that:*

$$\mathbb{P}[T = t_1] = 2 - \frac{a_{t_1}}{\cdot a_{t_2}}$$
$$\mathbb{P}[T = t_2] = \frac{a_{t_1}}{\cdot a_{t_2}} - 1$$

Proof. We separately analyze Case 2 and Case 3 in Chapter 8.

Case 2: $(5 - \sqrt{5})/2 \leq a_{t_1}/a_{t_2} \leq 1.5$**.** We observe that:

$$\begin{aligned} \mathbb{P}[T = t_2] &= 2a_{t_2} \cdot \left(\frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} - \frac{1}{a_{t_2}} \right) \cdot \left(2 - \frac{a_{t_1}}{a_{t_2}}\right) \\ &\quad + 2a_{t_2} \cdot \left(\frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right) - \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} \right) \\ &= \left(\frac{1}{\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} - 2 \right) \cdot \left(2 - \frac{a_{t_1}}{a_{t_2}}\right) + \left(4 - \frac{a_{t_1}}{\cdot a_{t_2}} - \frac{1}{\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)}\right) \\ &= \frac{a_{t_1}}{a_{t_2}} - \left(\frac{a_{t_1}}{a_{t_2}} - 1\right) \frac{1}{\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} \\ &= \frac{a_{t_1}}{\cdot a_{t_2}} - 1. \end{aligned}$$

This also implies that:

$$\mathbb{P}[T = t_1] = 1 - \mathbb{P}[T = t_2] = 2 - \frac{a_{t_1}}{a_{t_2}}$$

as desired.

Case 3: $1 \leq a_{t_1}/a_{t_2} \leq (5-\sqrt{5})/2$**.** We observe that:

$$\begin{aligned}
\mathbb{P}[T = t_2] &= 2a_{t_2} \cdot \left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2} \cdot \left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} - \frac{1}{a_{t_2}} \right) \cdot \left(2 - \frac{a_{t_1}}{a_{t_2}}\right) \\
&= \left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}} - 2 \right) \cdot \left(2 - \frac{a_{t_1}}{a_{t_2}}\right) \\
&= \left(\frac{\frac{a_{t_1}}{a_{t_2}} - 1}{2 - \frac{a_{t_1}}{a_{t_2}}} \right) \cdot \left(2 - \frac{a_{t_1}}{a_{t_2}}\right) \\
&= \frac{a_{t_1}}{a_{t_2}} - 1.
\end{aligned}$$

This also implies that:

$$\mathbb{P}[T = t_1] = 1 - \mathbb{P}[T = t_2] = 2 - \frac{a_{t_1}}{a_{t_2}}$$

as desired.

□

F.2 Proofs for Chapter 9.2

We prove Theorem 54. The proof follows similarly to existence of equilibrium proof in (Proposition 2, Jagadeesan et al., 2023a), and we similarly leverage equilibrium existence technology for discontinuous games (Reny, 1999).

The proof will use the following lemma, which is a simple fact about continuously differentiable functions that we reprove for completeness.

Lemma 206. *Let $Q \subseteq \mathbb{R}^2_{\geq 0}$ be a compact set. Any continuously differentiable function $f : Q \to \mathbb{R}_{\geq 0}$ is Lipschitz in the metric $d(x, y) = \|x - y\|_2$.*

Proof. By assumption, the gradient mapping $G : \mathbb{R}^2_{\geq 0} \to \mathbb{R}^2_{\geq 0}$ given by $G(w) = \nabla(w)$ is continuous. Any continuous function on a compact set is bounded, so we know that for some constant B it holds that $\|\nabla(w)\|_2 \leq B$ for all $w \in Q$. Since the gradient is bounded, this means that the function is Lipschitz as desired. □

We are now ready to prove Theorem 54.

Proof of Theorem 54. We leverage a standard result about the existence of symmetric, mixed strategy equilibria in discontinuous games (see Corollary 5.3 of (Reny, 1999)). We adopt the terminology of Reny (1999) and refer the reader to Reny (1999) for a formal definition of the conditions.

First, the game is symmetric by construction. In particular, the creators have symmetric utility functions. Even though the functions U_i as written are not explicitly symmetric, the fact that we break ties uniformly at random means that $U_i(w_i; \mathbf{w}_{-i}) = U_j(w_i; \mathbf{w}_{-i})$ for all $i, j \in [P]$. We thus let $U(w_1, w_{-1}) = U_1(w_1, w_{-1})$ denote this utility function for the remainder of the analysis.

To show the existence of a symmetric mixed equilibrium, it suffices to show that: (1) the action space is convex and compact and (2) the game is diagonally better-reply secure.

Creator action space is convex and compact. In the current game, the action space $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ is not compact. However, we show that we can define a modified game, where the action space is convex and compact, and where an equilibrium in this modified game is also an equilibrium in the original game. For the remainder of the proof, we analyze this modified game.

We define the modified action space as follows. Let $w_{\text{costly}}^{\max}$ be defined to be:

$$w_{\text{costly}}^{\max} = 1 + \sup \left\{ w_{\text{costly}} \geq 0 \mid c([w_{\text{costly}}, 0]) \leq 1 \right\}.$$

Let $w_{\text{cheap}}^{\max}$ be defined to be:

$$w_{\text{cheap}}^{\max} := 1 + \sup \left\{ w_{\text{cheap}} \geq 0 \mid \text{there exists } t \in \mathcal{T} \text{ such that } u([w_{\text{costly}}^{\max}, w_{\text{cheap}}], t) \geq 0 \right\}.$$

(We add an additive factor of 1 slack to guarantee that there exists a best-response by a creator will be in the *interior* of the action space and not on the boundary.) We take the action space to be

$$\mathcal{W} := [0, w_{\text{costly}}^{\max}] \times [0, w_{\text{cheap}}^{\max}],$$

which is compact and convex by construction.

We show that for any distribution μ over $\mathcal{W}$, there exists a best-response $w^* \in \mathbb{R}_{\geq 0}^2$ to:

$$\arg\max_{w \in \mathbb{R}_{\geq 0}^2} \mathbb{E}_{\mathbf{w}_{-i} \sim \mu^{P-1}} [U_i(w_i; \mathbf{w}_{-i})]$$

such that w^* is in the interior of $\mathcal{W}$. To show this, let w^* be any best-response to the above optimization program. First we show that $w_{\text{costly}}^* < w_{\text{costly}}^{\max}$. Assume for sake of contradiction that $w_{\text{costly}}^* \geq w_{\text{costly}}^{\max}$. Then it must hold that

$$c(w^*) \geq c([w_{\text{costly}}^*, 0]) \geq c([w_{\text{costly}}^{\max}, 0]) > 1.$$

This means that

$$\mathbb{E}_{\mathbf{w}_{-i} \sim \mu^{P-1}} [U_i(w^*; \mathbf{w}_{-i})] < 0 \leq \mathbb{E}_{\mathbf{w}_{-i} \sim \mu^{P-1}} [U_i([0, 0]; \mathbf{w}_{-i})],$$

which is a contradiction. This proves that $w^*_{\text{costly}} < w^{\max}_{\text{costly}}$ as desired. We next show that we can construct a best-response w' such that $w'_{\text{costly}} = w^*_{\text{costly}}$ and $w'_{\text{cheap}} < w^{\max}_{\text{cheap}}$. If $u(w', t) \geq 0$ for some $t \in \mathcal{T}$, then it must hold that $w^*_{\text{cheap}} < w^{\max}_{\text{cheap}}$, so we can take $w'_{\text{cheap}} = w^*_{\text{cheap}}$. If $u(w', t) < 0$ for all $t \in \mathcal{T}$, then we see that:

$$\mathbb{E}_{\mathbf{w}_{-i} \sim \mu^{P-1}}[U_i(w^*; \mathbf{w}_{-i})] \leq 0 \leq \mathbb{E}_{\mathbf{w}_{-i} \sim \mu^{P-1}}[U_i([w^*_{\text{costly}}, 0]; \mathbf{w}_{-i})],$$

so we can take $w'_{\text{cheap}} = 0 < w^{\max}_{\text{cheap}}$. Altogether, this proves that there exists a best-response w' satisfying $w'_{\text{costly}} < w^{\max}_{\text{costly}}$ and $w'_{\text{cheap}} < w^{\max}_{\text{cheap}}$, which means that w' is in the interior of $\mathcal{W}$

This proves that any symmetric mixed equilibrium of the game with restricted action space $\mathcal{W}$ will also be a symmetric mixed equilibrium of the new game.

Establishing diagonal better reply security. In this analysis, we slightly abuse notation and implicitly extend the definition of each utility function U to mixed strategies by considering expected utility.

First, we show the payoff function $U(\mu; [\mu, \ldots, \mu])$ (where μ is a distribution over the action space $\mathcal{W}$) is upper semi-continuous in μ with respect to the weak* topology. Using the fact that each creator receives a $1/P$ fraction of users in expectation at a symmetric solution, we see that:

$$U(\mu; [\mu, \ldots, \mu]) = \frac{1}{P} \cdot \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left(1 - \left(\int_w \mathbb{1}[u(w, t) < 0] d\mu\right)^{P-1}\right) - \int_w c(w) d\mu.$$

Since c is a continuous function, we see immediately that $\int_w c(w) d\mu$ is continuous in μ. Moreover, for each $t \in \mathcal{T}$, we see that $\int_w \mathbb{1}[u(w, t) < 0] d\mu$ is lower semi-continuous in μ. This proves that $U(\mu; [\mu, \ldots, \mu])$ is upper upper semi-continuous in μ as desired.

For each relevant payoff in the closure of the graph of the game's diagonal payoff function, we construct an action that secures that payoff along the diagonal. More formally, let (μ^*, α^*) be in the closure of the graph of the game's diagonal payoff function, and suppose that $(\mu^*, \ldots, \mu^*)$ is not an equilibrium; it suffices to show that a creator can secure a payoff of $\alpha > \alpha^*$ along the diagonal at $(\mu^*, \ldots, \mu^*)$. Since U is upper semi-continuous, it actually suffices to show the statement for (μ^*, α^*) where $\alpha^* = U(\mu^*; [\mu^*, \ldots, \mu^*])$ and $(\mu^*, \ldots, \mu^*)$ is not an equilibrium. For each such (μ^*, α^*), we construct μ^{sec} that secures a payoff of $\alpha > \alpha^*$ along the diagonal at $(\mu^*, \ldots, \mu^*)$ as follows.

Since $(\mu^*, \ldots, \mu^*)$ is not an equilibrium and since there exists a best-response in the interior of $\mathcal{W}$ as shown above, we know that there exists w in the interior of $\mathcal{W}$ such that:

$$U(w; [\mu^*, \ldots, \mu^*]) > U(w; [\mu^*, \ldots, \mu^*]) = \alpha^*.$$

Since we want to find w that achieves high profit in an open neighborhood of μ^*, we need to strengthen the above statement; we can achieve by this by appropriately perturbing w (which we can do since w is in the interior of $\mathcal{W}$). First, we can perturb w to $\tilde{w}$ such that the distribution $M^{\text{E}}(w')$ where $w' \sim \mu^*$ does not have a point mass at $M^{\text{E}}(\tilde{w})$, and such that:

$$U(\tilde{w}; [\mu^*, \ldots, \mu^*]) = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(\tilde{w}, t) \geq 0] \cdot \left(\mathbb{P}_{w' \sim \mu^*}[M^{\text{E}}(\tilde{w}) > M^{\text{E}}(w') \text{ or } u(w', t) < 0]\right)^{P-1} - c(\tilde{w}) \quad > \alpha^*.$$

Now, we construct w^{sec} as a perturbation of $\tilde{w}$ along the costly dimension w_{cheap} to add ε slack to the constraint $M^{\text{E}}(\tilde{w}) > M^{\text{E}}(w')$. Since M^{E} is strictly increasing in the expensive component and since $\mathbb{P}_{w' \sim \mu^*}[u(w', t) \in (0, \varepsilon)] \to 0$ as $\varepsilon \to 0$[1], we observe that there exists $\varepsilon^* > 0$ and $w^{\text{sec}} \in \mathcal{W}$ (constructed as $w^{\text{sec}} = \tilde{w} + [\varepsilon', 0]$ for some $\varepsilon' > 0$) such that

$$\begin{aligned}
&\alpha^* \\
&< \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\text{sec}}, t) \geq 0] \cdot \left(\mathbb{P}_{w' \sim \mu^*}[M^{\text{E}}(w^{\text{sec}}) > M^{\text{E}}(w') + \varepsilon^* \text{ or } u(w', t) < -\varepsilon^*]\right)^{P-1} - c(w^{\text{sec}}) \\
&\leq \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\text{sec}}, t) \geq 0] \cdot \left(\mathbb{P}_{w' \sim \mu^*}[M^{\text{E}}(w^{\text{sec}}) > M^{\text{E}}(w') \text{ or } u(w', t) < 0]\right)^{P-1} - c(w^{\text{sec}}) \\
&= U(w^{\text{sec}}; [\mu^*, \ldots, \mu^*]).
\end{aligned} \tag{F.3}$$

We claim that μ^{sec} taken to be the point mass at w^{sec} will secure a payoff of

$$\begin{aligned}
&\alpha \\
&= \frac{\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\text{sec}}, t) \geq 0] \cdot \left(\mathbb{P}_{w' \sim \mu^*}[M^{\text{E}}(w^{\text{sec}}) > M^{\text{E}}(w') + \varepsilon^* \text{ or } u(w', t) < -\varepsilon^*]\right)^{P-1} - c(w^{\text{sec}}) + \alpha^*}{2} \\
&> \alpha^*
\end{aligned}$$

along the diagonal at $(\mu^*, \ldots, \mu^*)$. For each $t \in \mathcal{T}$, we define the event A_t to be:

$$A_t = \left\{w' \mid M^{\text{E}}(w^{\text{sec}}) > M^{\text{E}}(w') \text{ or } u(w', t) < 0\right\}$$

and for $\varepsilon > 0$, we define the event A_t^{ε} as:

$$A_t^{\varepsilon} = \left\{w' \mid M^{\text{E}}(w^{\text{sec}}) > M^{\text{E}}(w') + \varepsilon \text{ or } u(w', t) < -\varepsilon\right\}.$$

In this notation, we can rewrite equation (F.3) as:

$$U(w^{\text{sec}}; [\mu^*, \ldots, \mu^*]) \geq \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\text{sec}}, t) \geq 0] \cdot \left(\mu^*(A_t^{\varepsilon^*})\right)^{P-1} - c(w^{\text{sec}}) > \alpha^*$$

and α as:

$$\alpha = \frac{\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\text{sec}}, t) \geq 0] \cdot \left(\mu^*(A_t^{\varepsilon^*})\right)^{P-1} - c(w^{\text{sec}}) + \alpha^*}{2} > \alpha^*$$

We define a metric d on $\mathbb{R}^2_{\geq 0}$ as follows. Using Lemma 206, we know that $M^{\text{E}}(\cdot)$ and $u(\cdot, t)$ for each $t \in \mathcal{T}$ are Lipschitz in $\|\cdot\|_2$. Let the Lipschitz constants be L_M and L_t for each $t \in \mathcal{T}$, respectively. Consider the metric on $\mathbb{R}^2_{\geq 0}$ given by

$$d(w, w') = \max(L_M, \max_{t \in \mathcal{T}} L_t) \cdot \|w - w'\|_2.$$

[1] To see this, let $S_i = (0, 2^{-n})$ and use that $0 = \mu^*(\cap_{i \geq 1} S_i) = \lim_{i \to \infty} \mu^*(S_i)$.

For $\varepsilon > 0$ let $B_\varepsilon(\mu^*)$ denote the ε-ball with respect to the Prohorov metric; using the definition of the weak* topology, we see that $B_\varepsilon(\mu^*)$ is an open set with respect to the weak* topology. For every $w' \in A_t^\varepsilon$, we see that A_t contains the open neighborhood $B_\varepsilon(w')$ with respect to d. By the definition of the Prohorov metric, we know that for all $\mu' \in B_\varepsilon(\mu^*)$, it holds that

$$\mu'(A_i) \geq \mu^*(A_i^\varepsilon) - \varepsilon.$$

This implies that

$$\begin{aligned}
U(w^{\mathrm{sec}}; [\mu', \ldots, \mu']) &\geq \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot (\mu'(A_t))^{P-1} - c(w^{\mathrm{sec}}) \\
&\geq \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot (\mu^*(A_t^\varepsilon) - \varepsilon)^{P-1} - c(w^{\mathrm{sec}}) \\
&\geq \left(\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot (\mu^*(A_t^\varepsilon))^{P-1} - c(w^{\mathrm{sec}}) \right) \\
&\quad - \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left(\mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot \underbrace{(\mu^*(A_t^\varepsilon))^{P-1} - (\mu^*(A_t^\varepsilon) - \varepsilon)^{P-1}}_{(A)} \right).
\end{aligned}$$

Using that (A) goes to 0 as ε goes to 0, we see that for sufficiently small ε, it holds that:

$$\begin{aligned}
&\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left(\mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot \underbrace{(\mu^*(A_t^\varepsilon))^{P-1} - (\mu^*(A_t^\varepsilon) - \varepsilon)^{P-1}}_{(A)} \right) \\
&\leq \frac{\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot \left(\mu^*(A_t^{\varepsilon^*})\right)^{P-1} - c(w^{\mathrm{sec}}) - \alpha^*}{3}.
\end{aligned}$$

As long as ε is also less than ε^*, this means that:

$$\begin{aligned}
U(w^{\mathrm{sec}}; [\mu', \ldots, \mu']) &\geq \left(\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot (\mu^*(A_t^\varepsilon))^{P-1} - c(w^{\mathrm{sec}}) \right) \\
&\quad - \frac{\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot \left(\mu^*(A_t^{\varepsilon^*})\right)^{P-1} - c(w^{\mathrm{sec}}) - \alpha^*}{3} \\
&\geq \left(\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot \left(\mu^*(A_t^{\varepsilon^*})\right)^{P-1} - c(w^{\mathrm{sec}}) \right) \\
&\quad - \frac{\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot \left(\mu^*(A_t^{\varepsilon^*})\right)^{P-1} - c(w^{\mathrm{sec}}) - \alpha^*}{3} \\
&= \frac{2 \left(\frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \mathbb{1}[u(w^{\mathrm{sec}}, t) \geq 0] \cdot \left(\mu^*(A_t^{\varepsilon^*})\right)^{P-1} - c(w^{\mathrm{sec}}) \right) + \alpha^*}{3} \\
&> \alpha
\end{aligned}$$

for all $\mu' \in B_\varepsilon(\mu^*)$, as desired. $\square$

F.3 Proofs for Chapter 9.3

To prove Proposition 55, we first show that the support $\mathcal{C}_t^{\text{aug}}$ of the equilibrium is contained within union of curves of the following form in $\mathbb{R}^2_{\geq 0}$:

$$\{[w_{\text{costly}}, w_{\text{cheap}}] \mid u([w_{\text{costly}}, w_{\text{cheap}}], t) = 0\} \cup \left\{(0, w_{\text{cheap}}) \mid w_{\text{cheap}} \in [0, \min_{w' \mid u(w',t)=0} w'_{\text{cheap}})]\right\},$$

as formalized in the following lemma.

Lemma 207. *Let $\mathcal{T} \subseteq \mathbb{R}_{\geq 0}$ be any finite type space. Suppose that $(\mu_1, \ldots, \mu_P)$ is an equilibrium in the game with $M = M^E$. Then supp$(\mu_i) \subseteq \cup_{t \in \mathcal{T}} \mathcal{C}_t^{aug} \cup \{w \mid c(w) = 0\}$ for all $i \in [P]$.*

We prove this lemma as a corollary to the following sublemma:

Lemma 208. *Let $\mathcal{T} \subseteq \mathbb{R}_{\geq 0}$ be any finite type space. Let $(\mu_1, \ldots, \mu_P)$ be any mixed strategy profile. Then, in the game with $M = M^E$, any best-response $w \in \mathbb{R}^2_{\geq 0}$ to:*

$$\arg\max_{w \in \mathbb{R}^2_{\geq 0}} \mathbb{E}_{\boldsymbol{w}_{-i} \sim \mu_{-i})}[U_i(w; \boldsymbol{w}_{-i})]$$

satisfies $w \in \cup_{t \in \mathcal{T}} \mathcal{C}_t^{aug} \cup \{w \mid c(w) = 0\}$.

Proof of Lemma 208. Assume for sake of contradiction that a best-response w' satisfies $w' \notin \cup_{t \in \mathcal{T}} \mathcal{C}_t^{\text{aug}} \cup \{w \mid c(w) = 0\}$. Let $m = M^{\text{E}}(w)$. We will show that w' is not a best response.

Let $T := \{t \in \mathcal{T} \mid u(w', t) \geq 0\}$ be the set of types for which w' incurs nonnegative user utility. For the remainder of the analysis, we split into two cases: $T = \emptyset$ and $T \neq \emptyset$.

First, suppose that $T = \emptyset$. Then no user will never consume content, so since $c(w') > 0$, it holds that $\mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w'; \mathbf{w}_{-i})] < 0$. However, if the creator were to instead choose $[0, 0]$ which incurs 0 cost, they would get nonnegative utility

$$\mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i([0, 0]; \mathbf{w}_{-i})] \geq 0 > \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w'; \mathbf{w}_{-i})].$$

Thus, w' is not a best response, which is a contradiction.

Next, suppose that $T \neq \emptyset$. By the assumptions on u, it holds that $t = \min_{t' \in \mathcal{T}} t' \in T$. Then, w is a feasible solution to the optimization program $\min_w c(w)$ s.t. $u(w, t) \geq 0, M^{\text{E}}(w) \geq m$. By Lemma 202, there exists $w^* \in \mathcal{C}_t^{\text{aug}}$ that is a feasible solution to $\min_w c(w)$ s.t. $u(w, t) \geq 0, M^{\text{E}}(w) \geq m$ such that $c(w^*) < c(w')$. Moreover, by assumption (A3), since $u(w, t) \geq 0$, we see that $u(w^*, t') \geq 0$ for all $t' \in T$. Thus,

$$\mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w^*; \mathbf{w}_{-i})] > \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w'; \mathbf{w}_{-i})],$$

so w' is not a best response which is a contradiction. $\square$

We can now deduce Lemma 207.

Proof of Lemma 207. Assume for sake of contradiction that $w \in \text{supp}(\mu_i)$ satisfies $w \notin \cup_{t\in\mathcal{T}}\mathcal{C}_t^{\text{aug}} \cup \{w \mid c(w) = 0\}$. By Lemma 208, w is not a best response, which is a contradiction. □

Proposition 55 follows immediately from Lemma 207 along with the lemmas in Chapter F.1.

Proof of Proposition 55. Applying Lemma 207 and Lemma 200, we know that:

$$\text{supp}(\mu_i) \subseteq \cup_{t\in\mathcal{T}}\mathcal{C}_t^{\text{aug}} \cup \{w \mid c(w) = 0\} = \cup_{t\in\mathcal{T}} \{[f_t(w_{\text{cheap}}), w_{\text{cheap}}] \mid w_{\text{cheap}} \geq 0\} \cup \{w \mid c(w) = 0\}\,.$$

Using the fact that gaming is not costless, we see that:

$$\begin{aligned}\text{supp}(\mu_i) &\subseteq \cup_{t\in\mathcal{T}} \{[f_t(w_{\text{cheap}}), w_{\text{cheap}}] \mid w_{\text{cheap}} \geq 0\} \cup \{w \mid c(w) = 0\} \\ &= \cup_{t\in\mathcal{T}} \{[f_t(w_{\text{cheap}}), w_{\text{cheap}}] \mid w_{\text{cheap}} \geq 0\} \cup \{[0, 0]\}\end{aligned}$$

The fact that the functions f_t are weakly increasing follows from Lemma 201. □

Theorem 56 follows as a consequence of Proposition 55.

Proof of Theorem 56. Let $\mu_1, \ldots, \mu_P$ be an equilibrium, and suppose that $w^1, w^2 \in \cup_{i\in[P]}\text{supp}(\mu_i)$. By Proposition 55, we see that:

$$\cup_{i\in[P]}\text{supp}(\mu_i) \subseteq \{[f_t(w_{\text{cheap}}), w_{\text{cheap}}] \mid w_{\text{cheap}} \geq 0\} \cup \{[0, 0]\}\,.$$

Using the fact that f_t is weakly increasing and $f_t(0) \geq 0$, we see that if $w^2_{\text{cheap}} \geq w^1_{\text{cheap}}$, then $w^2_{\text{costly}} \geq w^1_{\text{costly}}$. □

F.4 Proofs for Chapter 9.6

F.4.1 Proof of Chapter 67

Before proving Theorem 67, we prove the following properties of $\mu^{\text{e}}(P, c, u, \mathcal{T})$.

Lemma 209. *Let $\mathcal{T} = \{t\}$. The distribution $\mu^e(P, c, u, \mathcal{T})$ satisfies the following properties:*

(P1) The only possible atom in the distribution $\mu^e(P, c, u, \mathcal{T})$ is at $(0, 0)$, and moreover that $(0, 0)$ is an atom when $f_t(0) > 0$.

(P2) Suppose that $(w_{cheap}, w_{costly}) \in supp(\mu^e(P, c, u, \mathcal{T}))$. If $(w_{cheap}, w_{costly}) \neq (0, 0)$ or if $f_t(0) = 0$, then it holds that $u([w_{cheap}, w_{costly}], t) \geq 0$.

(P3) If $(0, 0)$ is an atom of $\mu^e(P, c, u, \mathcal{T})$, then $u([0, 0], t) < 0$.

Proof. To prove (P1), note that if $(w_{\text{cheap}}, w_{\text{costly}}) \in \text{supp}(\mu^{\text{e}}(P, c, u, \mathcal{T}))$ is an atom, then w_{cheap} must be an atom in the marginal distribution W_{cheap}. The specification of the cdf shows that the only possible atom is at $W_{\text{cheap}} = 0$. Moreover, 0 is an atom of W_{cheap} if and only if $c(f_t(0), 0) > 0$, which occurs if and only if $f_t(0) > 0$. When $f_t(0) > 0$, we further see that the conditional distribution W_{costly} is a point mass at 0, as desired.

To prove (P2), note that W_{costly} is a point mass at $f_t(w_{\text{cheap}})$. By the definition of f_t, it holds that $u([w_{\text{cheap}}, w_{\text{costly}}], t) \geq 0$.

To prove (P3), note that the first property showed that $(0, 0)$ is an atom if and only if $f_t(0) > 0$. By the definition of f_t, we see that $u([0, 0], t) < 0$ as desired.

$\square$

We prove Theorem 67.

Proof of Theorem 67. Let $\mu = \mu^{\text{e}}(P, c, u, \mathcal{T})$ for notational convenience. We analyze the expected utility of $H(w) = \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w; w_{-i})]$ of a content creator if all of the creators choose the strategy μ. It suffices to show that any $w^* \in \text{supp}(\mu)$ is a best response $w^* \in \arg\max_w H(w)$. We use the properties (P1)-(P4) in Lemma 209.

First, we observe that we can write $H(w)$ as:

$$\begin{aligned} H(w) &= \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w; w_{-i})] \\ &= \mathbb{1}[u(w, t) \geq 0] \cdot \mathbb{P}_{W_{\text{cheap}}}[M^{\text{E}}(w) > M^{\text{E}}([f_t(W_{\text{cheap}}), W_{\text{cheap}}])]^{P-1} - c(w), \end{aligned}$$

because (P1) implies that the only possible atom occurs at $[0, 0]$, (P3) implies that $u([0, 0], t) < 0$ if $[0, 0]$ is an atom, and (P2) implies that $\mathbb{1}[u(w, t) \geq 0] = 0$ for $(w_{\text{cheap}}, w_{\text{costly}}) \neq (0, 0)$.

If $w \in \text{supp}(\mu)$, then we claim that $H(w) = 0$. If $w_{\text{cheap}} = 0$ and $f_t(w_{\text{cheap}}) = 0$, it is immediate that $H(w) = 0$. If $w_{\text{cheap}} > 0$, then $w = [f_t(w_{\text{cheap}}), w_{\text{cheap}}]$. By (P2), it holds that $u(w, t) \geq 0$. This means that:

$$\begin{aligned} H(w) &= \mathbb{P}_{W_{\text{cheap}}}[M^{\text{E}}([f_t(w_{\text{cheap}}), w_{\text{cheap}}]) > M^{\text{E}}([f_t(W_{\text{cheap}}), W_{\text{cheap}}])]^{P-1} - c(w) \\ &=_{(1)} \mathbb{P}_{W_{\text{cheap}}}[w_{\text{cheap}} > W_{\text{cheap}}]^{P-1} - c(w) \\ &= 0, \end{aligned}$$

where (1) uses the fact that $M^{\text{E}}([f_t(w_{\text{cheap}}), w_{\text{cheap}}])$ is strictly increasing in w_{cheap} (Lemma 201).

The remainder of the proof boils down to showing that $H(w) \leq 0$ for any w. If $u(w, t) < 0$, then $H(w) \leq 0$. If $u(w, t) \geq 0$, then

$$H(w) = \mathbb{P}_{W_{\text{cheap}}}[M^{\text{E}}(w) > M^{\text{E}}([f_t(W_{\text{cheap}}), W_{\text{cheap}}])]^{P-1} - c(w).$$

It suffices to show that $H(w) \leq 0$ at any best-response w such that $u(w, t) \geq 0$. If w is a best response and $u(w, t) \geq 0$, then it must be true that w is a solution to (9.7). By Lemma 202, this means that $w \in \mathcal{C}_t^{\text{aug}}$, and by Lemma 200, this means that w is of the form $[f_t(w_{\text{cheap}}), w_{\text{cheap}}]$, which means that:

$$H(w) = \mathbb{P}_{W_{\text{cheap}}}[w_{\text{cheap}} > W_{\text{cheap}}]^{P-1} - c(w) \leq 0,$$

which proves the desired statement.

□

F.4.2 Proof of Theorem 68

We prove Theorem 68.

Proof of Theorem 68. The high-level idea of the proof is to define a new game with a restricted action space that is easier to analyze (we define slightly different variant games for Part 2 and Part 3). Suppose that μ is a symmetric mixed equilibrium. By Lemma 207 and Lemma 200, we see that:

$$\operatorname{supp}(\mu) \subseteq \mathcal{C}_t^{\text{aug}} \cup \{w \mid c(w) = 0\} = \{[w_{\text{cheap}}, f_t(w_{\text{cheap}})] \mid w_{\text{cheap}} \geq 0\} \cup \{w \mid c(w) = 0\}.$$

Using the assumption that gaming tricks are costly (i.e. $(\nabla c(w))_2 > 0$ for all $w \in \mathbb{R}^2_{\geq 0}$), we obtain that:

$$\operatorname{supp}(\mu) \subseteq \{[w_{\text{cheap}}, f_t(w_{\text{cheap}})] \mid w_{\text{cheap}} \geq 0\} \cup \{[0, 0]\}.$$

This simplification will ultimately enable us to convert the 2-dimensional action space to the 1-dimensional space specified by engagement. By Lemma 201 and since M^{E} is strictly increasing in both arguments, it holds that $M^{\text{E}}([f_t(w_{\text{cheap}}), w_{\text{cheap}}])$ is a strictly increasing function of w_{cheap}. Let

$$m_{\min} := \inf_{w_{\text{cheap}} \geq 0} M^{\text{E}}([f_t(w_{\text{cheap}}), w_{\text{cheap}}])$$

and

$$m_{\max} := \sup_{w_{\text{cheap}} \geq 0} M^{\text{E}}([f_t(w_{\text{cheap}}), w_{\text{cheap}}]).$$

(Note that $m_{\max}$ might be equal to ∞.) This means that for each value $m \in [m_{\min}, m_{\max})$, there is exactly one value $w = [f_t(w_{\text{cheap}}), w_{\text{cheap}}]$ such that $M^{\text{E}}(w) = m$.

We break into two cases: $f_t(w_{\text{cheap}}) = 0$ and $f_t(w_{\text{cheap}}) \geq 0$ for the remainder of the analysis.

Case 1: $f_t(w_{\text{cheap}}) = 0$. In the case that $f_t(w_{\text{cheap}}) = 0$, we can further simplify:

$$\operatorname{supp}(\mu) \subseteq \{[w_{\text{cheap}}, f_t(w_{\text{cheap}})] \mid w_{\text{cheap}} \geq 0\} \cup \{[0, 0]\} = \{[w_{\text{cheap}}, f_t(w_{\text{cheap}})] \mid w_{\text{cheap}} \geq 0\}.$$

Let's consider a different game where the action set is $A = [m_{\min}, m_{\max})$. The action $m \in [m_{\min}, \infty)$ corresponds to the unique value $w = [f_t(w_{\text{cheap}}), w_{\text{cheap}},]$ such that $M^{\text{E}}(w) = m$. In this new game, the utility function is

$$\tilde{U}(a_i; a_{-i}) := 1[a_i = \max_{j \in [P]} a_j] - \tilde{c}(a) \tag{F.4}$$

where ties are broken uniformly at random and where $\tilde{c}(m) = c(w)$ where w is the unique value in $\{[w_{\text{cheap}}, f_t(w_{\text{cheap}})] \mid w_{\text{cheap}} \geq 0\}$ such that $M^{\text{E}}(w) = m$. We see that $\tilde{c}$ is strictly increasing in m since $f_t(w_{\text{cheap}})$ is weakly increasing in w_{cheap} by Lemma 201 and since gaming tricks are costly by assumption. (In the remainder of the proof, we also slightly abuse notation and implicitly extend the definition of $\tilde{U}$ to mixed strategies by considering expected utility.)

We first show that there exists a symmetric equilibrium in the new game with action set A. We use the fact that we have constructed a symmetric equilibrium in the original game in Theorem 67. The transformed distributions $\tilde{\mu}$ is a symmetric equilibrium in the new game; thus, there exists an equilibrium in the new game.

We claim that it suffices to show that there is at most one symmetric equilibrium in the new game with action set A. Every action $a \in A$ corresponds to a unique w. Thus, uniqueness of equilibrium in the transformed game guarantees uniqueness of equilibrium in the original game.

The remainder of the proof of this case boils down to showing that there is at most one symmetric mixed equilibrium $\tilde{\mu}$ in the new game. Let $\tilde{\mu}$ be a symmetric mixed Nash equilibrium of the transformed game.

First, we claim that $\mathbb{P}_{M\sim\tilde{\mu}}[M = m'] = 0$ for all m' (no point masses). If $\mathbb{P}[M = m'] > 0$, then because of uniform-at-random tiebreaking, it holds that $m' + \varepsilon$ for some ε performs strictly better than m' for some $\varepsilon > 0$.

Next, we claim that $m_{\min} \in \operatorname{supp}(\tilde{\mu})$. Assume for sake of contradiction that $m' = \inf_{m\in\operatorname{supp}(\tilde{\mu})} > m_{\min}$. At $m = m'$, the creator gets a utility of $-\tilde{c}(m') < 0$, which means that the creator would get higher utility from the deviation $m = m_{\min}$ where the utility would be 0. (Since $f_t(w_{\text{cheap}}) = 0$, it holds that $\tilde{c}(m_{\min}) = 0$.) This is a contradiction.

For $m \in \operatorname{supp}(\tilde{\mu})$, we claim that $\mathbb{P}_{M\sim\tilde{\mu}}[M \leq m] = (\tilde{c}(m))^{1/(P-1)}$. This is because $m_{\min} \in \operatorname{supp}(\tilde{\mu})$ and the utility at $m_{\min}$ is 0, so the utility at any $m \in \operatorname{supp}(\tilde{\mu})$ must be 0. This implies that $\mathbb{P}_{M\sim\tilde{\mu}}[M \leq m] = (\tilde{c}(m))^{1/(P-1)}$.

Finally, it suffices to show that the support is a closed interval of the form $[m_{\min}, m^*]$ where m^* is the unique value such that $\tilde{c}(m^*) = 1$. We first see that since $\mathbb{P}_{M\sim\tilde{\mu}}[M \leq m] \leq 1$, it must hold that $m \leq m^*$ for any $m \in \operatorname{supp}(\tilde{\mu})$. To see this, let $Q = \operatorname{supp}(\tilde{\mu}) \cup [m^*, m_{\max}]$. Since Q is a finite union of closed sets, it is a closed set. This means that $\bar{Q} = [m_{\min}, m_{\max}] \setminus Q$ is an open set. It suffices to prove that $\bar{Q} = \emptyset$. Assume for sake of contradiction $\bar{Q} \neq \emptyset$. If $m' \in \bar{Q}$, let $m_1 = \inf\{m < m' \mid m \in \operatorname{supp}(\tilde{\mu})\}$ and $m_2 = \sup\{m > m' \mid m \in \operatorname{supp}(\tilde{\mu})\}$. Since $\bar{Q}$ is open, there is an open neighborhood such that $B_\varepsilon(m') \subseteq \bar{Q}$, which means that $m_2 > m_1$. However, this means that $\mathbb{P}_{M\sim\tilde{\mu}}[M = m_2] - = \mathbb{P}_{M\sim\tilde{\mu}}[M \leq m_2] - \mathbb{P}_{M\leq\tilde{\mu}}[M = m_2] > 0$, which is a contradiction.

This proves the desired statement for Case 1.

Case 2: $f_t(w_{\text{cheap}}) > 0$**.** For this case, let's consider a different game where the action set is $A = \{\bot\} \cup [m_{\min}, \infty)$. The action $m \in [m_{\min}, \infty)$ corresponds to the unique value $w = [f_t(w_{\text{cheap}}), w_{\text{cheap}},]$ such that $M^{\mathrm{E}}(w) = m$. In this new game, the utility function is

$$\tilde{U}(a_i; a_{-i}) := 1[a_i \neq \bot] \cdot 1[a_i = \max_{j\in[P]} a_j] - \tilde{c}(a) \tag{F.5}$$

where we use the ordering that $\bot < m_{\min}$, where ties are broken uniformly at random and where $\tilde{c}(\bot) = 0$ and $\tilde{c}(m) = c(w)$ where w is the unique value in $\{[f_t(w_{\text{cheap}}), w_{\text{cheap}}] \mid w_{\text{cheap}} \geq 0\}$

such that $M^{\mathrm{E}}(w) = m$. We similarly see that $\tilde{c}$ is strictly increasing in m since $f_t(w_{\text{cheap}})$ is weakly increasing in w_{cheap} by Lemma 201 and since gaming tricks are costly by assumption. (As before, in the remainder of the proof, we also slightly abuse notation and implicitly extend the definition of $\tilde{U}$ to mixed strategies by considering expected utility.)

By an analogous argument to the previous case, we know that there exists a symmetric equilibrium in the new game with action set A, and it suffices to show that there is at most one symmetric equilibrium in the new game with action set A. The remainder of the proof of this case boils down to showing that there is at most one symmetric mixed equilibrium $\tilde{\mu}$ in the new game. Let $\tilde{\mu}$ be a symmetric mixed Nash equilibrium of the transformed game. We split into several subcases: (1) $\tilde{c}(m_{\min}) > 1$, (2) $\tilde{c}(m_{\min}) = 1$, and (3) $0 < \tilde{c}(m_{\min}) > 1 < 1$.

Subcase 2a: $\tilde{c}(m_{\mathbf{min}}) > 1$**.** Choosing $\perp$ is a strictly dominant strategy. This means that the unique equilibrium is where each $\tilde{\mu}$ is a point mass at $\perp$.

Subcase 2b: $\tilde{c}(m_{\mathbf{min}}) = 1$**.** We claim that $\tilde{\mu}$ as a point mass $\perp$ is still the unique symmetric mixed Nash equilibrium. Assume for sake of contradiction that there is an equilibrium $\tilde{\mu}$ where $\mathbb{P}_{m \sim \tilde{\mu}}[a \neq \perp] > 0$. It must be true that $\text{supp}(\tilde{\mu}) \subseteq \{\perp, m_{\min}\}$ (otherwise, $\perp$ would be a better response). Thus, $\mathbb{P}_{a^* \sim \tilde{\mu}}[a \neq \perp] > 0$. However, this means that:

$$\tilde{U}(a^*; [\tilde{\mu}, \ldots, \tilde{\mu}]) < 0 = \tilde{U}(\perp; [\tilde{\mu}, \ldots, \tilde{\mu}])$$

because of uniform-at-random tiebreaking. This is a contradiction since a^* needs to be a best response.

Subcase 2b: $0 < \tilde{c}(m_{\mathbf{min}}) < 1$**.** First, we claim that $\mathbb{P}[m = m'] = 0$ for all $m' \in [m_{\min}, \infty)$. If $\mathbb{P}[m = m'] > 0$, then because of uniform-at-random tiebreaking, it holds that $m' + \varepsilon$ for some ε performs strictly better than m' for some $\varepsilon > 0$.

Next, we claim that $\mathbb{P}_{\tilde{\mu}}[\perp] = \tilde{c}(m_{\min})^{1/(P-1)}$. First, assume for sake of contradiction that $\mathbb{P}_{\tilde{\mu}}[\perp] > \tilde{c}(m_{\min})^{1/(P-1)}$: in this case,

$$U(m_{\min}; [\tilde{\mu}, \ldots, \tilde{\mu}]) = \alpha^{P-1} - \tilde{c}(m_{\min}) > 0 = U(\perp; [\tilde{\mu}, \ldots, \tilde{\mu}])$$

which is a contradiction. Next, assume for sake of contradiction that $\mathbb{P}_{\tilde{\mu}}[\perp] < \tilde{c}(m_{\min})^{1/(P-1)}$. Since $\tilde{c}(m_{\min})^{1/(P-1)} < 1$ by assumption, this means that that $\mathbb{P}_{\tilde{\mu}}[\perp] < 1$. Let $m' = \min\{m \geq m_{\min} \mid m \in \text{supp}(\mu)\}$. Using that $\mathbb{P}_{\tilde{\mu}}[m = m_{\min}] = 0$, it holds that:

$$\tilde{U}(m'; [\tilde{\mu}, \ldots, \tilde{\mu}]) = \alpha^{P-1} - \tilde{c}(m') \leq \alpha^{P-1} - \tilde{c}(m_{\min}) < 0 = U(\perp; [\tilde{\mu}, \ldots, \tilde{\mu}]),$$

which is a contradiction.

For $m \in \text{supp}(\tilde{\mu})$, we claim that $\mathbb{P}_{M \sim \tilde{\mu}}[M \leq m] = (\tilde{c}(m))^{1/(P-1)}$. This is because $\perp \in \text{supp}(\tilde{\mu})$ and the utility at $\perp$ is 0, so the utility at any $m \in \text{supp}(\tilde{\mu})$ must be 0. This implies that $\mathbb{P}_{M \sim \tilde{\mu}}[M \leq m] = (\tilde{c}(m))^{1/(P-1)}$.

Finally, it suffices to show that the support is a closed interval of the form $[m_{\min}, m^*]$ where m^* is the unique value such that $\tilde{c}(m^*) = 1$. We first see that since $\mathbb{P}_{M \sim \tilde{\mu}}[M \leq m] \leq 1$, it must hold that $m \leq m^*$ for any $m \in \text{supp}(\tilde{\mu})$. To see this, let $S = \text{supp}(\tilde{\mu}) \cup [m^*, m_{\max}]$.

Since Q is a finite union of closed sets, it is a closed set. This means that $\bar{Q} = [m_{\min}, m_{\max}] \setminus Q$ is an open set. It suffices to prove that $\bar{Q} = \emptyset$. Assume for sake of contradiction $\bar{Q} \neq \emptyset$. If $m' \in \bar{Q}$, let $m_1 = \inf \{m < m' \mid m \in \text{supp}(\tilde{\mu})\}$ and $m_2 = \sup \{m > m' \mid m \in \text{supp}(\tilde{\mu})\}$. Since $\bar{Q}$ is open, there is an open neighborhood such that $B_\varepsilon(m') \subseteq \bar{Q}$, which means that $m_2 > m_1$. However, this means that $\mathbb{P}_{M\sim\tilde{\mu}}[M = m_2] - = \mathbb{P}_{M\sim\tilde{\mu}}[M \leq m_2] - \mathbb{P}_{M\leq\tilde{\mu}}[M = m_2] > 0$, which is a contradiction.

□

F.4.3 Useful setup and lemmas for heterogeneous users

In our analysis, it is cleaner to work in the reparametrized space

$$S := \left\{(M^{\mathrm{E}}([w_{\text{costly}}, w_{\text{cheap}}]) - s, t) \mid t \in \mathcal{T}, u([w_{\text{costly}}, w_{\text{cheap}}], t) = 0\right\}$$

than directly over the content space $\mathbb{R}_{\geq 0}$. (This is the same reparameterized space described in Chapter 9.6.4.) Recall that we map each $(v, t) \in S$ to the unique content $h(v, t) \in \mathbb{R}^2_{\geq 0}$ of the form $h(v, t) = [f_t(w_{\text{cheap}}), w_{\text{cheap}}]$ such that $M^{\mathrm{E}}([f_t(w_{\text{cheap}}), w_{\text{cheap}}]) = v - s$. Conceptually, $h(v, t)$ captures content with engagement $v - s$ optimized for winning type t.

Using that the coefficients a_t are strictly decreasing as given by Assumption 5, it is easy to see that h is an one-to-one function mapping S to

$$\cup_{t\in\mathcal{T}}\mathcal{C}_t = \{[w_{\text{costly}}, w_{\text{cheap}}] \mid t \in \mathcal{T}, u([w_{\text{costly}}, w_{\text{cheap}}], t) = 0\}.$$

We let h^{-1} denote its inverse which is defined on the image $\cup_{t\in\mathcal{T}}\mathcal{C}_t$.

We show that if the support of μ is contained in $\cup_{t\in\mathcal{T}}\mathcal{C}_t$, then there exists a best response w in $\mathcal{C}_t$.

Lemma 210. *Suppose that gaming is costless (i.e., $\nabla(c(w))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$), and suppose that $u([0, 0], t) \geq 0$ for all $t \in \mathcal{T}$. Let μ be a distribution supported on $\cup_{t\in\mathcal{T}}\mathcal{C}_t$. Then, in the game with $M = M^E$, there exists a best response $w \in \mathbb{R}^2_{\geq 0}$ to:*

$$\arg\max_{w\in\mathbb{R}^2_{\geq 0}} \mathbb{E}_{\boldsymbol{w}_{-i}\sim\mu^{P-1}}[U_i(w; \boldsymbol{w}_{-i})]$$

such that $w \in \cup_{t\in\mathcal{T}}\mathcal{C}_t$.

Proof. Let w be any best-response. By Lemma 208, we know that $w \in \cup_{t\in\mathcal{T}}\mathcal{C}_t^{\text{aug}} \cup \{w \mid c(w) = 0\}$. We split into two cases: (1) $u([w_{\text{costly}}, w_{\text{cheap}}], t) < 0$ for all $t \in \mathcal{T}$, and (2) $u([w_{\text{costly}}, w_{\text{cheap}}], t) \geq 0$ for some $t \in \mathcal{T}$.

Case 1: $u([w_{\textbf{costly}}, w_{\textbf{cheap}}], t) < 0$ **for all** $t \in \mathcal{T}$**.** In this case, no users will consume the content w. Letting $t_{\min} = \min(\mathcal{T})$ and w'_{cheap} be such that $u([0, w'_{\text{cheap}}], t_{\min}) = 0$, since gaming is costless, it holds that:

$$\mathbb{E}_{\mathbf{w}_{-i}\sim(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^{P-1}}[U_i(w; \mathbf{w}_{-i})] \leq \mathbb{E}_{\mathbf{w}_{-i}\sim(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^{P-1}}[U_i([0, w'_{\text{cheap}}]; \mathbf{w}_{-i})].$$

This means that there exists a best-response in $\cup_{t \in \mathcal{T}} \mathcal{C}_t$.

Case 2: $u([w_{\textbf{costly}}, w_{\textbf{cheap}}], t) \geq 0$ **for some** $t \in \mathcal{T}$**.** For the second case, where $u([w_{\text{costly}}, w_{\text{cheap}}], t) \geq 0$ for some $t \in \mathcal{T}$, let:

$$w'_{\text{cheap}} = \inf \left\{ w''_{\text{cheap}} \mid u([w_{\text{costly}}, w''_{\text{cheap}}], t) = 0 \text{ for some } t \in \mathcal{T}, w''_{\text{cheap}} \geq w_{\text{cheap}} \right\}.$$

We see that by construction, it holds that $\{t \in \mathcal{T} \mid u(w, t) \geq 0\} = \left\{t \in \mathcal{T} \mid u([w_{\text{costly}}, w'_{\text{cheap}}]) \geq 0\right\}$ and it also holds that $M^{\text{E}}(w) \leq M^{\text{E}}([w_{\text{costly}}, w'_{\text{cheap}}])$. Since gaming is costless, we see that

$$\mathbb{E}_{\mathbf{w}_{-i} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^{P-1}}[U_i(w; \mathbf{w}_{-i})] \leq \mathbb{E}_{\mathbf{w}_{-i} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^{P-1}}[U_i([w_{\text{costly}}, w'_{\text{cheap}}]; \mathbf{w}_{-i})].$$

This means that there exists a best-response in $\cup_{t \in \mathcal{T}} \mathcal{C}_t$.

□

We translate the costs into the reparameterized space.

Lemma 211. *Let $P = 2$, suppose that gaming tricks are costless (that is, $(\nabla(c(w)))_2 = 0$ for all $w \in \mathbb{R}^2_{\geq 0}$), and suppose that $u([0, 0], t) \geq 0$ for all $t \in \mathcal{T}$. Suppose that Assumption 5 holds. Then for any $w \in \cup_{t \in \mathcal{T}} \mathcal{C}_t$, it holds that:*

$$c(w) = \max(0, a_t \cdot v - 1),$$

where $(v, t) = h^{-1}(w)$.

Proof. We first claim that

$$c(w) = C^E_t(M^{\text{E}}(w)).$$

To see this, observe that by Lemma 202, there exists $w' \in \mathcal{C}^{\text{aug}}_t$ such that $M^{\text{E}}(w') \geq M^{\text{E}}(w)$ and $C^E_t(M^{\text{E}}(w)) = c(w') \leq c(w)$. By Lemma 201, this implies that $w' = w$, so $c(w) = C^E_t(M^{\text{E}}(w))$ as desired.

Next, we observe that by Assumption 5, it holds that:

$$C^E_t(M^{\text{E}}(w)) = \max(0, a_t \cdot v - 1).$$

Putting this all together gives the desired statement.

□

F.4.4 Proof of Theorem 70 and Proposition 69

We prove Theorem 70 and then deduce Proposition 69 as a corollary.

We first apply the reparametrization from Chapter F.4.3 to the random variable (V, T) to Definition 7. Using the specification in in Definition 7, it is easy to see that every $w \in \text{supp}(\mu^{\text{e}}(P, c, u, \mathcal{T}))$ is $\mathcal{C}_t$, which means that the distribution $h^{-1}(W)$ for $W \sim \mu^{\text{e}}(P, c, u, \mathcal{T})$ is well-defined. In the following proposition, we simplify the distribution (V, T).

Proposition 212. *Consider the setup of Definition 7. Let $(V,T) = h^{-1}(W)$ where $W \sim \mu^e(P,c,u,\mathcal{T})$. Then it holds that $\mathbb{P}[T = t_i]$ satisfies:*

$$\mathbb{P}[T = t_i] = \begin{cases} \frac{1}{N-i+1} & \text{if } 1 \leq i \leq N'-1 \\ 1 - \sum_{i'=1}^{N'-1} \frac{1}{N-i'+1} & \text{if } i = N' \\ 0 & \text{if } i > N'. \end{cases}$$

Moreover, the conditional distribution $V \mid T = t_i$ is distributed as a uniform distribution over $\left[\frac{1}{a_{t_i}}, \frac{1}{a_{t_i}} \cdot \left(1 + \frac{1}{N}\right)\right]$ for $1 \leq i \leq N'-1$ and a uniform distribution over

$$\left[\frac{1}{a_{t_{N'}}}, \frac{1}{a_{t_{N'}}} \cdot \left(1 + \frac{N-N'+1}{N} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)\right)\right]$$

for $i = N'$.

Proof. It suffices to show that for each $1 \leq i \leq N'$, it holds that $V \mid T = t_i$ is distributed according to the uniform distribution specified in the proposition statement.

The remainder of the analysis boils down to analyzing the distribution $V \mid T = t_i$. Let the distributions $W^i = (W^i_{\text{cheap}}, W^i_{\text{costly}})$ for $1 \leq i \leq N'$, and the mixture weights α^i be defined as in Definition 7. Observe that $V \mid T = t_i$ is distributed as $(M^{\mathrm{E}}([W^i_{\text{costly}}, W^i_{\text{cheap}}] + s, t_i)$. Note that every $w \in \text{supp}(W^i_{\text{costly}}, W^i_{\text{cheap}})$ is in the image $\cup_{t \in \mathcal{T}} \mathcal{C}_t$, which means that it can be uniquely expressed as $w = h(v,t)$. To translate the costs, we observe that:

$$\begin{aligned} C_t(w_{\text{cheap}}) &= c([f_{t_i}(w_{\text{cheap}}), w_{\text{cheap}}]) \\ &=_{(A)} \max(0, a_{t_i} \cdot v - 1) \\ &=_{(B)} a_{t_i} \cdot v - 1 \end{aligned}$$

where (A) follows from Lemma 211, and (B) follows the fact that $w_{\text{cheap}} \in \text{supp}(W^i_{\text{cheap}})$ implies that $C_{t_i}(w_{\text{cheap}}) > 0$. Putting this all together, we see that $\mathbb{P}[V \leq v \mid T = t_i] = \mathbb{P}[W^i_{\text{cheap}} \leq w_{\text{cheap}}]$ for any $w_{\text{cheap}} \in \text{supp}(W^i_{\text{cheap}})$. Using the specification in Definition 7, this means that if $1 \leq i \leq N'-1$, then it holds that

$$\begin{aligned} \mathbb{P}[V \leq v \mid T = t_i] &= \mathbb{P}[W^i_{\text{cheap}} \leq w_{\text{cheap}}] \\ &= \min\left(N \cdot C_{t_i}(w_{\text{cheap}}), 1\right) \\ &= \min\left(N \cdot (a_{t_i} \cdot v - 1), 1\right), \end{aligned}$$

which implies that $V \mid T = t_i$ is a uniform distribution over $\left[\frac{1}{a_{t_i}}, \frac{1}{a_{t_i}} \cdot \left(1 + \frac{1}{N}\right)\right]$ as desired.

Similarly, if $i = N'$, then

$$\begin{aligned}\mathbb{P}[V \leq v \mid T = t_i] &= \mathbb{P}[W^i_{\text{cheap}} \leq w_{\text{cheap}}] \\ &= \min\left(\frac{N}{N-N'+1} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)^{-1} \cdot C_{t_i}(w_{\text{cheap}}), 1\right) \\ &= \min\left(\frac{N}{N-N'+1} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)^{-1} \cdot (a_{t_i} \cdot v - 1), 1\right),\end{aligned}$$

which implies that $V \mid T = t_i$ is a uniform distribution over

$$\left[\frac{1}{a_{t_{N'}}}, \frac{1}{a_{t_{N'}}} \cdot \left(1 + \frac{N-N'+1}{N} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)\right)\right]$$

as desired. □

We prove Theorem 70. The main ingredient of the proof is the following lemma which computes the expected utility for different choices of $(v,t) \in S$ when the other creator's actions are selected according to the distribution $(V,T) = h^{-1}(W)$ for $W \sim \mu^{\mathrm{e}}(P, c, u, \mathcal{T})$.

Lemma 213. *Consider the setup of Definition 7. Let $(V,T) = h^{-1}(W)$ be the reparameterized distribution of $W \sim \mu^{\mathrm{e}}(P, c, u, \mathcal{T})$. For any $(v,t) \in S$, the expected utility satisfies:*

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] \leq \frac{N-N'+1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N-j+1}.$$

Moreover, for $(v,t) \in supp((V,T))$, it holds that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] = \frac{N-N'+1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N-j+1}.$$

Proof of Lemma 213. We first simplify the expected utility.

$$\begin{aligned}&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\ &= \frac{1}{|\mathcal{T}|} \sum_{t' \in \mathcal{T}} (\mathbb{P}_{V,T}[v \geq V] \cdot \mathbb{1}[t \leq t'] + \mathbb{P}_{V,T}[V > v, T > t'] \cdot \mathbb{1}[t \leq t']) - c(h^{-1}(v,t)) \\ &=_{(A)} \frac{|\{t' \in \mathcal{T} \mid t' \geq t\}|}{|\mathcal{T}|} \cdot \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{|\mathcal{T}|} \sum_{t' \in \mathcal{T} \mid t' \geq t} \mathbb{P}_{V,T}[V > v, T > t'].\end{aligned}$$

where (A) uses Lemma 211.

To analyze this expression, we apply Proposition 212 to obtain the reparameterized joint distribution (V,T) of $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$. By the well-separatedness assumption on the type space, we see that the support of the distributions $\mathrm{supp}(V \mid T \mid t = i)$ are disjoint and that $\max(\mathrm{supp}(V)) = \frac{1}{a_{t_{N'}}} \cdot \left(1 + \frac{N-N'+1}{N} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)\right)$.

We split the remainder of the analysis into several cases: (1) $t \leq t_{N'-1}$ and $\frac{1}{a_t} \leq v < \frac{1}{a_t} \cdot (1 + \frac{1}{N})$, (2) $t \leq t_{N'-1}$, $v \geq \frac{1}{a_t} \cdot (1 + \frac{1}{N})$, and $v \leq \max(\mathrm{supp}(V))$, (3) $t = t_{N'}$ and $\frac{1}{a_{t_{N'}}} \leq v \leq \max(\mathrm{supp}(V))$, (4) $t \leq N'$ and $v \geq \max(\mathrm{supp}(V))$, and and $v \leq \max(\mathrm{supp}(V))$, and (5) $t > t_{N'}$.

Case 1: $t \leq t_{N'-1}$ and $\frac{1}{a_t} \leq v < \frac{1}{a_t} \cdot (1 + \frac{1}{N})$. We know that $t = t_i$ for some $1 \leq i \leq N' - 1$. This means that for $(v', t') \in \mathrm{supp}((V, T))$, if $t' > t$, then $v' > v$; moreover, if $t' < t$, then $v' < v$. This implies that

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))] \\
&= \frac{|\{t' \in \mathcal{T} \mid t' \geq t_i\}|}{|\mathcal{T}|} \cdot \mathbb{P}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{|\mathcal{T}|} \sum_{t' \in \mathcal{T} | t' \geq t_i} \mathbb{P}[V > v, T > t'] \\
&= \underbrace{\frac{N-i+1}{N} \cdot \mathbb{P}[V \leq v, T = t_i] - \max(0, a_t \cdot v - 1)}_{(A)} + \underbrace{\frac{N-i+1}{N} \cdot \mathbb{P}[T < t_i] + \frac{1}{N} \sum_{t' \in \mathcal{T} | t' \geq t_i} \mathbb{P}[T > t']}_{(B)}.
\end{aligned}$$

For term (A), we apply Proposition 212 to see that:

$$\begin{aligned}
&\frac{N-i+1}{N} \cdot \mathbb{P}[V \leq v, T = t_i] - \max(0, a_t \cdot v - 1) \\
&= \frac{N-i+1}{N} \cdot \mathbb{P}[V \leq v \mid T = t_i] \cdot \mathbb{P}[T = t_i] - \max(0, a_t \cdot v - 1) \\
&= \frac{(N-i+1) \cdot N \cdot (a_{t_i} \cdot v - 1)}{N \cdot (N-i+1)} - (a_t \cdot v - 1) \\
&= 0.
\end{aligned}$$

For term (B), we apply Proposition 212 to see that:

$$
\begin{aligned}
&\frac{N-i+1}{N} \cdot \mathbb{P}[T < t_i] + \frac{1}{|\mathcal{T}|} \sum_{t' \in \mathcal{T} | t' \geq t_i} \mathbb{P}[T > t'] \\
&= \frac{N-i+1}{N} \cdot \mathbb{P}[T < t_i] + \frac{1}{N} \sum_{t' \in \mathcal{T} | t' \geq t_i} (1 - \mathbb{P}[T \leq t']) \\
&= \frac{N-i+1}{N} \cdot \sum_{j=1}^{i-1} \frac{1}{N-j+1} + \frac{1}{N} \sum_{j=i}^{N'-1} \left(1 - \sum_{j'=1}^{j} \frac{1}{N-j'+1}\right) \\
&= \frac{1}{N} \left(\sum_{j=1}^{i-1} \frac{N-i+1}{N-j+1} - \sum_{j'=1}^{i-1} \frac{N'-i}{N-j'+1} + N' - i - \sum_{j'=i}^{N'-1} \frac{N'-j'}{N-j'+1} \right) \\
&= \frac{1}{N} \left(\sum_{j=1}^{i-1} \frac{N-N'+1}{N-j+1} + \sum_{j'=i}^{N'-1} \left(1 - \frac{N'-j'}{N-j'+1}\right) \right) \\
&= \frac{1}{N} \left(\sum_{j=1}^{i-1} \frac{N-N'+1}{N-j+1} + \sum_{j'=i}^{N'-1} \frac{N-N'+1}{N-j'+1} \right) \\
&= \frac{N-N'+1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N-j+1}.
\end{aligned}
$$

Altogether, this proves that:

$$
\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] = \frac{N-N'+1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N-j+1}
$$

as desired.

Case 2: $t \leq t_{N'-1}$**,** $v \geq \frac{1}{a_t} \cdot (1 + \frac{1}{N})$**, and** $v \leq \max(\mathbf{supp}(V))$**.** We know that $t = t_i$ for some $1 \leq i \leq N' - 1$. Let $i^* \in [i, N']$ be the *maximum* value such that $v > \frac{1}{a_{t_{i^*}}} \cdot (1 + \frac{1}{N})$. (We immediately see that $i^* \neq N'$, because $v \leq \max(\text{supp}(V)) \leq \frac{1}{a_{t_{N'}}} \cdot (1 + \frac{1}{N})$.) This means that for $(v', t') \in \text{supp}((V,T))$, if $t' > t_{i^*}$, then $v' > v$; moreover, if $t' < t_{i^*}$, then $v' < v$. This implies that

$$
\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))] \\
&= \frac{|\{t' \in \mathcal{T} \mid t' \geq t_i\}|}{|\mathcal{T}|} \cdot \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{|\mathcal{T}|} \sum_{t' \in \mathcal{T} | t' \geq t_i} \mathbb{P}_{V,T}[V > v, T > t'] \\
&= \underbrace{\frac{N-i+1}{N} \cdot \left(\mathbb{P}[V \leq v, T = t_{i^*+1}] + \sum_{j=i}^{i^*} \mathbb{P}[T = t_j] \right) - \frac{1}{N} \sum_{t' \in \mathcal{T} | t' \geq t_i} \mathbb{P}[T > t', V \leq v] - \max(0, a_t \cdot v - 1)}_{(A)} \\
&\quad \underbrace{+ \frac{N-i+1}{N} \cdot \mathbb{P}[T < t_i] + \frac{1}{N} \sum_{t' \in \mathcal{T} | t' \geq t_i} \mathbb{P}[T > t']}_{(B)}.
\end{aligned}
$$

For term (A), we apply Proposition 212 to see that:

$$
\begin{aligned}
&\frac{N-i+1}{N} \cdot \left(\mathbb{P}[V \leq v, T = t_{i^*+1}] + \sum_{j=i}^{i^*} \mathbb{P}[T = t_j] \right) - \frac{1}{N} \sum_{t' \in \mathcal{T} | t' \geq t_i} \mathbb{P}[T > t', V \leq v] - \max(0, a_t \cdot v - 1) \\
&= \frac{N-i+1}{N} \cdot \left(\mathbb{P}[V \leq v, T = t_{i^*+1}] + \sum_{j=i}^{i^*} \mathbb{P}[T = t_j] \right) \\
&- \frac{1}{N} \sum_{t' \in \mathcal{T} | t' \geq t_i} \left(\mathbb{P}[V \leq v, T = t_{i^*+1}, t_{i^*+1} > t'] + \sum_{j=i}^{i^*} \mathbb{P}[T = t_j, t_j > t'] \right) - (a_t \cdot v - 1) \\
&= \frac{N-i+1}{N} \cdot \left(\mathbb{P}[V \leq v, T = t_{i^*+1}] + \sum_{j=i}^{i^*} \mathbb{P}[T = t_j] \right) \\
&- \frac{i^*-i+1}{N} \cdot \mathbb{P}[V \leq v, T = t_{i^*+1}] - \sum_{j=i}^{i^*} \frac{j-i}{N} \cdot \mathbb{P}[T = t_j] - (a_t \cdot v - 1) \\
&= \frac{N-i^*}{N} \cdot \mathbb{P}[V \leq v, T = t_{i^*+1}] + \sum_{j=i}^{i^*} \frac{N-j+1}{N} \cdot \mathbb{P}[T = t_j] - (a_t \cdot v - 1) \\
&\leq \frac{N-i^*}{N} \cdot \mathbb{P}[T = t_{i^*+1}] \cdot \mathbb{P}[V \leq v | T = t_{i^*+1}] + \frac{i^*-i+1}{N} - (a_t \cdot v - 1) \\
&\leq_{(1)} (a_{t_{i^*+1}} \cdot v - 1) + \frac{i^*-i+1}{N} - (a_t \cdot v - 1) \\
&= -(a_t - a_{t_{i^*+1}}) \cdot v + \frac{i^*-i+1}{N} \\
&\leq -(a_t - a_{t_{i^*+1}}) \cdot \frac{1}{a_{t_{i^*}}} \cdot (1 + \frac{1}{N}) + \frac{i^*-i+1}{N} \\
&\leq -\frac{a_{t_i}}{a_{t_{i^*}}} \cdot (1 + \frac{1}{N}) + \frac{a_{t_{i^*+1}}}{a_{t_{i^*}}} \cdot (1 + \frac{1}{N}) + \frac{i^*-i+1}{N} \\
&\leq -\left(1 + \frac{1}{N}\right)^{i^*-i+1} + 1 + \frac{i^*-i+1}{N} \\
&\leq_{(2)} 0
\end{aligned}
$$

where (1) uses the fact that

$$\mathbb{P}[V \leq v \mid T = t_{i^*+1}] \leq \min\left(N \cdot \frac{1}{(N-i^*)\alpha^{i^*+1}} \cdot (a_{t_{i^*+1}} \cdot v - 1), 1\right)$$
$$\leq N \cdot \frac{1}{(N-i^*)\alpha^{i^*+1}} \cdot (a_{t_{i^*+1}} \cdot v - 1),$$

where (2) uses the fact that for $m \geq 1$ and $x \geq 0$, it holds that $(1+x)^m \geq 1 + m \cdot x$.

For term (B), we apply the same argument as in Case 1 to see that it is equal to $\frac{N-N'+1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N-j+1}$.

Altogether, this proves that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))] \leq \frac{N-N'+1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N-j+1}$$

as desired.

Case 3: $t = t_{N'}$ and $\frac{1}{a_{t_{N'}}} \leq v \leq \max(\mathbf{supp}(V))$. For $(v',t') \in \operatorname{supp}((V,T))$, if $t' < t$, then $v' < v$. Moreover, we see that $\max(\operatorname{supp}(V)) = \frac{1}{a_{t_{N'}}} \cdot \left(1 + \frac{N-N'+1}{N} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)\right)$. This implies that

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))] \\
&= \frac{|\{t' \in \mathcal{T} \mid t' \geq t_{N'}\}|}{|\mathcal{T}|} \cdot \mathbb{P}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{|\mathcal{T}|} \sum_{t' \in \mathcal{T} | t' \geq t_{N'}} \mathbb{P}[V > v, T > t'] \\
&= \frac{|\{t' \in \mathcal{T} \mid t' \geq t_{N'}\}|}{|\mathcal{T}|} \cdot \mathbb{P}[V \leq v] - \max(0, a_t \cdot v - 1) \\
&= \underbrace{\frac{N-N'+1}{N} \cdot \mathbb{P}[V \leq v, T = t_{N'}] - \max(0, a_t \cdot v - 1)}_{(A)} + \underbrace{\frac{N-N'+1}{N} \cdot \mathbb{P}[T < t_{N'}]}_{(B)}.
\end{aligned}$$

For term (A), we apply Proposition 212 to see that:

$$\begin{aligned}
&\frac{N-N'+1}{N} \cdot \mathbb{P}[V \leq v, T = t_{N'}] - \max(0, a_t \cdot v - 1) \\
&= \frac{N-N'+1}{N} \cdot \mathbb{P}[V \leq v \mid T = t_{N'}] \cdot \mathbb{P}[T = t_{N'}] \\
&= \frac{N-N'+1}{N} \cdot \frac{N}{N-N'+1} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right)^{-1} \cdot (a_{t_i} \cdot v - 1) \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N-j+1}\right) - (a_t \cdot v - 1) \\
&= 0.
\end{aligned}$$

For term (B), we apply Proposition 212 to see that:

$$\frac{N-N'+1}{N}\cdot\mathbb{P}[T<t_{N'}]=\frac{N-N'+1}{N}\cdot\sum_{j=1}^{N'-1}\mathbb{P}[T=t_j]=\frac{N-N'+1}{N}\cdot\sum_{j=1}^{N'-1}\frac{1}{N-j+1}.$$

Altogether, this proves that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))]=\frac{N-N'+1}{N}\cdot\sum_{j=1}^{N'-1}\frac{1}{N-j+1}$$

as desired.

Case 4: $t\le N'$ and $v\ge\max(\mathbf{supp}(V))$. We know that $t=t_i$ for some $1\le i\le N$. For $(v',t')\in\operatorname{supp}((V,T))$, we see that $v'<v$. This implies that

$$\begin{aligned}&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))]\\&=\frac{|\{t'\in\mathcal{T}\mid t'\ge t_i\}|}{|\mathcal{T}|}\cdot\mathbb{P}[V\le v]-\max(0,a_t\cdot v-1)+\frac{1}{|\mathcal{T}|}\sum_{t'\in\mathcal{T}|t'\ge t_{N'}}\mathbb{P}[V>v,T>t']\\&=\frac{N-i+1}{N}-(a_{t_i}\cdot v-1).\end{aligned}$$

This expression is upper bounded by the case where $v=\max(\operatorname{supp}(V))$. If $1\le i\le N'-1$, we can thus apply Case 2; if $i=N'$, we can apply Case 3. Altogether, this proves that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))]\le\frac{N-N'+1}{N}\cdot\sum_{j=1}^{N'-1}\frac{1}{N-j+1}$$

as desired.

Case 5: $t>t_{N'}$. In this case, $(v,t)\in S$ satisfies

$$v\ge\frac{1}{a_{N'+1}}\ge\frac{1}{a_{t_{N'}}}\cdot\left(1+\frac{1}{N}\right)\ge\frac{1}{a_{t_{N'}}}\cdot\left(1+\frac{N-N'+1}{N}\cdot\left(1-\sum_{j=1}^{N'-1}\frac{1}{N-j+1}\right)\right)=\max(\operatorname{supp}(V)).$$

This means that

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);(V,T))]=\frac{|\{t'\in\mathcal{T}\mid t'\ge t\}|}{|\mathcal{T}|}-\max(0,a_t\cdot v-1)\le\frac{|\{t'\in\mathcal{T}\mid t'\ge t\}|}{|\mathcal{T}|}\le\frac{N-N'}{N}.$$

It suffices to show that $\frac{N-N'}{N}\le\frac{N-N'+1}{N}\cdot\sum_{j=1}^{N'-1}\frac{1}{N-j+1}$, which we can be written as:

$$\frac{N-N'}{N-N'+1}=1-\frac{1}{N-N'+1}\le\sum_{j=1}^{N'-1}\frac{1}{N-j+1},$$

we know is true by the definition of N'.

$\square$

Using Lemma 213, we prove Theorem 70.

Proof of Theorem 70. We first claim that we can work over the reparameterized space (V, T) defined in Chapter F.4.3. As described above, every $w \in \text{supp}(\mu^{\text{e}}(P, c, u, \mathcal{T}))$ is in the image $\cup_{t \in \mathcal{T}} \mathcal{C}_t$, which means that it is associated with a unique value $h^{-1}(w) = (v, t) \in S$. It thus suffices to show that there exists a best response $w \in \mathbb{R}^2_{\geq 0}$ to:

$$\arg\max_{w \in \mathbb{R}^2_{\geq 0}} \mathbb{E}_{\mathbf{w}_{-i} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^{P-1}}[U_i(w; \mathbf{w}_{-i})]$$

that is also in the image $\cup_{t \in \mathcal{T}} \mathcal{C}_t$; this follows from Lemma 210.

We thus work over the reparameterized space for the remainder of the analysis. By Lemma 213, we see that for any $(v, t) \in S$, the expected utility satisfies:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] \leq \frac{N - N' + 1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N - j + 1}.$$

Moreover, for $(v, t) \in \text{supp}((V, T))$, it holds that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] = \frac{N - N' + 1}{N} \cdot \sum_{j=1}^{N'-1} \frac{1}{N - j + 1}.$$

This proves that (V, T) is an equilibrium in the reparameterized space S.

Putting this all together, this proves that $\mu^{\text{e}}(P, c, u, \mathcal{T})$ is an equilibrium in the original space $\mathbb{R}^2_{\geq 0}$. $\square$

We prove Chapter 69.

Proof of Chapter 69. We apply Chapter 70. It suffices to show that $N' = 2$. To see this, since $N = 2$, we see that $\sum_{i=1}^{m} \frac{1}{N-i+1}$ is equal to $\frac{1}{2}$ if $m = 1$ and $\sum_{i=1}^{m} \frac{1}{N-i+1}$ is equal to $1 + \frac{1}{2} \geq 1$ if $m = 2$. This means that $N' = 2$ as desired. $\square$

F.4.5 Proof of Theorem 71

Before diving into the proof of Chapter 71, we first verify that the distributions in Chapter 8 are indeed well-defined: in particular, it suffices to verify that the ordering of v values at the boundary points indeed proceeds in the order shown in Chapter 9.1b and that the total density of g is 1. We split into the three cases in Chapter 8.

Case 1: $a_{t_1}/a_{t_2} \geq 1.5$. We first show that:

$$\frac{1}{a_{t_1}} \leq_{(1)} \frac{3}{2 \cdot a_{t_1}} \leq_{(2)} \frac{1}{a_{t_2}} \leq_{(3)} \frac{5}{4 \cdot a_{t_2}}.$$

Inequalities (1) and (3) are trivial, and inequality (2) follows from the fact that $a_{t_1}/a_{t_2} \geq 1.5$.

We next observe that:

$$\begin{aligned}\int_v g(v)dv &= a_{t_1}\left(\frac{3}{2\cdot a_{t_1}} - \frac{1}{a_{t_1}}\right) + 2a_{t_2}\cdot\left(\frac{5}{4\cdot a_{t_2}} - \frac{1}{a_{t_2}}\right)\\ &= 0.5 + 0.5\\ &= 1.\end{aligned}$$

Case 2: $(5-\sqrt{5})/2 \le a_{t_1}/a_{t_2} \le 1.5$**.** We first show that:

$$\frac{1}{a_{t_1}} \le_{(1)} \frac{1}{a_{t_2}} \le_{(1)} \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} \le_{(3)} \frac{1}{a_{t_2}}\cdot\left(2 - \frac{a_{t_1}}{2\cdot a_{t_2}}\right).$$

Inequality (1) follows from the fact that $a_{t_1} > a_{t_2}$. Inequality (2) follows from the fact that $2\left(\frac{a_{t_1}}{a_{t_2}} - 1\right) \le 1$. Inequality (3) can be rewritten as:

$$2\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)\left(2 - \frac{a_{t_1}}{2\cdot a_{t_2}}\right) \ge 1,$$

which follows from the fact that $(5-\sqrt{5})/2 \le a_{t_1}/a_{t_2} \le 1.5$.

We next observe that:

$$\begin{aligned}\int_v g(v)dv &= a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + 2a_{t_2}\cdot\left(\frac{1}{a_{t_2}}\cdot\left(2 - \frac{a_{t_1}}{2\cdot a_{t_2}}\right) - \frac{1}{a_{t_2}}\right)\\ &= \frac{a_{t_1}}{a_{t_2}} - 1 + 4 - \frac{a_{t_1}}{\cdot a_{t_2}} - 2\\ &= 1.\end{aligned}$$

Case 3: $1 \le a_{t_1}/a_{t_2} \le (5-\sqrt{5})/2$**.** We first show that:

$$\frac{1}{a_{t_1}} \le_{(1)} \frac{1}{a_{t_2}} \le_{(2)} \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} \le_{(3)} \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right)$$

Inequality (1) follows from the fact that $a_{t_1} > a_{t_2}$. Inequality (2) can be written as:

$$2\left(2 - \frac{a_{t_1}}{a_{t_2}}\right) \le 3 - \frac{a_{t_1}}{a_{t_2}},$$

which can be written as:

$$1 \le \frac{a_{t_1}}{a_{t_2}},$$

which holds because $a_{t_1} > a_{t_2}$. Inequality (3) can be written as:

$$\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right) \le \frac{1}{a_{t_1}},$$

which can be rewritten as:

$$\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)\left(3 - \frac{a_{t_1}}{a_{t_2}}\right) \leq 2 - \frac{a_{t_1}}{a_{t_2}},$$

which follows from the fact that $1 \leq a_{t_1}/a_{t_2} \leq (5 - \sqrt{5})/2$.

We next observe that:

$$\begin{aligned}
&\int_v g(v)dv \\
&= a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + 2a_{t_2} \cdot \left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} - \frac{1}{a_{t_2}}\right) \\
&+ a_{t_1} \cdot \left(\frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right) - \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}\right) \\
&= \frac{a_{t_1}}{a_{t_2}} - 1 + \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} - 2 + 1 + \left(1 - \frac{a_{t_1}}{2a_{t_2}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right) - \frac{a_{t_1}}{2a_{t_2}} \cdot \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} \\
&= \frac{a_{t_1}}{a_{t_2}} - 2 + \left(2 - \frac{a_{t_1}}{a_{t_2}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right) \\
&= \frac{a_{t_1}}{a_{t_2}} - 2 + 3 - \frac{a_{t_1}}{a_{t_2}} \\
&= 1.
\end{aligned}$$

Now, we turn to proving Chapter 71. Like in the proof of Chapter 70, the main ingredient is computing the expected utility for different choices of $(v, t) \in S$ when the other creator's actions are selected according to the distribution $(V, T) = h^{-1}(W)$ for $W \sim \mu^e(P, c, u, \mathcal{T})$. Since we have already analyzed Case 1 ($a_{t_1}/a_{t_2} \geq 1.5$) as a consequence of Chapter 70, we can focus on Cases 2 and 3. We analyze the expected utility separately for Case 2 and Case 3 in the following lemmas.

Lemma 214. *Consider the setup of Definition 8, and and assume that $(5-\sqrt{5})/2 \leq a_{t_1}/a_{t_2} \leq 1.5$. Let $(V, T) = h^{-1}(W)$ be the reparameterized distribution of $W \sim \mu^e(P, c, u, \mathcal{T})$. For any $(v, t) \in S$, the expected utility satisfies:*

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v, t); (V, T))] \leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Moreover, for $(v, t) \in supp((V, T))$, it holds that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v, t); (V, T))] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Proof of Lemma 214. We first simplify the expected utility.

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\
&= \mathbb{P}_{V,T}[v \geq V] \cdot \mathbb{1}[t = t_1] + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \cdot \mathbb{1}[t = t_1] + \frac{1}{2} \cdot \mathbb{P}_{V,T}[v \geq V] \cdot \mathbb{1}[t = t_2] - c(h^{-1}(v,t)) \\
&=_{(1)} \left(\mathbb{1}[t = t_1] + \mathbb{1}[t = t_2] \cdot \frac{1}{2}\right) \cdot \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \cdot \mathbb{1}[t = t_1].
\end{aligned}$$

where (1) uses Lemma 211.

We split the remainder of the analysis into several cases: (A) $v \leq \frac{1}{a_{t_2}}$ and $t = t_1$, (B), $\frac{1}{a_{t_2}} \leq v \leq \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)}$ and $t = t_1$, (C) $\frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} \leq v \leq \frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right)$ and $t = t_1$, (D) $\frac{1}{a_{t_2}} \leq v \leq \frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right)$ and $t = t_2$, and (E) $v \geq \frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right)$.

Case A: $v \leq \frac{1}{a_{t_2}}$ and $t = t_1$. We observe that:

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\
&= \left(\mathbb{1}[t = t_1] + \mathbb{1}[t = t_2] \cdot \frac{1}{2}\right) \cdot \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \cdot \mathbb{1}[t = t_1] \\
&= \underbrace{\mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1)}_{(1)} + \underbrace{\frac{1}{2} \cdot \mathbb{P}_{V,T}[T = t_2]}_{(2)}.
\end{aligned}$$

For term (1), we observe that:

$$\mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1) = a_t\left(v - \frac{1}{a_t}\right) - (a_t \cdot v - 1) = 0.$$

For term (2), we apply Lemma 205 to see that

$$\frac{1}{2} \cdot \mathbb{P}[T = t_2] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}$$

as desired.

Putting this all together, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Case B: $\frac{1}{a_{t_2}} \leq v \leq \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}}-1\right)}$ and $t = t_1$. We observe that:

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);h^{-1}(V,T)]\\
&= \left(\mathbb{1}[t=t_1] + \mathbb{1}[t=t_2]\cdot\frac{1}{2}\right)\cdot \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t\cdot v - 1) + \frac{1}{2}\cdot \mathbb{P}_{V,T}[V > v, T = t_2]\cdot \mathbb{1}[t=t_1]\\
&= \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t\cdot v - 1) + \frac{1}{2}\cdot \mathbb{P}_{V,T}[V > v, T = t_2]\\
&= \mathbb{P}_{V,T}[V \leq v] - \frac{1}{2}\cdot \mathbb{P}_{V,T}[V \leq v, T = t_2] - \max(0, a_t\cdot v - 1) + \frac{1}{2}\cdot \mathbb{P}_{V,T}[T = t_2]\\
&= \underbrace{\mathbb{P}_{V,T}[V \leq v] - \frac{1}{2}\cdot \mathbb{P}_{V,T}[V \leq v, T = t_2] - \max(0, a_t\cdot v - 1)}_{(1)} + \underbrace{\frac{1}{2}\cdot \mathbb{P}_{V,T}[T = t_2]}_{(2)}.
\end{aligned}$$

For term (1), we observe that:

$$\begin{aligned}
&\mathbb{P}_{V,T}[V \leq v] - \frac{1}{2}\cdot \mathbb{P}_{V,T}[V \leq v, T = t_2] - \max(0, a_t\cdot v - 1)\\
&= \mathbb{P}_{V,T}[V \leq 1/a_{t_2}] + \mathbb{P}_{V,T}[1/a_{t_2} \leq V \leq v]\left(1 - \frac{1}{2}\cdot \mathbb{P}_{V,T}[T = t_2 \mid 1/a_{t_2} \leq V \leq v]\right) - (a_t\cdot v - 1)\\
&= a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + 2a_{t_2}\cdot\left(v - \frac{1}{a_{t_2}}\right)\cdot\left(1 - \frac{1}{2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)\right) - (a_{t_1}\cdot v - 1)\\
&= a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + a_{t_1}\cdot\left(v - \frac{1}{a_{t_2}}\right) - (a_{t_1}\cdot v - 1)\\
&= 0.
\end{aligned}$$

For term (2), we apply Lemma 205 to see that

$$\frac{1}{2}\cdot \mathbb{P}[T = t_2] = \frac{a_{t_1}}{2\cdot a_{t_2}} - \frac{1}{2}$$

as desired.

Putting this all together, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);h^{-1}(V,T)] = \frac{a_{t_1}}{2\cdot a_{t_2}} - \frac{1}{2}.$$

Case C: $\frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}}-1\right)} \leq v \leq \frac{1}{a_{t_2}}\cdot\left(2 - \frac{a_{t_1}}{2\cdot a_{t_2}}\right)$ and $t = t_1$. We use the same analysis as in Case B to see that:

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t);h^{-1}(V,T)]\\
&= \underbrace{\mathbb{P}_{V,T}[V \leq v] - \frac{1}{2}\cdot \mathbb{P}_{V,T}[V \leq v, T = t_2] - \max(0, a_t\cdot v - 1)}_{(1)} + \underbrace{\frac{1}{2}\cdot \mathbb{P}_{V,T}[T = t_2]}_{(2)}.
\end{aligned}$$

For term (1), we observe that:

$$\begin{aligned}
&\mathbb{P}_{V,T}[V \leq v] - \frac{1}{2} \cdot \mathbb{P}_{V,T}[V \leq v, T = t_2] - \max(0, a_t \cdot v - 1) \\
&= \mathbb{P}_{V,T}[V \leq 1/a_{t_2}] \\
&+ \mathbb{P}_{V,T}\left[1/a_{t_2} \leq V \leq \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)}\right]\left(1 - \frac{1}{2} \cdot \mathbb{P}_{V,T}\left[T = t_2 \mid 1/a_{t_2} \leq V \leq \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)}\right]\right) \\
&+ \mathbb{P}_{V,T}\left[\frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} \leq V \leq v\right]\left(1 - \frac{1}{2}\right) - (a_t \cdot v - 1) \\
&= a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + a_{t_1} \cdot \left(\frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)} - \frac{1}{a_{t_2}}\right) + 2a_{t_2} \cdot \frac{1}{2}\left(v - \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)}\right) - (a_{t_1} \cdot v - 1) \\
&= -1 \cdot (a_{t_1} - a_{t_2}) \cdot \left(v - \frac{1}{2a_{t_2}\left(\frac{a_{t_1}}{a_{t_2}} - 1\right)}\right) \\
&\leq 0.
\end{aligned}$$

For term (2), we apply Lemma 205 to see that

$$\frac{1}{2} \cdot \mathbb{P}[T = t_2] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}$$

as desired.

Putting this all together, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v, t); h^{-1}(V, T)] \leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Case D: $\frac{1}{a_{t_2}} \leq v \leq \frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right)$ **and** $t = t_2$**.** We observe that:

$$\begin{aligned}
\mathbb{E}_{V,T}[U_1(h^{-1}(v, t); h^{-1}(V, T)] &= \frac{1}{2} \cdot \mathbb{P}_{V,T}[V \leq v] - (a_{t_2} \cdot v - 1) \\
&= \frac{1}{2} \cdot a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + \frac{1}{2} \cdot 2a_{t_2} \cdot \left(v - \frac{1}{a_{t_2}}\right) - (a_{t_2} \cdot v - 1) \\
&= \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2} + a_{t_2} \cdot \left(v - \frac{1}{a_{t_2}}\right) - (a_{t_2} \cdot v - 1) \\
&= \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2},
\end{aligned}$$

as desired.

Case E: $v \geq \frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right)$**.** We observe that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v, t); h^{-1}(V, T)] = \left(\mathbb{1}[t = t_1] + \frac{1}{2} \cdot \mathbb{1}[t = t_2]\right) - \max(0, a_t \cdot v - 1).$$

For each value of t, since $\frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right) = \max(\text{supp}(V))$, we see that the above expression is upper bounded by the case of $v = \frac{1}{a_{t_2}} \cdot \left(2 - \frac{a_{t_1}}{2 \cdot a_{t_2}}\right)$. If $t = t_1$, we can apply the analysis from Case C; if $t = t_2$, we can apply the analysis from Case E. Altogether, this means that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}$$

as desired.

□

Lemma 215. *Consider the setup of Definition 8, and and assume that* $1 \leq a_{t_1}/a_{t_2} \leq (5 - \sqrt{5})/2$. *Let* $(V,T) = h^{-1}(W)$ *be the reparameterized distribution of* $W \sim \mu^e(P, c, u, \mathcal{T})$. *For any* $(v,t) \in S$, *the expected utility satisfies:*

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] \leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Moreover, for $(v,t) \in$ *supp*$((V,T))$, *it holds that:*

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Proof of Lemma 215. We first simplify the expected utility.

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\
&= \mathbb{P}_{V,T}[v \geq V] \cdot \mathbb{1}[t = t_1] + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \cdot \mathbb{1}[t = t_1] + \frac{1}{2} \cdot \mathbb{P}_{V,T}[v \geq V] \cdot \mathbb{1}[t = t_2] - c(h^{-1}(v,t)) \\
&=_{(1)} \left(\mathbb{1}[t = t_1] + \mathbb{1}[t = t_2] \cdot \frac{1}{2}\right) \cdot \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \cdot \mathbb{1}[t = t_1]
\end{aligned}$$

where (1) uses Lemma 211.

We split the remainder of the analysis into several cases: (A) $v \leq \frac{1}{a_{t_2}}$ and $t = t_1$, (B), $\frac{1}{a_{t_2}} \leq v \leq \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$ and $t = t_1$, (C) $\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} \leq v \leq \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right)$ and $t = t_1$, (D) $\frac{1}{a_{t_2}} \leq v \leq \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$ and $t = t_2$, (E) $\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} \leq v \leq \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right)$ and $t = t_2$, and (F) $v \geq \max(\text{supp}(V))$.

Case A: $v \leq \frac{1}{a_{t_2}}$ **and** $t = t_1$**.** We observe that:

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\
&= \left(\mathbb{1}[t = t_1] + \mathbb{1}[t = t_2] \cdot \frac{1}{2}\right) \cdot \mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1) + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \cdot \mathbb{1}[t = t_1] \\
&= \underbrace{\mathbb{P}_{V,T}[V \leq v] - \max(0, a_t \cdot v - 1)}_{(1)} + \underbrace{\frac{1}{2} \cdot \mathbb{P}_{V,T}[T = t_2]}_{(2)}
\end{aligned}$$

For term (1), we see that:

$$\mathbb{P}_{V,T}[V \le v] - \max(0, a_t \cdot v - 1) = a_{t_1}\left(v - \frac{1}{a_{t_1}}\right) - (a_{t_1} \cdot v - 1) = 0.$$

For term (2), we apply Lemma 205 to see that it is equal to $\frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}$. Putting this all together, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Case B: $\frac{1}{a_{t_2}} \le v \le \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$ **and** $t = t_1$**.** We observe that:

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\
&= \left(\mathbb{1}[t = t_1] + \mathbb{1}[t = t_2] \cdot \frac{1}{2}\right) \cdot \mathbb{P}_{V,T}[V \le v] - \max(0, a_t \cdot v - 1) + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \cdot \mathbb{1}[t = t_1] \\
&= \mathbb{P}_{V,T}[V \le v] - \max(0, a_t \cdot v - 1) + \frac{1}{2} \cdot \mathbb{P}_{V,T}[V > v, T = t_2] \\
&= \underbrace{\mathbb{P}_{V,T}[V \le v] - \frac{1}{2} \cdot \mathbb{P}_{V,T}[V \le v, T = t_2] - \max(0, a_t \cdot v - 1)}_{(1)} + \underbrace{\frac{1}{2} \cdot \mathbb{P}_{V,T}[T = t_2]}_{(2)}.
\end{aligned}$$

For term (1), we see that:

$$\begin{aligned}
&\mathbb{P}_{V,T}[V \le v] - \frac{1}{2} \cdot \mathbb{P}_{V,T}[V \le v, T = t_2] - \max(0, a_t \cdot v - 1) \\
&= a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + 2a_{t_2}\left(v - \frac{1}{a_{t_2}}\right) - \frac{1}{2} \cdot 2 \cdot a_{t_2}\left(v - \frac{1}{a_{t_2}}\right) \cdot \left(2 - \frac{a_{t_1}}{a_{t_2}}\right) - (a_{t_1} \cdot v - 1) \\
&= a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + 2a_{t_2}\left(v - \frac{1}{a_{t_2}}\right)\left(1 - \frac{1}{2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)\right) - (a_{t_1} \cdot v - 1) \\
&= a_{t_1}\left(v - \frac{1}{a_{t_1}}\right) - (a_{t_1} \cdot v - 1) \\
&= 0.
\end{aligned}$$

For term (2), we apply Lemma 205 to see that it is equal to $\frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}$. Putting this all together, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Case C: $\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} \le v \le \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2 - \frac{a_{t_1}}{a_{t_2}}}\right)$ **and** $t = t_1$**.** We use the same argument as in Case B to see that:

$$\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\
&= \underbrace{\mathbb{P}_{V,T}[V \le v] - \frac{1}{2} \cdot \mathbb{P}_{V,T}[V \le v, T = t_2] - \max(0, a_t \cdot v - 1)}_{(1)} + \underbrace{\frac{1}{2} \cdot \mathbb{P}_{V,T}[T = t_2]}_{(2)}.
\end{aligned}$$

For term (1), we see that:

$$\begin{aligned}
&\mathbb{P}_{V,T}[V \le v] - \frac{1}{2} \cdot \mathbb{P}_{V,T}[V \le v, T = t_2] - \max(0, a_t \cdot v - 1) \\
&= a_{t_1} \cdot \mathbb{P}\left[\frac{1}{a_{t_1}} \le V \le \frac{1}{a_{t_2}}\right] + 2a_{t_2} \cdot \mathbb{P}\left[\frac{1}{a_{t_2}} \le V \le \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}\right]\left(1 - \frac{1}{2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)\right) \\
&+ a_{t_1} \cdot \mathbb{P}\left[\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} \le V \le v\right] - (a_{t_1} \cdot v - 1) \\
&= a_{t_1} \cdot \left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + a_{t_1} \cdot \left(\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} - \frac{1}{a_{t_2}}\right) + a_{t_1}\left(v - \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}\right) - (a_{t_1} \cdot v - 1) \\
&= a_{t_1}\left(v - \frac{1}{a_{t_1}}\right) - (a_{t_1} \cdot v - 1) \\
&= 0.
\end{aligned}$$

For term (2), we apply Lemma 205 to see that it is equal to $\frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}$. Putting this all together, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T))] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Case D: $\frac{1}{a_{t_2}} \le v \le \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$ **and** $t = t_2$**.** We observe that:

$$\begin{aligned}
\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T))] &= \frac{1}{2}\mathbb{P}[V \le v] - (a_{t_2} \cdot v - 1) \\
&= \frac{1}{2} \cdot a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + \frac{1}{2} \cdot 2a_{t_2}\left(v - \frac{1}{a_{t_2}}\right) - (a_{t_2} \cdot v - 1) \\
&= \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2} + a_{t_2} \cdot \left(v - \frac{1}{a_{t_2}}\right) - (a_{t_2} \cdot v - 1) \\
&= \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.
\end{aligned}$$

Case E: $\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)} \leq v \leq \frac{1}{a_{t_1}} + \left(\frac{1}{a_{t_1}} - \frac{1}{2a_{t_2}}\right)\left(\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2-\frac{a_{t_1}}{a_{t_2}}}\right)$ **and** $t = t_2$**.** We observe that:

$$
\begin{aligned}
&\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \\
&= \frac{1}{2}\mathbb{P}[V \leq v] - (a_{t_2} \cdot v - 1) \\
&= \frac{1}{2} \cdot a_{t_1}\left(\frac{1}{a_{t_2}} - \frac{1}{a_{t_1}}\right) + \frac{1}{2} \cdot 2a_{t_2}\left(\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)} - \frac{1}{a_{t_2}}\right) + \frac{1}{2} \cdot a_{t_1} \cdot \left(v - \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right) - (a_{t_2} \cdot v - 1) \\
&= \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2} - a_{t_2}\left(v - \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right) + \frac{1}{2} \cdot a_{t_1} \cdot \left(v - \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right) \\
&= \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2} + \left(\frac{a_{t_1}}{2} - a_{t_2}\right)\left(v - \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right) \\
&\leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.
\end{aligned}
$$

Case F: $v \geq \max(\mathbf{supp}(V))$**.** In this case, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] = \left(\mathbb{1}[t = t_1] + \frac{1}{2} \cdot \mathbb{1}[t = t_2]\right) - (a_t \cdot v - 1).$$

For each value of t, this expression is maximized at $v = \max(\text{supp}(V))$. If $t = t_1$, we can thus apply Case C; if $t = t_2$, we can thus apply Case E. Putting this all together, we see that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); h^{-1}(V,T)] \leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

□

We now prove Chapter 71 using Lemma 214 and Lemma 215.

Proof of Chapter 71. For Case 1 (where $a_{t_1}/a_{t_2} \geq 1.5$), we directly obtain the result from the analysis for N well-separated types. In particular, we can apply Theorem 70 (or Proposition 69). Applying Proposition 212, we see that the reparameterization of the distribution specified in Definition 8 is identical to the distribution in Definition 6, which yields the desired statement.

The remainder of the proof boils down to analyzing Case 2 ($(5-\sqrt{5})/2 \leq a_{t_1}/a_{t_2} \leq 1.5$)and Case 3 ($1 < a_{t_1}/a_{t_2} \leq (5-\sqrt{5})/2$).

We first claim that we can work over the reparameterized space (V, T) defined in Chapter F.4.3. This follows from an analogous argument to the proof of Chapter 70 which we repeat for completeness. Note that every $w \in \text{supp}(\mu^{\text{e}}(P, c, u, \mathcal{T}))$ is in the image $\cup_{t \in \mathcal{T}} \mathcal{C}_t$, which means that it is associated with a unique value $h^{-1}(w) = (v, t) \in S$. It thus suffices to show that there exists a best response $w \in \mathbb{R}^2_{\geq 0}$ to:

$$\arg\max_{w \in \mathbb{R}^2_{\geq 0}} \mathbb{E}_{\mathbf{w}_{-i} \sim (\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^{P-1}}[U_i(w; \mathbf{w}_{-i})]$$

that is also in the image $\cup_{t \in \mathcal{T}} \mathcal{C}_t$; this follows from Lemma 210.

We thus work over the reparameterized space for the remainder of the analysis. For Case 2, by Lemma 214, we see that for any $(v, t) \in S$, the expected utility satisfies:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] \leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Moreover, for $(v, t) \in \mathrm{supp}((V, T))$, it holds that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

This proves that (V, T) is an equilibrium in the reparameterized space S. Similarly, for Case 3, by Lemma 215, we see that for any $(v, t) \in S$, the expected utility satisfies:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] \leq \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

Moreover, for $(v, t) \in \mathrm{supp}((V, T))$, it holds that:

$$\mathbb{E}_{V,T}[U_1(h^{-1}(v,t); (V,T))] = \frac{a_{t_1}}{2 \cdot a_{t_2}} - \frac{1}{2}.$$

This proves that (V, T) is an equilibrium in the reparameterized space S.

Putting this all together, this proves that $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ is an equilibrium in the original space $\mathbb{R}^2_{\geq 0}$. □

F.5 Proofs for Chapter 9.5

F.5.1 Proof of Theorem 65

We prove Theorem 65.

Proof of Theorem 65. Let $\mu = \mu^{\mathrm{i}}(P, c, u, \mathcal{T})$, and let $(W_{\mathrm{costly}}, W_{\mathrm{cheap}}) \sim \mu$. We analyze the expected utility of

$$H(w) = \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w; w_{-i})]$$

of a content creator if all of the creators choose the strategy μ. We show that $H(w) = 0$ if $w \in \mathrm{supp}(\mu)$ and $H(w) \leq 0$ for any $w \in \mathbb{R}^2_{\geq 0}$.

Let $w \in \mathbb{R}^2_{\geq 0}$ be any content vector, and let $w' = [w_{\text{costly}}, 0]$ be the vector with identical quality but no gaming tricks. Since

$$U_b^I(w_{\text{costly}}, t) = u(w', t) \geq u(w, t),$$

$M^{\text{I}}(w) = M^{\text{I}}(w')$, and $c(w) \geq c(w')$, it holds that $H(w) \leq H(w')$. Since all $w'' \in \text{supp}(\mu)$ also satisfy $w''_{\text{cheap}} = 0$, we can restrict the rest of our analysis to w such that $w_{\text{cheap}} = 0$.

We split the remainder of the analysis into two cases: (1) $\mathcal{T} = \{t\}$ and (2) $\beta_t = 0$ for all $t \in \mathcal{T}$.

Case 1: $\mathcal{T} = \{t\}$. It suffices to show that $H(w) = 0$ if $w \in \text{supp}(\mu)$ and $H(w) \leq 0$ for any $w = [w_{\text{costly}}, 0] \in \mathbb{R}_{\geq 0} \times \{0\}$.

First, we show that $H(w) \leq 0$ for any $w \in \mathbb{R}^2_{\geq 0}$ such that $w_{\text{cheap}} = 0$. If $w_{\text{costly}} < \beta_t$, then it follows immediately that $H(w) \leq 0$. If $w_{\text{costly}} \geq \beta_t$, then we see that:

$$\begin{aligned}
H(w) &= \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w'; w_{-i})] \\
&= \left(\min\left(1, C_b^I(w_{\text{costly}})\right)\right) \cdot \mathbb{1}[U_b^I(w_{\text{costly}}, t) \geq 0] - C_b^I(w_{\text{costly}}) \\
&\leq C_b^I(w_{\text{costly}}) - C_b^I(w_{\text{costly}}) \\
&= 0.
\end{aligned}$$

Next, we show that if $w \in \text{supp}(\mu)$, then it holds that $H(w) = 0$. If $w_{\text{costly}} = 0$, then it follows easily that $H(w) = 0$. Otherwise, we see that:

$$\begin{aligned}
H(w) &= \left(\min\left(1, C_b^I(w_{\text{costly}})\right)\right) \cdot \mathbb{1}[U_b^I(w_{\text{costly}}, t) \geq 0] - C_b^I(w_{\text{costly}}) \\
&= C_b^I(w_{\text{costly}}) \cdot \mathbb{1}[U_b^I(w_{\text{costly}}, t) \geq 0] - C_b^I(w_{\text{costly}}) \\
&= 0.
\end{aligned}$$

This proves that μ is an equilibrium as desired.

Case 2: $\beta_t = 0$ for all $t \in \mathcal{T}$. It suffices to show that $H(w) = 0$ if $w \in \text{supp}(\mu)$ and $H(w) \leq 0$ for any $w = [w_{\text{costly}}, 0] \in \mathbb{R}_{\geq 0} \times \{0\}$.

First, we show that $H(w) \leq 0$ for any $w \in \mathbb{R}^2_{\geq 0}$ such that $w_{\text{cheap}} = 0$. Then we see that:

$$\begin{aligned}
H(w) &= \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w'; w_{-i})] \\
&= \min\left(1, C_b^I(w_{\text{costly}})\right) - C_b^I(w_{\text{costly}}) \\
&\leq C_b^I(w_{\text{costly}}) - C_b^I(w_{\text{costly}}) \\
&= 0.
\end{aligned}$$

Next, we show that if $w \in \text{supp}(\mu)$, then it holds that $H(w) = 0$. We see that:

$$\begin{aligned}
H(w) &= C_b^I(w_{\text{costly}}) - C_b^I(w_{\text{costly}}) \\
&= 0.
\end{aligned}$$

This proves that μ is an equilibrium as desired.

$\square$

F.5.2 Proof of Theorem 66

We prove Theorem 66.

Proof. Let $\mu = \mu^{\mathrm{r}}(P, c, u, \mathcal{T})$ for notational convenience, and let $(W_{\mathrm{costly}}, W_{\mathrm{cheap}}) \sim \mu$. We analyze the expected utility of

$$H(w) = \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w; w_{-i})]$$

of a content creator if all of the creators choose the strategy μ. We show that $w \in \arg\max_{w'} H(w')$ for any $w \in \mathrm{supp}(\mu)$.

Let $w \in \mathbb{R}^2_{\geq 0}$ be any content vector, and let $w' = [w_{\mathrm{costly}}, 0]$ be the vector with identical quality but no gaming tricks. Since

$$U^I_b(w_{\mathrm{costly}}, t) = u(w', t) \geq u(w, t),$$

$M^{\mathrm{R}}(w) = M^{\mathrm{R}}(w')$, and $c(w) \geq c(w')$, it holds that $H(w) \leq H(w')$. Since all $w'' \in \mathrm{supp}(\mu)$ also satisfy $w''_{\mathrm{cheap}} = 0$, we can restrict the rest of our analysis to w such that $w_{\mathrm{cheap}} = 0$.

We split the remainder of the analysis into two cases: (1) $\mathcal{T} = \{t\}$ and (2) $\beta_t = 0$ for all $t \in \mathcal{T}$.

Case 1: $\mathcal{T} = \{t\}$. We split into two subcases: $\kappa \leq 1/P$ and $\kappa \in (1/P, 1]$.

If $\kappa \leq 1/P$, then W_{costly} is a point mass at β_t. Note that:

$$H(w) = \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w; w_{-i})] = \frac{\mathbb{1}[U^I_b(w_{\mathrm{costly}}, t) \geq 0]}{P} - C^I_b(w) \leq \frac{1}{P} - \kappa.$$

Moreover, for $w_{\mathrm{costly}} = w^*_{\mathrm{costly}}$, it holds that $H(w) = \frac{1}{P} - \kappa$. This proves that $w \in \arg\max_{w'} H(w')$ for any $w \in \mathrm{supp}(\mu)$, as desired.

If $\kappa \in (1/P, 1]$, then we see that ν is the unique value such $\sum_{i=1}^{P-1} \nu^i = P \cdot \kappa$. Note that W_{costly} is β_t with probability $1 - \nu$ and 0 with probability ν. Moreover, note that:

$$H(w) = \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w; w_{-i})] = \mathbb{1}[U^I_b(w_{\mathrm{costly}}, t) \geq 0] \cdot \mathbb{E}_Y \left[\frac{1}{1+Y}\right] - C^I_b(w),$$

where $Y \sim \mathrm{Bin}(P - 1, 1 - \nu)$ is distributed as a binomial random variable with success probability $1 - \nu$. (The second equality holds because Y is distributed as the number of creators $j \neq i$ who choose content generating nonnegative utility for the user.) A simple calculation shows that:

$$\mathbb{E}_Y \left[\frac{1}{1+Y}\right] = \frac{1}{P} \sum_{i=0}^{P-1} \nu^i = \kappa,$$

where the last equality follows from the definition of η. This means that $H(w) \leq 0$ for all w. For $w_{\mathrm{costly}} = \beta_t$ and $w_{\mathrm{costly}} = 0$, it holds that $H(w) = 0$. This means that $w \in \arg\max_{w'} H(w')$ for any $w \in \mathrm{supp}(\mu)$, as desired.

Case 2: $\beta_t = 0$ **for all** $t \in \mathcal{T}$**.** In this case, W_{costly} is a point mass at 0. Note that:

$$H(w) = \mathbb{E}_{\mathbf{w}_{-i} \sim \mu_{-i}}[U_i(w; w_{-i})] = \frac{\mathbb{1}[U_b^I(w_{\text{costly}}, t) \geq 0]}{P} - C_b^I(w) \leq \frac{1}{P}.$$

Moreover, for $w_{\text{costly}} = 0$, it holds that $H(w) = \frac{1}{P}$. This proves that $w \in \arg\max_{w'} H(w')$ for any $w \in \text{supp}(\mu)$, as desired. □

F.6 Proofs for Chapter 9.4.1

F.6.1 Proofs of Theorem 57 and Theorem 58

The main lemma is the following characterization of user consumption of utility as the maximum investment in quality across the content landscape. We slightly abuse notation and for a content landscape $\mathbf{w} = (w_1, \ldots, w_P)$, we use the notation $w \in \mathbf{w}$ to denote w_j for $j \in [P]$.

Lemma 216. *Consider the setup of Theorem 57. For* $\boldsymbol{w} \in supp(\mu^i(P, c, u, \mathcal{T})^P)$*, it holds that*

$$UCQ(M^I; \boldsymbol{w}) = \max_{w \in \boldsymbol{w}} w_{costly}$$

and for $\boldsymbol{w} \in supp(\mu^i(P, c, u, \mathcal{T})^P)$*, it holds that*

$$UCQ(M^E; \boldsymbol{w}) = \max_{w \in \boldsymbol{w}} w_{costly}.$$

We now prove Lemma 216.

Proof of Lemma 216. We observe that for $w \in \text{supp}(\mu^{\text{i}}(P, c, u, \mathcal{T}))$, it holds that if $\mathbb{1}[u(w, t)] < 0$, then $w = [0, 0]$. Thus, for $\mathbf{w} \in \text{supp}(\mu^{\text{i}}(P, c, u, \mathcal{T}))^P$, it holds that:

$$\text{UCQ}(M^{\text{I}}; \mathbf{w}) = \mathbb{E}\left[w_{i^*(M^{\text{I}}; \mathbf{w})}^{\text{costly}} \cdot \mathbb{1}[u(w_{i^*(M; \mathbf{w})}, t) \geq 0]\right] = \mathbb{E}\left[w_{i^*(M^{\text{I}}; \mathbf{w})}^{\text{costly}}\right].$$

Moreover, since $w_{\text{cheap}} = 0$ for all $w \in \text{supp}(\mu^{\text{i}}(P, c, u, \mathcal{T}))$ and by the definition of M^{I}, we see that $w_{i^*(M^{\text{I}}; \mathbf{w})}^{\text{costly}} = \max_{w \in \mathbf{w}} w_{\text{costly}}$. This means that:

$$\text{UCQ}(M^{\text{I}}; \mathbf{w}) = \mathbb{E}\left[\max_{w \in \mathbf{w}} w_{\text{costly}}\right].$$

Similarly, we observe that for $w \in \text{supp}(\mu^{\text{e}}(P, c, u, \mathcal{T}))$, it holds that if $\mathbb{1}[u(w, t)] < 0$, then $w = [0, 0]$. Thus, for $\mathbf{w} \in \text{supp}(\mu^{\text{e}}(P, c, u, \mathcal{T}))^P$, it holds that:

$$\text{UCQ}(M^{\text{E}}; \mathbf{w}) = \mathbb{E}\left[w_{i^*(M; \mathbf{w})}^{\text{costly}} \cdot \mathbb{1}[u(w_{i^*(M; \mathbf{w})}, t)] \geq 0\right] = \mathbb{E}\left[w_{i^*(M; \mathbf{w})}^{\text{costly}}\right].$$

Moreover, by the definition of $\text{supp}(\mu^{\text{e}}(P, c, u, \mathcal{T}))$ and by the definition of M^{E}, we see that $w^{\text{costly}}_{i^*(M^{\text{E}};\mathbf{w})} = \max_{w \in \mathbf{w}} w_{\text{costly}}$. This means that:

$$\text{UCQ}(M^{\text{E}}; \mathbf{w}) = \mathbb{E}\left[\max_{w \in \mathbf{w}} w_{\text{costly}}\right].$$

□

Using Lemma 216, we prove Theorem 57 and Theorem 58.

Proof of Theorem 57 and Theorem 58. By Lemma 216, it suffices to analyze

$$\mathbb{E}_{\mathbf{w} \sim \mu^P}\left[\max_{w \in \mathbf{w}} w_{\text{costly}}\right]$$

where $\mu \in \left\{\mu^{\text{e}}(P, c, u, \mathcal{T}), \mu^{\text{i}}(P, c, u, \mathcal{T})\right\}$. By Chapter 203, we see that $\mathbb{P}_{(W_{\text{costly}}, W_{\text{cheap}}) \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))}[W_{\text{costly}} \le w_{\text{costly}}]$ is equal to:

$$\begin{cases} (-\alpha)^{1/(P-1)} & \text{if } 0 \le w_{\text{costly}} \le -\alpha \\ (\min(1, w_{\text{costly}} + \gamma \cdot t \cdot (w_{\text{costly}} + \alpha)))^{1/(P-1)} & \text{if } w_{\text{costly}} \ge \max(0, -\alpha). \end{cases}$$

By Chapter 204, we see that $\mathbb{P}_{(W_{\text{costly}}, W_{\text{cheap}}) \sim (\mu^{\text{i}}(P,c,u,\mathcal{T}))}[W_{\text{costly}} \le w_{\text{costly}}]$ equals:

$$\begin{cases} (-\alpha)^{1/(P-1)} & \text{if } 0 \le w_{\text{costly}} \le -\alpha \\ (\min(1, w_{\text{costly}}))^{1/(P-1)} & \text{if } w_{\text{costly}} \ge \max(0, -\alpha). \end{cases}$$

Proof of Theorem 57. The marginal distribution of W_{costly} for $\mu^{\text{e}}(P, c, u, \mathcal{T})$ implies for engagement-based optimization, the distribution of W_{costly} for higher values of γ is stochastically dominated by the distribution of W_{costly} for lower values of γ. This implies that $\mathbb{E}_{\mathbf{w} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^P}[\max_{w \in \mathbf{w}} w_{\text{costly}}]$ is strictly increasing in γ, which implies that $\mathbb{E}_{\mathbf{w} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^P}\left[\text{UCQ}(M^{\text{E}}; \mathbf{w})\right]$ is strictly increasing in γ.

Proof of Theorem 58. Observe that the marginal distribution of W_{costly} for $\mu^{\text{i}}(P, c, u, \mathcal{T})$ stochastically dominates the distribution of W_{costly} for $\mu^{\text{e}}(P, c, u, \mathcal{T})$, with strict stochastic dominance for $\gamma > 0$. This implies that if $\gamma > 0$:

$$\mathbb{E}_{\mathbf{w} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^P}\left[\max_{w \in \mathbf{w}} w_{\text{costly}}\right] < \mathbb{E}_{\mathbf{w} \sim (\mu^{\text{i}}_{P,\alpha,t})^P}\left[\max_{w \in \mathbf{w}} w_{\text{costly}}\right],$$

which implies that

$$\mathbb{E}_{\mathbf{w} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^P}\left[\text{UCQ}(M^{\text{E}}; \mathbf{w})\right] < \mathbb{E}_{\mathbf{w} \sim (\mu^{\text{i}}_{P,\alpha,t})^P}\left[\text{UCQ}(M^{\text{I}}; \mathbf{w})\right].$$

Moreover, if $\gamma = 0$, the two distributions are equal, which implies that

$$\mathbb{E}_{\mathbf{w} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^P}\left[\max_{w \in \mathbf{w}} w_{\text{costly}}\right] = \mathbb{E}_{\mathbf{w} \sim (\mu^{\text{i}}_{P,\alpha,t})^P}\left[\max_{w \in \mathbf{w}} w_{\text{costly}}\right],$$

which implies that

$$\mathbb{E}_{\mathbf{w} \sim (\mu^{\text{e}}(P,c,u,\mathcal{T}))^P}\left[\text{UCQ}(M^{\text{E}}; \mathbf{w})\right] = \mathbb{E}_{\mathbf{w} \sim (\mu^{\text{i}}_{P,\alpha,t})^P}\left[\text{UCQ}(M^{\text{I}}; \mathbf{w})\right].$$

□

F.6.2 Proof of Proposition 59

We prove Chapter 59.

Proof of Chapter 59. We construct the following instantation of Example 4 with $N \geq 2$ types. Let the type space be $\mathcal{T}_{N,\varepsilon} = \{(1+\varepsilon)(1+1/N)^{i-1} - 1 \mid 1 \leq i \leq N\}$, and let $P = 2$, $\alpha = 1$, and $\gamma = 0$. It suffices to show that $\mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2}[\mathrm{UCQ}(M^{\mathrm{I}}; \mathbf{w})] = 2/3$ and $\mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2} \leq 1/N$.

Before analyzing these two expressions, we first compute f_t and C_t for this example: we observe that $f_t(w_{\text{cheap}}) = w_{\text{cheap}}/t$ and $C_t(w_{\text{cheap}}) = w_{\text{cheap}}/t$.

First, we show that $\mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2}[\mathrm{UCQ}(M^{\mathrm{I}}; \mathbf{w})] = 2/3$. We use the characterization in Theorem 65. It is easy to see that:

$$\mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2}[\mathrm{UCQ}(M^{\mathrm{I}}; \mathbf{w})] = \int_0^\infty (1 - \mathbb{P}[W_{\text{costly}} \leq w_{\text{costly}}]^2) dw_{\text{costly}}.$$

We see that:

$$\mathbb{P}[W_{\text{costly}} \leq w_{\text{costly}}] = \mathbb{P}[W_{\text{cheap}} \leq t \cdot w_{\text{costly}}] = \min(1, C_t(t \cdot w_{\text{costly}})) = \min(1, w_{\text{costly}}).$$

This means that:

$$\mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2}[\mathrm{UCQ}(M^{\mathrm{I}}; \mathbf{w})] = \int_0^1 (1 - w_{\text{costly}}^2) dw_{\text{costly}} = \frac{2}{3},$$

as desired.

Next, we show that $\mathbb{E}_{\mathbf{w} \sim \left(\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2} \leq 1/N$. We use the characterization in Theorem 70. We observe that $\max(\mathrm{supp}(W^i_{\text{cheap}})) \leq \frac{1}{N t_i}$. This follows from the fact that $\max(\mathrm{supp}(W^i_{\text{cheap}})) = \frac{t_i}{N}$ for $1 \leq i \leq N' - 1$ and

$$\begin{aligned}
\max(\mathrm{supp}(W^i_{\text{cheap}})) &= t_i \cdot \frac{N}{N - N' + 1} \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N - j + 1}\right)^{-1} \\
&\leq t_i \cdot \frac{N - N' + 1}{N} \cdot \left(1 - \sum_{j=1}^{N'-1} \frac{1}{N - j + 1}\right) \\
&\leq t_i \cdot \frac{N - N' + 1}{N} \cdot \frac{1}{N - N' + 1} \\
&= \frac{t_i}{N}
\end{aligned}$$

for $i = N'$. This implies that:

$$\max_{1 \leq i \leq N'} \max(\mathrm{supp}(W^i_{\text{costly}})) = \max_{1 \leq i \leq N'} t \cdot \frac{1}{N t_i} = \frac{1}{N}.$$

Thus, we see that:

$$\mathbb{E}_{\mathbf{w}\sim\left(\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})\right)^2}[\mathrm{UCQ}(M^{\mathrm{I}};\mathbf{w})] \leq \max_{w\in\mathrm{supp}(\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon}))} w_{\mathrm{costly}} \leq \frac{1}{N}$$

as desired. □

F.7 Proofs for Chapter 9.4.2

F.7.1 Proof of Theorem 60

The main ingredient of the proof of Theorem 60 is constructing and analyzing an instance where engagement-based optimization achieves a low realized engagement. We construct the instance to be Example 4 with costless gaming ($\gamma = 0$), baseline utility $\alpha = 1$, and $P = 2$ creators, and type space $\mathcal{T}_{N,\varepsilon} := \{(1+\varepsilon)(1+1/N)^{i-1} - 1 \mid 1 \leq i \leq N\}$ for sufficiently small $\varepsilon > 0$ and sufficiently large N.

In order to prove Theorem 60, we first show the following lemma that relates the realized engagement to the maximum engagement achieved by any content in the content landscape. We again slightly abuse notation and for a content landscape $\mathbf{w} = (w_1, \ldots, w_P)$, we use the notation $w \in \mathbf{w}$ to denote w_j for $j \in [P]$.

Lemma 217. *Consider Example 4 with costless gaming ($\gamma = 0$), baseline utility $\alpha = 1$, and $P = 2$ creators, and type space $\mathcal{T}_{N,\varepsilon} := \{(1+\varepsilon)(1+1/N)^{i-1} - 1 \mid 1 \leq i \leq N\}$ for some $N \geq 1$ and $\varepsilon > 0$. For $\boldsymbol{w} \in supp(\mu^i(2, c, u, \mathcal{T}_{N,\varepsilon})^2)$, it holds that*

$$RE(M^I; \boldsymbol{w}) = \max_{w\in\boldsymbol{w}} M^E(w)$$

and for $\boldsymbol{w} \in supp(\mu^e(2, c, u, \mathcal{T}_{N,\varepsilon})^2)$, it holds that

$$RE(M^E; \boldsymbol{w}) \leq \max_{w\in\boldsymbol{w}} M^E(w).$$

We defer the proof of Lemma 217 to Chapter F.7.1

Given Lemma 217, it suffices to analyze the engagement distribution at equilibrium for engagement-based optimization and investment-based optimization. More formally, within the instance constructed above, let $V^{I,\varepsilon,\mathcal{T}_{N,\varepsilon}}$ be the distribution of $M^{\mathrm{E}}(w) + s$ where $w \sim \mu^{\mathrm{i}}(2, c, u, \mathcal{T}_{N,\varepsilon})$. The distribution $V^{I,\varepsilon,\mathcal{T}_{N,\varepsilon}}$ can be characterized in closed-form as follows:

Lemma 218. *The distribution $V^{I,\varepsilon,\mathcal{T}_{N,\varepsilon}}$ has cdf equal to:*

$$\mathbb{P}[V^{I,\varepsilon,\mathcal{T}_{N,\varepsilon}} \leq v] = \begin{cases} 0 & \textit{if } v \leq 1 \\ v-1 & \textit{if } 1 \leq v \leq 2 \\ 1 & \textit{if } v \geq 2 \end{cases}$$

Moreover, let $V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}}$ be the distribution of $M^{\mathrm{E}}(w) + s$ where $w \sim \mu^{\mathrm{e}}(2, c, u, \mathcal{T})$. While the distribution $V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}}$ is somewhat messy, for each $\varepsilon > 0$, we show pointwise convergence to a simpler distribution $V^{e,\varepsilon,\infty}$ defined by:

$$\mathbb{P}[V^{e,\varepsilon,\infty} \leq v] = \begin{cases} 0 & \text{if } v \leq 1+\varepsilon \\ \ln\left(\frac{1}{1-\ln\left(\frac{v}{1+\varepsilon}\right)}\right) & \ln\left(\frac{1}{1-\ln\left(\frac{v}{1+\varepsilon}\right)}\right) \text{ if } v \in [1+\varepsilon, (1+\varepsilon)e^{1-e}] \\ 1 & \text{if } v \geq (1+\varepsilon)e^{1-e} \end{cases}$$

as formalized in the following lemma:

Lemma 219. *For each $\varepsilon > 0$, the cdf of $V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}}$ as $N \to \infty$ converges pointwise a.e. to the cdf of $V^{e,\varepsilon,\infty}$.*

We defer the proof of Lemma 219 to Chapter F.7.1

Using Lemma 217, Lemma 218, Lemma 220, and Lemma 219, we prove Theorem 60.

Proof of Theorem 60. By Lemma 217, it suffices to show that the following limits exist and that

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} \mathbb{E}_{\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{N,\varepsilon})}\left[\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)\right] < \mathbb{E}_{\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})}\left[\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)\right] \tag{F.6}$$

We analyze the left-hand side of (F.6), then analyze the right-hand side of (F.6), and then use these analyses to prove (F.6).

Analysis of left-hand side of (F.6). We first analyze the left-hand side of (F.6). We see that:

$$\begin{aligned} \mathbb{E}_{\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{N,\varepsilon})}\left[\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)\right] &= \mathbb{E}_{(V_1,V_2) \sim (V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}})^2}\left[\max(V_1, V_2)\right] \\ &= \int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}} \leq v]\right)^2\right) dv. \end{aligned}$$

We take a limit of this expression as $N \to \infty$. By Lemma 219, the function $\left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}} \leq v]\right)^2\right)$ pointwise approaches the function $\left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\infty} \leq v]\right)^2\right)$. Moreover, we see that

$$\left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}} \leq v]\right)^2\right) \leq g(v)$$

, where $g(v) = 1$ if $0 \leq v \leq 3$ and $g(v) = 0$ if $v \geq 3$. Applying dominated convergence with dominating function g, we see that:

$$\lim_{N \to \infty} \int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}} \leq v]\right)^2\right) dv = \int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\infty} \leq v]\right)^2\right) dv.$$

We next take a limit of $\int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\infty} \le v]\right)^2\right) dv$ as $\varepsilon \to 0$. We see that $V^{E,\varepsilon,\infty}$ pointwise a.e. approaches $V^{E,\infty}$ defined by:

$$\mathbb{P}[V^{E,\infty} \le v] = \begin{cases} 0 & \text{if } v \le 1 \\ \ln\left(\frac{1}{1-\ln(x)}\right) & \text{if } 1 \le v \le e^{1-1/e} \\ 1 & \text{if } v \ge e^{1-1/e} \end{cases}.$$

We again apply dominated convergence with the same function g as above to see that:

$$\int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\infty} \le v]\right)^2\right) dv = \lim_{\varepsilon \to 0} \int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\varepsilon,\infty} \le v]\right)^2\right) dv.$$

Putting this all together, we see that:

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} \mathbb{E}_{\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{N,\varepsilon})} \left[\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)\right] = \int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\infty} \le v]\right)^2\right) dv.$$

Analysis of right-hand side of (F.6). We next analyze the right-hand side of (F.6) as

$$\begin{aligned} \mathbb{E}_{\mu^{\mathrm{i}}(2,c,u,\mathcal{T}_{N,\varepsilon})} \left[\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)\right] &= \mathbb{E}_{(V_1,V_2) \sim (V^{I,\varepsilon,\mathcal{T}_{N,\varepsilon}})^2} [\max(V_1, V_2)] \\ &= \int_0^\infty \left(1 - \left(\mathbb{P}[V^{I,\varepsilon,\mathcal{T}_{N,\varepsilon}} \le v]\right)^2\right) dv \\ &= \int_0^\infty \left(1 - (\min(1, \max(0, v-1)))^2\right) dv. \end{aligned}$$

Proof of (F.6). We are now ready to prove (F.6). Using the above calculations, it suffices to show:

$$\int_0^\infty \left(1 - \left(\mathbb{P}[V^{E,\infty} \le v]\right)^2\right) dv < \int_0^\infty \left(1 - (\min(1, \max(0, v-1)))^2\right) dv.$$

To show this, it suffices to show that $\mathbb{P}[V^{E,\infty} \le v] \ge \min(1, \max(0, v-1))$ for all v and $\mathbb{P}[V^{E,\infty} \le v] > \min(1, \max(0, v-1))$ for $v \in (e^{1-1/e}, 1)$. The fact that $\mathbb{P}[V^{E,\infty} \le v] > \min(1, \max(0, v-1))$ for $v \in (e^{1-1/e}, 1)$ follows easily from the functional form of $\mathbb{P}[V^{E,\infty} \le v]$. Moreover, $\mathbb{P}[V^{E,\infty} \le v] = 0 = \min(1, \max(0, v-1))$ for $v \le 1$. To see that $\mathbb{P}[V^{E,\infty} \le v] \ge \min(1, \max(0, v-1))$ for $v \in [1, e^{1-1/e}]$, we apply Lemma 222 (see Chapter F.7.1 for the statement and proof).

□

Proof of Lemma 217

We prove Lemma 217.

Proof of Lemma 217. We first prove the lemma statement for investment-based optimization, and then we prove the statement for engagement-based optimization.

Proof for investment-based optimization. Recall that for the instance that we have constructed, the baseline utility is $\alpha = 1$ and the engagement metric is $M^{\mathrm{E}}(w) = w_{\mathrm{costly}} + w_{\mathrm{cheap}}$. By Chapter 204 and Chapter 65, there is a symmetric mixed Nash equilibrium $\mu^{\mathrm{i}}(2, c, u, \mathcal{T})$ given by the joint distribution $(W_{\mathrm{costly}}, W_{\mathrm{cheap}})$ where W_{cheap} is a point mass at 0 and W_{costly} is specified by:

$$\mathbb{P}[W_{\mathrm{costly}} \leq w_{\mathrm{costly}}] = \min(1, w_{\mathrm{costly}}).$$

This means that for $w \in \mathrm{supp}(\mu^{\mathrm{i}}(2, c, u, \mathcal{T}))$, it holds that

$$M^{\mathrm{E}}(w) = M^{\mathrm{I}}(w) = w_{\mathrm{costly}}.$$

This means that for $\mathbf{w} \in \mathrm{supp}(\mu^{\mathrm{i}}(2, c, u, \mathcal{T})^2)$, it holds that:

$$w_{i^*(M^{\mathrm{I}};\mathbf{w})} = \arg\max_{w \in \mathbf{w}} M^{\mathrm{I}}(w) = \arg\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w).$$

This means that:

$$\mathrm{RE}(M^{\mathrm{I}}; \mathbf{w}) = M^{\mathrm{E}}\left(\arg\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)\right) = \max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)$$

as desired.

Proof for engagement-based optimization. Suppose that $\mathbf{w} \in \mathrm{supp}(\mu^{\mathrm{e}}(2, c, u, \mathcal{T})^2)$. Since $w_{i^*(M^{\mathrm{E}};\mathbf{w})} \in \mathbf{w}$, it holds that:

$$\mathrm{RE}(M^{\mathrm{E}}; \mathbf{w}) = M^{\mathrm{E}}(w_{i^*(M^{\mathrm{E}};\mathbf{w})}) \leq \max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)$$

as desired.

□

Proof of Lemma 219

To prove Lemma 219, we first compute the cdf of the distribution $V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}}$.

Lemma 220. *Let N' be the minimum number such that $\sum_{i=1}^{N'} \frac{1}{N-i+1} \geq 1$. Let α^i be equal to $\frac{1}{N-i+1}$ for $1 \leq i \leq N'-1$ and be equal to $1 - \sum_{i'=1}^{N'-1} \frac{1}{N-i'+1}$ for $i = N'$. The distribution $V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}}$ has cdf $\mathbb{P}[V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}} \leq v]$ equal to:*

$$\begin{cases} 0 & \text{if } v \leq (1+\varepsilon) \\ \sum_{i'=1}^{i-1} \frac{1}{N-i'+1} + \frac{N}{(N-i+1)}\left(\frac{v}{(1+\varepsilon)(1+1/N)^{i-1}} - 1\right) & \text{if } v \in \left[(1+\varepsilon)\cdot(1+1/N)^{i-1}, (1+\varepsilon)\cdot(1+1/N)^{i}\right] \text{ for } i \in [N'-1] \\ \sum_{i'=1}^{N'-1} \frac{1}{N-i'+1} & \text{if } v \geq (1+\varepsilon)\cdot(1+1/N)^{N'-1}, \\ + \sum_{i'=1}^{N'-1} \frac{N}{(N-N'+1)}\left(\frac{v-(1+\varepsilon)(1+1/N)^{N'-1}}{(1+\varepsilon)(1+1/N)^{N'-1}}\right) & v \leq (1+\varepsilon)\cdot(1+1/N)^{N'-1}\cdot\left(1 + \frac{N-N'+1}{N}\cdot\left(1 - \sum_{j=1}^{N'-1}\frac{1}{N-j+1}\right)\right) \\ 1 & \text{if } v \geq (1+\varepsilon)\cdot(1+1/N)^{N'-1}\cdot\left(1 + \frac{N-N'+1}{N}\cdot\left(1 - \sum_{j=1}^{N'-1}\frac{1}{N-j+1}\right)\right). \end{cases}$$

Proof. We apply Proposition 212. The statement follows from this specification along with the fact that the $\mathrm{supp}(V \mid T = t_i)$ and $\mathrm{supp}(V \mid T = t_j)$ are disjoint for $i \neq j$ by the assumption of well-separated types. □

We next bound the value N' in Lemma 220.

Lemma 221. *Let N' be the minimum number such that $\sum_{i=1}^{N'} \frac{1}{N-i+1} \geq 1$. For sufficiently large N, it holds that:*

$$\frac{N+1}{e} - 1 < N - N' + 1 < \frac{N}{e^{1-\frac{3}{N+1}}}.$$

Proof. We first rewrite:

$$\sum_{i=1}^{M} \frac{1}{N-i+1} = \sum_{i=N-M+1}^{N} \frac{1}{i}.$$

Using an integral bound, we observe that

$$\int_{N-M+1}^{N+1} \frac{1}{x} dx \leq \sum_{i=N-M+1}^{N} \frac{1}{i} \leq \frac{1}{N-M+1} + \int_{N-M+1}^{N} \frac{1}{x} dx.$$

This implies that:

$$\ln\left(\frac{N+1}{N-M+1}\right) \leq \sum_{i=N-M+1}^{N} \frac{1}{i} \leq \frac{1}{N-M+1} + \ln\left(\frac{N}{N-M+1}\right).$$

Since N' is the minimum number such that $\sum_{i=N-N'+1}^{N} \frac{1}{i} \geq 1$, it must hold that: (1) $\sum_{i=N-N'+1}^{N} \frac{1}{i} \geq 1$, and (2) $\sum_{i=N-N'+2}^{N} \frac{1}{i} < 1$.

Condition (2) implies that:

$$\ln\left(\frac{N+1}{N-N'+2}\right) \leq \sum_{i=N-N'+2}^{N} \frac{1}{i} < 1,$$

which we can rewrite as:

$$N+1 < e \cdot (N - N' + 2),$$

which we can rewrite as:

$$N - N' + 1 > \frac{N+1}{e} - 1.$$

Condition (1) implies that:

$$\frac{1}{N-N'+1} + \ln\left(\frac{N}{N-N'+1}\right) \geq \sum_{i=N-N'+1}^{N} \frac{1}{i} \geq 1,$$

which we can rewrite as:

$$\ln\left(\frac{N}{N-N'+1}\right) \geq 1 - \frac{1}{N-N'+1}.$$

Using Condition (2), we see that $N - N' + 1 > \frac{N+1}{e} - 1 \geq \frac{N+1}{3}$ for sufficiently large N, so:

$$\ln\left(\frac{N}{N-N'+1}\right) \geq 1 - \frac{1}{N-N'+1} > 1 - \frac{3}{N+1}.$$

We can write this as:

$$N - N' + 1 < \frac{N}{e^{1-\frac{3}{N+1}}}$$

where the last inequality uses the upper bound on N' derived from Condition (2).

□

We prove Lemma 219.

Proof of Lemma 219. Fix $\varepsilon > 0$. We apply Lemma 220. Let $F_{N,\varepsilon}$ be the cdf of $V^{E,\varepsilon,\mathcal{T}_{N,\varepsilon}}$.

The first case is $v \leq 1 + \varepsilon$. It follows easily that $F_{N,\varepsilon}(x) = 0$ for all $v \leq (1+\varepsilon)$ for all $N \geq 2$, which means that $\lim_{N\to\infty} F_{N,\varepsilon}(v) = 0$ for all $x \leq (1+\varepsilon)$.

The next case is $v > e^{1-\frac{1}{e}}(1+\varepsilon)$. We see that for

$$v \geq (1+\varepsilon)\cdot(1+1/N)^{N'-1}\cdot\left(1+\frac{N-N'+1}{N}\cdot\left(1-\sum_{j=1}^{N'-1}\frac{1}{N-j+1}\right)\right),$$

it holds that $F_{N,\varepsilon}(x) = 1$. Observe that $\left(1-\sum_{j=1}^{N'-1}\frac{1}{N-j+1}\right) \leq \frac{1}{N-N'+1}$ by the definition of N', which means that:

$$(1+\varepsilon)\cdot(1+1/N)^{N'-1}\cdot\left(1+\frac{N-N'+1}{N}\cdot\left(1-\sum_{j=1}^{N'-1}\frac{1}{N-j+1}\right)\right) \leq (1+\varepsilon)\cdot(1+1/N)^{N'}.$$

For sufficiently large N, applying Lemma 221, we see that:

$$(1+\varepsilon)\cdot(1+1/N)^{N'} = (1+\varepsilon)\cdot\left((1+1/N)^N\right)^{N'/N} \geq (1+\varepsilon)\cdot\left((1+1/N)^N\right)^{\frac{N+1}{N}-\frac{1}{e^{1-\frac{3}{N+1}}}}.$$

This expression approaches $e^{1-\frac{1}{e}}\cdot(1+\varepsilon)$, which means that for any $v > e^{1-\frac{1}{e}}\cdot(1+\varepsilon)$, for sufficiently large N, it holds that $F_{N,\varepsilon}(v) = 1$ as desired. Thus, for any $v > e^{1-\frac{1}{e}}\cdot(1+\varepsilon)$, it holds that $\lim_{N\to\infty} F_{N,\varepsilon}(v) = 1$.

The next case is $(1+\varepsilon) < v < e^{1-\frac{1}{e}}(1+\varepsilon)$. In this case, for sufficiently large N, it holds that:

$$(1+\varepsilon)\left(1+\frac{1}{N}\right)^{N'-1} = (1+\varepsilon)\left(\left(1+\frac{1}{N}\right)^{N}\right)^{\frac{N'-1}{N}}$$
$$> (1+\varepsilon)\left((1+1/N)^N\right)^{1-\frac{1}{e^{1-\frac{3}{N+1}}}},$$

which approaches $e^{1-\frac{1}{e}} \cdot (1+\varepsilon)$ in the limit. This means that for sufficiently large N, it holds that $x < (1+\varepsilon)\left(1+\frac{1}{N}\right)^{N'-1}$. For $v \in [(1+\varepsilon)\cdot(1+1/N)^{i-1}, (1+\varepsilon)\cdot(1+1/N)^{i}]$, applying Lemma 220, we see that:

$$\sum_{i'=1}^{i-1} \frac{1}{N-i'+1} \le F_{N,\varepsilon}(v) \le \sum_{i'=1}^{i} \frac{1}{N-i'+1}.$$

Using an integral bound, we can lower bound the left-hand side as:

$$\sum_{i'=1}^{i-1} \frac{1}{N-i'+1} \ge \int_{N-i+2}^{N+1} \frac{1}{x}dx = \ln\left(\frac{N+1}{N-i+2}\right) = \ln\left(\frac{1+\frac{1}{N}}{1-\frac{i}{N}+\frac{2}{N}}\right).$$

We can also upper bound, for sufficiently large N, the right-hand side as:

$$\begin{aligned}\sum_{i'=1}^{i} \frac{1}{N-i'+1} &\le \frac{1}{N-i+1} + \int_{N-i+1}^{N} \frac{1}{x}dx \\ &= \frac{1}{N-i+1} + \ln\left(\frac{N}{N-i+1}\right) \\ &\le \frac{1}{N-N'+1} + \ln\left(\frac{N}{N-i+1}\right) \\ &\le_{(A)} \frac{1}{\frac{N+1}{e}-1} + \ln\left(\frac{1}{1-\frac{i}{N}+\frac{1}{N}}\right)\end{aligned}$$

where (A) follows from Lemma 221. Putting this all together, we see that

$$v \in \left[(1+\varepsilon)\cdot(1+1/N)^{i-1}, (1+\varepsilon)\cdot(1+1/N)^{i}\right],$$

then

$$\ln\left(\frac{1+\frac{1}{N}}{1-\frac{i}{N}+\frac{2}{N}}\right) \le F_{N,\varepsilon}(v) \le \frac{1}{\frac{N+1}{e}-1} + \ln\left(\frac{1}{1-\frac{i}{N}+\frac{1}{N}}\right).$$

We next reparameterize i as $\alpha = i/N$. We can rewrite
$v \in [(1+\varepsilon)\cdot(1+1/N)^{i-1}, (1+\varepsilon)\cdot(1+1/N)^{i}]$
as $v \in \left[(1+\varepsilon)\cdot((1+1/N)^N)^{\frac{i}{N}-\frac{1}{N}}, (1+\varepsilon)\cdot((1+1/N)^N)^{\frac{i}{N}}\right]$, or alternatively as

$$v \in \left[(1+\varepsilon)\cdot((1+1/N)^N)^{\alpha-\frac{1}{N}}, (1+\varepsilon)\cdot((1+1/N)^N)^{\alpha}\right],$$

and the bound as:

$$\ln\left(\frac{1+\frac{1}{N}}{1-\alpha+\frac{2}{N}}\right) \le F_{N,\varepsilon}(v) \le \frac{1}{\frac{N+1}{e}-1} + \ln\left(\frac{1}{1-\alpha+\frac{1}{N}}\right).$$

For any $v = (1+\varepsilon)\cdot e^{\beta}$ where $\beta < 1 - 1/e$, we see that for sufficiently large N, it holds that:

$$\lim_{N\to\infty} F_{N,\varepsilon}((1+\varepsilon)\cdot e^{\beta}) = \ln\left(\frac{1}{1-\beta}\right)$$

Reparameterizing $\beta = \ln\left(\frac{v}{1+\varepsilon}\right)$, we obtain:

$$\lim_{N\to\infty} F_{N,\varepsilon}(v) = \ln\left(\frac{1}{1-\ln\left(\frac{v}{1+\varepsilon}\right)}\right)$$

□

Statement and proof of Lemma 222

Lemma 222. *For $x \in [1, e^{1-1/e}]$, it holds that:*

$$\ln\left(\frac{1}{1-\ln(x)}\right) \ge x - 1 \tag{F.7}$$

Proof. In the proof, we will use the following two standard bounds: (F1) $\ln(1+z) \ge \frac{2z}{2+z}$ for $z \ge -1$ and (F2) $e^{2z} \le \frac{1+z}{1-z}$ for $z \in (0,1)$.

First, let's reparamterize and set $x_1 = \ln(x)$, so that the range of x_1 is now $(0, 1-1/e)$. Then we can rewrite (F.7) as:

$$\ln\left(\frac{1}{1-x_1}\right) \ge e^{x_1} - 1.$$

We can now apply (F2) to $z = x_1/2 \in (0, 0.5 - \frac{1}{2e})$ to see that it suffices to show that:

$$\ln\left(\frac{1}{1-x_1}\right) \ge \frac{1+x_1}{1-x_1} - 1,$$

which can be simplified to

$$\ln\left(\frac{1}{1-x_1}\right) \ge \frac{x_1}{1-x_1/2},$$

which can be simplified to

$$\ln\left(\frac{1}{1-x_1}\right) \ge \frac{2x_1}{2-x_1}$$

Let's reparameterize again to set $x_2 = \frac{x_1}{1-x_1}$ so that the range of x_2 is now $[0, e-1]$. Using that $x_1 = \frac{x_2}{1+x_2}$, it thus suffices to show:

$$\ln(1+x_2) \geq \frac{\frac{2x_2}{1+x_2}}{2 - \frac{x_2}{1+x_2}},$$

which can be simplified to:

$$\ln(1+x_2) \geq \frac{2x_2}{2+x_2}.$$

This follows from (F1).

□

F.7.2 Proof of Proposition 61

We prove Chapter 61.

Proof of Chapter 61. Observe that for the instance that we have constructed, the minimum investment level is $\beta_t = \max(0, -\alpha)$, the cost function is $c([w_{\text{costly}}, 0]) = w_{\text{costly}} + \gamma \cdot w_{\text{cheap}}$, and the engagement metric is $M^{\mathrm{E}}(w) = w_{\text{costly}} + w_{\text{cheap}}$. The function $f_t(w_{\text{cheap}})$ is equal to $\max(0, (w_{\text{cheap}}/t) - \alpha)$.

We define the following quantities:

$$T^{\mathrm{I}} := \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{i}}(P,c,u,\mathcal{T}))^P}[\mathrm{RE}(M^{\mathrm{I}}; \mathbf{w})]$$
$$T^{\mathrm{E}} := \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P}[\mathrm{RE}(M^{\mathrm{E}}; \mathbf{w})].$$

It suffices to show that $T^{\mathrm{E}} \geq T^{\mathrm{I}}$.

We first analyze each term separately.

For the term T^{I}, we apply Chapter 204 and Chapter 65. This means that $\mu^{\mathrm{i}}(P, c, u, \mathcal{T})$ is specified by joint distribution $(W_{\text{costly}}, W_{\text{cheap}})$ where W_{cheap} is a point mass at 0 and W_{costly} is distributed as:

$$\mathbb{P}_{(W_{\text{costly}}, W_{\text{cheap}}) \sim \mu^{\mathrm{i}}(P,c,u,\mathcal{T})}[W_{\text{costly}} \leq w_{\text{costly}}] = \begin{cases} (-\alpha)^{1/(P-1)} & \text{if } 0 \leq w_{\text{costly}} \leq -\alpha \\ (\min(1, w_{\text{costly}}))^{1/(P-1)} & \text{if } w_{\text{costly}} \geq \max(0, -\alpha). \end{cases}$$

To analyze T^I, we observe that for $w \in \operatorname{supp}(\mu^{\mathrm{i}}(P, c, u, \mathcal{T}))$, it holds that

$$M^{\mathrm{E}}(w) = M^{\mathrm{I}}(w) = w_{\text{costly}}.$$

This means that for $\mathbf{w} \in \operatorname{supp}(\mu^{\mathrm{i}}(P, c, u, \mathcal{T})^P)$, it holds that:

$$w_{i^*(M^{\mathrm{I}};\mathbf{w})} = \arg\max_{w \in \mathbf{w}} M^{\mathrm{I}}(w) = \arg\max_{w \in \mathbf{w}} w_{\text{costly}}.$$

This means that:

$$\mathrm{RE}(M^{\mathrm{I}}; \mathbf{w}) = M^{\mathrm{E}}\left(\arg\max_{w \in \mathbf{w}} w_{\mathrm{costly}}\right) = \max_{w \in \mathbf{w}} w_{\mathrm{costly}},$$

which means that:

$$T^{\mathrm{I}} = \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{i}}(P,c,u,\mathcal{T}))^P}[\mathrm{RE}(M^{\mathrm{I}}; \mathbf{w})] = \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{i}}(P,c,u,\mathcal{T}))^P}\left[\max_{w \in \mathbf{w}} w_{\mathrm{costly}}\right].$$

We rewrite this expression as follows. Let Z^{I} be a random variable given by the maximum of P i.i.d. realizations of a random variable distributed as W_{costly}. This random variable has cdf:

$$\mathbb{P}[Z^{\mathrm{I}} \leq z] = \begin{cases} (-\alpha)^{P/(P-1)} & \text{if } 0 \leq z \leq -\alpha \\ (\min(1, z))^{P/(P-1)} & \text{if } z \geq \max(0, -\alpha). \end{cases}$$

This means that:

$$T^{\mathrm{I}} = \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{i}}(P,c,u,\mathcal{T}))^P}\left[Z^{\mathrm{I}}\right].$$

For the term T^{E}, we apply Chapter 203 and Chapter 67. This means that $\mu^{\mathrm{e}}(P, c, u, \mathcal{T})$ is specified by joint distribution $(W_{\mathrm{costly}}, W_{\mathrm{cheap}})$ where $\mathbb{P}_{(W_{\mathrm{costly}}, W_{\mathrm{cheap}}) \sim \mu^{\mathrm{e}}(P,c,u,\mathcal{T})}$ is equal to:

$$\begin{cases} (-\alpha)^{1/(P-1)} & \text{if } 0 \leq w_{\mathrm{costly}} \leq -\alpha \\ (\min(1, w_{\mathrm{costly}} + \gamma \cdot t \cdot (w_{\mathrm{costly}} + \alpha)))^{1/(P-1)} & \text{if } w_{\mathrm{costly}} \geq \max(0, -\alpha). \end{cases}$$

Moreover, this means that the distribution $W_{\mathrm{cheap}} \mid W_{\mathrm{costly}} = w_{\mathrm{costly}}$ for $w_{\mathrm{costly}} \in \mathrm{supp}(W_{\mathrm{costly}})$ takes the following form. If $w_{\mathrm{costly}} > 0$, the distribution $W_{\mathrm{cheap}} \mid W_{\mathrm{costly}} = w_{\mathrm{costly}}$ is a point mass at $f_t^{-1}(w_{\mathrm{costly}}) = t \cdot (w_{\mathrm{costly}} + \alpha)$. If $w_{\mathrm{costly}} = 0$, then $W_{\mathrm{cheap}} \mid W_{\mathrm{costly}} = w_{\mathrm{costly}}$ is a point mass at 0 if $\alpha \leq 0$, $W_{\mathrm{cheap}} \mid W_{\mathrm{costly}} = w_{\mathrm{costly}}$ is distributed according to the cdf $\min\left(1, \left(\frac{w_{\mathrm{cheap}}}{t \cdot \alpha}\right)^{1/(P-1)}\right)$ if $\alpha > 0$ and $\gamma > 0$, and $W_{\mathrm{cheap}} \mid W_{\mathrm{costly}} = w_{\mathrm{costly}}$ is distributed as a point mass at $t \cdot \alpha$ if $\alpha > 0$ and $\gamma = 0$. To analyze T^{E}, we observe that:

$$w_{i^*(M^{\mathrm{E}}; \mathbf{w})} = \arg\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w).$$

This means that:

$$\mathrm{RE}(M^{\mathrm{E}}; \mathbf{w}) = M^{\mathrm{E}}\left(\arg\max_{w \in \mathbf{w}} M^{\mathrm{E}}(w)\right) = \max_{w \in \mathbf{w}}(w_{\mathrm{costly}} + w_{\mathrm{cheap}}),$$

which means that:

$$T^{\mathrm{E}} = \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P}[\mathrm{RE}(M^{\mathrm{I}}; \mathbf{w})] = \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P}\left[\max_{w \in \mathbf{w}}(w_{\mathrm{costly}} + w_{\mathrm{cheap}})\right].$$

We rewrite this expression as follows. Let Z^{E} be a random variable given by the maximum of P i.i.d. realizations of a random variable distributed as $W_{\mathrm{costly}} + W_{\mathrm{cheap}}$. To formalize the cdf of Z^{E}, we need to rewrite $w_{\text{costly}} + \gamma \cdot t \cdot (w_{\text{costly}} + \alpha)$ in terms of $z := w_{\text{costly}} + w_{\text{cheap}}$. Using the fact that for $w \in \operatorname{supp}(\mu^{\mathrm{e}}(P, c, u, \mathcal{T}))$ such that $w_{\text{costly}} > 0$, it holds that:

$$z := w_{\text{costly}} + w_{\text{cheap}} = w_{\text{costly}} + t \cdot (w_{\text{costly}} + \alpha) = w_{\text{costly}} \cdot (1 + t) + t \cdot \alpha.$$

and

$$\begin{aligned}
w_{\text{costly}} + \gamma \cdot t \cdot (w_{\text{costly}} + \alpha) &= w_{\text{costly}}(1 + \gamma \cdot t) + \gamma \cdot t \cdot \alpha \\
&= \frac{1 + \gamma \cdot t}{1 + t}(1 + t) w_{\text{costly}} + \gamma \cdot t \cdot \alpha \\
&= \frac{1 + \gamma \cdot t}{1 + t}\left((1 + t) w_{\text{costly}} + t \cdot \alpha\right) - t \cdot \alpha \cdot \frac{1 + \gamma \cdot t}{1 + t} + \gamma \cdot t \cdot \alpha \\
&= \frac{1 + \gamma \cdot t}{1 + t} \cdot z + t \cdot \alpha \cdot \left(\gamma - \frac{1 + \gamma \cdot t}{1 + t}\right) \\
&= \frac{1 + \gamma \cdot t}{1 + t} \cdot z - t \cdot \alpha \cdot \frac{1 - \gamma}{1 + t}.
\end{aligned}$$

If $\alpha \leq 0$, then this random variable has cdf:

$$\mathbb{P}[Z^{\mathrm{E}} \leq z] = \begin{cases} (-\alpha)^{P/(P-1)} & \text{if } 0 \leq z \leq -\alpha \\ \left(\min\left(1, \frac{1+\gamma \cdot t}{1+t} \cdot z - t \cdot \alpha \cdot \frac{1-\gamma}{1+t}\right)\right)^{P/(P-1)} & \text{if } z \geq \max(0, -\alpha). \end{cases}$$

Otherwise, if $\alpha > 0$, then this random variable has cdf:

$$\mathbb{P}[Z^{\mathrm{E}} \leq z] = \begin{cases} (z \cdot \gamma)^{1/(P-1)} & \text{if } z \leq t \cdot \alpha \\ \left(\min\left(1, \frac{1+\gamma \cdot t}{1+t} \cdot z - t \cdot \alpha \cdot \frac{1-\gamma}{1+t}\right)\right)^{P/(P-1)} & \text{if } z > t \cdot \alpha. \end{cases}$$

This means that:

$$T^{\mathrm{E}} = \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{i}}(P, c, u, \mathcal{T}))^P}\left[Z^{\mathrm{E}}\right].$$

We now combine these expressions and compare T^{E} and T^{I}. First, we see that Z^{E} stochastically dominates Z^{I}, since:

$$\frac{1 + \gamma \cdot t}{1 + t} \cdot z - t \cdot \alpha \cdot \frac{1 - \gamma}{1 + t} \leq z.$$

This implies that:

$$T^{\mathrm{E}} = \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{i}}(P, c, u, \mathcal{T}))^P}\left[Z^{\mathrm{E}}\right] \geq \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{i}}(P, c, u, \mathcal{T}))^P}\left[Z^{\mathrm{I}}\right] = T^{\mathrm{I}}$$

as desired. $\square$

F.8 Proofs for Chapter 9.4.3

F.8.1 Proof of Theorem 62

The proof of Theorem 62 follows from the following characterizations of the realized user utility for engagement-based optimization (Lemma 224) and random recommendations (Lemma 223) , stated and proved below.

Lemma 223. *Consider the setup of Theorem 62. Then it holds that:*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^r(P,c,u,\mathcal{T}))^P}[UW(M^R;\mathbf{w})] = \begin{cases} \alpha & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha \leq 0. \end{cases}$$

Proof. If $\alpha > 0$, then we see that $U_b^I(0,t) = \alpha$. This means that:

$$\min_{w_{\text{costly}}} \left\{ C_b^I(w_{\text{costly}}) \mid U_b^I(w_{\text{costly}}, t) \geq 0 \right\} = 0$$

and moreover the min is achieved at $w = [0,0]$. This means that $\nu = 0$ and $\mu^{\mathrm{r}}(P,c,u,\mathcal{T})$ is a point mass at $[0,0]$. This means that:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{r}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(M^{\mathrm{R}};\mathbf{w})] = U_b^I(0,t) = \alpha.$$

If $\alpha \leq 0$, then

$$w_{\text{costly}}^* := \arg\min_{w'_{\text{costly}}} \left\{ C_b^I(w'_{\text{costly}}) \mid U_b^I(w'_{\text{costly}}, t) \geq 0 \right\}$$

satisfies $U_b^I(w_{\text{costly}}^*, t) = 0$. This means that if $(W_{\text{costly}}, W_{\text{cheap}}) \sim \mu^{\mathrm{r}}(P,c,u,\mathcal{T})$, it holds that $\mathrm{supp}(W_{\text{costly}}) \subseteq \left\{ w_{\text{costly}}^*, 0 \right\}$. Moreover, for any content landscape $\mathbf{w} \in \mathrm{supp}(\mu^{\mathrm{r}}(P,c,u,\mathcal{T}))^P$, we see that:

$$\mathrm{UW}(M^{\mathrm{R}};\mathbf{w}) := \mathbb{E}[u(w_{i^*(M^{\mathrm{R}};\mathbf{w})}, t) \cdot \mathbb{1}[u(w_{i^*(M^{\mathrm{R}};\mathbf{w})}, t) \geq 0]] = 0.$$

This means that:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{r}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(M^{\mathrm{R}};\mathbf{w})] = U_b^I(w_{\text{costly}}^*) = 0.$$

□

Lemma 224. *Consider the setup of Theorem 62. If $\alpha > 0$, then it holds that:*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(P,c,u,\mathcal{T}))^P}[UW(M^E;\mathbf{w})] < \alpha.$$

If $\alpha \leq 0$, then it holds that:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(P,c,u,\mathcal{T}))^P}[UW(M^E;\mathbf{w})] = 0.$$

Proof. First, suppose that $\alpha \leq 0$. In this case, we see that $u(w,t) \leq 0$ for all $w \in \operatorname{supp}(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P)$. This implies that for any content landscape $\mathbf{w} \in \operatorname{supp}(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P$, we see that:

$$\mathrm{UW}(M^{\mathrm{E}};\mathbf{w}) := \mathbb{E}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t) \cdot \mathbb{1}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t) \geq 0]] = 0.$$

This means that:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(M^{\mathrm{E}};\mathbf{w})] = 0.$$

For $\alpha > 0$, we see that $w = [0,0]$ is the unique value such that $w \in \mathcal{C}_t^{\mathrm{aug}}$ and $u(w,t) \geq \alpha$. Moreover, by Lemma 200, we know that $\{[f_t(w_{\mathrm{cheap}}), w_{\mathrm{cheap}}] \mid w_{\mathrm{cheap}} \geq 0\} = \mathcal{C}_t^{\mathrm{aug}}$. We observe that $\operatorname{supp}(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P$ is contained in $\{[f_t(w_{\mathrm{cheap}}), w_{\mathrm{cheap}}] \mid w_{\mathrm{cheap}} \geq 0\} = \mathcal{C}_t^{\mathrm{aug}}$. This means that $u(w,t) < \alpha$ for all $w \in \operatorname{supp}(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P$ such that $w \neq [0,0]$. Since there is no point mass at 0, this means that the probability $[0,0]$ shows up in the content landscape is 0, so

$$\mathbb{P}[\mathrm{UW}(M^{\mathrm{E}};\mathbf{w}) < \alpha] = \mathbb{P}\left[\mathbb{E}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t) \cdot \mathbb{1}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t) \geq 0]] < \alpha\right] = 1.$$

This means that:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(M^{\mathrm{E}};\mathbf{w})] < \alpha.$$

□

Using Lemma 224 and Lemma 223, we prove Theorem 62.

Proof of Theorem 62. We apply Lemma 224 and Lemma 223. When $\alpha > 0$, we see that:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(\mu^{\mathrm{e}};\mathbf{w})] < \alpha = \mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{r}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(M^{\mathrm{R}};\mathbf{w})].$$

When $\alpha \leq 0$, we see that:

$$\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{e}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(\mu^{\mathrm{e}};\mathbf{w})] = 0 = \mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{r}}(P,c,u,\mathcal{T}))^P}[\mathrm{UW}(M^{\mathrm{R}};\mathbf{w})]$$

□

F.8.2 Proofs of Proposition 63 and Proposition 64

Both Proposition 63 and Proposition 64 leverage instantiations of Chapter 5 with $P = 2$ and $\gamma = 0$, where the type space is

$$\mathcal{T}_{2,\varepsilon,c} = \{\varepsilon, c(1+\varepsilon) - 1\} \tag{F.8}$$

for some $\varepsilon > 0$ and $c > 1$. We first analyze the user utility for engagement-based optimization and investment-based optimization for instantations of this form.

Lemma 225. *Consider Chapter 5 with $P = 2$, $\gamma = 0$, and type space $\mathcal{T}_{2,\varepsilon,c}$ defined by (F.8). If $c \geq 1.5$, then*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[UW(M^E;\mathbf{w})] \leq \frac{3}{16} \cdot W \cdot (c-1) \cdot (1+\varepsilon).$$

If $1 < c \leq (5-\sqrt{5})/2$, then $\mathbb{E}_{\mathbf{w}\sim(\mu^e(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}$ is at least

$$\frac{1}{2} \cdot \left(1 - \left(\frac{(3-c)(c-1)}{2-c}\right)^2\right) \cdot W \cdot \frac{c \cdot (3-c) \cdot (c-1) \cdot (1+\varepsilon)}{2(2-c)}.$$

Proof. We apply Chapter 71. Let $t_1 = \varepsilon$ and $t_2 = c(1+\varepsilon) - 1$. We see that $a_{t_1} = \frac{1}{1+\varepsilon}$, $a_{t_2} = \frac{1}{c(1+\varepsilon)}$, and $s = 0$ in Chapter 5. This implies that $c = \frac{a_{t_1}}{a_{t_2}}$. Moreover, we see that $f_t(w_{\text{cheap}}) = \max(0, (w_{\text{cheap}})/t) - 1)$. Recall that our goal is to analyze

$$\mathbb{E}_{\mathbf{w}\sim(\mu^e(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}};\mathbf{w})] = \mathbb{E}_{\mathbf{w}\sim(\mu^e(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathbb{E}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t)\cdot\mathbb{1}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t) > 0]]].$$

To analyze this expression, we consider the reparameterization of the equilibrium given by the random vector (V, T) defined in Chapter 9.6.4. As described in Chapter 9.6.4, the function $h(v, t)$ corresponds to unique content over the form

$$h(v, t) = [f_t(w_{\text{cheap}}), w_{\text{cheap}}] = [\max(0, (w_{\text{cheap}})/t) - 1), w_{\text{cheap}}]$$

that satisfies

$$M^{\mathrm{E}}(h(v,t)) = M^{\mathrm{E}}([\max(0,(w_{\text{cheap}})/t)-1), w_{\text{cheap}}]) = \max(0,(w_{\text{cheap}})/t)-1)+w_{\text{cheap}}+1 = v.$$

We first show that if $u(h(v,t),t') > 0$ for $(v,t) \in \text{supp}((V,T))$ and $i \in \{1,2\}$, it must hold that $t' = t_2$ and $t = t_1$. We first observe that $u(h(v,t),t_1) \leq 0$. Moreover, it holds that $u(h(v,t),t_2) = 0$ if $t = t_2$. This proves the desired statement.

We next show that $(v, t_1) \in \text{supp}((V,T))$, it holds that:

$$u(h(v,t_1),t_2) = \frac{W}{t_1} \cdot w_{\text{cheap}} \cdot (c-1) \cdot (1+\varepsilon).$$

Let $w = h(v, t_1)$. Observe that the utility function satisfies:

$$u(w, t_2) = W \cdot t_2 \cdot (w_{\text{costly}} - w_{\text{cheap}}/t_2 + 1)$$

by definition. Since $(v, t_1) \in \text{supp}((V,T))$, the equilibrium structure tells us that $w = h(v, t_1)$ satisfies $w_{\text{cheap}} \geq t_1$ and $w_{\text{costly}} = (w_{\text{cheap}}/t_1) - 1$. This implies that

$$\begin{aligned}
u(h(v,t_1),t_2) &= u(w,t_2) \\
&= W \cdot t_2 \cdot (w_{\text{costly}} - w_{\text{cheap}}/t_2 + 1) \\
&= W \cdot t_2 \cdot w_{\text{cheap}} \cdot \left(\frac{1}{t_1} - \frac{1}{t_2}\right) = \frac{W}{t_1} \cdot w_{\text{cheap}} \cdot (t_2 - t_1) \\
&= \frac{W}{t_1} \cdot w_{\text{cheap}} \cdot (c-1) \cdot (1+\varepsilon)
\end{aligned}$$

as desired.

For the remainder of the analysis, we split into two cases: $c \geq 1.5$ and $1 < c \leq (5-\sqrt{5})/2$.

Case 1: $c \geq 1.5$**.** This corresponds to the first case of Chapter 71. To analyze the user welfare expression, it suffices to restrict to the cases where the winning content $w_{i^*(M^{\mathrm{E}};\mathbf{w})}$ satisfies $u(w_{i^*(M^{\mathrm{E}};\mathbf{w})}, t') > 0$. By the above argument, it suffices to restrict to the case where the user has type $t' = t_2$ which occurs with probability $1/2$.

We next show that $(v, t_1) \in \mathrm{supp}((V, T))$, it holds that:

$$u(h(v, t_1), t_2) = \frac{3}{2} \cdot W \cdot (c-1) \cdot (1+\varepsilon).$$

We see that

$$v = w_{\text{cheap}}(1 + 1/t_1) = w_{\text{cheap}} \cdot \frac{t_1+1}{t_1} \leq \frac{3}{2a_{t_1}} = \frac{3(t_1+1)}{2}$$

which means that:

$$w_{\text{cheap}} \leq \frac{3t_1}{2}.$$

This implies that:

$$\begin{aligned} u(h(v, t_1), t_2) &= \frac{W}{t_1} \cdot w_{\text{cheap}} \cdot (c-1) \cdot (1+\varepsilon) \\ &\leq \frac{3}{2} \cdot W \cdot (c-1) \cdot (1+\varepsilon) \end{aligned}$$

as desired.

Let the content landscape $\mathbf{w} = [w_1, w_2]$ in the reparameterized space be such that $w_1 = h(v^1, t^1)$ and $w_2 = h(v^2, t^2)$, and let $w_{i^*(M^{\mathrm{E}};\mathbf{w})}$ be $h(v^*, t^*)$. Since the user has type $t' = t_2$, the structure of the equilibrium in Chapter 71 implies that $t^* = t_1$ only if $t^1 = t^2 = t_1$. This implies that if $u(w_{i^*(M^{\mathrm{E}};\mathbf{w})}, t') > 0$, then it must hold that $t^1 = t^2 = t_1$. Moreover, in this case, the user utility is at most:

$$u(w_{i^*(M^{\mathrm{E}};\mathbf{w})}, t') \leq \frac{3}{2} \cdot W \cdot (c-1) \cdot (1+\varepsilon).$$

Putting this all together, we see that:

$$\begin{aligned} &\mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}}; \mathbf{w})] \\ &= \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathbb{E}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})}, t) \cdot \mathbb{1}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})}, t) > 0]]] \\ &\leq \mathbb{P}_{t' \sim T}[t' = t_2] \cdot \mathbb{P}_{(v^1,t^1) \sim (V,T)}[t^1 = t_1] \cdot \mathbb{P}_{(v^2,t^2) \sim (V,T)}[t^2 = t_1] \cdot \frac{3}{2} \cdot W \cdot (c-1) \cdot (1+\varepsilon) \\ &= \frac{3}{16} \cdot W \cdot (c-1) \cdot (1+\varepsilon). \end{aligned}$$

Case 2: $1 < c \leq (5-\sqrt{5})/2$**.** This corresponds to the third case of Chapter 71. To analyze the user welfare expression, it suffices to restrict to the cases where the winning content

$w_{i^*(M^{\mathrm{E}};\mathbf{w})}$ satisfies $u(w_{i^*(M^{\mathrm{E}};\mathbf{w})}, t') > 0$. By the above argument, it suffices to restrict to the case where the user has type $t' = t_2$ which occurs with probability $1/2$.

Let the content landscape $\mathbf{w} = [w_1, w_2]$ in the reparameterized space be such that $w_1 = h(v^1, t^1)$ and $w_2 = h(v^2, t^2)$, and let $w_{i^*(M^{\mathrm{E}};\mathbf{w})}$ be $h(v^*, t^*)$. Since the user has type $t' = t_2$, the structure of the equilibrium in Chapter 71 implies that if $v^1 > \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$ or if $v^2 > \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$, then it holds that $t^* = t_1$ and $v^* > \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$.

Moreover, we claim that if $w = h(v, t)$ is such that $v > \frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)}$ and $t = t_1$, then the user utility is at least:

$$u(w, t') \geq W \cdot \frac{c \cdot (3 - c) \cdot (c - 1) \cdot (1 + \varepsilon)}{2(2 - c)}.$$

Note that:

$$\frac{3 - \frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2 - \frac{a_{t_1}}{a_{t_2}}\right)} = \frac{c \cdot (1 + \varepsilon) \cdot (3 - c)}{2(2 - c)}.$$

Moreover, we see that

$$v = w_{\text{cheap}}(1 + 1/t_1) = w_{\text{cheap}} \cdot \frac{1 + \varepsilon}{\varepsilon} \geq \frac{c \cdot (1 + \varepsilon) \cdot (3 - c)}{2(2 - c)}.$$

This implies that:

$$w_{\text{cheap}} \geq \frac{c \cdot \varepsilon \cdot (3 - c)}{2(2 - c)}.$$

This implies that:

$$\begin{aligned} u(w, t') &\geq \frac{W}{t_1} \cdot w_{\text{cheap}} \cdot (c - 1) \cdot (1 + \varepsilon) \\ &\geq \frac{W}{\varepsilon} \cdot \frac{c \cdot \varepsilon \cdot (3 - c)}{2(2 - c)} \cdot (c - 1) \cdot (1 + \varepsilon) \\ &= W \cdot \frac{c \cdot (3 - c) \cdot (c - 1) \cdot (1 + \varepsilon)}{2(2 - c)}. \end{aligned}$$

Putting this all together, we see that:

$$\begin{aligned}
&\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}};\mathbf{w})] \\
&= \mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathbb{E}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t)\cdot \mathbb{1}[u(w_{i^*(M^{\mathrm{E}};\mathbf{w})},t)>0]]] \\
&\geq \mathbb{P}_{t'\sim T}[t'=t_2]\cdot \mathbb{P}_{(v^1,t^1)\sim(V,T),(v^2,t^2)\sim(V,T)}\left[v^1 \text{ or } v^2 > \frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right]\cdot W\cdot \frac{c\cdot(3-c)\cdot(c-1)\cdot(1+\varepsilon)}{2(2-c)}. \\
&= \frac{1}{2}\cdot\left(1-\left(1-\mathbb{P}_{(v,t)\sim(V,T)}\left[v>\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right]\right)^2\right)\cdot W\cdot\frac{c\cdot(3-c)\cdot(c-1)\cdot(1+\varepsilon)}{2(2-c)}.
\end{aligned}$$

To simplify this expression, we see that:

$$\begin{aligned}
1-\mathbb{P}_{(v,t)\sim(V,T)}\left[v>\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2a_{t_2}\left(2-\frac{a_{t_1}}{a_{t_2}}\right)}\right] &= 1-a_{t_1}\cdot\left(\frac{1}{a_{t_1}}-\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2-\frac{a_{t_1}}{a_{t_2}}}\left(\frac{1}{a_{t_2}}-\frac{1}{a_{t_1}}\right)\right) \\
&= 1-a_{t_1}\cdot\frac{1}{a_{t_1}}\left(1-\frac{3-\frac{a_{t_1}}{a_{t_2}}}{2-\frac{a_{t_1}}{a_{t_2}}}\left(\frac{a_{t_1}}{a_{t_2}}-1\right)\right) \\
&= 1-\left(1-\frac{(3-c)(c-1)}{2-c}\right) \\
&= \frac{(3-c)(c-1)}{2-c}.
\end{aligned}$$

Plugging this into the above expression, we obtain:

$$\frac{1}{2}\cdot\left(1-\left(\frac{(3-c)(c-1)}{2-c}\right)^2\right)\cdot W\cdot\frac{c\cdot(3-c)\cdot(c-1)\cdot(1+\varepsilon)}{2(2-c)}.$$

□

Lemma 226. *Consider Chapter 5 with $P=2$, $\gamma=0$, and type space $\mathcal{T}_{2,\varepsilon,c}$ defined by (F.8). If $c\geq 1.5$, then*

$$\mathbb{E}_{\mathbf{w}\sim(\mu^{r}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[UW(M^R;\mathbf{w})] = \frac{W}{2}\cdot(\varepsilon+c(1+\varepsilon)-1)\,.$$

Proof. We apply Theorem 66. In the construction, we see that the minimum investment level $\beta_t=0$. This means that $\kappa=0$ so $\nu=0$. Thus we see that the distribution $\mu^{\mathrm{r}}(P,c,u,\mathcal{T})$ is a point mass at $w=[0,0]$. Observe that $u([0,0],t)=W\cdot t$, which means that:

$$\begin{aligned}
\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{r}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{R}};\mathbf{w})] &= \frac{1}{2}\left(u([0,0],\varepsilon)+u([0,0],c(1+\varepsilon)-1)\right) \\
&= \frac{W}{2}\cdot(\varepsilon+c(1+\varepsilon)-1)\,.
\end{aligned}$$

□

Using Lemma 225 and Lemma 226, we prove Chapter 63.

Proof of Chapter 63. We construct an instances with two types where the user welfare of engagement-based optimization exceeds the user welfare of random recommendations. Interestingly, users are *nearly homogeneous* in these instances. Consider Chapter 5 with $P = 2$, $\gamma = 0$, and type space $\mathcal{T}_{2,\varepsilon,c}$ defined by (F.8). Let $\mu^{\mathrm{e}}(2, c, u, \mathcal{T}_{2,\varepsilon,c})$ be the symmetric mixed equilibrium specified in Definition 8, and let $\mu^{\mathrm{i}}_{2,c,u} := \mu^{\mathrm{i}}(P, c, u, \mathcal{T}_{2,\varepsilon,c})$ be the symmetric mixed equilibrium specified in Theorem 65.[2] It suffices to show that there exists $c > 1$ such that

$$\limsup_{\varepsilon \to 0} \frac{\mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}}; \mathbf{w})]}{\mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{r}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}}; \mathbf{w})]} > 1.$$

We apply Lemma 225 and Lemma 226 to see that:

$$\begin{aligned}
&\limsup_{\varepsilon \to 0} \frac{\mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}}; \mathbf{w})]}{\mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{r}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}}; \mathbf{w})]}[\mathrm{UW}(M^{\mathrm{E}}; \mathbf{w})] \\
&\geq \limsup_{\varepsilon \to 0} \frac{\frac{1}{2} \cdot \left(1 - \left(\frac{(3-c)(c-1)}{2-c}\right)^2\right) \cdot W \cdot \frac{c \cdot (3-c) \cdot (c-1) \cdot (1+\varepsilon)}{2(2-c)}}{\frac{W}{2} \cdot (\varepsilon + c(1+\varepsilon) - 1)} \\
&= \limsup_{\varepsilon \to 0} \frac{\left(1 - \left(\frac{(3-c)(c-1)}{2-c}\right)^2\right) \cdot \frac{c \cdot (3-c) \cdot (c-1) \cdot (1+\varepsilon)}{2(2-c)}}{(\varepsilon + c(1+\varepsilon) - 1)} \\
&= \frac{\left(1 - \left(\frac{(3-c)(c-1)}{2-c}\right)^2\right) \cdot \frac{c \cdot (3-c) \cdot (c-1)}{2(2-c)}}{c-1} \\
&= \left(1 - \left(\frac{(3-c)(c-1)}{2-c}\right)^2\right) \cdot \frac{c \cdot (3-c)}{2(2-c)}.
\end{aligned}$$

It is easy to see that there exists $c > 1$ such that the above expression is strictly greater than 1, as desired. □

Using Lemma 225 and Lemma 226, we prove Chapter 64.

Proof of Chapter 64. We construct instances with two types where user welfare of random recommendations exceeds the user welfare of engagement-based optimization. Interestingly, users are *well-separated* in these instances. Consider Chapter 5 with $P = 2$, $\gamma = 0$, and type space $\mathcal{T}_{2,\varepsilon,c}$ defined by (F.8). It suffices to show that if $c \geq 1.5$, then it holds that:

$$\mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}}; \mathbf{w})] < \mathbb{E}_{\mathbf{w} \sim (\mu^{\mathrm{r}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{R}}; \mathbf{w})].$$

[2] A simple consequence of the formulation in Chapter 9.5.1 is that $\mu^{\mathrm{i}}(P, c, u, \mathcal{T}_{2,\varepsilon,c})$ is independent of c and ε for the type spaces that we have constructed.

We apply Lemma 225 and Lemma 226 to see that:

$$\begin{aligned}\mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{e}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{E}};\mathbf{w})] &\leq \frac{3}{16}\cdot W\cdot(c-1)\cdot(1+\varepsilon)\\ &< \frac{W}{2}\cdot(c-1)\cdot(1+\varepsilon)\\ &\leq \frac{W}{2}\cdot(\varepsilon+c(1+\varepsilon)-1)\\ &= \mathbb{E}_{\mathbf{w}\sim(\mu^{\mathrm{r}}(2,c,u,\mathcal{T}_{2,\varepsilon,c}))^2}[\mathrm{UW}(M^{\mathrm{R}};\mathbf{w})],\end{aligned}$$

as desired. $\square$

Appendix G

Appendix for Chapter 10

This chapter is based on *"Flattening Supply Chains: When do Technology Improvements lead to Disintermediation?"*, which is joint work with S. Nageeb Ali, Nicole Immorlica and Brendan Lucier.

G.1 Useful lemmas

Lemma 227. *Suppose that g is continuously differentiable, strictly convex, and satisfies $g(0) = g'(0) = 0$. Then it holds that:*

$$\lim_{w \to^+ 0} \frac{g(w)}{g'(w)} = 0.$$

Proof. Since $g(w)$ and $g'(w)$ are both positive for $w > 0$, it suffices to show that:

$$\lim_{w \to^+ 0} \frac{g(w)}{g'(w)} \leq 0.$$

Using convexity, we know that:

$$0 = g(0) \geq g(w) + g'(w)(0 - w) = g(w) - wg'(w),$$

which means that $g(w) \leq w \cdot g'(w)$. This means that:

$$\lim_{w \to^+ 0} \frac{g(w)}{g'(w)} \leq \lim_{w \to^+ 0} w = 0$$

as desired. □

G.1.1 Properties of $\max_{w \geq 0}(w - \nu g(w)))$

Lemma 228. *Suppose that g is continuously differentiable, strictly convex, satisfies $g(0) = g'(0) = 0$, and satisfies $\lim_{w \to \infty} g(w) = \lim_{w \to \infty} g'(w) = \infty$. For any $\nu > 0$, then $\max_{w \geq 0}(w - \nu g(w))$ has a unique optima, which is in $(0, \infty)$.*

Proof. Using that g is convex, we see that w^* is a maximum of $\max_{w \geq 0}(w - \nu g(w))$ if and only if:

$$g'(w) = \frac{1}{\nu}.$$

Using the other conditions on g, we see that this occurs at a unique value $w^* \in (0, \infty)$. $\square$

Lemma 229. *Suppose that g is continuously differentiable, strictly convex, satisfies $g(0) = g'(0) = 0$, and satisfies $\lim_{w \to \infty} g(w) = \lim_{w \to \infty} g'(w) = \infty$. Let $w^*(\nu)$ be an optima of $\max_{w \geq 0}(w - \nu g(w))$. Then, it holds that:*

$$\frac{\partial w^*(\nu)}{\partial \nu} < 0.$$

Moreover, for any $w \in (0, \infty)$, there exists a unique value $\nu > 0$ such that $w^*(\nu) = w$.

Proof. By Lemma 228, we know that $\max_{w \geq 0}(w - \nu g(w))$ has a unique optima, so $w^*(\nu)$ is uniquely defined. Using that g is convex, we see that:

$$g'(w^*(\nu)) = \frac{1}{\nu}.$$

This, coupled with the other conditions on g, give us the desired result. $\square$

Lemma 230. *Suppose that g is continuously differentiable, strictly convex, satisfies $g(0) = g'(0) = 0$, and satisfies $\lim_{w \to \infty} g(w) = \lim_{w \to \infty} g'(w) = \infty$. The derivative of $\max_{w \geq 0}(w - \nu g(w)))$ with respect to ν is equal to $-g(\arg\max_{w \geq 0}(w - \nu g(w)))$.*

Proof. We apply Lemma 228 and let $w^*(\nu)$ be the unique maximizer of $\max_{w \geq 0}(w - \nu g(w))$. By the envelope theorem, we see that

$$\frac{\partial}{\partial \nu}\left(\max_{w \geq 0}(w - \nu g(w))\right) = -g(w^*),$$

as desired. $\square$

G.1.2 Properties of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$

Lemma 231. *Suppose that g is continuously differentiable, strictly convex, satisfies $g(0) = g'(0) = 0$, and satisfies $\lim_{w \to \infty} g(w) = \lim_{w \to \infty} g'(w) = \infty$. The derivative of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ with respect to ν is equal to:*

$$g'\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right)\left(\frac{g\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right)}{g'\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right)} - \frac{g(\arg\max_{w \geq 0}(w - \nu g(w)))}{g'(\arg\max_{w \geq 0}(w - \nu g(w)))}\right).$$

Moreover, the sign of the derivative is:

$$\begin{cases} 0 & \textit{if} \quad \frac{g(w^*)}{g'(w^*)} = \alpha \\ \textit{positive} & \textit{if} \quad \frac{g(w^*)}{g'(w^*)} < \alpha \\ \textit{negative} & \textit{if} \quad \frac{g(w^*)}{g'(w^*)} > \alpha, \end{cases}$$

where w^* is the maximizer of $\max_{w \geq 0}(w - \nu g(w)))$.

Proof. We apply Lemma 228 and let w^* be the unique maximizer of $\max_{w \geq 0}(w - \nu g(w)))$ (note that this depends on ν). Using the first-order condition, we observe that:

$$g'(w^*) = \frac{1}{\nu}.$$

By Lemma 230, we observe that the derivative of $\max_{w \geq 0}(w - \nu g(w)))$ with respect to ν is $g(w^*)$.

Now, we can take a derivative of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ to obtain:

$$\begin{aligned}
&\frac{\partial}{\partial \nu}\left(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))\right) \\
&= g\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right) - \nu g'\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right) g(w^*). \\
&= g\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right) - \frac{g'\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right) g(w^*)}{g'(w^*)} \\
&= g'\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right)\left(\frac{g\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right)}{g'\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right)} - \frac{g(w^*)}{g'(w^*)}\right) \\
&= g'\left(\alpha + (w^* - \nu g(w^*))\right)\left(\frac{g\left(\alpha + (w^* - \nu g(w^*))\right)}{g'\left(\alpha + (w^* - \nu g(w^*))\right)} - \frac{g(w^*)}{g'(w^*)}\right)
\end{aligned}$$

Observe that the derivative has the same sign as

$$\frac{g\left(\alpha + (w^* - \nu g(w^*))\right)}{g'\left(\alpha + (w^* - \nu g(w^*))\right)} - \frac{g(w^*)}{g'(w^*)}.$$

Since g is strictly log-concave, we know that $\frac{g(w)}{g'(w)}$ is strictly increasing in w. This means that the sign of the derivative is the same as the sign of

$$\alpha + (w^* - \nu g(w^*)) - w^* = \alpha - \nu \cdot g(w^*).$$

Using the first-order condition, this is equal to:

$$\alpha - \frac{g(w^*)}{g'(w^*)}.$$

This proves the desired statement. □

Lemma 232. *Consider the setup of Theorem 76. Then, it holds that $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ is U-shaped in ν. Moreover, it holds that*

$$\min_{\nu \geq 0} \left(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))) \right) = \alpha.$$

Furthermore, there is a unique global optimum $\nu \in [0, \infty)$, and this value is the unique solution to

$$\frac{g(\arg\max(w - \nu g(w)))}{g'(\arg\max(w - \nu g(w)))} = \alpha.$$

Proof. We apply Lemma 228 and let w^* be the unique maximizer of $\max_{w \geq 0}(w - \nu g(w)))$ (note that this depends on ν). We use Lemma 231 to see that the sign of the derivative of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ with respect to ν is:

$$\begin{cases} 0 & \text{if } \frac{g(w^*)}{g'(w^*)} = \alpha \\ \text{positive} & \text{if } \frac{g(w^*)}{g'(w^*)} < \alpha \\ \text{negative} & \text{if } \frac{g(w^*)}{g'(w^*)} > \alpha, \end{cases}$$

We next show there is a unique value of ν such that $\frac{g(w^*)}{g'(w^*)} = \alpha$. Since g is strictly log-concave, we know that $\frac{g(w)}{g'(w)}$ is strictly increasing in w. By Lemma 227, we know that $\lim_{w \to 0} \frac{g(w)}{g'(w)} = 0$ and by the assumed condition in the theorem statement, we know that:

$$\lim_{w \to \infty} \frac{g(w)}{g'(w)} \geq \lim_{w \to \infty} \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g'(w)} = \infty.$$

Since $\frac{g(w)}{g'(w)}$ is continuous, this means that there exists $w > 0$ such that $\frac{g(w)}{g'(w)} = \alpha$. Using Lemma 229, this means that there exists a unique value of $\nu > 0$ such that $w^* = w$.

Next, we show that $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))))$ is U-shaped with a unique global minimum. Let $w^*(\nu')$ be the unique optimum of $\max_{w \geq 0}(w - \nu' g(w))$ (Lemma 228).

- For prices $\nu' < \nu$ below this threshold, by Lemma 229, we observe that $w^*(\nu') > w^*(\nu)$. Using log-concavity of g, this means that:

$$\frac{g(w^*(\nu'))}{g'(w^*(\nu'))} > \frac{g(w^*)}{g'(w^*)} = \alpha,$$

which means that the derivative is negative. Applying this for every $\nu' < \nu$ means that

$$\left(\nu' \cdot g(\alpha + \max_{w \geq 0}(w - \nu' g(w))) \right) > \alpha.$$

- For prices $\nu' > \nu$ below this threshold, by Lemma 229, we observe that $w^*(\nu') < w^*(\nu)$. Using log-concavity of g, this means that:

$$\frac{g(w^*(\nu'))}{g'(w^*(\nu'))} < \frac{g(w^*)}{g'(w^*)} = \alpha,$$

which means that the derivative is positive. Applying this for every $\nu' < \nu$ means that

$$\left(\nu' \cdot g(\alpha + \max_{w \geq 0}(w - \nu' g(w)))\right) > \alpha.$$

Finally, at the value of ν such that $\frac{g(w^*)}{g'(w^*)} = \alpha$, it holds that:

$$\begin{aligned}\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))) &= \nu \cdot g(\alpha + w^* - \nu g(w^*))) \\ &= \nu \cdot g(w^*) \\ &= \frac{g(w^*)}{g'(w^*)} \\ &= \alpha.\end{aligned}$$

□

G.2 Proofs for Chapter 10.2

The main lemma is the following characterization of the equilibria in the subgame between the intermediary and consumers.

Lemma 233. *Suppose that suppliers choose prices $\nu_1, \ldots, \nu_P$ and consider the subgame between the intermediary and consumers (Stages 2-3). Under the tiebreaking assumptions discussed in Chapter 10.2.2, there exists a unique pure strategy equilibrium in this subgame which takes the following form. Let $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$, and consider the condition*

$$\nu \cdot g\left(\alpha + \max_{w \geq 0}(w - \nu g(w))\right) > \alpha C. \tag{G.1}$$

- *If (G.1) holds, then $w_m = 0$. Moreover, for all $j \in [C]$, it holds that $a_j = D$, $w_{c,j} = \arg\max_{w \geq 0}(w - \nu g(w))$. Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the consumer chooses $i_j = 0$. Otherwise, the consumer chooses $i_j = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index).*

- *If (G.1) does not hold, then $w_m = w_{c,j} = \alpha + \max_{w \geq 0}(w - \nu g(w))$. Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the intermediary chooses $i_m = 0$; otherwise, the intermediary chooses $i_m = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index). Finally, it holds that $a_j = M$ and $w_{c,j} = w_m$ for all $j \in [C]$.*

Proof. Recall that when consumers or the intermediary produce content, they choose the option that minimizes their production costs. If $\nu^H + \min_{i \in [P]} \nu_i < \nu_0$, they leverage the technology of the supplier who offers the lowest price, and otherwise, they produce content without using the technology. This means that they face production costs $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$.

When consumer j chooses $a_j = D$, then they maximize their utility and thus produce content $w^*(\nu) = \arg\max(w - \nu g(w))$ and achieve utility $\max(w - \nu g(w))$. Since the consumer pays the intermediary a fee of α, the intermediary must produce content satisfying $w' \geq \alpha + \max_{w \geq 0}(w - \nu g(w))$ to incentivize the consumer to choose $a_j = M$. Producing content $w' \geq \alpha + \max_{w \geq 0}(w - \nu g(w))$ would incentivize all of the consumers to choose the intermediary, so the intermediary would earn utility

$$\alpha \cdot C - \nu \cdot g(w').$$

This also means that the intermediary prefers producing content $\alpha + \max_{w \geq 0}(w - \nu g(w))$ over any $w' > \alpha + \max_{w \geq 0}(w - \nu g(w))$ in order to minimize costs. The intermediary prefers producing this content over producing content $w = 0$ which would not attract any consumers if and only if:

$$\alpha \cdot C - \nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))) \geq 0.$$

This, coupled with the tiebreaking rules, proves the desired statement. □

Using this lemma, we can characterize pure strategy equilibria in our game.

Lemma 234. *Under the tiebreaking assumptions in Chapter 10.2.2, there exists a pure strategy equilibrium which takes the following form. All suppliers choose the price $\nu_i = \nu^*$ for $i \in [P]$, and the intermediary and consumers choose actions according to the subgame equilibrium constructed in Lemma 233.*

Proof. If $\nu_0 < \nu^* + \nu^H$, then by Lemma 233, then consumers and the intermediary produce content without the technology, so suppliers all have zero utility regardless of what price they choose.

If $\nu_0 \geq \nu^* + \nu^H$, then consumers and the intermediary choose manual production if $\min_{i \in [P]} \nu_i > \nu_0$ and otherwise choose supplier $\text{argmin}_{i \in [P]} \nu_i$. We show that $\nu_i = \nu^*$ for $i \in [P]$ is an equilibrium. At this equilibrium, note that all of the suppliers earn zero utility. If a supplier deviates to $\nu_i < \nu^*$, then by Lemma 233, production would be done through the supplier which would result in negative utility. Deviating to $\nu > \nu^*$ would result in zero utility. Thus, there are no profitable deviations for the suppliers. □

Lemma 234 implies both Theorem 72 and Theorem 73.

Proof of Theorem 72. This follows from Lemma 234. □

Proof of Theorem 73. We show that the actions of the intermediary and consumers, as well as the production cost $\min(\nu^H + \min_{i\in[P]\nu_i}, \nu_0)$ is the same at every pure strategy equilibrium. To do this, we show that these values are the same as at the pure strategy equilibrium constructed in Lemma 234.

Suppose that $\nu_0 < \nu^* + \nu^H$. If $\min_{i\in[P]\nu_i} > \nu_0$, then production is done without the technology, and the intermediary and consumers choose the same actions as in the equilibrium in Lemma 234. If $\min_{i\in[P]\nu_i} \leq \nu_0$, then by Lemma 234, production would done through supplier $\mathrm{argmin}_{i\in[P]\nu_i} \leq \nu_0$, and that supplier would earn negative utility. This is not possible because the supplier could deviate to $\nu_i = \nu^*$ and earn zero utility.

Now, suppose that $\nu_0 \geq \nu^* + \nu^H$. In this case, assume for sake of contradiction that $\min_{i\in[P]\nu_i} \neq \nu^*$. If $\min_{i\in[P]\nu_i} > \nu^*$, then using Lemma 233, a supplier could earn higher profit by choosing $\nu = \min(\min_{i\in[P]\nu_i}, \nu_0) - \varepsilon$ for sufficiently small ε, which is a contradiction. If $\min_{i\in[P]\nu_i} < \nu^*$, then using Lemma 233, the supplier $\mathrm{argmin}_{i\in[P]\nu_i} < \nu^*$ with lowest index could earn higher utility by instead choosing $\nu = \nu^*$, which is a contradiction. This means that $\min_{i\in[P]\nu_i} = \nu^*$, so by Lemma 233 the intermediary and the consumers take the same actions as in the equilibrium in Lemma 234. □

G.3 Proofs for Chapter 10.3

G.3.1 Analysis of specific cost function families

We first analyze the derivative and log-derivative of several families of cost functions.

Lemma 235. *The following statements hold:*

1. *For $g(w) = w^\beta$ for $\beta > 1$, the derivative is $g'(w) = \beta \cdot w^{\beta-1}$, and the log-derivative is $\frac{g'(w)}{g(w)} = \frac{\beta}{w}$*

2. *For $g(w) = w^\beta \cdot e^{\sqrt{w}}$ for $\beta \geq 1$, the derivative is*
$$g'(w) = \beta \cdot w^{\beta-1} \cdot e^{\sqrt{w}} + 0.5 \cdot w^{\beta-\frac{1}{2}} \cdot e^{\sqrt{w}}$$
and the log-derivative is:
$$\frac{g'(w)}{g(w)} = \frac{\beta + 0.5 \cdot w^{\frac{1}{2}}}{w} = \frac{\beta}{w} + \frac{0.5}{w^{\frac{1}{2}}}.$$

3. *For $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$, the derivative is:*
$$g'(w) = \beta \cdot w^{\beta-1} \cdot (\log(w+1)^\gamma) + \frac{w^\beta}{w+1} \cdot \gamma(\log(w+1)^{\gamma-1})$$
and the log-derivative is:
$$\frac{g'(w)}{g(w)} = \frac{\beta \cdot \log(w+1) + \frac{w}{w+1} \cdot \gamma}{w \cdot \log(w+1)} = \frac{\beta}{w} + \frac{\gamma}{(w+1)(\log(w+1))}$$

4. *For $g(w) = w^\beta \cdot e^w$ for $\beta \geq 1$, the derivative is $g'(w) = \beta \cdot w^{\beta-1} \cdot e^w + w^\beta \cdot e^w$ and the log-derivative is:*
$$\frac{g'(w)}{g(w)} = 1 + \frac{\beta}{w}.$$

Proof. We analyze each family of cost functions separately.

Family 1: $g(w) = w^\beta$ for $\beta > 1$. The derivative is $g'(w) = \beta \cdot w^{\beta-1}$. The log derivative is:

$$\frac{g'(w)}{g(w)} = \frac{\beta}{w}.$$

Family 2: $g(w) = w^\beta \cdot e^{\sqrt{w}}$. The derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot e^{\sqrt{w}} + w^\beta \cdot e^{\sqrt{w}} \cdot 0.5 \cdot w^{-\frac{1}{2}} = \beta \cdot w^{\beta-1} \cdot e^{\sqrt{w}} + 0.5 \cdot w^{\beta-\frac{1}{2}} \cdot e^{\sqrt{w}}.$$

The log-derivative is:

$$\frac{g'(w)}{g(w)} = \frac{\beta \cdot w^{\beta-1} \cdot e^{\sqrt{w}} + 0.5 \cdot w^{\beta-\frac{1}{2}} \cdot e^{\sqrt{w}}}{w^\beta \cdot e^{\sqrt{w}}} = \frac{\beta + 0.5 \cdot w^{\frac{1}{2}}}{w} = \frac{\beta}{w} + \frac{0.5}{w^{\frac{1}{2}}}.$$

Family 3: $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$. The derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot (\log(w+1)^\gamma) + \frac{w^\beta}{w+1} \cdot \gamma(\log(w+1)^{\gamma-1}).$$

The log-derivative is:

$$\begin{aligned} \frac{g'(w)}{g(w)} &= \frac{\beta \cdot w^{\beta-1} \cdot (\log(w+1)^\gamma) + \frac{w^\beta}{w+1} \cdot \gamma(\log(w+1)^{\gamma-1})}{w^\beta \cdot (\log(w+1)^\gamma)} \\ &= \frac{\beta \cdot \log(w+1) + \frac{w}{w+1} \cdot \gamma}{w \cdot \log(w+1)} = \frac{\beta}{w} + \frac{\gamma}{(w+1)(\log(w+1))}. \end{aligned}$$

Family 4: $g(w) = w^\beta \cdot e^w$. The derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot e^w + w^\beta \cdot e^w.$$

The log-derivative is:

$$\frac{g'(w)}{g(w)} = \frac{\beta \cdot w^{\beta-1} \cdot e^w + w^\beta \cdot e^w}{w^\beta \cdot e^w} = \frac{\beta + w}{w} = 1 + \frac{\beta}{w}.$$

□

Using Lemma 235, we prove that several families of cost functions satisfy the assumptions for Theorem 75.

Proposition 236. *The following cost functions satisfy the assumptions of Theorem 75: (1) $g(w) = w^\beta$ for $\beta > 1$, (2) $g(w) = w^\beta \cdot e^{\sqrt{w}}$ for $\beta \geq 1$, (3) $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$, and (4) $g(w) = w^\beta \cdot e^w$ for $\beta \geq 1$.*

Proof. We analyze each family of cost functions separately. It suffices to prove that these functions are strictly increasing, continuously differentiable, strictly convex, satisfy $g(0) = g'(0) = 0$ and $\lim_{w\to\infty} g(w) = \lim_{w\to\infty} g'(w) = \infty$, and strictly log-concave.

Family 1: $g(w) = w^\beta$ for $\beta > 1$. By Lemma 235, the derivative is $g'(w) = \beta \cdot w^{\beta-1}$. This means that g is strictly increasing and continuously differentiable. Moreover, the derivative is increasing, so the function is strictly convex. We also see that $g(0) = g'(0) = 0$ and $\lim_{w\to\infty} g(w) = \lim_{w\to\infty} g'(w) = \infty$. To show that g is strictly log-concave, it suffices to show that the log-derivative is strictly decreasing. By Lemma 235, the log-derivative is

$$\frac{g'(w)}{g(w)} = \frac{\beta}{w},$$

which is strictly decreasing as desired.

Family 2: $g(w) = w^\beta \cdot e^{\sqrt{w}}$. By Lemma 235, the derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot e^{\sqrt{w}} + 0.5 \cdot w^{\beta-\frac{1}{2}} \cdot e^{\sqrt{w}}.$$

This means that g is strictly increasing and continuously differentiable. Moreover, the derivative is increasing, so the function is strictly convex. We also see that $g(0) = g'(0) = 0$ and $\lim_{w\to\infty} g(w) = \lim_{w\to\infty} g'(w) = \infty$. To show that g is strictly log-concave, it suffices to show that the log-derivative is strictly decreasing. By Lemma 235, the log-derivative is

$$\frac{g'(w)}{g(w)} = \frac{\beta + 0.5 \cdot w^{\frac{1}{2}}}{w} = \frac{\beta}{w} + \frac{0.5}{w^{\frac{1}{2}}}$$

which is strictly decreasing as desired.

Family 3: $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$. By Lemma 235, the derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot (\log(w+1)^\gamma) + \frac{w^\beta}{w+1} \cdot \gamma(\log(w+1)^{\gamma-1}).$$

This means that g is strictly increasing and continuously differentiable. Since $\frac{w^\beta}{w+1}$ is increasing, the derivative is increasing, so the function is strictly convex. We also see that $g(0) = g'(0) = 0$ and $\lim_{w\to\infty} g(w) = \lim_{w\to\infty} g'(w) = \infty$. To show that g is strictly log-concave, it suffices to show that the log-derivative is strictly decreasing. By Lemma 235, the log-derivative is and the log-derivative is:

$$\frac{g'(w)}{g(w)} = \frac{\beta \cdot \log(w+1) + \frac{w}{w+1} \cdot \gamma}{w \cdot \log(w+1)} = \frac{\beta}{w} + \frac{\gamma}{(w+1)(\log(w+1))}$$

which is strictly decreasing as desired.

Family 4: $g(w) = w^\beta \cdot e^w$**.** By Lemma 235, the derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot e^w + w^\beta \cdot e^w.$$

This means that g is strictly increasing and continuously differentiable. Moreover, the derivative is increasing, so the function is strictly convex. We also see that $g(0) = g'(0) = 0$ and $\lim_{w\to\infty} g(w) = \lim_{w\to\infty} g'(w) = \infty$. To show that g is strictly log-concave, it suffices to show that the log-derivative is strictly decreasing. By Lemma 235, the log-derivative is

$$\frac{g'(w)}{g(w)} = \frac{\beta + w}{w} = 1 + \frac{\beta}{w},$$

which is strictly decreasing as desired. □

We next identify several cost functions which satisfy the assumptions for Theorem 76.

Proposition 237. *The following cost functions satisfy the assumptions of Theorem 76: (1) $g(w) = w^\beta$ for $\beta > 1$, (2) $g(w) = w^\beta \cdot e^{\sqrt{w}}$ for $\beta \geq 1$, and (3) $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$.*

Proof. By Proposition 236, all three of these families satisfy the assumptions of Theorem 75. Thus, it suffices to show that

$$\lim_{w\to\infty} \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g'(w)} = \infty.$$

Family 1: $g(w) = w^\beta$ **for** $\beta > 1$**.** By Lemma 235, the derivative is $g'(w) = \beta \cdot w^{\beta-1}$ and the log-derivative is

$$\frac{g'(w)}{g(w)} = \frac{\beta}{w}.$$

This means that:

$$\lim_{w\to\infty} \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g'(w)} = \lim_{w\to\infty} \frac{\left(w - \frac{w}{\beta}\right)^\beta}{\beta \cdot w^{\beta-1}} = (1 - \frac{1}{\beta})^\beta \cdot \lim_{w\to\infty} \frac{w^\beta}{\beta \cdot w^{\beta-1}} = (1 - \frac{1}{\beta})^\beta \cdot \lim_{w\to\infty} \frac{w}{\beta} = \infty,$$

as desired.

Family 2: $g(w) = w^\beta \cdot e^{\sqrt{w}}$**.** By Lemma 235, the derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot e^{\sqrt{w}} + 0.5 \cdot w^{\beta-\frac{1}{2}} \cdot e^{\sqrt{w}}$$

and the log-derivative is

$$\frac{g'(w)}{g(w)} = \frac{\beta + 0.5 \cdot w^{\frac{1}{2}}}{w} = \frac{\beta}{w} + \frac{0.5}{w^{\frac{1}{2}}}.$$

This means that:

$$\frac{g(w)}{g'(w)} = \frac{1}{\frac{\beta}{w} + \frac{0.5}{w^{\frac{1}{2}}}} = \frac{w}{\beta + 0.5\sqrt{w}}$$

This means that:

$$\begin{aligned}
&\lim_{w\to\infty} \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g'(w)} \\
&= \lim_{w\to\infty} \frac{g(w)}{g'(w)} \cdot \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g(w)} \\
&= \lim_{w\to\infty} \frac{w}{\beta + 0.5\sqrt{w}} \cdot \frac{\left(w - \frac{g(w)}{g'(w)}\right)^{\beta} \cdot e^{\sqrt{w - \frac{g(w)}{g'(w)}}}}{w^{\beta} e^{\sqrt{w}}} \\
&= \lim_{w\to\infty} \frac{w}{\beta + 0.5\sqrt{w}} \cdot \left(\frac{w - \frac{g(w)}{g'(w)}}{w}\right)^{\beta} \cdot e^{\sqrt{w - \frac{g(w)}{g'(w)}} - \sqrt{w}} \\
&= \lim_{w\to\infty} \frac{w}{\beta + 0.5\sqrt{w}} \cdot \left(1 - \frac{1}{\beta + 0.5\sqrt{w}}\right)^{\beta} \cdot e^{\left(\sqrt{w - \frac{w}{\beta+0.5\sqrt{w}}} - \sqrt{w}\right)}.
\end{aligned}$$

Since $\lim_{w\to\infty} \frac{w}{\beta+0.5\sqrt{w}} = \infty$ and $\lim_{w\to\infty} \left(1 - \frac{1}{\beta+0.5\sqrt{w}}\right)^{\beta} = 1$, it suffices to show that

$$\lim_{w\to\infty} e^{\left(\sqrt{w - \frac{w}{\beta+0.5\sqrt{w}}} - \sqrt{w}\right)} = e^{-1}.$$

It suffices to show that

$$\lim_{w\to\infty} \left(\sqrt{w - \frac{w}{\beta + 0.5\sqrt{w}}} - \sqrt{w}\right) = -1,$$

which can be rewritten as:

$$\lim_{w\to\infty} \frac{\sqrt{1 - \frac{1}{\beta+0.5\sqrt{w}}} - 1}{w^{-1/2}} = -1.$$

Using L'Hôpital's rule, we see that this is equal to:

$$\begin{aligned}
\lim_{w\to\infty} \frac{\sqrt{1 - \frac{1}{\beta+0.5\sqrt{w}}} - 1}{w^{-1/2}} &= \lim_{w\to\infty} \frac{\frac{0.125}{\sqrt{w}\sqrt{1-\frac{1}{c+0.5\sqrt{w}}}\cdot(\beta+0.5\sqrt{w})^2}}{-0.5 \cdot w^{-3/2}} \\
&= \lim_{w\to\infty} -\frac{0.25w}{\sqrt{1 - \frac{1}{c+0.5\sqrt{w}}} \cdot (\beta + 0.5\sqrt{w})^2} \\
&= \lim_{w\to\infty} -\frac{0.25w}{\beta^2 + 0.25w + \beta\sqrt{w}} \\
&= -1.
\end{aligned}$$

Family 3: $g(w) = w^\beta \cdot (\log(w+1)^\gamma)$ for any $\beta, \gamma > 1$. By Lemma 235, the derivative is

$$g'(w) = \beta \cdot w^{\beta-1} \cdot (\log(w+1)^\gamma) + \frac{w^\beta}{w+1} \cdot \gamma(\log(w+1)^{\gamma-1})$$

and the log-derivative is:

$$\frac{g'(w)}{g(w)} = \frac{\beta \cdot \log(w+1) + \frac{w}{w+1} \cdot \gamma}{w \cdot \log(w+1)} = \frac{\beta}{w} + \frac{\gamma}{(w+1)(\log(w+1))} = \frac{1}{w} \cdot \left(\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))} \right).$$

This means that:

$$\frac{g(w)}{g'(w)} = \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}}.$$

This means that:

$$\begin{aligned}
&\lim_{w\to\infty} \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g'(w)} \\
&= \lim_{w\to\infty} \frac{g\,(w)}{g'(w)} \cdot \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g(w)} \\
&= \lim_{w\to\infty} \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}} \cdot \frac{g\left(w - \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}}\right)}{g(w)} \\
&= \lim_{w\to\infty} \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}} \cdot \left(\frac{w - \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}}}{w} \right)^\beta \cdot \left(\frac{\log\left(1 + w - \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}}\right)}{\log(1+w)} \right)^\gamma.
\end{aligned}$$

We analyze each term separately. Note that:

$$\lim_{w\to\infty} \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}} = \infty$$

and

$$\lim_{w\to\infty} \left(\frac{w - \frac{w}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}}}{w} \right)^\beta = \lim_{w\to\infty} \left(1 - \frac{1}{\beta + \frac{\gamma \cdot w}{(w+1)(\log(w+1))}} \right)^\beta = \lim_{w\to\infty} \left(1 - \frac{1}{\beta}\right)^\beta$$

and

$$\lim_{w\to\infty}\left(\frac{\log\left(1+w-\frac{w}{\beta+\frac{\gamma\cdot w}{(w+1)(\log(w+1))}}\right)}{\log(1+w)}\right)^{\gamma} \geq \lim_{w\to\infty}\left(\frac{\log\left(1+w-\frac{w}{\beta}\right)}{\log(1+w)}\right)^{\gamma}$$
$$\geq \lim_{w\to\infty}\left(\frac{\log\left(1-\frac{1}{\beta}+w-\frac{w}{\beta}\right)}{\log(1+w)}\right)^{\gamma}$$
$$\geq \lim_{w\to\infty}\left(\frac{\log\left(1-\frac{1}{\beta}\right)+\log\left(1+w\right)}{\log(1+w)}\right)^{\gamma}$$
$$= 1.$$

This proves the desired statement. □

G.3.2 Proof of Theorem 74

We prove Theorem 74 as a corollary of Theorem 76.

Proof. By Proposition 237, we know that $g(w) = w^{\beta}$ for $\beta > 1$ satisfies the conditions of Theorem 76. This implies the existence of thresholds $0 < T_L(C,\alpha,\beta) < T_U(C,\alpha,\beta) < \infty$ such that the intermediary usage satisfies

$$\sum_{j=1}^{C}\mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) < T_L(C,\alpha,\beta) \\ C & \text{if } \min(\nu^* + \nu^H, \nu_0) \in [T_L(C,\alpha,\beta), T_U(C,\alpha,\beta)] \\ 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) > T_U(C,\alpha,\beta) \end{cases},$$

where $T_L(C,\alpha,\beta)$ and $T_U(C,\alpha,\beta)$ are the two unique solutions to

$$\nu \cdot g(\alpha + \max_w(w - \nu g(w)) - \alpha C = 0.$$

Using the structure of $g(w) = w^{\beta}$, we see that

$$\max_w(w - \nu g(w)) = \nu^{-\frac{1}{\beta-1}}\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right).$$

When we plug this into the above expression, we obtain:

$$\nu \cdot \left(\alpha + \nu^{-\frac{1}{\beta-1}}\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right)\right)^{\beta} - \alpha C = 0.$$

This can be rewritten as:

$$\nu^{\frac{1}{\beta}} \cdot \left(\alpha + \nu^{-\frac{1}{\beta-1}}\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right)\right) - \alpha^{\frac{1}{\beta}} C^{\frac{1}{\beta}} = 0.$$

This can be rewritten as:

$$\nu^{\frac{1}{\beta}}\alpha^{\frac{\beta-1}{\beta}} + \nu^{-\frac{1}{\beta(\beta-1)}}\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right)\alpha^{-\frac{1}{\beta}} - C^{\frac{1}{\beta}} = 0,$$

as desired. □

G.3.3 Proof of Theorem 75

First, we prove the following lemma which characterizes the sign of the derivative of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$.

Lemma 238. *Consider the setup of Theorem 75. Then, there exist a (possibly infinite or negative) threshold ν^T such that the sign of the derivative of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ satisfies:*

$$\begin{cases} 0 & \textit{if} \quad \nu = \nu^T \\ \textit{positive} & \textit{if} \quad \nu > \nu^T \\ \textit{negative} & \textit{if} \quad \nu < \nu^T. \end{cases}$$

Proof. We apply Lemma 228 and let $w^*(\nu)$ be the unique maximizer of $\max_{w \geq 0}(w - \nu g(w)))$. We use Lemma 231 to see that the sign of the derivative of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ with respect to ν is:

$$\begin{cases} 0 & \text{if} \quad \frac{g(w^*(\nu))}{g'(w^*(\nu))} = \alpha \\ \text{positive} & \text{if} \quad \frac{g(w^*(\nu))}{g'(w^*(\nu))} < \alpha \\ \text{negative} & \text{if} \quad \frac{g(w^*(\nu))}{g'(w^*(\nu))} > \alpha, \end{cases}$$

Since g is strictly log-concave, we know that $\frac{g(w)}{g'(w)}$ is strictly increasing in w. Using Lemma 229, we see that $\frac{g(w^*(\nu))}{g'(w^*(\nu))}$ is strictly decreasing in ν. This guarantees that there exists a (possibly infinite or negative) threshold ν^T such that the sign of the derivative is:

$$\begin{cases} 0 & \text{if} \quad \nu = \nu^T \\ \text{positive} & \text{if} \quad \nu > \nu^T \\ \text{negative} & \text{if} \quad \nu < \nu^T \end{cases}$$

as desired. □

We now prove Theorem 75.

Proof of Theorem 75. By Lemma 234, disintermediation occurs if and only if $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))) > \alpha C$. By Lemma 238, there exist a (possibly infinite or negative) threshold ν^T such that the sign of the derivative of $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ satisfies:

$$\begin{cases} 0 & \text{if} \quad \nu = \nu^T \\ \text{positive} & \text{if} \quad \nu > \nu^T \\ \text{negative} & \text{if} \quad \nu < \nu^T \end{cases}$$

This implies that there exist (possibly infinite or negative) thresholds $T_L(C, \alpha, g)$ and $T_U(C, \alpha, g)$ such that

$$\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) < T_L(C, \alpha, \beta) \\ C & \text{if } \min(\nu^* + \nu^H, \nu_0) \in [T_L(C, \alpha, \beta), T_U(C, \alpha, \beta)] \\ 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) > T_U(C, \alpha, \beta) \end{cases}$$

as desired. □

G.3.4 Proof of Theorem 76

First, we prove the following lemma that shows that the disintermediation boundary has exactly two solutions.

Lemma 239. *Consider the setup of Theorem 76. The equation*

$$\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))) = \alpha C$$

has exactly two solutions.

Proof. We apply Lemma 228 and let $w^*(\nu)$ be the unique maximizer of $\max_{w \geq 0}(w - \nu g(w)))$. We use Lemma 232 to see that $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$ is U-shaped and has global minimum

$$\min_{\nu > 0}(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))) = \alpha < \alpha C.$$

To show that $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))) = \alpha C$ has exactly two solutions, it suffices to show that

$$\lim_{\nu \to \infty}(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))) = \infty = \lim_{\nu \to 0}(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))).$$

First, we take a limit as $\nu \to \infty$. Observe that:

$$\lim_{\nu \to \infty}(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))) \geq \lim_{\nu \to \infty}(\nu \cdot g(\alpha)) = \infty.$$

Next, we take a limit as $\nu \to 0$. Observe that

$$\lim_{\nu \to 0}(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))) \geq \lim_{\nu \to 0}(\nu \cdot g(\max_{w \geq 0}(w - \nu g(w)))).$$

Using the first-order condition for $\max_{w \geq 0}(w - \nu g(w))$, we see this is equal to:

$$\lim_{\nu \to 0}(\nu \cdot g(\max_{w \geq 0}(w - \nu g(w)))) = \lim_{\nu \to 0} \frac{g(w^*(\nu) - \nu \cdot g(w^*(\nu)))}{g'(w^*(\nu))} = \lim_{\nu \to 0} \frac{g(w^*(\nu) - \frac{g(w^*(\nu)))}{g'(w^*(\nu)))}}{g'(w^*(\nu))}.$$

Using Lemma 229, we can reparameterize and see that this is equal to:

$$\lim_{w \to \infty} \frac{g\left(w - \frac{g(w)}{g'(w)}\right)}{g'(w)}.$$

This is equal to ∞ by the assumption in the theorem statement. □

Now we prove Theorem 76

Proof of Theorem 76. By Lemma 234, disintermediation occurs if and only if $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)) > \alpha C$. By Lemma 232, we know that $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))$ is U-shaped and by Lemma 239 we know that $\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)) = \alpha C$ has exactly two solutions. This means that $0 < T_L(C, \alpha, g) < T_U(C, \alpha, g) < \infty$ can be taken to be equal to these two solutions. This also means that the lower threshold is decreasing as a function of the number of consumers C, and the upper threshold is increasing as a function of C, as desired. ☐

G.3.5 Proof of Proposition 77

We prove Proposition 77.

Proof. Using Lemma 234, it suffices to show that

$$\lim_{\nu \to 0}(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w))) < \alpha C.$$

Using the first-order condition for $\max_{w \geq 0}(w - \nu g(w))$, we see this is equal to:

$$\lim_{\nu \to 0}(\nu \cdot g(\alpha + \max_{w \geq 0}(w - \nu g(w)))) = \lim_{\nu \to 0} \frac{g(\alpha + w^*(\nu) - \nu \cdot g(w^*(\nu)))}{g'(w^*(\nu))} = \lim_{\nu \to 0} \frac{g(\alpha + w^*(\nu) - \frac{g(w^*(\nu)))}{g'(w^*(\nu)))}}{g'(w^*(\nu))}.$$

Using Lemma 229, we can reparameterize and see that this is equal to:

$$\lim_{w \to \infty} \frac{g\left(\alpha + w - \frac{g(w)}{g'(w)}\right)}{g'(w)}.$$

Using Lemma 235, we see that this is equal to:

$$\begin{aligned}
\lim_{w \to \infty} \frac{g\left(\alpha + w - \frac{g(w)}{g'(w)}\right)}{g'(w)} &= \lim_{w \to \infty} \frac{g\left(\alpha + w - \frac{g(w)}{g'(w)}\right)}{g\left(w\right)} \frac{g\left(w\right)}{g'(w)} \\
&= \lim_{w \to \infty} \frac{g\left(\alpha + w - \frac{1}{1+\frac{\beta}{w}}\right)}{g\left(w\right)} \frac{1}{1 + \frac{\beta}{w}} \\
&\leq \lim_{w \to \infty} \frac{g\left(\alpha + w\right)}{g\left(w\right)} \\
&= e^{\alpha} \cdot \lim_{w \to \infty} \frac{(\alpha + w)^{\beta}}{w^{\beta}} \\
&= e^{\alpha}.
\end{aligned}$$

Since $e^{\alpha} < \alpha \cdot C$, this proves the desired statement. ☐

G.4 Proofs for Chapter 10.4

G.4.1 Proof of Proposition 78

We prove Proposition 78. This result follows easily from Lemma 234.

Proof of Proposition 78. We split into two cases: $\nu < T_L(C, \alpha, g)$ or $\nu > T_U(C, \alpha, g)$, and $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$.

Case 1: $\nu < T_L(C, \alpha, g)$ **or** $\nu > T_U(C, \alpha, g)$**.** In this case, disintermediation occurs (Theorem 75). By Lemma 234, the consumer creates the content $\arg\max_{w \geq 0}(w - \nu \cdot g(w))$ that maximizes their utility.

Case 2: $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$**.** In this case, the intermediary survives (Theorem 75). By Lemma 234, the intermediary produces content

$$w_m = \alpha + \max_{w \geq 0}(w - \nu g(w)).$$

□

G.4.2 Proof of Theorem 79

We prove Theorem 79.

Proof of Theorem 79. First, we show that the quality is continuous in ν when $\nu \neq T_L(C, \alpha, g)$ and $\nu \neq T_U(C, \alpha, g)$. This follows from the functional forms from Proposition 78.

Next, we show that the quality is decreasing in ν when $\nu \neq T_L(C, \alpha, g)$ and $\nu \neq T_U(C, \alpha, g)$. We again use Proposition 78. For $\nu < T_L(C, \alpha, g)$ or $\nu > T_U(C, \alpha, g)$, the content quality is $\arg\max_{w \geq 0}(w - \nu g(w))$. This is equal to $w^*(\nu)$ such that $g'(w^*(\nu)) = 1/\nu$. Since g' is increasing in its argument, this is decreasing in ρ. For $\nu \in (T_L(C, \alpha, g), T_U(C, \alpha, g))$, we the content quality is $\alpha + \max_{w \geq 0}(w - \nu g(w))$. By Lemma 230, this is decreasing in ν.

Next, we analyze the content quality at the thresholds $\nu = T_L(C, \alpha, g)$ and $\nu = T_U(C, \alpha, g)$. We again use Proposition 78. It suffices to show that:

$$\lim_{\nu \to^- T_L(C,\alpha,g)} \arg\max_{w \geq 0}(w - \nu g(w)) > \alpha + \max_{w \geq 0}(w - (T_L(C, \alpha, g))g(w))$$

and

$$\lim_{\nu \to^+ T_U(C,\alpha,g)} \arg\max_{w \geq 0}(w - \nu g(w)) < \alpha + \max_{w \geq 0}(w - (T_U(C, \alpha, g)) \cdot g(w)).$$

For the first limit, we can rewrite the desired inequality as $\alpha < T_L(C, \alpha, g) \cdot g(w^*(T_L(C, \alpha, g)))$, where $w^*(\nu) = \arg\max_w(w - \nu g(w))$. This holds because by Lemma 232 we know that at $\nu = T_L(C, \alpha, g)$, the sign of the derivative of $\nu g(\alpha + \max_w(w - \nu g(w))$ is negative, and by Lemma 231, we know that this means that $g(w^*(\nu)) > \alpha g'(w^*(\nu))$, which means that $\nu g(w^*(\nu)) > \alpha$ as desired.

For the second limit, we can rewrite the desired inequality as $\alpha > T_U(C,\alpha,g) \cdot g(w^*(T_U(C,\alpha,g)))$, where $w^*(\nu) = \arg\max_w(w - \nu g(w))$. This holds because by Lemma 232 we know at $\nu = T_U(C,\alpha,g)$, that the sign of the derivative of $\nu g(\alpha + \max_w(w - \nu g(w))$ is positive, and by Lemma 231, we know that this means that $g(w^*(\nu)) < \alpha g'(w^*(\nu))$, which means that $\nu g(w^*(\nu)) < \alpha$ as desired.

□

G.4.3 Proof of Proposition 80

We prove Proposition 80. This result follows from easily Lemma 234.

Proof of Proposition 80. We split into two cases: $\nu < T_L(C,\alpha,g)$ or $\nu > T_U(C,\alpha,g)$, and $\nu \in [T_L(C,\alpha,g), T_U(C,\alpha,g)]$.

Case 1: $\nu < T_L(C,\alpha,g)$ or $\nu > T_U(C,\alpha,g)$. In this case, disintermediation occurs (Theorem 75). This means that the intermediary has utility zero.

Case 2: $\nu \in [T_L(C,\alpha,g), T_U(C,\alpha,g)]$. In this case, the intermediary survives (Theorem 75). By Lemma 234, the intermediary produces content at equilibrium

$$w_m = \alpha + \max_{w \geq 0}(w - \nu g(w)),$$

and their utility is their revenue $\alpha \cdot C$ minus their costs $\nu g\left(\alpha + \max_{w\geq 0}\left(w - \nu \cdot g(w)\right)\right)$. □

G.4.4 Proof of Theorem 81

We prove Theorem 81.

Proof of Theorem 81. First, we show that the intermediary utility is inverse U-shaped. By Proposition 80, it suffices to show that $\nu \cdot g(\alpha + \max_{w\geq 0}(w - \nu \cdot g(w)))$ is U-shaped as a function of ν. This follows from Lemma 232.

Next, we compute the maximum intermediary utility. By Proposition 80, it suffices to find the minimum value of $\nu \cdot g(\alpha + \max_{w\geq 0}(w - \nu \cdot g(w)))$. We again apply Lemma 232 to see that this is equal to α, which means that the maximum intermediary utility is equal to $\alpha(C-1)$ Since $\nu \cdot g(\alpha + \max_{w\geq 0}(w - \nu \cdot g(w)))$ is U-shaped and using Theorem 75, we also know that this optima is attained for ν in the range where intermediation occurs.

We now turn to content w_m produced at this optima. Using Lemma 232 again, we also see that the optima is attained at ν such that $g(\arg\max(w - \nu g(w)) = \alpha \cdot g'(\arg\max(w - \nu g(w))$. Using that g is convex, we know that $g'(\arg\max(w - \nu g(w)) = \frac{1}{\nu}$, so this implies that:

$$\nu \cdot g(\arg\max(w - \nu g(w)) = \alpha.$$

This means that:

$$\begin{aligned} w_m &= \alpha + \max_{w \geq 0}(w - \nu \cdot g(w)) \\ &= \alpha + \arg\max(w - \nu g(w)) - \nu g(\arg\max(w - \nu g(w))) \\ &= \arg\max(w - \nu g(w)) \end{aligned}$$

as desired. □

G.4.5 Proof of Theorem 82

We prove Theorem 82. This follows easily from Lemma 234.

Proof of Theorem 82. We split into two cases: $\nu < T_L(C, \alpha, g)$ or $\nu > T_U(C, \alpha, g)$, and $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$.

Case 1: $\nu < T_L(C, \alpha, g)$ or $\nu > T_U(C, \alpha, g)$. In this case, disintermediation occurs (Theorem 75). By Lemma 234, the consumer produces content $\arg\max_{w \geq 0}(w - \nu g(w))$ and their utility is thus $\max_{w \geq 0}(w - \nu g(w))$.

Case 2: $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$. In this case, the intermediary survives (Theorem 75). By Lemma 234, the intermediary produces content at equilibrium

$$w_m = \alpha + \max_{w \geq 0}(w - \nu g(w)).$$

The consumer utility is thus $\max_{w \geq 0}(w - \nu g(w))$. □

G.4.6 Proof of Corollary 83

We prove Corollary 83.

Proof of Corollary 83. We apply Theorem 82 to see that the consumer utility is $\max_{w \geq 0}(w - \nu g(w))$. We apply Lemma 230 to see that the derivative of $\max_{w \geq 0}(w - \nu g(w))$ with respect to ν is negative. This proves that $\max_{w \geq 0}(w - \nu g(w))$ is continuous and decreasing in ν. □

G.4.7 Proof of Proposition 84

We prove Proposition 84.

Proof of Proposition 84. We add up the utility of consumers, suppliers, and the intermediary. By Theorem 82, the total utility of consumers is equal to $C \cdot (\max_{w \geq 0} (w - \nu \cdot g(w)))$ regardless of whether disintermediation occurs (Theorem 82). By Lemma 234, the suppliers choose $\nu = \nu^*$ and thus have have zero profit.

We split into two cases: $\nu < T_L(C, \alpha, g)$ or $\nu > T_U(C, \alpha, g)$, and $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$.

Case 1: $\nu < T_L(C, \alpha, g)$ **or** $\nu > T_U(C, \alpha, g)$**.** In this case, disintermediation occurs (Theorem 75). When disintermediation occurs, the social welfare is thus equal to the total utility of consumers, which is $C \cdot (\max_{w \geq 0} (w - \nu \cdot g(w)))$.

Case 2: $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$**.** In this case, the intermediary survives (Theorem 75). When the intermediary survives, the social welfare is equal to the intermediary's utility plus the total consumer utility. By Proposition 80, the intermediary's utility is $\alpha C - \nu g(\alpha + \max_{w \geq 0}(w - \nu g(w)))$. This means that the social welfare is equal to $\alpha C - \nu g(\alpha + \max_{w \geq 0}(w - \nu g(w))) + C \cdot \max_{w \geq 0}(w - \nu g(w))$ as desired.

□

G.4.8 Proof of Proposition 85

We prove Proposition 85.

Proof of Proposition 85. Let $\nu = \min(\nu^* + \nu^H, \nu_0)$. Production is done through the suppliers if $\nu^* + \nu^H \leq \nu_0$ and using manual content creation if $\nu^* + \nu^H > \nu_0$.

If the intermediary does not exist, then the market outcome that maximizes the social welfare is that each consumer produces content $\arg\max_{w \geq 0}(w - \nu g(w))$. The social welfare is $C \cdot \max_{w \geq 0}(w - \nu g(w))$.

For the case where the intermediary exists, we construct a market outcome that maximizes the social welfare: the suppliers set prices $\nu_1 = \ldots \nu_P = \nu^*$ equal to the supply-side costs, the intermediary produces content

$$w_m = \arg\max_{w \geq 0}(Cw - \nu \cdot g(w)).$$

and all consumers $j \in [C]$ all choose consumption mode $a_j = M$ and consume the content $w_{c,j} = w_m$ created by the intermediary. The social welfare is $\max_{w \geq 0}(C \cdot w - \nu g(w))$.

□

G.4.9 Proof of Theorem 86

We prove Theorem 86.

Proof of Theorem 86. We apply Proposition 84 to see that the social welfare is equal to:

$$\begin{cases} C \cdot (\max_{w \geq 0} (w - \nu \cdot g(w))) & \text{if } \nu < T_L(C, \alpha, g) \\ C\alpha - \nu g (\alpha + \max_{w \geq 0} (w - \nu \cdot g(w))) + C \max_{w \geq 0} (w - \nu \cdot g(w)) & \text{if } \nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)] \\ C \cdot (\max_{w \geq 0} (w - \nu \cdot g(w))) & \text{if } \nu > T_U(C, \alpha, g). \end{cases}$$

Throughout this proof, let $w^*(\nu')$ be the unique solution to $\max_{w \geq 0}(w - \nu' \cdot g(w))$ (Lemma 228).

First, we show that the social welfare is continuous in the production costs. This follows immediately within each of the three regimes, and at the boundaries, it follows from the fact that $\alpha \cdot C - \nu g(\alpha + \max_{w \geq 0}(w - \nu g(w))) = 0$ so

$$C\alpha - \nu g\left(\alpha + \max_{w \geq 0}\left(w - \nu \cdot g(w)\right)\right) + C \max_{w \geq 0}\left(w - \nu \cdot g(w)\right) = C \cdot \max_{w \geq 0}\left(w - \nu \cdot g(w)\right).$$

Next, we show that the social welfare is decreasing in ν. For the first and third regime, the social welfare is equal to the total consumer utility. This is C times the utility of any given consumer. So by Theorem 82, this is decreasing in ν. For the second regime, we take a derivative of $C\alpha - -\nu g\left(\alpha + \max_{w \geq 0}\left(w - \nu \cdot g(w)\right)\right) + C \max_{w \geq 0}\left(w - \nu \cdot g(w)\right)$ with respect to ν. By Lemma 231 and Lemma 230, this is equal to:

$$-g(\alpha + \max_{w \geq 0}(w - \nu g(w)) + g'(\alpha + \max_{w \geq 0}(w - \nu g(w)) \frac{g(w^*(\nu))}{g'(w^*(\nu))} - C \cdot g(w^*(\nu)).$$

where $w^*(\nu) = \arg\max_{w \geq 0}(w - \nu g(w))$. It suffices to show that:

$$g(\alpha + \max_{w \geq 0}(w - \nu g(w)) + C \cdot g(w^*(\nu)) > g'(\alpha + \max_{w \geq 0}(w - \nu g(w)) \frac{g(w^*(\nu))}{g'(w^*(\nu))}.$$

We split into two cases: (1) $\alpha + \max_{w \geq 0}(w - \nu g(w) > w^*(\nu)$ and (2) $\alpha + \max_{w \geq 0}(w - \nu g(w) \leq w^*(\nu)$.

Case 1: $\alpha + \max_{w \geq 0}(w - \nu g(w) > w^*(\nu)$. It suffices to show that

$$g(\alpha + \max_{w \geq 0}(w - \nu g(w)) > g'(\alpha + \max_{w \geq 0}(w - \nu g(w)) \frac{g(w^*(\nu))}{g'(w^*(\nu))}.$$

We can write this:

$$\frac{g(\alpha + \max_{w \geq 0}(w - \nu g(w))}{g'(\alpha + \max_{w \geq 0}(w - \nu g(w))} > \frac{g(w^*(\nu))}{g'(w^*(\nu))}.$$

Using that g is log-concave, we know that $g(x)/g'(x)$ is increasing in x, so we know that this holds.

Case 2: $\alpha + \max_{w \geq 0}(w - \nu g(w) \leq w^*(\nu)$. It suffices to show that

$$C \cdot g(w^*(\nu)) > g'(\alpha + \max_{w \geq 0}(w - \nu g(w)) \frac{g(w^*(\nu))}{g'(w^*(\nu))}.$$

We can write this as:

$$Cg'(w^*(\nu)) > g'(\alpha + \max_{w \geq 0}(w - \nu g(w)).$$

Since $g'(x)$ is increasing in x, this means that $g'(w^*(\nu)) \geq g'(\alpha + \max_{w \geq 0}(w - \nu g(w))$. This coupled with $C > 0$ implies the desired statement.

Next, we show that the social welfare is strictly greater than the social planner's optimal without the intermediary when $\nu \in (T_L(C, \alpha, g), T_U(C, \alpha, g))$. Using Proposition 85, the social planner's optimal welfare without the intermediary is equal to $C \cdot \max_{w \geq 0}(w - \nu g(w))$. This means that it suffices to show that:

$$\alpha \cdot C - \nu g(\alpha + \max_w(w - \nu g(w)) > 0.$$

This follows from the fact that

$$\alpha \cdot C - \nu g(\alpha + \max_w(w - \nu g(w)) \geq 0$$

when $\nu \in (T_L(C, \alpha, g), T_U(C, \alpha, g))$. To obtain a strict inequality, it suffices to show that the derivative of $\nu g(\alpha + \max_w(w - \nu g(w))$ is positive at the boundaries. This is because by Lemma 232, the optimum occurs when $\nu g(\alpha + \max_w(w - \nu g(w)) = \alpha < \alpha C$.

Finally, we show that it is strictly below the social planner's optimal with the intermediary except at at most one bliss point. We know that the social planner's optimal is always at least as large as the market with intermediary. Thus, it suffices to show that the social welfare is *not equal to* the social planner's optima $\max_{w \geq 0} (C \cdot w - \nu \cdot g(w))$ except at at most one bliss point. We split into cases depending on the value of ν and depending on whether $\alpha + \max_w(w - \nu g(w)) = \arg\max_{w \geq 0}(C \cdot w - \nu g(w))$ holds, and show that the social welfare is not equal to the social optima unless $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$ and $\alpha + \max_w(w - \nu g(w)) = \arg\max_{w \geq 0}(C \cdot w - \nu g(w))$.

Case 1: $\nu < T_L(C, \alpha, g)$ or $\nu > T_U(C, \alpha, g)$. The social welfare is equal to $C \cdot (\max_{w \geq 0} (w - \nu \cdot g(w)))$. We see that:

$$C \cdot \left(\max_{w \geq 0} (w - \nu \cdot g(w)) \right) = \max_{w \geq 0} (C \cdot w - C \cdot \nu \cdot g(w)) \neq_{(A)} \max_{w \geq 0} (C \cdot w - \nu \cdot g(w)),$$

where (A) holds because both optima occur at $w > 0$.

Case 2: $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$ and $\alpha + \max_w(w - \nu g(w)) \neq \arg\max_{w \geq 0}(C \cdot w - \nu g(w))$. In this case, we see that the social welfare is equal to:

$$C \cdot (\alpha + \max_w(w - \nu g(w))) - \nu g(\alpha + \max_w(w - \nu g(w)))$$

and the social planner's optima is equal to $\max_{w \geq 0}(C \cdot w - \nu g(w))$. Since the function $f(x) = C \cdot x - \nu g(x)$ has a unique global optima, this means that

$$C \cdot (\alpha + \max_w(w - \nu g(w))) - \nu g(\alpha + \max_w(w - \nu g(w))) \neq \max_{w \geq 0}(C \cdot w - \nu g(w)).$$

Case 3: $\nu \in [T_L(C, \alpha, g), T_U(C, \alpha, g)]$ and $\alpha + \max_w(w - \nu g(w)) = \arg\max_{w \geq 0}(C \cdot w - \nu g(w))$. In this case, we see that the social welfare is equal to:

$$C \cdot (\alpha + \max_w(w - \nu g(w))) - \nu g(\alpha + \max_w(w - \nu g(w)))$$

and the social planner's optima is equal to $\max_{w\geq 0}(C\cdot w - \nu g(w))$. These expressions are equal because $\alpha + \max_w(w - \nu g(w)) = \arg\max_{w\geq 0}(C\cdot w - \nu g(w))$.

Now, it suffices to show that there is at most one value of ν such that $\nu \in [T_L(C,\alpha,g), T_U(C,\alpha,g)]$ and $\alpha + \max_w(w - \nu g(w)) = \arg\max_{w\geq 0}(C\cdot w - \nu g(w))$. It suffices to show that the derivative of

$$\alpha + \max_w(w - \nu g(w)) - \arg\max_{w\geq 0}(C\cdot w - \nu g(w))$$

with respect to ν is always positive. First, let's simplify this expression. Note that $\arg\max_{w\geq 0}(C\cdot w - \nu g(w))$ occurs when $g'(w) = \frac{C}{\nu}$, which means that $\arg\max_{w\geq 0}(C\cdot w - \nu g(w)) = w^*(\nu/C)$. This means that the expression is equal to:

$$\alpha + \max_w(w - \nu g(w)) - w^*\left(\frac{\nu}{C}\right).$$

Now, taking a derivative and applying Lemma 230, we obtain:

$$-g(w^*(\nu)) - (w^*)'\left(\frac{\nu}{C}\right)\cdot\frac{1}{C}.$$

To compute $(w^*)'(z)$, we use the fact that $g'(w^*(z)) = \frac{1}{z}$, so $w^*(z) = (g')^{-1}(1/z)$. By the inverse function theorem, this means that:

$$(w^*)'(z) = -\frac{1}{z^2\cdot g''(w^*(z))} = -\frac{(g'(w^*(z)))^2}{g''(w^*(z))}.$$

Plugging this in, we obtain:

$$-g(w^*(\nu)) + \frac{(g'(w^*\left(\frac{\nu}{C}\right)))^2}{g''(w^*\left(\frac{\nu}{C}\right))}\cdot\frac{1}{C},$$

This is positive if and only if:

$$(g'(w^*\left(\frac{\nu}{C}\right)))^2 \geq Cg(w^*(\nu))\cdot g''(w^*\left(\frac{\nu}{C}\right)).$$

By log-concavity, we know that:

$$(g'(w^*\left(\frac{\nu}{C}\right)))^2 \geq g(w^*\left(\frac{\nu}{C}\right))\cdot g''(w^*\left(\frac{\nu}{C}\right)).$$

Using log-concavity again and that $g'(w^*(z)) = \frac{1}{z}$, we know that:

$$\frac{\nu}{C}\cdot g(w^*\left(\frac{\nu}{C}\right)) = \frac{g(w^*\left(\frac{\nu}{C}\right))}{g'(w^*\left(\frac{\nu}{C}\right))} > \frac{g(w^*(\nu))}{g'(w^*(\nu))} = \nu\cdot g(w^*(\nu)).$$

This means that:

$$g(w^*\left(\frac{\nu}{C}\right)) > C\cdot g(w^*(\nu)).$$

Putting this all together, this implies that:

$$(g'(w^*\left(\frac{\nu}{C}\right)))^2 \geq g(w^*\left(\frac{\nu}{C}\right))\cdot g''(w^*\left(\frac{\nu}{C}\right)) > Cg(w^*(\nu))\cdot g''(w^*\left(\frac{\nu}{C}\right)).$$

as desired.

$\square$

G.4.10 Statement and Proof of Proposition 240

We state and prove Proposition 240.

Proposition 240. *Consider the setup of Theorem 74, and suppose that $\beta \geq 2$. Then, there exists a bliss point ν where the social welfare of the market equals the social planner's optima.*

Proof. Following the proof of Theorem 86, we know the social welfare of the market equals the social planner's optima if and only if $\nu \in [T_L(C,\alpha,g), T_U(C,\alpha,g)]$ and $\alpha + \max_w(w - \nu g(w)) = \arg\max_{w\geq 0}(C \cdot w - \nu g(w))$. Observe that:

$$\alpha + \max_w(w - \nu g(w)) = \alpha + \nu^{-\frac{1}{\beta-1}}\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right)$$

and:

$$\arg\max_{w\geq 0}(C^{\frac{1}{\beta-1}} \cdot w - \nu g(w)) = \nu^{-\frac{1}{\beta-1}} C^{\frac{1}{\beta-1}} \beta^{-\frac{1}{\beta-1}}.$$

These two expressions are equal when:

$$\nu^{-\frac{1}{\beta-1}}\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right) = \alpha,$$

which can be written as:

$$\nu^{-\frac{1}{\beta-1}} = \frac{\alpha}{\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right)},$$

To show this occurs for $\nu \in [T_L(C,\alpha,g), T_U(C,\alpha,g)]$, we observe that:

$$\begin{aligned}
&\nu g(\alpha + \max_w(w - \nu g(w))) \\
&= \nu\left(\alpha + \frac{\alpha\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right)}{\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right)}\right)^{\beta} \\
&= \left(\frac{\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right)}{\alpha}\right)^{\beta-1} \cdot \left(\alpha + \frac{\alpha\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right)}{\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right)}\right)^{\beta} \\
&= \alpha \cdot \frac{\left(\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right) + \left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right)\right)^{\beta}}{\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right)} \\
&= \alpha \cdot \frac{\left(C^{\frac{1}{\beta-1}} \cdot \beta^{-\frac{1}{\beta-1}}\right)^{\beta}}{\left((C^{\frac{1}{\beta-1}} - 1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right)}.
\end{aligned}$$

We want to show this is at most $\alpha \cdot C$. It suffices to show that:

$$\left(C^{\frac{1}{\beta-1}} \cdot \beta^{-\frac{1}{\beta-1}}\right)^{\beta} \leq C\left((C^{\frac{1}{\beta-1}}-1)\beta^{-\frac{1}{\beta-1}} + \beta^{-\frac{\beta}{\beta-1}}\right).$$

This simplifies to

$$C^{\frac{\beta}{\beta-1}} \cdot \beta^{-\frac{\beta}{\beta-1}} \leq C^{\frac{\beta}{\beta-1}}\beta^{-\frac{1}{\beta-1}} - C\beta^{-\frac{1}{\beta-1}} + C\beta^{-\frac{\beta}{\beta-1}},$$

which simplifies to:

$$C \geq 1,$$

which holds. □

G.5 Proofs for Chapter 10.5

G.5.1 Proof of Lemma 88 and Theorem 87

For the purposes of this result, we slightly modify the tiebreaking rules: we assume that each consumer j tiebreaks in favor of direct usage (i.e., $a_j = D$) rather than in favor of the intermediary (i.e., $a_j = M$) when $\nu < T_U(C, \alpha, \beta)$, but tiebreaks in favor of the intermediary when $\nu \geq T_U(C, \alpha, \beta)$.

First, we prove the following modified version of Lemma 233 for this modified tiebreaking rule.

Lemma 241. *Consider the setup of Chapter 10.5.1 and Theorem 87. Suppose that the supplier chooses price ν_1 and consider the subgame between the intermediary and consumers (Stages 2-3). Let $\nu = \min(\nu^H + \nu_1, \nu_0)$, and let $T_L(C, \alpha, \beta)$ and $T_U(C, \alpha, \beta)$ be defined as in Theorem 74. Under the tiebreaking assumptions above, there exists a unique pure strategy where:*

- *Suppose that $\nu \leq T_L(C, \alpha, \beta)$ or $\nu > T_U(C, \alpha, \beta)$. Then $w_m = 0$. Moreover, for all $j \in [C]$, it holds that $a_j = D$,*

  $$w_{c,j} = \arg\max_{w \geq 0} \left(w - \nu(1+\gamma)g(w)\right).$$

  Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the consumer chooses $i_j = 0$. Otherwise, the consumer chooses $i_j = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index).

- *Suppose that $\nu \in (T_L(C, \alpha, \beta), T_U(C, \alpha, \beta)]$. Then*

  $$w_m = w_{c,j} = \alpha + \max_{w \geq 0} \left(w - \nu(1+\gamma)g(w)\right).$$

  Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the intermediary chooses $i_m = 0$; otherwise, the intermediary chooses $i_m = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index). Finally, it holds that $a_j = M$ and $w_{c,j} = w_m$ for all $j \in [C]$.

Proof. The proof follows similarly to the proof of Lemma 233, but additionally uses the analysis of the condition $\nu g(\alpha + \max_{w \geq 0}(w - \nu g(w)) - \alpha C$ from Lemma 239 and Theorem 75. $\square$

We prove an analogue of Lemma 234 for this setting, which strengthens Lemma 88 (this result directly implies Lemma 88).

Lemma 242. *Consider the setup of Chapter 10.5.1 and Theorem 87. Under the tiebreaking assumptions described above, there exists a unique pure strategy equilibrium which takes the following form: the supplier chooses the price ν_1 as specified in Lemma 88, and the intermediary and consumers choose actions according to the subgame equilibrium constructed in Lemma 241.*

To prove Lemma 242, a key technical challenge is that the supplier can influence whether intermediation occurs in terms of how it sets its prices. To capture these effects, we separately analyze the optimal price for the suppler in each regime in the following intermediate lemmas.

First, we bound the optimal price for the supplier in the range which induces disintermediation.

Lemma 243. *Consider the setup of Theorem 87. Let ν^M be the value of ν' that attains the maximum*

$$\max_{\nu' \geq \nu, \nu \in [0, T_L(C,\alpha,\beta)] \cup [T_U(C,\alpha,\beta), \infty)} \left(C \cdot (\nu' - \nu) \cdot g \left(\arg\max_{w \geq 0} (w - \nu' g(w)) \right) . \right)$$

Then, $\nu^M = \beta \cdot \nu$ if $\beta \cdot \nu \leq T_L(C,\alpha,\beta)$ or $\beta \cdot \nu \geq T_U(C,\alpha,\beta)$. Otherwise, $\nu^M \in \{T_L(C,\alpha,\beta), T_U(C,\alpha,\beta)\}$, and the optimal value is upper bounded by the value of $C \cdot (\nu' - \nu) \cdot g\left(\arg\max_{w \geq 0}(w - \nu' g(w))\right)$ when $\nu' = \beta \cdot \nu$.

Proof. Throughout the proof, let $w^*(\nu) = \arg\max_w (w - \nu g(w))$. Using the structure of $g(w) = w^\beta$, we know that:

$$w^*(\nu') = (\nu')^{-\frac{1}{\beta-1}} \beta^{-\frac{1}{\beta-1}}.$$

This means that the objective can be simplified to:

$$C(\nu' - \nu)(\nu')^{-\frac{\beta}{\beta-1}} \beta^{-\frac{\beta}{\beta-1}}.$$

We split into two cases: (1) $\beta \cdot \nu \leq T_L(C,\alpha,\beta)$ or $\beta \cdot \nu \geq T_U(C,\alpha,\beta)$, and (2) $\beta \cdot \nu \in (T_L(C,\alpha,\beta), T_U(C,\alpha,\beta))$.

Case 1: $\beta \cdot \nu \leq T_L(C,\alpha,\beta)$ or $\beta \cdot \nu \geq T_U(C,\alpha,\beta)$. We take a first-order condition to obtain that $\nu^M = \beta \cdot \nu$ as desired.

Case 2: $\beta \cdot \nu \in (T_L(C,\alpha,\beta), T_U(C,\alpha,\beta))$. In this case, the function is increasing on $[T_L(C,\alpha,\beta), \nu)$ and decreasing on $(\nu, T_U(C,\alpha,\beta]$. This means that the optima for the constrained domain is attained at $\nu^M \in \{T_L(C,\alpha,\beta), T_U(C,\alpha,\beta)\}$, and the optimal value is upper bounded by the optimal for the unconstrained domain, which is equal to $C \cdot (\nu' - \nu) \cdot g\left(\arg\max_{w \geq 0}(w - \nu' g(w))\right)$ when $\nu' = \beta \cdot \nu$.

□

Next,we bound the optimal price in the range which induces intermediation.

Lemma 244. *Let $g(w) = w^\beta$ for $\beta > 1$. Let $T_L(C, \alpha, \beta)$ and $T_U(C, \alpha, \beta)$ be defined as in Theorem 74, and let $\nu \leq T_U(C, \alpha, \beta)$. Then, the maximum*

$$\max_{\nu' \in [\max(\nu^*, T_L(C,\alpha,\beta)), T_U(C,\alpha,\beta)]} \left((\nu' - \nu) \cdot g\left(\alpha + \max_{w \geq 0}(w - \nu' g(w))\right)\right)$$

is uniquely attained at $\nu = T_U(C, \alpha, \beta)$.

Proof. We know that

$$\max_{\nu' \in [\max(\nu, T_L(C,\alpha,\beta)), T_U(C,\alpha,\beta)]} \left((\nu' - \nu) \cdot g\left(\alpha + \max_{w \geq 0}(w - \nu' g(w))\right)\right)$$

is equal to:

$$\max_{\nu' \in [\max(\nu, T_L(C,\alpha,\beta)), T_U(C,\alpha,\beta)]} \left(\left(1 - \frac{\nu}{\nu'}\right) \cdot \nu' \cdot g\left(\alpha + \max_{w \geq 0}(w - \nu' g(w))\right)\right).$$

This is at most:

$$\left(\max_{\nu \in [\max(\nu, T_L(C,\alpha,\beta)), T_U(C,\alpha,\beta)]} \left(1 - \frac{\nu}{\nu'}\right)\right) \cdot \max_{\nu' \in [\max(\nu, T_L(C,\alpha,\beta)), T_U(C,\alpha,\beta)]} \left(\nu' \cdot g\left(\alpha + \max_{w \geq 0}(w - \nu' g(w))\right)\right).$$

This is at most:

$$\left(1 - \frac{\nu}{T_U(C, \alpha, \beta)}\right) \cdot \alpha \cdot C.$$

This value is uniquely attained at $\nu = T_U(C, \alpha, \beta)$ as desired.

□

Now, we use Lemma 243 and Lemma 244 to prove Lemma 242.

Proof of Lemma 242. Let $\nu = \min(\nu^* + \nu^H, \nu_0)$. Let $w^*(\nu')$ be the unique solution to $\max_{w \geq 0}(w - \nu' \cdot g(w))$ (Lemma 228).

The supplier can choose to induce disintermediation or intermediation, which affects how they set their optimal price. We use Lemma 243 and Lemma 244 to narrow down the set of possible optimal prices in each regime. When disintermediation is induced, the supplier earns profit $(\nu' - \nu) \cdot C \cdot g(w^*(\nu'))$ from choosing price ν'. By Lemma 243, the optimal price in the disintermediation range is $\nu \cdot \beta$ if that price induces disintermediation, or otherwise is in the set $\{T_L(C, \alpha, \beta), \lim_{\varepsilon \to +0}(T_U(C, \alpha, \beta) + \varepsilon)\}$, where we use $\lim_{\varepsilon \to +0}(T_U(C, \alpha, \beta) + \varepsilon)$ to denote that the optimum in that range doesn't exist and the supplier would want to set their price arbitrarily close to $T_U(C,\alpha,\beta)$. At these prices, the supplier's profit is upper bounded by $(\nu'-\nu)\cdot C\cdot g(w^*(\nu'))$ where $\nu'=\nu\cdot\beta$. The realized profit is:

$$C\left(1-\frac{\nu}{\nu'}\right)\cdot\nu' g(w^*(\nu')) = C\left(1-\frac{\nu}{\nu'}\right)\cdot(\nu')^{-\frac{1}{\beta-1}}\cdot\beta^{-\frac{\beta}{\beta-1}}.$$

When intermediation is induced, the profit is $(\nu'-\nu)\cdot g(\alpha+\max_{w\geq 0}(w-\nu' g(w)))$. By Lemma 244, the optimal price in the intermediation range is $\nu = T_U(C,\alpha,\beta)$. The obtained profit is:

$$\left(1-\frac{\nu}{\nu'}\right)\alpha C.$$

We claim that prices of the form $T_U(C,\alpha,\beta)+\varepsilon$ for sufficiently small ε are dominated by $T_U(C,\alpha,\beta)$. Based on the above analysis, it suffices to show that $\nu' g(w^*(\nu'))<\alpha$. This holds by Lemma 231 and Lemma 232.

We split into several cases depending on the value of $\nu\cdot\beta$ and ν.

Case 1: $\nu \geq T_U(C,\alpha,\beta)$. In this case, we know that the supplier will set a price of $\nu > \nu^* \geq T_U(C,\alpha,\beta)$ to earn positive profit. This means that disintermediation occurs regardless of the price that they set. They thus set the price to $\nu_1=\nu\cdot\beta$ to maximize their profit.

Case 2: $\nu\cdot\beta > T_U(C,\alpha,\beta)$ and $\nu < T_U(C,\alpha,\beta)$. By the above analysis, we know that the supplier will either set the price to be $\nu\cdot\beta$ or to be $T_U(C,\alpha,\beta)$. We show that there is a threshold value $T_U^{\text{mon}}(C,\alpha,\beta)$ in this range such that the supplier sets the price to be $T_U(C,\alpha,\beta)$ if $\nu\leq T_U^{\text{mon}}(C,\alpha,\beta)$ and sets the price to be $\nu\cdot\beta$ otherwise. Note that the realized profit at $\nu\cdot\beta$ is

$$C\left(1-\frac{1}{\beta}\right)(\nu')^{-\frac{1}{\beta-1}}\beta^{-\frac{\beta}{\beta-1}} = C\left(1-\frac{1}{\beta}\right)\nu^{-\frac{1}{\beta-1}}\beta^{-\frac{\beta+1}{\beta-1}}$$

and at $T_U(C,\alpha,\beta)$ is

$$C\left(1-\frac{\nu}{T_U(C,\alpha,\beta)}\right)\alpha.$$

Let's consider the ratio:

$$\frac{C\left(1-\frac{1}{\beta}\right)\nu^{-\frac{1}{\beta-1}}\beta^{-\frac{\beta+1}{\beta-1}}}{C\left(1-\frac{\nu}{T_U(C,\alpha,\beta)}\right)\alpha} = \frac{\left(1-\frac{1}{\beta}\right)\beta^{-\frac{\beta+1}{\beta-1}}}{\alpha}\frac{\nu^{-\frac{1}{\beta-1}}}{\left(1-\frac{\nu}{T_U(C,\alpha,\beta)}\right)}.$$

The derivative is:

$$\frac{\left(1-\frac{1}{\beta}\right)\beta^{-\frac{\beta+1}{\beta-1}}}{\alpha}\cdot\frac{T_U(C,\alpha,\beta)\nu^{-\frac{\beta}{\beta-1}}\left(\beta\cdot\nu - T_U(C,\alpha,\beta)\right)}{(\beta-1)(T_U(C,\alpha,\beta)-\nu)^2},$$

which is positive in this regime, so the ratio is increasing in this regime. The ratio approaches ∞ as $\nu\to T_U(C,\alpha,\beta)$. As $\nu\cdot\beta\to T_U(C,\alpha,\beta)$, the price $\nu\cdot\beta$ is of the form $T_U(C,\alpha,\beta)+\varepsilon$

which we already proved to be dominated by $T_U(C,\alpha,\beta)$, meaning that the ratio is less than 1. This proves the desired statement.

Case 3: $T_L(C,\alpha,\beta) < \nu\cdot\beta < T_U(C,\alpha,\beta)$ **and** $(\nu\cdot\beta)\cdot g(w^*(\nu\cdot\beta)) < \alpha$**.** We know that the supplier will either set the price to be $T_L(C,\alpha,\beta)$ or $T_U(C,\alpha,\beta)$. The realized profit at $T_L(C,\alpha,\beta)$ is upper bounded by:

$$C(\beta\cdot\nu-\nu)\cdot g(w^*(\beta\cdot\nu)) = C\left(1-\frac{\nu}{\beta\cdot\nu}\right)\cdot(\beta\cdot\nu)g(w^*(\beta\cdot\nu)) < C\left(1-\frac{\nu}{T_U(C,\alpha,\beta)}\right)\cdot\alpha,$$

which is the profit at $T_U(C,\alpha,\beta)$. This means that $\nu_1 = T_U(C,\alpha,\beta)$.

Case 4: $T_L(C,\alpha,\beta) < \nu\cdot\beta < T_U(C,\alpha,\beta)$ **and** $(\nu\cdot\beta)\cdot g(w^*(\nu\cdot\beta)) > \alpha$**.** We know that the supplier will either set the price to be $T_L(C,\alpha,\beta)$ or $T_U(C,\alpha,\beta)$. We show that there is a threshold value $T_L^{\text{mon}}(C,\alpha,\beta)$ in this range such that the supplier sets the price to be $T_L(C,\alpha,\beta)$ if $\nu \le T_L^{\text{mon}}(C,\alpha,\beta)$ and sets the price to be $T_U(C,\alpha,\beta)$ otherwise. Note that the realized profit at $T_L(C,\alpha,\beta)$ is

$$C\left(1-\frac{\nu}{T_L(C,\alpha,\beta)}\right)T_L(C,\alpha,\beta)g(w^*(T_L(C,\alpha,\beta)))$$

and at $T_U(C,\alpha,\beta)$ is

$$C\left(1-\frac{\nu}{T_U(C,\alpha,\beta)}\right)\alpha.$$

It is easy to see that the ratio is decreasing in ν. At $\nu\cdot\beta = T_L(C,\alpha,\beta)$, we see that the ratio is at least:

$$(1-\frac{1}{\beta})C\cdot T_L(C,\alpha,\beta)g(w^*(T_L(C,\alpha,\beta))) > (1-\frac{1}{\beta})C\alpha > 1,$$

using our assumption that $C > \frac{\beta}{\beta-1}$. This proves the desired statement.

Case 5: $\nu^*\cdot\beta < T_L(C,\alpha,\beta)$**.** We know that the supplier will either set the price to be $\nu\cdot\beta$ or $T_U(C,\alpha,\beta)$. The above analysis for Case 4 shows that:

$$C\left(1-\frac{\nu}{T_U(C,\alpha,\beta)}\right)\alpha < C\left(1-\frac{\nu}{T_L(C,\alpha,\beta)}\right)T_L(C,\alpha,\beta)g(w^*(T_L(C,\alpha,\beta))).$$

We also know that:

$$C\left(1-\frac{\nu}{T_L(C,\alpha,\beta)}\right)T_L(C,\alpha,\beta)g(w^*(T_L(C,\alpha,\beta))) < C(\nu\cdot\beta-\nu)g(w^*(\nu\cdot\beta)).$$

This proves that $\nu_1 = \nu\cdot\beta$ as desired. □

We prove Theorem 87 from Lemma 242.

Proof. By Lemma 242, we know that the supplier's price satisfies:

$$\nu_1 = \begin{cases} \beta \cdot \nu^* & \text{if } \nu^* < \beta^{-1} \cdot T_L(C, \alpha, \beta) \\ T_L(C, \alpha, \beta) & \text{if } \nu^* \geq \beta^{-1} \cdot T_L(C, \alpha, \beta) \text{ and } \nu^* \leq T_L^{\text{mon}}(C, \alpha, \beta), \\ T_U(C, \alpha, \beta) & \text{if } \nu^* \in (T_L^{\text{mon}}(C, \alpha, \beta), T_U^{\text{mon}}(C, \alpha, \beta)] \\ \beta \cdot \nu^* & \text{if } \nu^* > T_U^{\text{mon}}(C, \alpha, \beta). \end{cases}$$

This, coupled with Lemma 241, gives us:

$$\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) \leq T_L^{\text{mon}}(C, \alpha, \beta) \\ C & \text{if } \min(\nu^* + \nu^H, \nu_0) \in (T_L^{\text{mon}}(C, \alpha, \beta), T_U^{\text{mon}}(C, \alpha, \beta)] \\ 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) > T_U^{\text{mon}}(C, \alpha, \beta). \end{cases}$$

as desired. □

G.5.2 Proof of Theorem 89

We prove Theorem 89. First, we prove the following analogue of Lemma 233.

Lemma 245. *Consider the setup of Chapter 10.5.2 and Theorem 89. Suppose that suppliers choose prices $\nu_1, \ldots, \nu_P$ and consider the subgame between the intermediary and consumers (Stages 2-3). Let $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$, and consider the condition*

$$\nu(1 + \gamma C) \cdot g\left(\alpha + \max_{w \geq 0} (w - \nu(1 + \gamma)g(w))\right) > \alpha C. \tag{G.2}$$

Under the tiebreaking assumptions discussed in Chapter 10.2.2, there exists a unique pure strategy where:

- *If (G.2) holds, then $w_m = 0$. Moreover, for all $j \in [C]$, it holds that $a_j = D$,*

$$w_{c,j} = \arg\max_{w \geq 0} (w - \nu(1 + \gamma)g(w)).$$

  Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the consumer chooses $i_j = 0$. Otherwise, the consumer chooses $i_j = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index).

- *If (G.1) does not hold, then*

$$w_m = w_{c,j} = \alpha + \max_{w \geq 0} (w - \nu(1 + \gamma)g(w)).$$

  Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the intermediary chooses $i_m = 0$; otherwise, the intermediary chooses $i_m = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index). Finally, it holds that $a_j = M$ and $w_{c,j} = w_m$ for all $j \in [C]$.

Proof. Like in the proof of Lemma 233, recall that when consumers or the intermediary produce content, they choose the option that minimizes their production costs. If $\nu^H + \min_{i \in [P]} \nu_i < \nu_0$, they leverage the technology of the supplier who offers the lowest price, and otherwise, they produce content without using the technology. This means that they face production costs $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$.

The main difference from Lemma 233 is that the consumers and the intermediary face marginal costs. Taking into account these marginal costs, when consumer j chooses $a_j = D$, then they maximize their utility and thus produce content $w^*(\nu) = \arg\max(w - \nu(1+\gamma)g(w))$ and achieve utility $\max(w - (1+\gamma)\nu g(w))$. Since the consumer pays the intermediary a fee of α, the intermediary must produce content satisfying $w' \geq \alpha + \max_{w \geq 0}(w - \nu(1+\gamma)g(w))$ to incentivize the consumer to choose $a_j = M$. Producing content $w' \geq \alpha + \max_{w \geq 0}(w - \nu(1 + \gamma)g(w))$ would incentivize all of the consumers to choose the intermediary, so the intermediary would earn utility

$$\alpha \cdot C - \nu \cdot (1 + \gamma C) \cdot g(w').$$

This also means that the intermediary prefers producing content $\alpha + \max_{w \geq 0}(w - \nu(1+\gamma)g(w))$ over any $w' > \alpha + \max_{w \geq 0}(w - \nu(1 + \gamma)g(w))$ in order to minimize costs. The intermediary prefers producing this content over producing content $w = 0$ which would not attract any consumers if and only if:

$$\alpha \cdot C - \nu \cdot (1 + \gamma C) \cdot g(\alpha + \max_{w \geq 0}(w - \nu(1 + \gamma)g(w))) \geq 0.$$

This, coupled with the tiebreaking rules, proves the desired statement. $\square$

Using this lemma, we can characterize the pure strategy equilibria in this extended model.

Lemma 246. *Consider the setup of Chapter 10.5.2 and Theorem 89. Under the tiebreaking assumptions in Chapter 10.2.2, there exists a pure strategy equilibrium which takes the following form: all suppliers choose the price $\nu_i = \nu^*$ for $i \in [P]$, and the intermediary and consumers choose actions according to the subgame equilibrium constructed in Lemma 245. The actions of the intermediary and consumers are the same at every pure strategy equilibrium; moreover, the production cost $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$ is the same at every pure strategy equilibrium.*

Proof. The proof of the equilibrium construction in the first sentence is analogous to the proof of Lemma 246. The proof of the second sentence is analogous to the proof of Theorem 73. $\square$

Using these characterization results, we prove Theorem 89.

Proof of Theorem 89. By Lemma 246, we know that the intermediary survives if and only if

$$\nu(1 + \gamma C) \cdot g\left(\alpha + \max_{w \geq 0}(w - \nu(1 + \gamma)g(w))\right) \leq \alpha C.$$

Let us change variables and let $\nu' = \nu(1+\gamma)$ and let $C' = \frac{C(1+\gamma)}{1+\gamma C}$. Then we can write the condition as:

$$\nu' \cdot g\left(\alpha + \max_{w \geq 0}\left(w - \nu' g(w)\right)\right) \leq \alpha C'.$$

This means that

$$\sum_{j=1}^{C} \mathbb{E}[1[a_j = M]] = \begin{cases} 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) < T_L^{\text{marg}}(C, \alpha, \beta, \gamma) \\ C & \text{if } \min(\nu^* + \nu^H, \nu_0) \in [T_L^{\text{marg}}(C, \alpha, \beta, \gamma), T_U^{\text{marg}}(C, \alpha, \beta, \gamma)] \\ 0 & \text{if } \min(\nu^* + \nu^H, \nu_0) > T_U^{\text{marg}}(C, \alpha, \beta, \gamma) \end{cases}$$

where $T_L^{\text{marg}}(C, \alpha, \beta, \gamma) = (1+\gamma)^{-1} \cdot T_L(C', \alpha, g)$ and $T_U^{\text{marg}}(C, \alpha, \beta, \gamma) = (1+\gamma)^{-1} \cdot T_U(C', \alpha, \beta)$ as desired. □

G.5.3 Proof of Theorem 90

We prove Theorem 90. First, we prove the following analogue of Lemma 233.

Lemma 247. *Consider the setup of Chapter 10.5.3 and Theorem 90. Suppose that suppliers choose prices $\nu_1, \ldots, \nu_P$ and consider the subgame between the intermediary and consumers (Stages 2-3). Let $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$, and consider the condition*

$$\alpha^{\frac{1}{\beta-1}}(1-\alpha) < C^{-\frac{1}{\beta-1}}\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right) \tag{G.3}$$

Under the tiebreaking assumptions discussed in Chapter 10.2.2, there exists a unique pure strategy where:

- *If (G.3) holds, then $w_m = 0$. Moreover, for all $j \in [C]$, it holds that $a_j = D$,*

  $$w_{c,j} = \arg\max_{w \geq 0}\left(w - \nu g(w)\right).$$

  Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the consumer chooses $i_j = 0$. Otherwise, the consumer chooses $i_j = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index).

- *If (G.3) does not hold, then*

  $$w_m = w_{c,j} = (1-\alpha)^{-1} \max_{w \geq 0}\left(w - \nu g(w)\right).$$

  Moreover, if $\nu_0 < \nu^H + \min_{i \in [P]} \nu_i$, then the intermediary chooses $i_m = 0$; otherwise, the intermediary chooses $i_m = \text{argmin}_{i \in [P]} \nu_i$ (tie-breaking in favor of suppliers with a lower index). Finally, it holds that $a_j = M$ and $w_{c,j} = w_m$ for all $j \in [C]$.

Proof. Like in the proof of Lemma 233, recall that when consumers or the intermediary produce content, they choose the option that minimizes their production costs. If $\nu^H + \min_{i \in [P]} \nu_i < \nu_0$, they leverage the technology of the supplier who offers the lowest price, and otherwise, they produce content without using the technology. This means that they face production costs $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$.

The main difference from Lemma 233 is in the fee structure. Like before, when consumer j chooses $a_j = D$, then they maximize their utility and thus produce content $w^*(\nu) = \arg\max(w - \nu g(w))$ and achieve utility $\max(w - \nu g(w))$. Since the consumer pays the intermediary a fee of $\alpha \cdot w$, the intermediary must produce content satisfying $w' \geq \alpha w' + \max_{w \geq 0}(w - \nu g(w))$ to incentivize the consumer to choose $a_j = M$. Producing content $w' \geq (1-\alpha)^{-1} \cdot \max_{w \geq 0}(w - \nu g(w))$ would incentivize all of the consumers to choose the intermediary. We can use the structure of $g(w)$ to simplify this condition to:

$$w' \geq (1-\alpha)^{-1} \cdot \max_{w \geq 0}(w - \nu g(w)) = (1-\alpha)^{-1} \cdot \nu^{-\frac{1}{\beta-1}} \left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right) \tag{G.4}$$

If the intermediary produces content w', they would earn utility

$$\alpha \cdot w' \cdot C - \nu \cdot g(w') = \alpha \cdot w' \cdot C - \nu \cdot (w')^\beta.$$

The intermediary prefers producing this content over producing content $w = 0$ which would not attract any consumers if and only if:

$$\alpha \cdot w' \cdot C - \nu \cdot (w')^\beta \geq 0.$$

We can solve this to obtain:

$$w' \leq \nu^{-\frac{1}{\beta-1}} (\alpha \cdot C)^{\frac{1}{\beta-1}} \tag{G.5}$$

Because of the structure of the tiebreaking rules, the intermediary survives in the market if and only if there exist w' satisfying both (G.5) and (G.4). This happens if and only if:

$$\nu^{-\frac{1}{\beta-1}} (\alpha \cdot C)^{\frac{1}{\beta-1}} \geq (1-\alpha)^{-1} \cdot \nu^{-\frac{1}{\beta-1}} \left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right).$$

This simplifies to:

$$\alpha^{\frac{1}{\beta-1}} (1-\alpha) \geq C^{-\frac{1}{\beta-1}} \left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right).$$

□

Using this lemma, we can characterize the pure strategy equilibria in this extended model.

Lemma 248. *Consider the setup of Chapter 10.5.3 and Theorem 90. Under the tiebreaking assumptions in Chapter 10.2.2, there exists a pure strategy equilibrium which takes the following form: all suppliers choose the price $\nu_i = \nu^*$ for $i \in [P]$, and the intermediary and consumers choose actions according to the subgame equilibrium constructed in Lemma 247. The actions of the intermediary and consumers are the same at every pure strategy equilibrium; moreover, the production cost $\nu = \min(\nu^H + \min_{i \in [P]} \nu_i, \nu_0)$ is the same at every pure strategy equilibrium.*

Proof. The proof of the equilibrium construction in the first sentence is analogous to the proof of Lemma 246. The proof of the second sentence is analogous to the proof of Theorem 73. □

Using these characterization results, we prove Theorem 90.

Proof of Theorem 90. By Lemma 248, we know that the intermediary survives if and only if

$$\alpha^{\frac{1}{\beta-1}}(1-\alpha) \geq C^{-\frac{1}{\beta-1}}\left(\beta^{-\frac{1}{\beta-1}} - \beta^{-\frac{\beta}{\beta-1}}\right).$$

This condition is independent of ν as desired. Moreover, the condition becomes weaker as C gets larger. □

Appendix H

Appendix for Chapter 12

H.1 Classical Results for Matching with Transferable Utilities

To be self-contained, we briefly state and prove the key results from Shapley and Shubik (1971) we need.

First, we explicitly relate the primal-dual formulation in Chapter 12.5.1 to stable matchings.

Theorem 249 (Shapley and Shubik (1971))**.** *If (X, τ) is stable, then (Z, p) is an optimal primal-dual pair to (P) and (D), where $p_a = \tau_a + u_a(X(a))$ and Z is the indicator matrix in $\mathrm{R}^{\mathcal{I}\times\mathcal{J}}$ corresponding to X.*

Moreover, if (Z, p) is an optimal primal-dual pair to (P) and (D) such that Z lies at an extreme point of the feasible set, then (X, τ) is stable where $\tau_a = p_a - u_a(X(a))$ and X is the matching corresponding to the nonzero entries of Z.

Proof. Both statements follow from the complementary slackness conditions and the definition of stability in Chapter 9. The complementary slackness conditions are:

- If $Z_{i,j} > 0$, then $p_i + p_j = u_i(j) + u_j(i)$.
- If $p_i > 0$, then $\sum_j Z_{i,j} = 1$.
- If $p_j > 0$, then $\sum_i Z_{i,j} = 1$.

Suppose that (X, τ) is stable. Let us first show that (Z, p) is feasible. We see that Z is primal feasible by definition. For dual feasibility, since there are no blocking pairs, we know that

$$\big(u_i(\mu_X(i)) + \tau_i\big) + \big(u_j(\mu_X(j)) + \tau_j\big) \geq u_i(j) + u_j(i),$$

which implies

$$p_i + p_j \geq u_i(j) + u_j(i).$$

The individual rationality condition $u_a(\mu_X(a))+\tau_a \geq 0$ tells us $p_a \geq 0$. Hence p is dual feasible. Next, we show that (Z, p) is an optimal primal-dual pair by checking the Karush–Kuhn–Tucker conditions. We have already shown primal and dual feasibility, so it suffices to show complementary slackness. The first condition follows from zero-sum transfers. To see the second and third conditions, we show the contrapositive: If $i \in \mathcal{I}$ is such that $\sum_j Z_{i,j} < 1$, then $\sum_j Z_{i,j} = 0$ by our assumption on Z. Hence i is unmatched (i.e., $u_i(\mu_X(i)) = 0$ and $\tau_i = 0$) which implies $p_i = 0$. The analogous argument applies for $j \in \mathcal{J}$.

We now prove the second part of the theorem. Suppose (Z, p) is an optimal solution to (P) and (D) such that Z is at a vertex. By the Birkhoff-von Neumann theorem, since Z is a vertex, it corresponds to a matching. We wish to show that (X, τ) has no blocking pairs, is individually rational, and has zero-sum transfers. Dual feasibility tells us that:

$$p_i + p_j \geq u_i(j) + u_j(i)$$

which means that:

$$\big(u_i(\mu_X(i)) + \tau_i\big) + \big(u_j(\mu_X(j)) + \tau_j\big) \geq u_i(j) + u_j(i),$$

so there are no blocking pairs. Dual feasibility also tells us that $p_a \geq 0$, which means that $u_a(\mu_X(a)) + \tau_a \geq 0$, so individual rationality is satisfied. To show that there are zero-sum transfers, we use complementary slackness. The first complementary slackness condition tells us that if $Z_{i,j} > 0$, then $p_i + p_j = u_i(j) + u_j(i)$. Using the fact that Z corresponds to a matching, this in particular means that if $(i, j) \in X$, we know $\tau_i + \tau_j = 0$. To show that agents who are unmatched receive 0 transfers, let's use the second and third complementary slackness conditions. The contrapositive tells us that if a is unmatched, then $p_a = 0$, which implies $\tau_a = 0$. □

Since (P) is exactly the maximum weight matching linear program, Chapter 249 immediately tells us that if (X, p) is stable, then X is a maximum weight matching. This means that stable matchings with transferable utilities maximize social welfare.

H.2 Proofs for Chapter 12.4

This section contains further exposition (including proofs) for Chapter 12.4.

H.2.1 Limitations of utility difference as an instability measure

To illustrate why utility difference fails to be a good measure of instability, we describe a matching with transfers that (i) is far from stable, and (ii) has zero utility difference (but large Subset Instability).

Example 14. *Consider the following market with two agents: $\mathcal{I} = \{i\}$ and $\mathcal{J} = \{j\}$. Suppose that $u_i(j) = 2$ and $u_j(i) = -1$. Consider the matching $X = \{(i, j)\}$ with transfers $\tau_i = -\xi$*

and $\tau_j = \xi$ for some $\xi > 0$. We will show that this matching with transfers will have the properties stated above when ξ is large.

This matching with transfers has a utility difference equal to zero (for any ξ) since it maximizes the sum of utilities. Indeed, it is stable for any $\xi \in [1, 2]$. However, when $\xi > 2$, this matching with transfers is no longer stable, since the individual rationality condition $u_i(j) + \tau_i \geq 0$ fails. (Intuitively, the larger ξ is, the further we are from stability.) But its utility difference remains at zero.

On the other hand, the Subset Instability of this matching with transfers is $\xi - 2 > 0$ when $\xi > 2$. In particular, Subset Instability increases with ξ in this regime, which is consistent with the intuition that outcomes with larger ξ should be more unstable.

H.2.2 Proof of Chapter 95

Proposition 95. *Minimum stabilizing subsidy equals Subset Instability for any market outcome.*

Proof of Chapter 95. We can take the dual of the linear program (12.5) to obtain:

$$\max_{\substack{S \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{J}|} \\ Z \in \mathbb{R}^{|\mathcal{A}|}}} \quad \sum_{(i,j) \in \mathcal{I} \times \mathcal{J}} S_{i,j} \Big(\big(u_i(j) - u_i(\mu_X(i)) - \tau_i\big) + \big(u_j(i) - u_j(\mu_X(j)) - \tau_j\big) \Big) - \sum_{a \in \mathcal{A}} Z_a (u_a(\mu_X(a)) + \tau_a) \tag{‡}$$

$$\begin{aligned} \text{s.t.} \quad & Z_i + \sum_{j \in \mathcal{J}} S_{i,j} \leq 1 \quad \forall i \in \mathcal{I}; \qquad && Z_j + \sum_{i \in \mathcal{I}} S_{i,j} \leq 1 \quad \forall j \in \mathcal{J}; \\ & S_{i,j} \geq 0 \quad \forall (i,j) \in \mathcal{I} \times \mathcal{J}; && Z_a \geq 0 \quad \forall a \in \mathcal{A}. \end{aligned}$$

By strong duality, the optimal values of (12.5) and (‡) are equal. Thus, it suffices to show that Subset Instability is equal to (‡). By Chapter 96, we know that Subset Instability is equal to the maximum unhappiness of any coalition. Thus it suffices to show that (‡) is equal to the maximum unhappiness of any coalition.

To interpret (‡), observe that there exist optimal S^* and Z^* all of whose entries lie in $\{0, 1\}$ because this linear program can be embedded into a maximum weight matching linear program. Take such a choice of optimal S^* and Z^*. Then, S^* is an indicator vector corresponding to a (partial) matching on a subset of the agents such that all pairs in this matching are blocking with respect to (X, τ). Similarly, Z^* is an indicator vector of agents who would rather be unmatched than match according to (X, τ).

We first prove the claim that $I(X, \tau; u, \mathcal{A})$ is at least (‡). Based on the above discussion, the optimal objective of (‡) is obtained through S^* and Z^* that represent a matching and a subset of agents respectively. Let S be the union of agents participating in S^* and Z^*. We see that the objective of (‡) is equal to the utility difference at S, i.e.:

$$\left(\max_{X' \in \mathscr{X}_S} \sum_{a \in S} u_a(\mu_{X'}(a)) \right) - \left(\sum_{a \in S} u_a(\mu_X(a)) + \tau_a \right).$$

This is no larger than Subset Instability by definition.

We next prove the claim that $I(X, \tau; u, \mathcal{A})$ is at most (‡). Let's consider S^* that maximizes:

$$\max_{S \subseteq \mathcal{A}} \left(\max_{X' \in \mathscr{X}_S} \sum_{a \in S} u_a(\mu_{X'}(a)) \right) - \left(\sum_{a \in S} u_a(\mu_X(a)) + \tau_a \right).$$

Let's take the maximum weight matching of S^*. Let S be given by the matched agents in this matching and let Z be given by the unmatched agents in this matching (using the interpretation of (‡) described above). We see that the objective at (‡) for (S, Z) is equal to Subset Instability which proves the desired statement.

□

H.2.3 Proof of Chapter 96

We first formally define the *unhappiness of a coalition*, as follows. In particular, the unhappiness with respect to (X, τ) of a coalition $\mathcal{S} \subseteq \mathcal{A}$ is defined to be:

$$\begin{aligned} \sup_{\substack{X' \in \mathscr{X}_{\mathcal{S}} \\ \tau' \in \mathrm{R}^{|\mathcal{S}|}}} \; & \sum_{a \in \mathcal{S}} \big(u_a(\mu_{X'}(a)) + \tau'_a\big) - \sum_{a \in \mathcal{S}} \big(u_a(\mu_X(a)) + \tau_a\big) \\ \text{s.t.} \; & u_a(\mu_{X'}(a)) + \tau'_a \geq u_a(\mu_X(a)) + \tau_a \qquad \forall a \in \mathcal{S} \\ & \tau'_a + \tau'_{\mu_{X'}(a)} = 0 \qquad \forall a \in \mathcal{S}, \end{aligned} \tag{H.1}$$

with unhappiness being 0 if there are no feasible X' and τ'. In the optimization program, (X', τ') represents a matching with transfers over $\mathcal{S}$, with the constraint $\tau'_a + \tau'_{\mu_{X'}(a)} = 0$ ensuring that it is zero-sum. The objective measures the difference between (X, τ) and (X', τ') of the total utility of the agents in $\mathcal{S}$. The constraint $u_a(\mu_{X'}(a)) + \tau'_a \geq u_a(\mu_X(a)) + \tau_a$ encodes the requirement that all agents be at least as well off under (X', τ') as they were under (X, τ). This optimization program therefore captures the objective of $\mathcal{S}$ to maximize their total payoff while ensuring that no member of the coalition is worse off than they were according to (X, τ).

Recall that, in terms of unhappiness, Chapter 96 is as follows:

Proposition 96. *The maximum unhappiness of any coalition $\mathcal{S} \subseteq \mathcal{A}$ with respect to (X, τ) equals the Subset Instability $I(X, \tau; u, \mathcal{A})$.*

Proof of Proposition 96. By Chapter 95, we know that Subset Instability is equal to (12.5). Moreover, by strong duality, we know that Subset Instability is equal to (‡) (the dual linear program of (12.5)). Thus, it suffices to prove that the maximum unhappiness of any coalition is equal to (‡).

We first prove the claim that (‡) is at most the maximum unhappiness of any coalition with respect to (X, τ). To do this, it suffices to construct a coalition $\mathcal{S} \subseteq \mathcal{A}$ such that (‡) is at most the unhappiness of $\mathcal{S}$. We construct $\mathcal{S}$ as follows: Recall that there exist optimal solutions S^* and Z^* to (‡) such that S^* corresponds to a (partial) matching on $\mathcal{I} \times \mathcal{J}$ and Z^* corresponds to a subset of $\mathcal{A}$. We may take $\mathcal{S}$ to be the union of the agents involved in S^* and in Z^*. Now, we upper bound the unhappiness of $\mathcal{S}$ by constructing X' and τ' that are feasible for (H.1). We can take X' to be the matching that corresponds to the indicator vector S^*. Because (S^*, Z^*) is optimal for (‡),

$$u_i(j) + u_j(i) \geq (u_i(\mu_X(i)) + \tau_i) + (u_j(\mu_X(j)) + \tau_j)$$

for all $(i, j) \in X'$. Thus, we can find a vector τ' of transfers that is feasible for (H.1). Then, since $\sum_{a \in \mathcal{S}} \tau'_a = 0$, the objective of (H.1) at (X', τ') is

$$\sum_{a \in \mathcal{S}} \big(u_a(\mu_{X'}(a)) - u_a(\mu_X(a)) - \tau_a\big).$$

This equals to the objective of (‡) at (S^*, Z^*), which equals (‡), as desired.

We now show the inequality in the other direction, that (‡) is at least the maximum unhappiness of any coalition with respect to (X, τ). It suffices to construct a feasible solution (S, Z) to (‡) that achieves at least the maximum unhappiness of any coalition. Let $\mathcal{S}$ be a coalition with maximum unhappiness, and let (X', τ') be an optimal solution for (H.1). Moreover, let S be the indicator vector corresponding to agents who are matched in X' and Z be the indicator vector corresponding to agents in $\mathcal{S}$ who are unmatched. The objective of (H.1) at (X', τ') is

$$\sum_{a \in \mathcal{S}} \big(u_a(\mu_{X'}(a)) - u_a(\mu_X(a)) - \tau_a\big),$$

which equals the objective of (‡) at the (S, Z) that we constructed.

□

H.2.4 Proof of Chapter 97

Proposition 97. *Subset Instability satisfies the following properties:*

1. *Subset Instability is always nonnegative and is zero if and only if (X, τ) is stable.*
2. *Subset Instability is Lipschitz continuous with respect to agent utilities. That is, for any possible market outcome (X, τ), and any pair of utility functions u and uii it holds that:*

$$|I(X, \tau; u, \mathcal{A}) - I(X, \tau; uii, \mathcal{A})| \leq 2 \sum_{a \in \mathcal{A}} \|u_a - uii_a\|_\infty.$$

3. *Subset Instability is always at least the utility difference.*

Proof of Chapter 97. We first prove the third part of the Proposition statement, then the first part of the Proposition statement, and finally the second part.

Proof of part (c). Because $\sum_{a\in\mathcal{A}}\tau_a = 0$, Subset Instability satisfies the following:

$$\begin{aligned} I(X,\tau;u,\mathcal{A}) &\geq \left(\max_{X'\in\mathscr{X}_\mathcal{A}}\sum_{a\in\mathcal{A}} u_a(\mu_{X'}(a))\right) - \left(\sum_{a\in\mathcal{A}} u_a(\mu_X(a)) + \tau_a\right) \\ &= \left(\max_{X'\in\mathscr{X}_\mathcal{A}}\sum_{a\in\mathcal{A}} u_a(\mu_{X'}(a))\right) - \left(\sum_{a\in\mathcal{A}} u_a(\mu_X(a))\right). \end{aligned}$$

The second line is exactly the utility difference.

Proof of part (a). From above, we have that Subset Instability is lower bounded by the utility difference, which is always nonnegative. Hence Subset Instability is also always nonnegative.

To see that Subset Instability is 0 if and only if (X,τ) is stable, first suppose (X,τ) is unstable. Then, there exists a blocking pair (i,j), in which case

$$I(X,\tau;u,\mathcal{A}) \geq u_i(j) + u_j(i) - (u_i(\mu_X(i)) + u_j(\mu_X(j)) + \tau_i + \tau_j) > 0$$

by the definition of blocking. Now, suppose $I(X,\tau;u,\mathcal{A}) > 0$. Then, there exists a subset $S\subseteq\mathcal{A}$ such that

$$\left(\max_{X'\in\mathscr{X}_S}\sum_{a\in S} u_a(\mu_{X'}(a))\right) - \left(\sum_{a\in S} u_a(\mu_X(a)) + \tau_a\right) > 0.$$

Let X' be a maximum weight matching on S. We can rewrite the above as

$$\sum_{(i,j)\in X'} \big(u_i(j) + u_j(i) - (u_i(\mu_X(i)) + u_j(\mu_X(j)) + \tau_i + \tau_j\big) > 0.$$

Some term in the sum on the left-hand side must be positive, so there exists a blocking pair $(i,j)\in X'$. In particular, (X,τ) is not stable.

Proof of part (b). We prove that

$$|I(X,\tau;u,\mathcal{A}) - I(X,\tau;\tilde{u},\mathcal{A})| \leq 2\sum_{a\in\mathcal{A}}\|u_a - uii_a\|_\infty.$$

The supremum of L-Lipschitz functions is L-Lipschitz, so it suffices to show that

$$\left(\max_{X'\in\mathscr{X}_S}\sum_{a\in S} u_a(\mu_{X'}(a))\right) - \sum_{a\in S}(u_a(\mu_X(a)) + \tau_a)$$

satisfies the desired Lipschitz condition for any $S\subseteq\mathcal{A}$. In particular, it suffices to show that

$$\left|\sum_{a\in S}(u_a(\mu_X(a)) + \tau_a) - \sum_{a\in S}(uii_a(\mu_X(a)) + \tau_a)\right| \leq \sum_{a\in\mathcal{A}}\|u_a - uii_a\|_\infty \tag{H.2}$$

and

$$\left|\left(\max_{X'\in\mathscr{X}_S}\sum_{a\in S}u_a(\mu_{X'}(a))\right)-\left(\max_{X'\in\mathscr{X}_S}\sum_{a\in S}uii_a(\mu_{X'}(a))\right)\right|\leq\sum_{a\in\mathcal{A}}\|u_a-uii_a\|_\infty. \tag{H.3}$$

For (H.2), we have

$$\begin{aligned}\left|\sum_{a\in S}(u_a(\mu_X(a))+\tau_a)-\sum_{a\in S}(uii_a(\mu_X(a))+\tau_a)\right| &= \left|\sum_{a\in S}\big(u_a(\mu_X(a))-uii_a(\mu_X(a))\big)\right| \\ &\leq \sum_{a\in\mathcal{A}}\|u_a-uii_a\|_\infty.\end{aligned}$$

For (H.3), this boils down to showing that the total utility of the maximum weight matching is Lipschitz. Using again the fact that the supremum of Lipschitz functions is Lipschitz, this follows from the total utility of any fixed matching being Lipschitz. □

H.3 Proofs for Chapter 12.5

H.3.1 Proof of Chapter 98

Theorem 98. *For preference class $\mathcal{U}_{\text{unstructured}}$ (see Chapter 12.3),* MatchUCB *(defined in Chapter 12.5.3) incurs expected regret $\mathbb{E}(R_T)=O\big(|\mathcal{A}|\sqrt{nT\log(|\mathcal{A}|T)}\big)$, where $n=\max_t|\mathcal{A}_t|$.*

Proof of Theorem 98. The starting point for our proof of Theorem 98 is the typical approach in multi-armed bandits and combinatorial bandits Gai et al. (2012); Chen et al. (2013); Lattimore and Szepesvári (2020) of bounding regret in terms of the sizes of the confidence interval of the chosen arms. However, rather than using the sizes of confidence intervals to bound the utility difference (as in the incentive-free maximum weight matching setting), we bound Subset Instability through Chapter 101. From here on, our approach composes cleanly with existing bandits analyses; in particular, we can follow the typical combinatorial bandits approach Gai et al. (2012); Chen et al. (2013) to get the desired upper bound.

For completeness, we present the full proof. We divide into two cases, based on the event E that all of the confidence sets contain their respective true utilities at every time step $t\leq T$. That is, $u_i(j)\in C_{i,j}$ and $u_j(i)\in C_{j,i}$ for all $(i,j)\in\mathcal{I}\times\mathcal{J}$ at all t.

Case 1: E holds. By Chapter 101, we may bound

$$I(X^t,\tau^t;u,\mathcal{A}^t)\leq\sum_{a\in\mathcal{A}^t}\Big(\max\big(C_{a,\mu_{X^t}(a)}\big)-\min\big(C_{a,\mu_{X^t}(a)}\big)\Big)=O\left(\sum_{(i,j)\in X^t}\sqrt{\frac{\log(|\mathcal{A}|T)}{n^t_{ij}}}\right),$$

where n^t_{ij} is the number of times that the pair (i,j) has been matched at the start of round t. Let $w^t_{i,j}=\frac{1}{\sqrt{n^t_{ij}}}$ be the size of the confidence set (with the log factor scaled out) for (i,j) at the start of round t.

At each time step t, let's consider the list consisting of $w^t_{i_t,j_t}$ for all $(i_t, j_t) \in X^t$. Let's now consider the overall list consisting of the concatenation of all of these lists over all rounds. Let's order this list in decreasing order to obtain a list $\tilde{w}_1, \ldots, \tilde{w}_L$ where $L = \sum_{t=1}^T |X^t| \le nT$. In this notation, we observe that:

$$\sum_{t=1}^T I(X^t, \tau^t; u, \mathcal{A}) \le \sum_{t=1}^T \sum_{a \in \mathcal{A}^t} \Big(\max\big(C_{a,\mu_{X^t(a)}}\big) - \min\big(C_{a,\mu_{X^t(a)}}\big)\Big) = \log(|\mathcal{A}|T) \sum_{l=1}^L \tilde{w}_l.$$

We claim that $\tilde{w}_l \le O\Big(\min(1, \frac{1}{\sqrt{(l/|\mathcal{A}|^2)-1}})\Big)$. The number of rounds that a pair of agents can have their confidence set have size at least $\tilde{w}_l$ is upper bounded by $1 + \frac{1}{\tilde{w}_l^2}$. Thus, the total number of times that any confidence set can have size at least $\tilde{w}_l$ is upper bounded by $(|\mathcal{A}|^2)(1 + \frac{1}{\tilde{w}_l^2})$.

Putting this together, we see that:

$$\begin{aligned}
\log(|\mathcal{A}|T) \sum_{l=1}^L \tilde{w}_l &\le O\left(\sum_{l=1}^L \min(1, \frac{1}{\sqrt{(l/|\mathcal{A}|^2) - 1}})\right) \\
&\le O\left(\log(|\mathcal{A}|T) \sum_{l=1}^{nT} \min(1, \frac{1}{\sqrt{(l/|\mathcal{A}|^2) - 1}})\right) \\
&\le O\Big(|\mathcal{A}|\sqrt{nT}\log(|\mathcal{A}|T)\Big).
\end{aligned}$$

Case 2: E does not hold. Since each $n_{ij}(\hat{u}_i(j) - u_i(j))$ is mean-zero and 1-subgaussian, and we have $O(|\mathcal{I}||\mathcal{J}|T)$ such random variables over T rounds, the probability that any of them exceeds

$$2\sqrt{\log(|\mathcal{I}||\mathcal{J}|T/\delta)} \le 2\sqrt{\log(|\mathcal{A}|^2T/\delta)}$$

is at most δ by a standard tail bound for the maximum of subgaussian random variables. It follows that E fails to hold with probability at most $|\mathcal{A}|^{-2}T^{-2}$. In the case that E fails to hold, our regret in any given round would be at most $4|\mathcal{A}|$ by the Lipschitz property in Chapter 97. (Recall that our upper confidence bound for any utility is wrong by at most 2 due to clipping each confidence interval to lie in $[-1, 1]$.) Thus, the expected regret from this scenario is at most

$$|\mathcal{A}|^{-2}T^{-2} \cdot 4|\mathcal{A}|T \le 4|\mathcal{A}|^{-1}T^{-1},$$

which is negligible compared to the regret bound from when E does occur. □

H.3.2 Proof of Chapter 99

Theorem 99. *For preference class $\mathcal{U}_{\text{typed}}$ (see Chapter 12.3), MatchTypedUCB (defined in Chapter 12.5.3) incurs expected regret $\mathbb{E}(R_T) = O\big(|\mathcal{C}|\sqrt{nT\log(|\mathcal{A}|T)}\big)$, where $n = \max_t |\mathcal{A}_t|$.*

Proof of Chapter 99. Like in the proof of Chapter 98, we divide into two cases, based on the event E that all of the confidence sets contain their respective true utilities at every time step $t \leq T$. That is, $u_a(a') \in C_{a,a'}$ for all pairs of agents at all t.

Case 1: E holds. By Chapter 101, we may bound

$$I(X^t, \tau^t; u, \mathcal{A}^t) \leq \sum_{a \in \mathcal{A}^t} \Big(\max\big(C_{c_a, c_{\mu_{X^t(a)}}}\big) - \min\big(C_{c_a, c_{\mu_{X^t(a)}}}\big) \Big) = O\left(\sum_{(i,j) \in X^t} \sqrt{\frac{\log(|\mathcal{A}|T)}{n^t_{c_i c_j}}} \right),$$

where $n^t_{c_1 c_2}$ is the number of times that the an agent of type c_1 has been matched with an agent of context c_2 at the start of round t. (We define $n^0_{c_1,c_2} = 0$ by default.) Let $w^t_{c_1,c_2} = \frac{1}{\sqrt{n^t_{c_1,c_2}}}$ be the size of the confidence set (with the log factor scaled out) for (c_1, c_2) at the start of round t.

At each time step t, let's consider the list consisting of $w^t_{c_{i_t}, c_{j_t}}$ for all $(i_t, j_t) \in X^t$. Let's now consider the overall list consisting of the concatenation of all of these lists over all rounds. Let's order this list in decreasing order to obtain a list $\tilde{w}_1, \ldots, \tilde{w}_L$ where $L = \sum_{t=1}^T |X^t| \leq nT$. In this notation, we observe that:

$$\sum_{t=1}^T I(X^t, \tau^t; u, \mathcal{A}^t) \leq \sum_{t=1}^T \sum_{a \in \mathcal{A}^t} \Big(\max\big(C_{c_a, c_{\mu_{X^t(a)}}}\big) - \min\big(C_{c_a, c_{\mu_{X^t(a)}}}\big) \Big) = \log(|\mathcal{A}|T) \sum_{l=1}^L \tilde{w}_l.$$

We claim that $\tilde{w}_l \leq O\Big(\min(1, \frac{1}{\sqrt{(l/|\mathcal{C}|^2) - 1}}) \Big)$. The number of instances that a pair of contexts can have their confidence set have size at least $\tilde{w}_l$ is upper bounded by $2n + \frac{1}{\tilde{w}_l^2}$. Thus, the total number of times that any confidence set can have size at least $\tilde{w}_l$ is upper bounded by $(|\mathcal{C}|)(2n + \frac{1}{\tilde{w}_l^2})$.

Putting this together, we see that:

$$\begin{aligned}
\log(|\mathcal{A}|T) \sum_{l=1}^L \tilde{w}_l &\leq O\left(\sum_{l=1}^L \min(1, \frac{1}{\sqrt{(l/|\mathcal{A}|^2) - 1}}) \right) \\
&\leq O\left(\log(|\mathcal{A}|T) \sum_{l=1}^{nT} \min(1, \frac{1}{\sqrt{(l/|\mathcal{C}|^2) - 1}}) \right) \\
&\leq O\Big(|\mathcal{C}| \sqrt{nT} \log(|\mathcal{C}|^2 T) \Big).
\end{aligned}$$

Case 2: E does not hold. Since each $n_{ij}(\hat{u}_i(j) - u_i(j))$ is mean-zero and 1-subgaussian, and we have $O(|\mathcal{I}||\mathcal{J}|T)$ such random variables over T rounds, the probability that any of them exceeds

$$2\sqrt{\log(|\mathcal{I}||\mathcal{J}|T/\delta)} \leq 2\sqrt{\log(|\mathcal{A}|^2 T/\delta)}$$

is at most δ by a standard tail bound for the maximum of subgaussian random variables. It follows that E fails to hold with probability at most $|\mathcal{A}|^{-2}T^{-2}$. In the case that E fails to hold, our regret in any given round would be at most $4|\mathcal{A}|$ by the Lipschitz property in Chapter 97. (Recall that our upper confidence bound for any utility is wrong by at most two due to clipping each confidence interval to lie in $[-1,1]$.) Thus, the expected regret from this scenario is at most

$$|\mathcal{A}|^{-2}T^{-2} \cdot 4|\mathcal{A}|T \leq 4|\mathcal{A}|^{-1}T^{-1},$$

which is negligible compared to the regret bound from when E does occur. □

H.3.3 Proof of Chapter 100

Theorem 100. *For preference class $\mathcal{U}_{\text{linear}}$ (see Chapter 12.3),* MatchLinUCB *(defined in Chapter 12.5.3) incurs expected regret $\mathbb{E}(R_T) = O\big(d\sqrt{|\mathcal{A}|}\sqrt{nT\log(|\mathcal{A}|T)}\big)$, where $n = \max_t |\mathcal{A}_t|$.*

To prove Chapter 100, it suffices to (a) show that the confidence sets contain the true utilities with high probability, and (b) bound the sum of the sizes of the confidence sets.

Part (a) follows from fact established in existing analysis of LinUCB in the classical linear contextual bandits setting Russo and Van Roy (2013).

Lemma 250 ((Russo and Van Roy, 2013, Proposition 2))**.** *Let the confidence sets be defined as above (and in* MatchLinUCB*). For each $a \in \mathcal{A}$, it holds that:*

$$\mathbb{P}[\phi(a) \in C_{\phi(a)} \quad \forall 1 \leq t \leq T] \geq 1 - 1/(|\mathcal{A}|^3 T^2).$$

Lemma 251. *Let the confidence sets be defined as above (and in* MatchLinUCB*). For each $a \in \mathcal{A}$ and for any $\varepsilon > 0$, it holds that:*

$$\sum_{t \mid a \in \mathcal{A}^t, \mu_{X^t}(a) \neq a} \mathbf{1}\left[\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big) > \varepsilon\right] \leq O\left(\left(\frac{4\beta_T}{\varepsilon^2} + 1\right) d \log(1/\varepsilon)\right).$$

Proof. We follow the same argument as the proof of Proposition 3 in Russo and Van Roy (2013).

We first recall the definition of ε-dependence and ε-eluder dimension: We say that an agent a' is *ε-dependent* on $a'_1, \ldots, a'_s$ if for all $\phi(a), \tilde{\phi}(a) \in \mathcal{B}^d$ such that

$$\sum_{k=1}^{s} \langle c_{a'_k}, \tilde{\phi}(a) - \phi(a)\rangle^2 \leq \varepsilon^2,$$

we also have $\langle c_{a'}, \tilde{\phi}(a) - \phi(a)\rangle^2 \leq \varepsilon^2$. The ε-eluder dimension $d_{\varepsilon\text{-eluder}}$ of $\mathcal{B}^d$ is the maximum length of a sequence $a'_1, \ldots, a'_s$ such that no element is ε-dependent on a prefix.

Consider the subset S_a of $\{t \mid a \in \mathcal{A}^t, \mu_{X^t}(a) \neq a\}$ such that

$$\mathbf{1}\left[\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big) > \varepsilon\right].$$

Suppose for the sake of contradiction that

$$|S_a| > \left(\frac{4\beta_T}{\varepsilon^2} + 1\right) d_{\varepsilon\text{-eluder}}.$$

Then, there exists an element t^* that is ε-dependent on $\frac{4\beta_T}{\varepsilon^2} + 1$ disjoint subsets of S_a: One can repeatedly remove sequences $a'_{\mu_{X^{t_1}}(a)}, \ldots, a'_{\mu_{X^{t_s}}(a)}$ of maximal length such that no element is ε-dependent on a prefix; note that $s \leq d_{\varepsilon\text{-eluder}}$ always. Let the subsets be $S_a^{(q)}$ for $q = 1, \ldots, \frac{4\beta_T}{\varepsilon^2} + 1$, and let $\phi(a), \tilde{\phi}(a)$ be such that $\langle c_{\mu_{X^{t^*}}(a)}, \tilde{\phi}(a) - \phi(a)\rangle > \varepsilon$. The above implies that

$$\sum_{q=1}^{\frac{4\beta_T}{\varepsilon^2}+1} \sum_{t \in S_a^{(q)}} \langle c\mu_{X^t}(a), \tilde{\phi}(a) - \phi(a)\rangle^2 > 4\beta_T$$

by the definition of ε-dependence. But this is impossible, since the left-hand side is upper bounded by

$$\sum_{t=1}^{T} \langle c\mu_{X^t}(a), \tilde{\phi}(a) - \phi(a)\rangle^2 \leq 4\beta_T$$

by the definition of the confidence sets. Hence it must hold that

$$|S_a| \leq \left(\frac{4\beta_T}{\varepsilon^2} + 1\right) d_{\varepsilon\text{-eluder}}.$$

Now, it follows from the bound on the eluder dimension for linear bandits (Proposition 6 in Russo and Van Roy (2013)) that the bound of $\tilde{O}\left(\left(\frac{4\beta_T}{\varepsilon^2} + 1\right) d \log(1/\varepsilon).\right)$ holds. □

Lemma 252. *Let the confidence sets be defined as above (and in* MatchLinUCB*). For any $a \in \mathcal{A}$, it holds that:*

$$\sum_{t \mid a \in \mathcal{A}^t, \mu_{X^t}(a) \neq a} \Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big)\Big) \leq O(d(\log(T|\mathcal{A}|))\sqrt{T_a}),$$

where T_a is the number of times that agents is matched.

Proof. Let's consider the set of confidence set sizes $\Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big)\Big)$ for t such that $a \in \mathcal{A}^t, \mu_{X^t}$. Let's sort these confidence set sizes in decreasing order and label them $w_1 \geq \ldots \geq w_{T_a}$. Restating Chapter 251, we see that

$$\sum_{t=1}^{T_a} w_t \mathbf{1}[w_t > \varepsilon] \leq O\left(\left(\frac{4\beta_T}{\varepsilon^2} + 1\right) d \log(1/\varepsilon)\right). \tag{H.4}$$

for all $\varepsilon > 0$.

We see that:

$$\begin{aligned}
\sum_{t \mid a \in \mathcal{A}^t, \mu_{X^t}(a) \neq a} \Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big)\Big) &= \sum_{t=1}^{T_a} w_t \\
&\leq \sum_{t=1}^{T_a} w_t \mathbf{1}[w_t > 1/T_a^2] + \sum_{t=1}^{T_a} w_t \mathbf{1}[w_t \leq 1/T_a^2] \\
&\leq \frac{1}{T_a} + \sum_{t=1}^{T_a} w_t \mathbf{1}[w_t > 1/T_a^2].
\end{aligned}$$

We claim that $w_i \leq 2$ if $i \geq d\log(T_a)$ and $w_i \leq \min(2, \frac{4\beta_T(d\log T_a)}{i - d\log T_a})$ if $i > d\log T_a$. The first part follows from the fact that we truncate the confidence sets to be within $[-1, 1]$. It thus suffices to show that $w_i \leq \frac{4\beta_T(d\log T_a)}{i - d\log T_a}$ for $t \leq d\log T$. If $w_i \geq \varepsilon > 1/T_a^2$, then we see that $\sum_{t=1}^{T_a} \mathbf{1}[w_t > \varepsilon] \geq i$, which means by (H.4) that $i \leq O\left(\left(\frac{4\beta_T}{\varepsilon^2} + 1\right) d\log(1/\varepsilon)\right) \leq O\left(\left(\frac{4\beta_T}{\varepsilon^2} + 1\right) d\log(T_a)\right)$ which means that $\varepsilon \leq \frac{4\beta_T(d\log T_a)}{i - d\log T_a}$. This proves the desired statement.

Now, we can plug this into the above expression to obtain:

$$\begin{aligned}
&\sum_{t \mid a \in \mathcal{A}^t, \mu_{X^t}(a) \neq a} \Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big)\Big) \\
&\leq \frac{1}{T_a} + \sum_{t=1}^{T_a} w_t \mathbf{1}[w_t > 1/T_a^2] \\
&\leq \frac{1}{T_a} + 2d\log(T_a) + \sum_{i > d\log T_a}^{T_a} \min\left(2, \frac{4\beta_T(d\log T_a)}{i - d\log T_a}\right) \\
&\leq \frac{1}{T_a} + 2d\log(T_a) + 2\sqrt{d\log T_a \beta_T} \int_{t=0}^{T_a} t^{-1/2} dt \\
&= \frac{1}{T_a} + 2d\log(T_a) + 4\sqrt{dT_a \log T_a \beta_T}.
\end{aligned}$$

We now use the fact that:

$$\beta_T = O(d\log T + \frac{1}{T}\sqrt{\log(T^2|A|)}).$$

Plugging this into the above expression, we obtain the desired result. □

We are now ready to prove Theorem 100.

Proof of Theorem 100. Like in the proof of Chapter 98, we divide into two cases, based on the event E that all of the confidence sets contain their respective true utilities at every time step $t \leq T$. That is, $u_{c_1}(c_2) \in C_{c_1,c_2}$ for all $c_1, c_2 \in \mathcal{C}$ at all t.

Case 1: E **holds.** By Chapter 101, we know that the cumulative regret is upper bounded by

$$\begin{aligned} R_T &\leq \sum_{t=1}^{T} \sum_{a \in \mathcal{A}^t} \Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big) \Big) \\ &= \sum_{a \in \mathcal{A}} \sum_{t \mid a \in \mathcal{A}^t, \mu_{X^t}(a) \neq a} \Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big)\big) \Big) \\ &\leq \sum_{a \in \mathcal{A}} O(d \log(T|\mathcal{A}|)\sqrt{T_a}), \end{aligned}$$

where the last inequality applies Chapter 252 to the inner summand. We see that $\sum_{a \in \mathcal{A}} T_a = \sum_t |\mathcal{A}_t| \leq nT$ by definition, since at most n agents show up at every round. Let's now observe that:

$$\sum_{a \in \mathcal{A}} \sqrt{T_a} \leq \sqrt{|\mathcal{A}|} \sqrt{\sum_{a \in \mathcal{A}} T_a} \leq \sqrt{|\mathcal{A}| nT},$$

as desired.

Case 2: E **does not hold.** From Chapter 250, it follows that:

$$\mathbb{P}[\phi(a) \in C_{\phi(a)} \quad \forall 1 \leq t \leq T] \geq 1 - 1/(|\mathcal{A}|^3 T^2).$$

Union bounding, we see that

$$\mathbb{P}[\phi(a) \in C_{\phi(a)} \quad \forall 1 \leq t \leq T \forall a \in \mathcal{A}] \geq 1 - 1/(|\mathcal{A}|^2 T^2).$$

By the definition of the confidence sets for the utilities, we see that:

$$\mathbb{P}[u(a, a') \in C_{a,a'} \quad \forall 1 \leq t \leq T, \forall a, a' \in \mathcal{A}] \geq 1/(|\mathcal{A}|^2 T^2). \tag{H.5}$$

Thus, the probability that event E does not hold is at most $|\mathcal{A}|^{-2} T^{-2}$. In the case that E fails to hold, our regret in any given round would be at most $4|\mathcal{A}|$ by the Lipschitz property in Chapter 97. Thus, the expected regret is at most $4|\mathcal{A}|^{-1} T^{-1}$ which is negligible compared to the regret bound from when E does occur.

□

H.3.4 Proof of Chapter 102

Lemma 102. *For any algorithm that learns a stable matching with respect to unstructured preferences, there exists an instance on which it has expected regret $\widetilde{\Omega}(|A|^{3/2}\sqrt{T})$ (where regret is given by Subset Instability).*

Proof of Chapter 102. Recall that, by Chapter 97, the problem of learning a maximum weight matching with respect to utility difference is no harder than that of learning a stable matching with respect to Subset Instability. In the remainder of our proof, we reduce a standard "hard instance" for stochastic multi-armed bandits to our setting of learning a maximum weight matching.

Step 1: Constructing the hard instance for stochastic MAB. Consider the following family of stochastic multi-armed bandits instances: for a fixed K, let $\mathcal{I}_\alpha$ for $\alpha \in \{1, \ldots, K\}$ denote the stochastic multi-armed bandits problem where all arms have 0-1 rewards, and the k-th arm has mean reward $\frac{1}{2} + \rho$ if $k = \alpha$ and $\frac{1}{2}$ otherwise, where $\rho > 0$ will be set later. A classical lower bound for stochastic multi-armed bandits is the following:

Lemma 253 (Auer et al. (2002b)). *The expected regret of any stochastic multi-armed bandit algorithm on an instance $\mathcal{I}_\alpha$ for α selected uniformly at random from $\{1, \ldots, K\}$ is $\Omega(\sqrt{KT})$.*

Step 2: Constructing a (random) instance for the maximum weight matching problem. We will reduce solving the above distribution over stochastic multi-armed bandits problems to a distribution over instances of learning a maximum weight matching. Let us now construct this random instance of the maximum weight matching problem. Let $|\mathcal{I}| = K$ and $|\mathcal{J}| = 10K \log(KT)$. Specifically, we sample inputs for learning a maximum weight matching as follows: For each man $i \in \mathcal{I}$, select $\alpha_i \in \{1, \ldots, K\}$ uniformly at random, and define $u_i(j)$ to be $\frac{1}{2} + \rho$ if $\lfloor (j-1)/\log K \rfloor = \alpha_i$ and $\frac{1}{2}$ otherwise. Furthermore, let $u_j(i) = 0$ for all $(i, j) \in \mathcal{I} \times \mathcal{J}$. Finally, suppose observations are always in $\{0, 1\}$ (but are unbiased).

The key property of the above setup that we will exploit for our reduction is the fact that, due to the imbalance in the market, the maximum weight matching for these utilities has with high probability each i matched with some j whom they value at $\frac{1}{2} + \rho$. Indeed, by a union bound, the probability that more than $10 \log(KT)$ different i have the same α_i is at most

$$K \cdot \binom{K}{10\log(KT)} K^{-10\log(KT)} = O\big(K^{-4}T^{-4}\big).$$

Thus, with probability $1 - O(K^{-4}T^{-4})$, this event holds. The case where this event does not hold contributes negligibly to regret, so we do not consider it further.

Step 3: Establishing the reduction. Now, suppose for the sake of contradiction that some algorithm could solve our random instance of learning a maximum weight matching problem with expected regret $o(K^{3/2}\sqrt{T})$. We can obtain a stochastic multi-armed bandits that solves the instances in Chapter 253 as follows: Choose a random $i^* \in \mathcal{I}$ and set $\alpha_{i^*} = \alpha$. Simulate the remaining i by choosing α_i for all $i \neq i^*$ uniformly at random. Run the algorithm on this instance of learning a maximum weight matching, "forwarding" arm pulls to the true instance when matching i^*.

To analyze the regret of this algorithm when faced with the distribution from Chapter 253, we first note that with high probability, all the agents $i \in \mathcal{I}$ can simultaneously be matched to a set of $j \in \mathcal{J}$ such that each i is matched to some j whom they value at $\frac{1}{2} + \rho$. Then, the regret of any matching is ρ times the number of $i \in \mathcal{I}$ who are not matched to a j whom they value at $\frac{1}{2} + \rho$. Thus, we can define the cumulative regret for an agent $i \in \mathcal{I}$ as ρ times the number of rounds they were not matched to someone whom they value at $\frac{1}{2} + \rho$. For i^*, this regret is just the regret for the distribution from Chapter 253. Since i^* was chosen uniformly at random, their expected cumulative regret is at most

$$\frac{1}{K} \cdot o(K^{3/2}\sqrt{T}) = o(\sqrt{KT}),$$

in violation of Chapter 253.

Step 4: Concluding the lower bound. This contradiction implies that no algorithm can hope to obtain $o(K^{3/2}\sqrt{T})$ expected regret on this distribution over instances of learning a maximum weight matching. Since there are $O(K\log(KT)) = \widetilde{O}(K)$ agents in the market total, the desired lower bound follows. □

H.4 Proof of Theorem 103

Theorem 103 (Instance-Dependent Regret)**.** *Suppose that* $\mathcal{A}_t = \mathcal{A}$ *for all* t*. Let* $u \in \mathcal{U}_{\text{unstructured}}$ *be any utility function, and put*

$$\Delta := \inf_{X \neq X^*} \left\{ \sum_{a\in\mathcal{A}} u_a(\mu_{X^*}(a)) - \sum_{a\in\mathcal{A}} u_a(\mu_X(a)) \right\}.$$

Then MatchUCB′ *incurs expected regret* $\mathbb{E}(R_T) = O(|\mathcal{A}|^5 \cdot \log(|\mathcal{A}|T)/\Delta^2)$.

H.4.1 MatchUCB′

MatchUCB′ is the same as MatchUCB, except we call ComputeMatch′ instead of ComputeMatch. The idea behind ComputeMatch′ is that we compute an optimal primal-dual solution for both the original confidence sets C as well as expanded confidence sets C', which we define to be twice the width of the original confidence sets. More formally, we define

$$C'_{a,a'} := \left[\min(C_{a,a'}) - \frac{\max(C_{a,a'}) - \min(C_{a,a'})}{2}, \max(C_{a,a'}) + \frac{\max(C_{a,a'}) - \min(C_{a,a'})}{2}\right].$$

We will adaptively explore (following UCB) according to both C and C'. Doing extra exploration according to the more pessimistic confidence sets C' is necessary for us to be able to find "robust" dual solutions for setting transfers.

We define (X^*, p^*), which will be an optimal primal-dual solution for the upper confidence bounds of C as follows. Let X^* be a maximum weight matching with respect to u^{UCB}. We next compute the gap

$$\Delta^{\text{UCB}} = \min_{X \neq X^*} \left\{ \sum_{a\in\mathcal{A}} u_a^{\text{UCB}}(\mu_{X^*}(a)) - \sum_{a\in\mathcal{A}} u_a^{\text{UCB}}(\mu_X(a)) \right\}$$

with respect to u^{UCB}. We can compute this gap by computing the maximum weight matching and the second-best matching with respect to u^{UCB}.[1] Next, define utility functions u'_a such that

$$u'_a(a') = \begin{cases} u_a^{\text{UCB}}(a') - \frac{\Delta^{\text{UCB}}}{|\mathcal{A}|} & \text{if } \mu_{X^*}(a) = a' \text{ and } a \neq a' \\ u_a^{\text{UCB}}(a') & \text{otherwise} \end{cases}$$

for all $a \in \mathcal{A}$. (We show in Chapter 254 that X^* is still a maximum weight matching for u'.) Now, compute an optimal dual solution p' for utility function u'. To get p^*, we add $\Delta^{\text{UCB}}/|\mathcal{A}|$ to p'_a for each matched agent a in X^*. (See Chapter 255 for a proof that (X^*, p^*) is an optimal primal-dual pair with respect to u^{UCB}.)

Finally, let $(X^{*,2}, p^{*,2})$ be any optimal primal-dual pair for the utility function $u^{\text{UCB},2}$ given by the upper confidence bounds $\max(C'_{a,a'})$ of C'.

With this setup, we define ComputeMatch$'$ as follows: If $X^* \neq X^{*,2}$, return $(X^{*,2}, \tau^{*,2})$, where $\tau^{*,2}$ is given by $\tau_a^{*,2} = p_a^{*,2} - u_a^{\text{UCB},2}(\mu_{X^{*,2}}(a))$ if a is matched and $\tau_a^{*,2} = 0$ if a is unmatched. Otherwise, return (X^*, τ^*), where τ^* is given by $\tau_a^* = p_a^* - u_a^{\text{UCB}}(\mu_X(a))$ if a is matched and $\tau_a^* = 0$ if a is unmatched.

H.4.2 Proof of Chapter 103

We first verify (as claimed above) that X^* is a maximum weight matching with respect to u'.

Lemma 254. *Matching X^* is a maximum weight matching with respect to u'.*

Proof. Consider any matching $X \neq X^*$. Since

$$\sum_{a \in \mathcal{A}} u_a^{\text{UCB}}(\mu_X(a)) \leq -\Delta^{\text{UCB}} + \sum_{a \in \mathcal{A}} u_a^{\text{UCB}}(\mu_{X^*}(a))$$

by the definition of Δ^{UCB}, we have

$$\sum_{a \in \mathcal{A}} u'_a(\mu_X(a)) \leq \sum_{a \in \mathcal{A}} u_a^{\text{UCB}}(\mu_X(a)) \leq \sum_{a \in \mathcal{A}} \left(u_a^{\text{UCB}}(\mu_{X^*}(a)) - \frac{\Delta^{\text{UCB}}}{|\mathcal{A}|} \right) \leq \sum_{a \in \mathcal{A}} u'_a(\mu_{X^*}(a)). \quad \square$$

We now prove the main lemma for this analysis, restated below. Lemma 104 shows that if the confidence sets are small enough, then the selected matching will be stable with respect to the true utilities.

Lemma 104. *Suppose* ComputeMatch$'$ *is run on a collection $\mathscr{C}$ of confidence sets $C_{i,j}$ and $C_{j,i}$ over the agent utilities that satisfy*

$$\max(C_{i,j}) - \min(C_{i,j}) \leq 0.05 \frac{\Delta}{|\mathcal{A}|} \quad \textit{and} \quad \max(C_{j,i}) - \min(C_{j,i}) \leq 0.05 \frac{\Delta}{|\mathcal{A}|}$$

[1] See Chegireddy and Hamacher (1987) for efficient algorithms for to compute the second-best matching.

for all (i,j) in the matching returned by ComputeMatch′. *Suppose also that the confidence sets $\mathscr{C}$ contain the true utilities for all pairs of agents. Then the market outcome returned by* ComputeMatch′ *is stable with respect to the true utilities u.*

Proof of Chapter 104. The proof proceeds in five steps, which we now outline. We first show the matching returned by ComputeMatch′ is the maximum weight matching X^{opt} with respect to u. We next show that X^* as defined in ComputeMatch′ also equals X^{opt}. These facts let us conclude that ComputeMatch′ returns (X^*, τ^*). We then show Δ^{UCB} is at least 0.1Δ. We then show that (X^*, τ^*) is stable with respect to u'. We finish by showing that this implies (X^*, τ^*) is a stable with respect to u.

Throughout the proof, we will use the following observation about the expanded confidence sets:

$$\max\bigl(C'_{i,j}\bigr) - \min\bigl(C'_{i,j}\bigr) \leq 0.1\frac{\Delta}{|\mathcal{A}|} \quad \text{and} \quad \max\bigl(C'_{j,i}\bigr) - \min\bigl(C'_{j,i}\bigr) \leq 0.1\frac{\Delta}{|\mathcal{A}|} \tag{H.6}$$

for all (i,j) in the matching returned by ComputeMatching′. This follows from the assumptions in the lemma statement.

Proving ComputeMatch′ **returns X^{opt} as the matching.** ComputeMatch′ by definition returns $X^{*,2}$ always, so it suffices to show that $X^{*,2} = X^{\mathrm{opt}}$. Note that $X^{*,2}$ is a maximum weight matching with respect to $u^{\mathrm{UCB},2}$. This means that

$$\begin{aligned}
\sum_{a\in\mathcal{A}} u_a(\mu_{X^{*,2}}(a)) &\geq -\sum_{a\in\mathcal{A}}\Bigl(\max\bigl(C'_{a,\mu_{X^{*,2}}(a)}\bigr) - \min\bigl(C'_{a,\mu_{X^{*,2}}(a)}\bigr)\Bigr) + \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB},2}(\mu_{X^{*,2}}(a)) \\
&\geq -0.1\Delta + \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB},2}(\mu_{X^{*,2}}(a)) \\
&\geq -0.1\Delta + \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB},2}(\mu_{X^{\mathrm{opt}}}(a)) \\
&\geq -0.1\Delta + \sum_{a\in\mathcal{A}} u_a(\mu_{X^{\mathrm{opt}}}(a)).
\end{aligned}$$

By the definition of the gap Δ, we conclude that $X^{*,2} = X^{\mathrm{opt}}$.

Proving $X^* = X^{\mathrm{opt}}$. Suppose for sake of contradiction that $X^* \neq X^{\mathrm{opt}}$. Then

$$\sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^*}(a)) \geq \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^{\mathrm{opt}}}(a)) \geq \sum_{a\in\mathcal{A}} u_a(\mu_{X^{\mathrm{opt}}}(a)),$$

since X^* is a maximum weight matching with respect to u^{UCB}. Moreover, by the definition of the gap, we know that $\sum_{a\in\mathcal{A}} u_a(\mu_{X^*}(a)) \leq \sum_{a\in\mathcal{A}} u_a(\mu_{X^{\mathrm{opt}}}(a)) - \Delta$. Putting this all together, we see that

$$\begin{aligned}
\sum_{a\in\mathcal{A}} \Bigl(\max\bigl(C_{a,\mu_{X^*}(a)}\bigr) - \min\bigl(C_{a,\mu_{X^*}(a)}\bigr)\Bigr) &\geq \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^*}(a)) - \sum_{a\in\mathcal{A}} u_a(\mu_{X^*}(a)) \\
&\geq \Delta.
\end{aligned}$$

We now use this to lower bound the utility of X^* on $u^{\mathrm{UCB},2}$. By the definition of the confidence sets, we see that

$$\begin{aligned}\sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB},2}(\mu_{X^*}(a)) &\geq \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^*}(a)) + \frac{1}{2}\sum_{a\in\mathcal{A}}\left(\max\left(C_{a,\mu_{X^*}(a)}\right) - \min\left(C_{a,\mu_{X^*}(a)}\right)\right) \\ &\geq \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^*}(a)) + 0.5\Delta.\end{aligned}$$

However, X^{opt} only achieves a utility of

$$\begin{aligned}\sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB},2}(\mu_{X^{\mathrm{opt}}}(a)) &\leq \sum_{a\in\mathcal{A}} u_a(\mu_{X^{\mathrm{opt}}}(a)) + \sum_{a\in\mathcal{A}}\left(\max\left(C'_{a,\mu_{X^{\mathrm{opt}}}(a)}\right) - \min\left(C'_{a,\mu_{X^{\mathrm{opt}}}(a)}\right)\right) \\ &\leq \sum_{a\in\mathcal{A}} u_a(\mu_{X^{\mathrm{opt}}}(a)) + 0.1\Delta.\end{aligned}$$

But this contradicts the fact (from above) that $X^{\mathrm{opt}} = X^{*,2}$ is a maximum weight matching with respect to $u^{\mathrm{UCB},2}$. Therefore, it must be that $X^* = X^{\mathrm{opt}}$.

Putting the above two arguments together, we conclude ComputeMatch′ returns (X^*, τ^*) in this case.

Bounding the gap Δ^{UCB}. We next show that $\Delta^{\mathrm{UCB}} \geq 0.1\Delta$. We proceed by assuming

$$\sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_X(a)) \geq -0.1\Delta + \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^*}(a)) \tag{H.7}$$

for some $X \neq X^*$ and deriving a contradiction.

We first show that (H.7) implies a lower bound on

$$S = \sum_{a\in\mathcal{A}}\left(\max\left(C_{a,\mu_X(a)}\right) - \min\left(C_{a,\mu_X(a)}\right)\right)$$

in terms of Δ. Because the confidence sets contain the true utilities and u_a^{UCB} upper bounds u_a pointwise, (H.7) implies

$$S + \sum_{a\in\mathcal{A}} u_a(\mu_X(a)) \geq \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_X(a)) \geq -0.1\Delta + \sum_{a\in\mathcal{A}} u_a(\mu_{X^*}(a)).$$

Applying the definition of Δ, we obtain the lower bound

$$S \geq -0.1\Delta + \sum_{a\in\mathcal{A}} u_a(\mu_{X^*}(a)) - \sum_{a\in\mathcal{A}} u_a(\mu_X(a)) \geq (1-0.1)\Delta.$$

Now, we apply the fact that $X^* = X^{*,2} = X^{\mathrm{opt}}$. We establish the following contradiction:

$$0.1\Delta + \sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^*}(a)) \geq 0.1\Delta + \sum_{a\in\mathcal{A}} u_a(\mu_{X^*}(a))$$

$$
\begin{aligned}
&= \sum_{a \in \mathcal{A}} (u_a(\mu_{X^*}(a)) + 0.1\Delta/|\mathcal{A}|) \\
&\overset{\text{(i)}}{\geq} \sum_{a \in \mathcal{A}} u_a^{\text{UCB},2}(\mu_{X^*}(a)) \\
&\overset{\text{(ii)}}{\geq} \sum_{a \in \mathcal{A}} u_a^{\text{UCB},2}(\mu_{X}(a)) \\
&\overset{\text{(iii)}}{\geq} \frac{S}{2} + \sum_{a \in \mathcal{A}} u_a^{\text{UCB}}(\mu_{X}(a)) \\
&\overset{\text{(iv)}}{\geq} \left(\frac{1}{2}(1-0.1)\right)\Delta + \sum_{a \in \mathcal{A}} u_a^{\text{UCB}}(\mu_{X}(a)) \\
&\overset{\text{(v)}}{\geq} \left(\frac{1}{2}(1-0.1) - 0.1\right)\Delta + \sum_{a \in \mathcal{A}} u_a^{\text{UCB}}(\mu_{X^*}(a)).
\end{aligned}
$$

Here, (i) comes from (H.6) in the lemma statement; (ii) holds because $X^* = X^{*,2}$ is a maximum weight matching with respect to $u^{\text{UCB},2}$; (iii) is by the definition of $u^{\text{UCB},2}$; (iv) follows from our lower bound on S; and (v) follows from (H.7).

Proving that (X^*, τ^*) is stable with respect to u'. By Chapter 254, (X^*, p') is an optimal primal-dual pair with respect to u'. Now, it suffices to show that the primal-dual solution corresponds to the market outcome (X^*, τ^*) for u'. To see this, notice that $p'_a = 0$ for unmatched agents and

$$
p'_a = p^*_a - \frac{\Delta^{\text{UCB}}}{2|\mathcal{A}|} = \tau^*_a + u'_a(\mu_{X^*}(a))
$$

for matched agents.

Proving that (X^*, τ^*) is stable with respect to u. We show the stability (X^*, τ^*) with respect to u by checking that individual rationality holds and that there are no blocking pairs.

The main fact that we will use is that

$$
u_a(\mu_{X^*}(a)) \geq u'_a(\mu_{X^*}(a)).
$$

To prove this, we split into two cases: (i) agent a is matched in X^* (i.e., $\mu_{X^*}(a) \neq a$), and (ii) agent a is not matched by X^*. For (i), if a is matched by X^*, then

$$
u_a(\mu_{X^*}(a)) \geq u_a^{\text{UCB}}(\mu_{X^*}(a)) - 0.1\frac{\Delta}{|\mathcal{A}|} \geq u_a^{\text{UCB}}(\mu_{X^*}(a)) - \frac{\Delta^{\text{UCB}}}{|\mathcal{A}|} = u'_a(\mu_{X^*}(a)).
$$

For (ii), if a is not matched by X^*, then $u_a(\mu_{X^*}(a)) \geq u'_a(\mu_{X^*}(a))$ because both sides are 0.

For individual rationality, we thus have

$$u_a(\mu_{X^*}(a)) + \tau_a^* \geq u'_a(\mu_{X^*}(a)) + \tau_a^* \geq 0,$$

where the second inequality comes from the individual rationality of (X^*, τ^*) with respect to u'.

Let's next show that there are no blocking pairs. If $(i, j) \in X^*$, then we see that:

$$u_i(\mu_{X^*}(i)) + \tau_i^* + u_j(\mu_{X^*}(j)) + \tau_j^* = u_i(\mu_{X^*}(i)) + u_j(\mu_{X^*}(j)),$$

as desired. Next, consider any pair $(i, j) \notin X^*$. Then,

$$u_i(j) + u_j(i) \leq u_i^{\mathrm{UCB}}(j) + u_j^{\mathrm{UCB}}(i) = u'_i(j) + u'_j(i).$$

It follows that

$$\begin{aligned} u_i(\mu_{X^*}(i)) + \tau_i^* + u_j(\mu_{X^*}(j)) + \tau_j^* &\geq u'_i(\mu_{X^*}(i)) + \tau_i^* + u_j(\mu_{X^*}(j)) + \tau_j^* \\ &\geq u'_i(j) + u'_j(i) \\ &\geq u_i(j) + u_j(i), \end{aligned}$$

where the second inequality comes from the fact that (X^*, τ^*) has no blocking pairs with respect to u'.

This completes our proof that (X^*, τ^*) is stable with respect to u. □

Now, we are ready to prove Theorem 103.

Proof of Theorem 103. As in the proof of Theorem 98, the starting point for our proof is the typical approach in multi-armed bandits and combinatorial bandits Gai et al. (2012); Chen et al. (2013); Lattimore and Szepesvári (2020) of bounding regret in terms of the sizes of the confidence interval of the chosen arms. Our approach does not quite compose cleanly with these proofs, since we need to handle the transfers in addition to the matching.

We divide in two cases, based on the event E that all of the confidence sets contain their respective true utilities at every time step $t \leq T$. That is, $u_i(j) \in C_{i,j}$ and $u_j(i) \in C_{j,i}$ for all $(i, j) \in \mathcal{I} \times \mathcal{J}$ at all t.

Case 1: E holds. Let n_{ij}^t be the number of times that the pair (i, j) has been matched by round t. For each pair (i, j), we maintain a "blame" counter b_{ij}^t. We will ultimately bound the total number of time steps where the algorithm chooses a matching that is not stable by $\sum_{(i,j)} b_{i,j}^T$.

We increment the blame counters as follows. First, suppose that

$$\max\left(C_{a,\mu_{X^t}(a)}\right) - \min\left(C_{a,\mu_{X^t}(a)}\right) \leq 0.1 \frac{\Delta}{|\mathcal{A}|}$$

for every matched agent $a \in \mathcal{A}$. By Chapter 104 and since the event E holds, we know the chosen matching is stable and thus incurs 0 regret. We do not increment any of the blame counters in this case. Now, suppose that

$$\max\left(C_{a,\mu_{X^t}(a)}\right) - \min\left(C_{a,\mu_{X^t}(a)}\right) > 0.1\frac{\Delta}{|\mathcal{A}|}$$

for some matched agent a. We increment the counter of the least-blamed pair $(i,j) \in X^t$.

We now bound the blame counter b_{ij}^T. We use the fact that the blame counter is only incremented when the corresponding confidence set is sufficiently large, and that a new sample of the utilities is received whenever the blame counter is incremented. This means that:

$$b_{ij}^T = O\left(\frac{|\mathcal{A}|^2 \log(|\mathcal{A}|T))}{\Delta^2}\right).$$

The maximum regret incurred by any matching is at most $12|\mathcal{A}|$ which means that the regret incurred by this case is at most:

$$12|\mathcal{A}| \sum_{(i,j)} b_{ij}^T \leq 12|\mathcal{A}| \sum_{(i,j)} O\left(\frac{|\mathcal{A}|^2 \log(|\mathcal{A}|T))}{\Delta^2}\right) = O\left(\frac{|\mathcal{A}|^5 \log(|\mathcal{A}|T))}{\Delta^2}\right).$$

Case 2: E does not hold. Since each $n_{ij}(\hat{u}_i(j) - u_i(j))$ is mean-zero and 1-subgaussian and we have $O(|\mathcal{I}||\mathcal{J}|T)$ such random variables over T rounds, the probability that any of them exceeds

$$2\sqrt{\log(|\mathcal{I}||\mathcal{J}|T/\delta)} \leq 2\sqrt{\log(|\mathcal{A}|^2 T/\delta)}$$

is at most δ by a standard tail bound for the maximum of subgaussian random variables. It follows that E fails to hold with probability at most $|\mathcal{A}|^{-2}T^{-2}$. In the case that E fails to hold, our regret in any given round would be at most $12|\mathcal{A}|$ by the Lipschitz property in Chapter 97. (Recall that our upper confidence bound is off by at most 6 due to clipping the confidence interval to lie in $[-1, 1]$, so that the expanded confidence sets also necessarily lie in $[-3, 3]$.) Thus, the expected regret from this scenario is at most

$$|\mathcal{A}|^{-2}T^{-2} \cdot 12|\mathcal{A}|T \leq 12|\mathcal{A}|^{-1}T^{-1},$$

which is negligible compared to the regret bound from when E does occur. □

H.4.3 Instance-independent regret bounds for MatchUCB′

To establish *instance-independent* regret bounds for MatchUCB′, we show that (X^*, p^*) is indeed optimal with respect to u^{UCB}; the remainder then follows the same argument as Chapter 98.

Lemma 255. *The pair (X^*, p^*) is an optimal primal-dual pair with respect to u^{UCB}.*

Proof. It suffices to verify feasibility and, by weak duality, check that X^* and p^* achieve the same objective value. It is clear that X^* is primal feasible. For dual feasibility, if $(i,j) \notin X^*$, then

$$p_i^* + p_j^* \geq p_i' + p_j' \geq u_i'(j) + u_j'(i) = u_i^{\mathrm{UCB}}(j) + u_j^{\mathrm{UCB}}(i);$$

and if $(i,j) \in X^*$, then

$$p_i^* + p_j^* = p_i' + p_j' + 2\frac{\Delta^{\mathrm{UCB}}}{|\mathcal{A}|} \geq u_i'(j) + u_j'(i) + 2\frac{\Delta^{\mathrm{UCB}}}{|\mathcal{A}|} = u_i^{\mathrm{UCB}}(j) + u_j^{\mathrm{UCB}}(i).$$

Finally, we check that they achieve the same objective value with respect to u^{UCB}. By Chapter 254 and strong duality, X^* achieves the same objective value as p' with respect to u'. Hence

$$\sum_{a\in\mathcal{A}} u_a^{\mathrm{UCB}}(\mu_{X^*}(a)) = 2|X^*|\frac{\Delta^{\mathrm{UCB}}}{|\mathcal{A}|} + \sum_{a\in\mathcal{A}} u_a'(\mu_{X^*}(a)) = 2|X^*|\frac{\Delta^{\mathrm{UCB}}}{|\mathcal{A}|} + \sum_{a\in\mathcal{A}} p_a' = \sum_{a\in\mathcal{A}} p_a^*. \quad \square$$

H.5 Proofs for Chapter 12.6.2

Theorem 105. *For preference class $\mathcal{U}_{\mathrm{unstructured}}$ (see Chapter 12.3), there exists an algorithm giving the platform*

$$\varepsilon T \sum_{t=1}^{T} |\mathcal{A}_t| - O\Big(|\mathcal{A}|\sqrt{nT}\sqrt{\log(|\mathcal{A}||T|)}\Big)$$

revenue in the presence of search frictions while maintaining stability with high probability.

Proof of Chapter 105. The algorithm is defined as follows. We set confidence sets according to MatchUCB and run essentially that algorithm, but with a modified ComputeMatch. Instead of ComputeMatch, we use the following algorithm. The platform first computes a matching with transfers (X^*, τ^*) according to the UCB estimates u^{UCB}, like before. Then, the platform chooses X^* to be the selected matching, and sets the transfers according to:

$$\tau_a = \tau_a^* - \varepsilon + \max\big(C_{a,\mu_X(a))}\big) - \min\big(C_{a,\mu_X(a)}\big).$$

This choice of transfers has a clean economic intuition: agents should be compensated based on the platform's uncertainty about their utilities with ε of their transfer shaved off as revenue for the platform.

First, we show that if the confidence sets contain the true utilities, then (X^*, τ) is ε-stable. It suffices to show that (X^*, τ') where:

$$\tau_a' = \tau_a^* + \max\big(C_{a,\mu_X(a))}\big) - \min\big(C_{a,\mu_X(a)}\big)$$

is stable. First, we see that

$$u_a(\mu_{X^{\mathrm{UCB}}}(a)) + \tau_a' = u_a^{\mathrm{UCB}}(\mu_{X^{\mathrm{UCB}}}(a)) + \tau_a^* \geq 0,$$

since (X, τ^*) is stable with respect to u^{UCB}. Furthermore, we see that:

$$\begin{aligned}\big(u_i(\mu_X(i)) + \tau_i'\big) + \big(u_j(\mu_X(j)) + \tau_j'\big) &\geq \big(u_i^{\text{UCB}}(\mu_X(i)) + \tau_i^*\big) + \big(u_j^{\text{UCB}}(\mu_X(j)) + \tau_j^*\big) \\ &\geq u_i^{\text{UCB}}(j) + u_j^{\text{UCB}}(i) \\ &\geq u_i(j) + u_j(i),\end{aligned}$$

where the second to last line follows from the fact that (X, τ^*) is stable with respect to u^{UCB}.

We first show that s is a feasible solution to (†):

$$\begin{aligned}&\min\big(u_i(j) - u_i(\mu_{X^{\text{UCB}}}(i)) - s_i, uii_j(i) - u_j(\mu_{X^{\text{UCB}}}(j)) - s_j\big) \\ &= \min\big(u_i(j) - u_i^{\text{UCB}}(\mu_{X^{\text{UCB}}}(i)), uii_j(i) - u_j^{\text{UCB}}(\mu_{X^{\text{UCB}}}(j))\big) \\ &\leq \min\big(u_i^{\text{UCB}}(j) - u_i^{\text{UCB}}(\mu_{X^{\text{UCB}}}(i)), u_j^{\text{UCB}}(i) - u_j^{\text{UCB}}(\mu_{X^{\text{UCB}}}(j))\big) \\ &\leq 0,\end{aligned}$$

where the last step uses the fact that $\mu_{X^{\text{UCB}}}$ is stable with respect to u^{UCB} by definition. Moreover, we see that

$$u_a(\mu_{X^{\text{UCB}}}(a)) + s_a = u_a^{\text{UCB}}(\mu_{X^{\text{UCB}}}(a)) \geq 0,$$

where the last inequality uses that $\mu_{X^{\text{UCB}}}$ is stable with respect to u^{UCB} by definition. This implies that s is feasible.

We see that the platform's revenue is equal to:

$$\begin{aligned}-\sum_{t=1}^{T}\sum_{a\in\mathcal{A}_t}\tau_a &= -\sum_{t=1}^{T}\sum_{a\in\mathcal{A}_t}\tau_a^* + \sum_{t=1}^{T}\sum_{a\in\mathcal{A}_t}\varepsilon + \sum_{t=1}^{T}\sum_{a\in\mathcal{A}_t}\big(\max\big(C_{a,\mu_X(a))}\big) - \min\big(C_{a,\mu_X(a)}\big)\big) \\ &= \varepsilon\sum_{t=1}^{T}|\mathcal{A}_t| - \sum_{t=1}^{T}\sum_{a\in\mathcal{A}_t}\big(\max\big(C_{a,\mu_X(a))}\big) - \min\big(C_{a,\mu_X(a)}\big)\big).\end{aligned}$$

Using the proof of Chapter 98, we see that

$$\sum_{t=1}^{T}\sum_{a\in\mathcal{A}_t}\big(\max\big(C_{a,\mu_X(a))}\big) - \min\big(C_{a,\mu_X(a)}\big)\big) \leq O(|\mathcal{A}|\sqrt{nT}\log(|\mathcal{A}|T)),$$

as desired.

□

H.6 Proofs for Chapter 12.6.3

H.6.1 Proof of Proposition 106

Proof of Proposition 106. We first prove the first part of the statement, and then the second part of the statement.

Proof of part (a). We note that it follows immediately from Definition 11 that NTU Subset Instability is nonnegative. Let's now show that $I(X; u, \mathcal{A})$ is zero if and only if (X, τ) is stable. It is not difficult to see that the infimum of (†) is attained at some s^*.

If $I(X; u, \mathcal{A}) = 0$, then we know that $s^*_a = 0$ for all $a \in \mathcal{A}$. The constraints in the optimization problem imply that X has no blocking pairs and individually rationality is satisfied, as desired.

If X is stable, then we see that $s = \vec{0}$ is a feasible solution to (†), which means that the optimum of (†) is at most zero. This coupled with the fact that $I(X; u, \mathcal{A})$ is always nonnegative means that $I(X; u, \mathcal{A}) = 0$ as desired.

Proof of part (b). Consider two utility functions u and uii. To show Lipchitz continuity, it suffices to show that for any matching X:

$$|I(X; u, \mathcal{A}) - I(X; uii, \mathcal{A})| \leq 2 \sum_{a \in \mathcal{A}} \|u_a - uii_a\|_\infty.$$

We show that:

$$I(X; uii, \mathcal{A}) \leq I(X; u, \mathcal{A}) + 2 \sum_{a \in \mathcal{A}} \|u_a - uii_a\|_\infty,$$

noting that the other direction follows from an analogous argument. Let s^* be an optimal solution to (†) for the utilities u. Consider the solution $s_a = s^*_a + 2\|u_a - u_a\|_\infty$. We first verify that s is a feasible solution to (†) for uii. We see that:

$$\begin{aligned}
&\min\bigl(uii_i(j) - uii_i(\mu_X(i)) - s_i, uii_j(i) - uii_j(\mu_X(j)) - s_j\bigr) \\
&= \min\bigl(uii_i(j) - uii_i(\mu_X(i)) - s^*_i - 2\|u_i - uii_i\|_\infty, uii_j(i) - uii_j(\mu_X(j)) - s^*_j - 2\|u_j - uii_j\|_\infty\bigr) \\
&\leq \min\bigl(u_i(j) - u_i(\mu_X(i)) - s^*_i, u_j(i) - u_j(\mu_X(j)) - s^*_j\bigr) \\
&\leq 0,
\end{aligned}$$

as desired. Moreover, we see that

$$uii_a(\mu_X(a)) + s_a = uii_a(\mu_X(a)) + s^*_a + 2\|u_a - u_a\|_\infty \leq u_a(\mu_X(a)) + s^*_a \geq 0.$$

Thus we have demonstrated that s is feasible. This means that:

$$I(X; uii, \mathcal{A}) \leq \sum_{a \in \mathcal{A}} s_a = \sum_{a \in \mathcal{A}} s^*_a + 2 \sum_{a \in \mathcal{A}} \|u_a - uii_a\|_\infty = [I(X; u, \mathcal{A}) + 2 \sum_{a \in \mathcal{A}} \|u_a - uii_a\|_\infty,$$

as desired.

□

H.6.2 Proof of Theorem 107

We show that the algorithmic approach from Chapter 12.5 can be adapted to the setting of matching with non-transferable utilities.

Drawing intuition from Chapter 12.5, at each round, we compute a stable matching for utilities given by the upper confidence bounds. More precisely, suppose we have a collection $\mathscr{C}$ of confidence sets $C_{i,j}, C_{j,i} \subseteq \mathrm{R}$ such that $u_i(j) \in C_{i,j}$ and $u_j(i) \in C_{j,i}$ for all $(i, j) \in \mathcal{I} \times \mathcal{J}$. Our algorithm uses $\mathscr{C}$ to get an upper confidence bound for each agent's utility function and then computes a stable matching with transfers as if these upper confidence bounds were the true utilities (see ComputeMatchNTU). This can be implemented efficiently if we use, e.g., the Gale-Shapley algorithm (either the customer-proposing algorithm or the provider-proposing algorithm will work).

Algorithm 16: ComputeMatchNTU: Compute matching with transfers from confidence sets

---

Input: Confidence sets $\mathscr{C}$

for $(i, j) \in \mathcal{I} \times \mathcal{J}$ **do**

  $u_i^{\mathrm{UCB}}(j) \leftarrow \max\big(C_{i,j}\big)$;

  $u_j^{\mathrm{UCB}}(i) \leftarrow \max\big(C_{j,i}\big)$; `// UCB estimates of utilities.`

Run any version of the Gale-Shapley algorithm Gale and Shapley (1962) on u^{UCB} to obtain a matching X^*;

return X^*;

The core property of ComputeMatchNTU is that we can upper bound NTU Subset Instability by the sum of the sizes of the relevant confidence sets, assuming that the confidence sets contain the true utilities.

Proposition 256. *Consider a collection of confidence sets $\mathscr{C}$ such that $u_i(j) \in C_{i,j}$ and $u_j(i) \in C_{j,i}$ for all $(i, j) \in \mathcal{I} \times \mathcal{J}$. The instability of the output X^{UCB} of ComputeMatch satisfies*

$$I(X^{\mathrm{UCB}}; u, \mathcal{A}) \leq \sum_{a \in \mathcal{A}^t} \Big(\max\big(C_{a,\mu_{X^{\mathrm{UCB}}}(a)}\big) - \min\big(C_{a,\mu_{X^{\mathrm{UCB}}}(a)}\big)\Big). \tag{H.8}$$

Proof. We construct subsidies for this setting to be:

$$s_a = \max\big(C_{a,\mu_X(a)}\big) - u_a(\mu_X(a)) \leq \max\big(C_{a,\mu_X(a)}\big) - \min\big(C_{a,\mu_X(a)}\big).$$

Step 1: Verifying feasibility. We first show that s is a feasible solution to (†).

$$\begin{aligned}
&\min\big(u_i(j) - u_i(\mu_{X^{\mathrm{UCB}}}(i)) - s_i, uii_j(i) - u_j(\mu_{X^{\mathrm{UCB}}}(j)) - s_j\big) \\
&= \min\big(u_i(j) - u_i^{\mathrm{UCB}}(\mu_{X^{\mathrm{UCB}}}(i)), uii_j(i) - u_j^{\mathrm{UCB}}(\mu_{X^{\mathrm{UCB}}}(j))\big) \\
&\leq \min\big(u_i^{\mathrm{UCB}}(j) - u_i^{\mathrm{UCB}}(\mu_{X^{\mathrm{UCB}}}(i)), u_j^{\mathrm{UCB}}(i) - u_j^{\mathrm{UCB}}(\mu_{X^{\mathrm{UCB}}}(j))\big) \\
&\leq 0,
\end{aligned}$$

where the last step uses the fact that $\mu_{X^{\mathrm{UCB}}}$ is stable with respect to u^{UCB} by definition. Moreover, we see that

$$u_a(\mu_{X^{\mathrm{UCB}}}(a)) + s_a = u_a^{\mathrm{UCB}}(\mu_{X^{\mathrm{UCB}}}(a)) \geq 0,$$

where the last inequality uses that $\mu_{X^{\mathrm{UCB}}}$ is stable with respect to u^{UCB} by definition. This implies that s is feasible.

Step 2: Computing the objective. We next compute the objective of (†) at s and use this to bound $I(X^*; u, \mathcal{A})$. A simple calculation shows that:

$$I(X^*; u, \mathcal{A}) \leq \sum_a s_a = \sum_{a \in \mathcal{A}} \Big(\max\big(C_{a, \mu_{X^{\mathrm{UCB}}}(a)}\big) - \min\big(C_{a, \mu_{X^{\mathrm{UCB}}}(a)}\big) \Big),$$

as desired.

□

Explicit algorithm and regret bounds

Using the same intuition as Chapter 12.5, the regret bound of Chapter 256 hints at an algorithm: each round, select the matching with transfers returned by ComputeMatchNTU and update confidence sets accordingly. To instantiate this approach, it remains to construct confidence intervals that contain the true utilities with high probability.

We showcase this algorithm in the simple setting of unstructured preferences. For this setting, we can construct our confidence intervals following the classical UCB approach. That is, for each utility value involving the pair (i, j), we take a length $O(\sqrt{\log(|\mathcal{A}|T)/n_{ij}})$ confidence interval centered around the empirical mean, where n_{ij} is the number of times the pair has been matched before. We describe this construction precisely in Chapter 3 (MatchNTUUCB).

Algorithm 17: MatchNTUUCB: A bandit algorithm for matching with non-transferable utilities.

Input: Time horizon T

for $(i, j) \in \mathcal{I} \times \mathcal{J}$ **do**
  ; `// Initialize confidence intervals and empirical mean.`
  $C_{i,j} \leftarrow [-1, 1]$;
  $C_{j,i} \leftarrow [-1, 1]$;
  $\hat{u}_i(j) \leftarrow 0$;
  $\hat{u}_j(i) \leftarrow 0$;
for $t \leftarrow 1$ **to** T **do**
  $X^t \leftarrow$ **ComputeMatch**(C);
  for $(i, j) \in X^t$ **do**
    ; `// Set confidence intervals and update means.`
    Update $\hat{u}_i(j)$ and $\hat{u}_j(i)$ from feedback; increment counter n_{ij};
    $C_{i,j} \leftarrow \big[\hat{u}_i(j) - 8\sqrt{\log(|\mathcal{A}|T)/n_{ij}}, \hat{u}_i(j) + 8\sqrt{\log(|\mathcal{A}|T)/n_{i,j}}\,\big] \cap [-1, 1]$;
    $C_{j,i} \leftarrow \big[\hat{u}_j(i) - 8\sqrt{\log(|\mathcal{A}|T)/n_{ij}}, \hat{u}_j(i) + 8\sqrt{\log(|\mathcal{A}|T)/n_{i,j}}\,\big] \cap [-1, 1]$;

To analyze MatchNTUUCB, recall that Chapter 101 bounds the regret at each step by the lengths of the confidence intervals of each pair in the selected matching. Like in Chapter 12.5, this yields the following instance-independent regret bound:

Theorem 257. *MatchNTUUCB incurs expected regret* $\mathbb{E}(R_T) \leq O\big(|\mathcal{A}|^{3/2}\sqrt{T}\sqrt{\log(|\mathcal{A}|T)}\big)$.

Proof. This proof proceeds very similarly to the proof of Theorem 98. We consider the event E that all of the confidence sets contain their respective true utilities at every time step $t \leq T$. That is, $u_i(j) \in C_{i,j}$ and $u_j(i) \in C_{j,i}$ for all $(i,j) \in \mathcal{I} \times \mathcal{J}$ at all t.

Case 1: E **holds.** By Chapter 101, we may bound

$$I(X^t; u, \mathcal{A}^t) \leq \sum_{a \in \mathcal{A}^t} \Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big)\Big) = O\left(\sum_{(i,j)\in X^t} \sqrt{\frac{\log(|\mathcal{A}|T)}{n_{ij}^t}}\right),$$

where n_{ij}^t is the number of times that the pair (i,j) has been matched at the start of round t. Let $w_{i,j}^t = \frac{1}{\sqrt{n_{ij}^t}}$ be the size of the confidence set (with the log factor scaled out) for (i,j) at the start of round t.

At each time step t, let's consider the list consisting of w_{i_t,j_t}^t for all $(i_t, j_t) \in X^t$. Let's now consider the overall list consisting of the concatenation of all of these lists over all rounds. Let's order this list in decreasing order to obtain a list $\tilde{w}_1, \dots, \tilde{w}_L$ where $L = \sum_{t=1}^T |X^t| \leq nT$. In this notation, we observe that:

$$\sum_{t=1}^T I(X^t; u, \mathcal{A}^t) \leq \sum_{t=1}^T \sum_{a \in \mathcal{A}^t} \Big(\max\big(C_{a,\mu_{X^t}(a)}\big) - \min\big(C_{a,\mu_{X^t}(a)}\big)\Big) = \log(|\mathcal{A}|T) \sum_{l=1}^L \tilde{w}_l.$$

We claim that $\tilde{w}_l \leq O\left(\min(1, \frac{1}{\sqrt{(l/|\mathcal{A}|^2)-1}})\right)$. The number of rounds that a pair of agents can have their confidence set have size at least $\tilde{w}_l$ is upper bounded by $1 + \frac{1}{\tilde{w}_l^2}$. Thus, the total number of times that any confidence set can have size at least $\tilde{w}_l$ is upper bounded by $(|\mathcal{A}|^2)(1 + \frac{1}{\tilde{w}_l^2})$.

Putting this together, we see that:

$$\begin{aligned}
\log(|\mathcal{A}|T) \sum_{l=1}^L \tilde{w}_l &\leq O\left(\sum_{l=1}^L \min(1, \frac{1}{\sqrt{(l/|\mathcal{A}|^2) - 1}})\right) \\
&\leq O\left(\log(|\mathcal{A}|T) \sum_{l=1}^{nT} \min(1, \frac{1}{\sqrt{(l/|\mathcal{A}|^2) - 1}})\right) \\
&\leq O\Big(|\mathcal{A}|\sqrt{nT}\log(|\mathcal{A}|T)\Big).
\end{aligned}$$

Case 2: E **does not hold.** Since each $n_{ij}(\hat{u}_i(j) - u_i(j))$ is mean-zero and 1-subgaussian, and we have $O(|\mathcal{I}||\mathcal{J}|T)$ such random variables over T rounds, the probability that any of them exceeds

$$2\sqrt{\log(|\mathcal{I}||\mathcal{J}|T/\delta)} \leq 2\sqrt{\log(|\mathcal{A}|^2 T/\delta)}$$

is at most δ by a standard tail bound for the maximum of subgaussian random variables. It follows that E fails to hold with probability at most $|\mathcal{A}|^{-2}T^{-2}$. In the case that E fails to hold, our regret in any given round would be at most $4|\mathcal{A}|$ by the Lipschitz property in Chapter 106. (Recall that our upper confidence bound for any utility is wrong by at most two due to clipping each confidence interval to lie in $[-1, 1]$.) Thus, the expected regret from this scenario is at most

$$|\mathcal{A}|^{-2}T^{-2} \cdot 4|\mathcal{A}|T \leq 4|\mathcal{A}|^{-1}T^{-1},$$

which is negligible compared to the regret bound from when E does occur. $\square$

Appendix I

Appendix for Chapter 13

I.1 Proofs from Chapter 13.2 and Chapter 13.3

I.1.1 Proof of Lemma 109

Notice that $\mathrm{PR}(\theta) - \mathrm{PR}(\theta') = (\mathrm{R}(\theta, \theta) - \mathrm{R}(\theta, \theta')) + (\mathrm{R}(\theta, \theta') - \mathrm{R}(\theta', \theta'))$. We bound the first difference using Lipschitzness of ℓ in θ as $|\mathrm{R}(\theta, \theta) - \mathrm{R}(\theta, \theta')| = |\mathrm{E}_{z\sim\mathcal{D}(\theta)}[\ell(z; \theta) - \ell(z; \theta')]| \leq L_\theta\|\theta - \theta'\|$. For the second term we combine Assumption 6 and Lipschitzness of ℓ in z via the Kantorovich-Rubinstein duality theorem. In particular, we get $|\mathrm{R}(\theta, \theta') - \mathrm{R}(\theta', \theta')| = |\mathrm{E}_{z\sim\mathcal{D}(\theta)}\ell(z; \theta') - \mathrm{E}_{z\sim\mathcal{D}(\theta')}\ell(z; \theta')| \leq \varepsilon L_z\|\theta - \theta'\|$. Putting both bounds together, we obtain the claimed Lipschitz bound.

I.1.2 Proof of Proposition 111

We construct a γ-cover of the parameter space, denoted $\mathcal{S}_\gamma$, and deploy all models in this cover. This gives us access to the distributions $\{\mathcal{D}(\theta) : \theta \in \mathcal{S}_\gamma\}$. Using this information, for any $\theta \in \Theta$ we can compute

$$\widehat{\mathrm{PR}}(\theta) = \mathrm{R}(\Pi_{\mathcal{S}_\gamma}(\theta), \theta) = \mathop{\mathbb{E}}_{z\sim\mathcal{D}(\Pi_{\mathcal{S}_\gamma}(\theta))} \ell(z; \theta),$$

where $\Pi_{\mathcal{S}_\gamma}(\theta) := \arg\min_{\theta'\in\mathcal{S}_\gamma} \|\theta' - \theta\|$ is the projection onto $\mathcal{S}_\gamma$. Note that $\|\theta - \Pi_{\mathcal{S}_\gamma}(\theta)\| \leq \gamma$ all $\theta \in \Theta$ since $\mathcal{S}_\gamma$ is a cover. Therefore, for any $\theta \in \Theta$, we can bound $\mathrm{PR}(\theta)$ as

$$\begin{aligned}
\mathrm{PR}(\theta) &\leq \mathrm{R}(\Pi_{\mathcal{S}_\gamma}(\theta), \theta) + L_z\varepsilon\|\Pi_{\mathcal{S}_\gamma}(\theta) - \theta\| \\
&\leq \mathrm{R}(\Pi_{\mathcal{S}_\gamma}(\theta), \theta) + L_z\varepsilon\gamma \\
&= \widehat{\mathrm{PR}}(\theta) + L_z\varepsilon\gamma.
\end{aligned}$$

Similarly we obtain $\mathrm{PR}(\theta) \geq \widehat{\mathrm{PR}}(\theta) - L_z\varepsilon\gamma$, which completes the proof.

I.1.3 Proof of Proposition 112

We will show that $\mathrm{PR}_{\mathrm{LB}}(\theta) \le \mathrm{PR}(\theta)$ and $\mathrm{PR}_{\min} \ge \mathrm{PR}(\theta_{\mathrm{PO}})$; these two facts immediately imply $\Delta(\theta) := \mathrm{PR}(\theta) - \mathrm{PR}(\theta_{\mathrm{PO}}) \ge \mathrm{PR}_{\mathrm{LB}}(\theta) - \mathrm{PR}_{\min}$.

The first bound follows because $\mathrm{PR}(\theta) = \mathrm{R}(\theta, \theta) \ge \mathrm{R}(\theta', \theta) - L_z\varepsilon\|\theta' - \theta\|$ for all θ', where we use $(L_z\varepsilon)$-Lipschitzness of R in the first argument. Similarly, the second bound follows because

$$\mathrm{PR}(\theta_{\mathrm{PO}}) = \min_{\theta} \mathrm{R}(\theta, \theta) \le \min_{\theta}(\mathrm{R}(\theta', \theta) + L_z\varepsilon\|\theta - \theta'\|),$$

for all θ'.

I.2 Regret analysis of Algorithm 6

In this section, we prove a regret bound for Algorithm 6. At a high level, Theorem 113 combines bounds specific to performative prediction with ingredients from the analysis of successive elimination (Even-Dar et al., 2002). First, using a finite-sample analogue of Proposition 112, we show that after phase p all models $\theta \in \mathcal{A}$ have suboptimality $\Delta(\theta) \le 8\gamma_p$. We then upper bound the number of models in each suboptimality band $\{\theta : 16L_z\varepsilon r \le \Delta(\theta) < 32L_z\varepsilon r\}$, for fixed r, that are deployed in each phase, by leveraging the definition of sequential zooming dimension. The remainder of the proof separately analyzes the regret incurred from the first $\log_2(1/(L_z\varepsilon))$ phases, in which the finite-sample error dominates the discretization error, and the regret from the later phases, in which the finite-sample error and the discretization error are of the same order.

We use $\mathrm{Reg}_{\mathrm{ph}}(p_1 : p_2)$ to denote the regret incurred from phase p_1 to phase p_2:

$$\mathrm{Reg}_{\mathrm{ph}}(p_1 : p_2) = \mathbb{E}\sum_{p=p_1}^{p_2} \Delta(\theta_p).$$

We let $\mathrm{Reg}_{\mathrm{ph}}(0 : p) \equiv \mathrm{Reg}_{\mathrm{ph}}(p)$. For phases p that happen after the time horizon T, we assume that the incurred regret is 0; for example, if phases $p_1 \le p_2$ happen after T, then $\mathrm{Reg}_{\mathrm{ph}}(p_1 : p_2) = 0$.

I.2.1 Clean event

First, we define a clean event that guarantees that the estimates $\widehat{\mathrm{DPR}}(\theta, \theta')$ are close to the true values $\mathrm{DPR}(\theta, \theta')$ at all phases. The clean event essentially guarantees uniform convergence over $\widehat{\mathrm{DPR}}(\theta, \cdot)$ for every $\theta \in \mathcal{P}_p$.

Definition 21 (Clean event)**.** *Denote the "clean event" by*

$$E_{\mathrm{clean}} = \left\{\forall p : \sup_{\theta\in\mathcal{P}_p} \sup_{\theta'\in\Theta} \left|\widehat{\mathrm{DPR}}(\theta, \theta') - \mathrm{R}(\theta, \theta')\right| \le \frac{2\mathfrak{C}^*(\ell) + 3\sqrt{\log(T)}}{\sqrt{n_p m_0}}\right\}, \tag{I.1}$$

where $\mathcal{P}_p$ is the set of all models deployed in phase p during time horizon T.

We show that the clean event occurs with high probability.

Lemma 258. *The clean event holds with high probability,*

$$\mathbb{P}\left\{E_{\text{clean}}\right\} \geq 1 - T^{-2}.$$

Proof. We consider each interval of length n_p in phase p, during which the same model is deployed, separately, and then take a union bound over these intervals across all phases. Therefore, we will say interval s in phase p to refer to steps $(s-1)n_p + 1, \ldots, sn_p$ in phase p. For the sake of this proof, we consider a "counterfactual" set of samples for each model θ that augments the set of actually observed samples. In particular, for interval s in phase p, we let $\{z_j^{\theta,s}\}_{j=1}^{n_p m_0}$ denote i.i.d. samples from $\mathcal{D}(\theta)$. The samples for different time intervals and different phases are independent. When model θ is deployed, we observe the samples corresponding to the interval in which θ is deployed.

For each phase p and each time interval s within phase p, let $E_{\text{end}}^{s,p}$ denote the event that phase p terminates strictly before interval s is reached. Let $E_{\text{clean}}^{s,p}$ denote the event that one of the following two holds:

(E1) $E_{\text{end}}^{s,p}$ occurs;

(E2) $E_{\text{end}}^{s,p}$ does not occur, and for the model θ_s deployed in time interval s it holds that:

$$\sup_{\theta' \in \Theta} \left| \widehat{\mathrm{DPR}}(\theta_s, \theta') - \mathrm{R}(\theta_s, \theta') \right| \leq \frac{2\mathfrak{C} + 3\sqrt{\log(T)}}{\sqrt{n_p m_0}},$$

where θ_s is a random variable.

The probability that $E_{\text{clean}}^{s,p}$ does not occur is at most:

$$\begin{aligned}
&\mathbb{P}\left[\neg E_{\text{end}}^{s,p} \;\;\&\;\; \sup_{\theta' \in \Theta} \left| \widehat{\mathrm{DPR}}(\theta_s, \theta') - \mathrm{R}(\theta_s, \theta') \right| > \frac{2\mathfrak{C} + 3\sqrt{\log(T)}}{\sqrt{n_p m_0}}\right] \\
&= \mathbb{P}\left[\neg E_{\text{end}}^{s,p}\right] \cdot \mathbb{P}\left[\sup_{\theta' \in \Theta} \left| \widehat{\mathrm{DPR}}(\theta_s, \theta') - \mathrm{R}(\theta_s, \theta') \right| > \frac{2\mathfrak{C} + 3\sqrt{\log(T)}}{\sqrt{n_p m_0}} \middle| \neg E_{\text{end}}^{s,p}\right] \\
&\leq \mathbb{P}\left[\sup_{\theta' \in \Theta} \left| \widehat{\mathrm{DPR}}(\theta_s, \theta') - \mathrm{R}(\theta_s, \theta') \right| > \frac{2\mathfrak{C} + 3\sqrt{\log(T)}}{\sqrt{n_p m_0}} \middle| \neg E_{\text{end}}^{s,p}\right].
\end{aligned}$$

We can equivalently write this as

$$\mathbb{P}\left[\sup_{\theta' \in \Theta} \left| \frac{1}{n_p m_0} \sum_{j=1}^{n_p m_0} \ell(z_j^{\theta_s,s}; \theta') - \mathrm{R}(\theta_s, \theta') \right| > \frac{2\mathfrak{C} + 3\sqrt{\log(T)}}{\sqrt{n_p m_0}} \middle| \neg E_{\text{end}}^{s,p}\right]$$

$$= \mathbb{E}_{\theta \sim \theta_s}\left[\mathbb{P}\left[\sup_{\theta' \in \Theta} \left| \frac{1}{n_p m_0} \sum_{j=1}^{n_p m_0} \ell(z_j^{\theta,s}; \theta') - \mathrm{R}(\theta, \theta') \right| > \frac{2\mathfrak{C} + 3\sqrt{\log(T)}}{\sqrt{n_p m_0}} \middle| \neg E_{\text{end}}^{s,p}, \theta_s = \theta\right]\right].$$

To upper bound this expression, it suffices to show an upper bound on

$$\mathbb{P}\left[\sup_{\theta'\in\Theta}\left|\frac{1}{n_p m_0}\sum_{j=1}^{n_p m_0}\ell(z_j^{\theta,s};\theta') - \mathrm{R}(\theta,\theta')\right| > \frac{2\mathfrak{C}+3\sqrt{\log(T)}}{\sqrt{n_p m_0}}\middle|\neg E^{s,p}_{\text{end}}, \theta_s = \theta\right]$$

that holds for every θ. The first observation is that for any θ, the samples $\{z_j^{\theta,s}\}_{j=1}^{n_p m_0}$ are independent of the event $\{\theta_s = \theta, \neg E^{s,p}_{\text{end}}\}$, since the event depends only on the samples collected in previous time intervals and phases. This means that the above probability is equal to:

$$\mathbb{P}\left[\sup_{\theta'\in\Theta}\left|\frac{1}{n_p m_0}\sum_{j=1}^{n_p m_0}\ell(z_j^{\theta,s};\theta') - \mathrm{R}(\theta,\theta')\right| > \frac{2\mathfrak{C}+3\sqrt{\log(T)}}{\sqrt{n_p m_0}}\right].$$

Let ε_j denote i.i.d. Rademacher random variables. Then, we can observe that with probability $1 - T^{-3}$, it holds that:

$$\begin{aligned}
&\sup_{\theta'\in\Theta}\left|\frac{1}{n_p m_0}\sum_{j=1}^{n_p m_0}\ell(z_j^{\theta,s};\theta') - \mathrm{R}(\theta,\theta')\right| \\
&\leq \mathbb{E}\left[\sup_{\theta'\in\Theta}\left|\frac{1}{n_p m_0}\sum_{j=1}^{n_p m_0}\ell(z_j^{\theta,s};\theta') - \mathrm{R}(\theta,\theta')\right|\right] + \sqrt{\frac{6\log(T)}{n_p m_0}} \\
&\leq 2\cdot\mathbb{E}\left[\sup_{\theta'\in\Theta}\left|\frac{1}{n_p m_0}\sum_{j=1}^{n_p m_0}\ell(z_j^{\theta,s};\theta')\cdot\varepsilon_j\right|\right] + \sqrt{\frac{6\log(T)}{n_p m_0}} \\
&\leq \frac{2}{\sqrt{n_p m_0}}\cdot\sup_{n\geq 1}\sqrt{n}\,\mathbb{E}\left[\sup_{\theta'\in\Theta}\left|\frac{1}{n}\sum_{j=1}^{n}\ell(z_j^{\theta};\theta')\cdot\varepsilon_j\right|\right] + \sqrt{\frac{6\log(T)}{n_p m_0}} \\
&\leq \frac{2\mathfrak{C}^*(\ell)+3\sqrt{\log(T)}}{\sqrt{n_p m_0}},
\end{aligned}$$

where the first step follows from the bounded differences inequality and the second step follows from a classical symmetrization argument. In the penultimate step we let $\{z_j^{\theta}\}_{j\in\mathbb{N}}$ denote an infinite sequence of samples from $\mathcal{D}(\theta)$. Putting this all together, we have that:

$$1 - \mathbb{P}\left[E^{s,p}_{\text{clean}}\right] \leq T^{-3}.$$

Finally, using that there are at most T intervals before time horizon T (across all phases), by a union bound we see that:

$$1 - \mathbb{P}\left[E_{\text{clean}}\right] \leq T^{-2},$$

as desired.

□

I.2.2 Suboptimality of the active set

We show that the elimination strategy in Algorithm 6 will never eliminate any performatively optimal point.

Lemma 259. *On the clean event* (I.1)*, any performatively optimal point* $\theta_{\mathrm{PO}} \in \arg\min_\theta \mathrm{PR}(\theta)$ *will always remain in* $\mathcal{A}$*.*

Proof. It suffices to show that θ_{PO} cannot be eliminated in Step 14 of Algorithm 6. Fix any phase p and denote by $\mathcal{P}_p$ the running set of deployed points at any point during phase p. Then, we have:

$$\begin{aligned}
\mathrm{PR}_{\mathrm{LB}}(\theta_{\mathrm{PO}}) &= \max_{\theta' \in \mathcal{P}_p} \left(\widehat{\mathrm{R}}(\theta', \theta_{\mathrm{PO}}) - L_z \varepsilon \|\theta_{\mathrm{PO}} - \theta'\| \right) \\
&\leq \max_{\theta' \in \mathcal{P}_p} \left(\mathrm{R}(\theta', \theta_{\mathrm{PO}}) - L_z \varepsilon \|\theta_{\mathrm{PO}} - \theta'\| \right) + \gamma_p \\
&\leq \mathrm{PR}(\theta_{\mathrm{PO}}) + \gamma_p \\
&= \min_\theta \mathrm{R}(\theta, \theta) + \gamma_p \\
&\leq \min_\theta \min_{\theta' \in \mathcal{P}_p} \mathrm{R}(\theta', \theta) + L_z \varepsilon \|\theta - \theta'\| + \gamma_p \\
&\leq \min_\theta \min_{\theta' \in \mathcal{P}_p} \widehat{\mathrm{R}}(\theta', \theta) + L_z \varepsilon \|\theta' - \theta\| + 2\gamma_p \\
&= \mathrm{PR}_{\min} + 2\gamma_p.
\end{aligned}$$

Therefore, $\mathrm{PR}_{\mathrm{LB}}(\theta_{\mathrm{PO}}) \leq \mathrm{PR}_{\min} + 2\gamma_p$, implying that θ_{PO} cannot be removed from $\mathcal{A}$ during phase p. Since this is true for any phase p, that completes the proof of the lemma. □

We next show that the elimination strategy is sufficiently effective that all models that remain active after a given phase p have suboptimality at most $8\gamma_p$.

Lemma 260. *On the clean event* (I.1)*, after phase* p *all models* $\theta \in \mathcal{A}$ *satisfy* $\Delta(\theta) \leq 8\gamma_p$*.*

Proof. Fix a phase p. We will analyze $\mathcal{P}_p$ at the *end* of phase p. The proof relies on two key facts:

(F1) If θ is active after phase p, then $\|\theta - \Pi_{\mathcal{P}_p}(\theta)\| \leq r_p$, where $\Pi_{\mathcal{P}_p}(\theta) = \arg\min_{\theta' \in \mathcal{P}_p} \|\theta - \theta'\|$.

(F2) θ_{PO} is active after phase p.

The first fact follows since during phase p net points cannot be eliminated from $\mathcal{S}_p$ in Step 13 while some parameter within an r_p-neighborhood is active. The second fact is proved in Lemma 259. Note that from fact (F1) it further follows that there is always a model in $\mathcal{P}_p$ within the r_p-neighborhood of θ_{PO}.

Now suppose that θ is active after phase p. Then, we have:

$$\begin{aligned}\mathrm{PR}(\theta) &\leq \mathrm{R}(\Pi_{\mathcal{P}_p}(\theta), \theta) + L_z\varepsilon\|\Pi_{\mathcal{P}_p}(\theta) - \theta\| \\ &\leq \widehat{\mathrm{R}}(\Pi_{\mathcal{P}_p}(\theta), \theta) + L_z\varepsilon\|\Pi_{\mathcal{P}_p}(\theta) - \theta\| + \gamma_p \\ &\leq \min_{\theta'} \left(\widehat{\mathrm{R}}(\Pi_{\mathcal{P}_p}(\theta'), \theta') + L_z\varepsilon\|\Pi_{\mathcal{P}_p}(\theta') - \theta'\|\right) + 2L_z\varepsilon\|\Pi_{\mathcal{P}_p}(\theta) - \theta\| + 3\gamma_p,\end{aligned}$$

where we used the definitions of $\mathrm{PR}_{\min}$ and $\mathrm{PR}_{\mathrm{LB}}(\theta)$, together with the fact that $\mathrm{PR}_{\mathrm{LB}}(\theta) \leq \mathrm{PR}_{\min} + 2\gamma_p$ for active models. Now choosing $\theta' = \theta_{\mathrm{PO}}$, applying (F1), (F2), and accounting for finite-sample uncertainty we find

$$\begin{aligned}\mathrm{PR}(\theta) &\leq \widehat{\mathrm{R}}(\Pi_{\mathcal{P}_p}(\theta_{\mathrm{PO}}), \theta_{\mathrm{PO}}) + L_z\varepsilon\|\Pi_{\mathcal{P}_p}(\theta_{\mathrm{PO}}) - \theta_{\mathrm{PO}}\| + 2L_z\varepsilon\|\Pi_{\mathcal{P}_p}(\theta) - \theta\| + 3\gamma_p \\ &\leq \widehat{\mathrm{R}}(\Pi_{\mathcal{P}_p}(\theta_{\mathrm{PO}}), \theta_{\mathrm{PO}}) + 3L_z\varepsilon r_p + 3\gamma_p \\ &\leq \mathrm{R}(\theta_{\mathrm{PO}}, \theta_{\mathrm{PO}}) + L_z\varepsilon\|\Pi_{\mathcal{P}_p}(\theta_{\mathrm{PO}}) - \theta_{\mathrm{PO}}\| + 3L_z\varepsilon r_p + 4\gamma_p \\ &\leq \mathrm{PR}(\theta_{\mathrm{PO}}) + 4(L_z\varepsilon r_p + \gamma_p) \\ &= \mathrm{PR}(\theta_{\mathrm{PO}}) + 8\gamma_p,\end{aligned}$$

where we use the fact that $r_p = \frac{\gamma_p}{L_z\varepsilon}$. Rearranging the terms we obtain $\Delta(\theta) = \mathrm{PR}(\theta) - \mathrm{PR}(\theta_{\mathrm{PO}}) \leq 8\gamma_p$ as claimed in Lemma 260. $\square$

I.2.3 Bounding the number of suboptimal deployments

For $i \geq 1$, we consider the suboptimality bands

$$\mathcal{E}_i = \left\{\theta : \Delta(\theta) \in [8 \cdot 2^{-i} L_z\varepsilon, 16 \cdot 2^{-i} L_z\varepsilon)\right\}.$$

In the following lemma, we bound the number of times that models in $\mathcal{E}_i$ can be deployed in a given phase.

Lemma 261. *Suppose that the clean event (I.1) holds. For $i \geq 1$, in phase $\log_2(1/(L_z\varepsilon)) \leq p \leq \log_2(1/(L_z\varepsilon)) + i + 1$, the number of models in $\mathcal{E}_i$ that are deployed is at most $\mathcal{O}\left((3/r_p)^d\right)$ in expectation, where d is the $(L_z\varepsilon)$-sequential zooming dimension.*

To provide intuition for Lemma 261, it is informative to consider a weaker version of the lemma where d is taken to be the $(L_z\varepsilon)$-zooming dimension rather than the $(L_z\varepsilon)$-sequential zooming dimension. To see why this weaker version of the lemma is true, notice that at the beginning of phase p, the set of active models $\mathcal{A}$ is a subset of $\{\theta : \Delta(\theta) \leq 8\gamma_{p-1}\} = \{\theta : \Delta(\theta) \leq 16\gamma_p\} = \{\theta : \Delta(\theta) \leq 16 L_z\varepsilon r_p\}$. The set of models deployed in phase p is contained in a minimal r_p-net of $\mathcal{A}$. Notice that $r_p \geq 2^{-(i+1)}$. By the definition of zooming dimension, we know that at most a multiple of $\left(\frac{3}{r_p}\right)^d$ elements from the set $\{\theta : \Delta(\theta) \in [8 \cdot 2^{-i} L_z\varepsilon, 16 \cdot 2^{-i} L_z\varepsilon)\} = \left\{\theta : \Delta(\theta) \in [16 \cdot 2^{-(i+1)} L_z\varepsilon, 32 \cdot 2^{-(i+1)} L_z\varepsilon)\right\}$ are deployed, as desired.

The proof of Lemma 261 boils down to refining this proof sketch to account for the sequential elimination aspect of Algorithm 6.

Proof. For the purposes of this analysis, we condition on the clean event.

Fix a phase $\log_2(1/(L_z\varepsilon)) \leq p \leq \log_2(1/(L_z\varepsilon)) + i + 1$. Let $\mathcal{S}_p^0$ be the covering of $\mathcal{A}$ chosen at the beginning of phase p, and let π be an ordering of $\mathcal{S}_p^0$ chosen uniformly at random. It is not difficult to see that Algorithm 6 is equivalent to drawing π at the beginning of the phase, and deploying models in the order given by π (naturally, skipping those that get eliminated). For technical convenience, we analyze this reformulation of the algorithm.

Condition on a realization π, and let $\mathcal{P}_p \subseteq \mathcal{S}_p^0$ be the set of models that are ultimately get deployed. Note that $\mathcal{P}_p$ depends on the randomness arising from finite-sample noise at each step of the phase. We will show a bound on $|\mathcal{P}_p|$ that deterministically holds on the clean event. In particular, consider the models $\theta \in \mathcal{S}_p^0 \cap \{\theta : 8L_z\varepsilon r_i \leq \Delta(\theta) < 16L_z\varepsilon r_i\}$ such that:

$$\mathrm{PR}_{\mathrm{LB}}^{r_p}(\pi(\theta)) \leq \mathrm{PR}_{\min}(\pi(\theta)) + 4L_z\varepsilon r_p = \mathrm{PR}_{\min}(\pi(\theta)) + 4\gamma_p. \tag{I.2}$$

We will show that $\mathcal{P}_p$ is a subset of such models.

Suppose that $\theta_{\mathrm{net}} \in \mathcal{S}_p^0$ is deployed in phase p. Then, that means that there exists $\theta'' \in \mathrm{Ball}_{r_p}(\theta_{\mathrm{net}})$ that remains active after the first $\pi(\theta_{\mathrm{net}}) - 1$ deployments; that is:

$$\begin{aligned}\max_{\theta':\pi(\theta')<\pi(\theta_{\mathrm{net}})}(\widehat{\mathrm{R}}(\theta',\theta'') - L_z\varepsilon\|\theta'-\theta''\|) &= \mathrm{PR}_{\mathrm{LB}}(\theta'')\\ &\leq \mathrm{PR}_{\min} + 2\gamma_p\\ &= \min_{\theta}\min_{\theta':\pi(\theta')<\pi(\theta_{\mathrm{net}})}(\widehat{\mathrm{R}}(\theta',\theta) + L_z\varepsilon\|\theta'-\theta\|) + 2\gamma_p.\end{aligned}$$

Since the clean event holds, we know that:

$$\max_{\theta':\pi(\theta')<\pi(\theta_{\mathrm{net}})}(\mathrm{DPR}(\theta',\theta'') - L_z\varepsilon\|\theta'-\theta''\|) - \gamma_p \leq \min_{\theta}\min_{\theta':\pi(\theta')<\pi(\theta_{\mathrm{net}})}(\mathrm{DPR}(\theta',\theta) + L_z\varepsilon\|\theta'-\theta\|) + 3\gamma_p.$$

Rearranging, this means that:

$$\begin{aligned}&\max_{\theta':\pi(\theta')<\pi(\theta_{\mathrm{net}})}(\mathrm{DPR}(\theta',\theta'') - L_z\varepsilon\|\theta'-\theta''\|)\\ &\qquad\leq \min_{\theta}\min_{\theta':\pi(\theta')<\pi(\theta_{\mathrm{net}})}(\mathrm{DPR}(\theta',\theta) + L_z\varepsilon\|\theta'-\theta\|) + 4\gamma_p = \mathrm{PR}_{\min}(\pi(\theta_{\mathrm{net}})) + 4\gamma_p.\end{aligned}$$

This further implies that:

$$\mathrm{PR}_{\mathrm{LB}}^{r_p}(\pi(\theta_{\mathrm{net}})) = \min_{\theta''\in\mathrm{Ball}_{r_p}(\theta_{\mathrm{net}})}\max_{\theta':\pi(\theta')<\pi(\theta_{\mathrm{net}})}(\mathrm{DPR}(\theta',\theta'') - L_z\varepsilon\|\theta'-\theta''\|) \leq \mathrm{PR}_{\min}(\pi(\theta_{\mathrm{net}})) + 4\gamma_p.$$

We see that any $\theta_{\mathrm{net}} \in \mathcal{P}_p$ must satisfy condition (I.2). By the definition of sequential zooming dimension, we know that the expected number of models in $\mathcal{E}_i$ that satisfy (I.2), where the expectation is taken over the randomness of π, is at most a multiple of $\left(\frac{3}{r_p}\right)^d$, hence $\mathbb{E}\,|\mathcal{P}_p \cap \mathcal{E}_i| \leq \mathcal{O}\left(\left(\frac{3}{r_p}\right)^d\right)$, as desired. $\square$

I.2.4 Regret bound on the clean event

To bound the regret on the clean event, we break the analysis into two cases: (a) the first $\log_2(1/(L_z\varepsilon))$ phases, and (b) all remaining phases.

Lemma 262. *Suppose that the clean event* (I.1) *holds. In the first* $\lfloor \log_2(1/(L_z\varepsilon)) \rfloor$ *phases, the algorithm has incurred regret at most*

$$\mathrm{Reg}_{\mathrm{ph}}\left(\lfloor \log_2(1/(L_z\varepsilon)) \rfloor\right) = \mathcal{O}\left(\sqrt{\frac{T}{m_0}}\left(\sqrt{\log T} + \mathfrak{C}\right)\right).$$

Proof. During phases $p \leq \log_2(1/(L_z\varepsilon))$, we deploy a single model since $r_p \geq 1$ and Θ is assumed to have radius 1.

We break the first $\lfloor \log_2(1/(L_z\varepsilon)) \rfloor$ phases into two cases. For a value of $N \geq 0$ specified later, we consider cases $p < N$ and $p \geq N$ separately.

Case 1: phases $N \leq p \leq \lfloor \log_2(1/(L_z\varepsilon)) \rfloor$. By Lemma 260, we see that the model deployed in phase N must have suboptimality at most $8 \cdot 2^{-N+1} = 2^{-N+4}$. Since the algorithm runs for at most T time steps, this means that the total regret incurred in these phases is at most $T \cdot 2^{-N+4}$.

Case 2: phases $0 \leq p < \min\{N, \lfloor \log_2(1/(L_z\varepsilon)) \rfloor\}$.

By Lemma 260, we know that the model deployed in phase p must have suboptimality at most $8 \cdot 2^{-p+1} = 2^{-p+4}$. Moreover, this model is deployed for $n_p = \left\lceil \frac{\left(2\mathfrak{C}+3\sqrt{\log T}\right)^2}{\gamma_p^2 m_0} \right\rceil$ steps. The regret incurred up to phase N can thus be bounded as:

$$\begin{aligned}
\mathrm{Reg}_{\mathrm{ph}}(N) &\leq \sum_{p=0}^{N-1} n_p 2^{-p+4} \\
&\leq 16 \sum_{p=0}^{N-1} 2^{-p} \left\lceil \frac{2^{2p}(2\mathfrak{C} + 3\sqrt{\log T})^2}{m_0} \right\rceil .
\end{aligned}$$

Since we assume $m_0 = o((\mathfrak{C} + \sqrt{\log T})^2)$, for a large enough T we have $n_p \geq 1$ and thus $\lceil n_p \rceil \leq 2 n_p$. Therefore,

$$\begin{aligned}
\mathrm{Reg}_{\mathrm{ph}}(N) &\leq C \sum_{p=0}^{N-1} 2^{-p} \frac{2^{2p}(2\mathfrak{C} + 3\sqrt{\log T})^2}{m_0} \\
&\leq C \frac{(2\mathfrak{C} + 3\sqrt{\log T})^2}{m_0} \left(\sum_{p=0}^{N-1} 2^p \right) \\
&\leq C \cdot 2^N \frac{(2\mathfrak{C} + 3\sqrt{\log T})^2}{m_0},
\end{aligned}$$

for some large enough constant $C > 0$.

Putting the two cases together, on the clean event, the total regret incurred in phases $p = 0, \ldots, \lfloor \log_2(1/(L_z\varepsilon)) \rfloor$ can be upper bounded by

$$C \cdot 2^N \frac{(2\mathfrak{C} + 3\sqrt{\log T})^2}{m_0} + T \cdot 2^{-N+4}.$$

We can also trivially upper bound the regret by T, using the fact that the loss incurred at each step is at most 1. This means that we obtain a regret bound of:

$$\mathcal{O}\left(\min\left\{T, 2^N \frac{(2\mathfrak{C} + 3\sqrt{\log T})^2}{m_0} + T \cdot 2^{-N+4}\right\}\right).$$

We now choose N to minimize this bound. We let $\eta = 2^{-N}$ and optimize over $\eta \in (0, 1)$. Optimizing over η instead of an integral value of N changes the bound by constant factors at most. This means that we can upper bound the regret by:

$$\mathcal{O}\left(\min_{0<\eta\leq 1} \min\left\{T, \eta^{-1} \frac{\left(2\mathfrak{C} + 3\sqrt{\log T}\right)^2}{m_0} + T\eta\right\}\right).$$

If $\eta > 1$, then the minimum of the two terms would be T, which is at least as big as the above expression. Therefore, we can upper bound the above expression by:

$$\mathcal{O}\left(\min_{\eta>0}\left(\eta^{-1} \frac{\left(2\mathfrak{C} + 3\sqrt{\log T}\right)^2}{m_0} + T\eta\right)\right).$$

We set $\eta = \frac{3\sqrt{\log T} + 2\mathfrak{C}}{\sqrt{m_0 T}}$ and obtain a regret bound of:

$$\mathrm{Reg}_{\mathrm{ph}}\left(\lfloor \log_2(1/(L_z\varepsilon)) \rfloor\right) = \mathcal{O}\left(\sqrt{\frac{T}{m_0}}\left(\sqrt{\log T} + \mathfrak{C}\right)\right),$$

as desired. □

Lemma 263. *Suppose that the clean event (I.1) holds. Let $d \geq 0$ be such that for every $i \geq 0$ and every phase $p \in [\log_2(1/(L_z\varepsilon)), \log_2(1/(L_z\varepsilon)) + i + 1]$, the number of models in $\mathcal{E}_i = \{\theta : \Delta(\theta) \in [2^{-i+3}L_z\varepsilon, 2^{-i+4}L_z\varepsilon]\}$ that are deployed in phase p is upper bounded by $\mathcal{O}\left(\left(\frac{3}{r_p}\right)^d\right)$ in expectation. Then, the regret incurred in phases $p \geq \log_2(1/(L_z\varepsilon))$, within time horizon T, can be upper bounded as*

$$\mathrm{Reg}_{\mathrm{ph}}\left(\lceil \log_2(1/(L_z\varepsilon)) \rceil : \infty\right) \leq \mathcal{O}\left(T^{\frac{d+1}{d+2}} (L_z\varepsilon)^{\frac{d}{d+2}} \left(\frac{(\sqrt{\log T} + \mathfrak{C})^2}{m_0}\right)^{\frac{1}{d+2}}\right).$$

Proof. By Lemma 260, we see that all models θ that are active in phase $p = \lceil \log_2(1/L_z\varepsilon)) \rceil$ or later have $\Delta(\theta) \leq 8L_z\varepsilon r_p \leq 8L_z\varepsilon$. We split these models into suboptimality bands and define, for each $i \geq 1$, the set:

$$\mathcal{E}_i = \left\{\theta : \Delta(\theta) \in [8 \cdot 2^{-i} L_z\varepsilon, 16 \cdot 2^{-i} L_z\varepsilon)\right\}.$$

Note that all models deployed starting with phase $\lceil \log_2(1/(L_z\varepsilon)) \rceil$ are in $\cup_{i\geq 1}\mathcal{E}_i$. For a value of N specified later, we break the analysis into two cases.

Case 1: models in $\cup_{i>N}\mathcal{E}_i$. Since the algorithm runs for at most T time steps, the total regret incurred due to deploying models in $\cup_{i>N}\mathcal{E}_i$ is at most

$$T \cdot 16 \cdot 2^{-N-1} L_z\varepsilon \leq 8T2^{-N} L_z\varepsilon.$$

Case 2: models in $\cup_{1\leq i\leq N}\mathcal{E}_i$. By Lemma 260, we know that all models θ that are active is phases $p \geq N + \log_2(1/L_z\varepsilon)$ have $\Delta(\theta) \leq 82^{-p} = 8 \cdot 2^{-N} L_z\varepsilon = 16 \cdot 2^{-N-1} L_z\varepsilon$. This means that all models that are active after phase $N + \log_2(1/L_z\varepsilon)$ are in $\cup_{i>N}\mathcal{E}_i$. Thus, to bound the regret incurred by deploying models in $\cup_{1\leq i\leq N}\mathcal{E}_i$ in phase $\lceil \log_2(1/L_z\varepsilon) \rceil$ or later, we only need to consider phases $p = \lceil \log_2(1/L_z\varepsilon) \rceil, \ldots, N + \log_2(1/L_z\varepsilon)$.

For $1 \leq i \leq N$, consider $\mathcal{E}_i$. By Lemma 260, we know that any $\theta \in \mathcal{E}_i$ can only be active during phases $p \leq \log_2(1/L_z\varepsilon) + i + 1$. By assumption, in phase p, the number of points in $\mathcal{E}_i$ that are deployed is at most of the order $\left(\frac{3}{r_p}\right)^d$ in expectation. Moreover, each point is deployed n_p times. Putting this all together, the expected number of points in $\mathcal{E}_i$ deployed in phase p is at most:

$$\mathcal{O}\left(\left(\frac{3}{r_p}\right)^d n_p\right) = \mathcal{O}\left(\left(\frac{3}{r_p}\right)^d \frac{(2\mathfrak{C} + 3\sqrt{\log T})^2}{L_z^2\varepsilon^2 r_p^2 m_0}\right),$$

where we use the fact that, given the condition $m_0 = o((\mathfrak{C} + \sqrt{\log T})^2)$, $n_p \geq 1$ for large enough T and hence we can bound $\lceil n_p \rceil \leq 2n_p$. Take $p = j + \log_2(1/L_z\varepsilon))$; then, $r_p = 2^{-j}$. We sum over phases $\log_2(1/(L_z\varepsilon)) \leq p \leq \log_2(1/(L_z\varepsilon)) + i + 1$ to obtain that in expectation, the total number of times that these models are deployed is at most:

$$\mathcal{O}\left(\frac{3^d(2\mathfrak{C} + 3\sqrt{\log T})^2}{L_z^2\varepsilon^2 m_0} \sum_{j=0}^{i+1} 2^{j(d+2)}\right) = \mathcal{O}\left(\frac{3^d(2\mathfrak{C} + 3\sqrt{\log T})^2}{L_z^2\varepsilon^2 m_0} 2^{(i+1)(d+2)}\right).$$

Using the fact that the models have suboptimality at most $16 \cdot 2^{-i} L_z\varepsilon = 32 \cdot 2^{-(i+1)} L_z\varepsilon$, we see that the regret incurred by deploying models in $\mathcal{E}_i$ is upper bounded by:

$$\mathcal{O}\left(\frac{3^d(2\mathfrak{C} + 3\sqrt{\log T})^2}{L_z\varepsilon m_0} 2^{(i+1)(d+1)}\right).$$

We sum over $1 \leq i \leq N$ to obtain the total regret incurred due to deploying models in $\cup_{1\leq i\leq N}\mathcal{E}_i$:

$$\mathcal{O}\left(\frac{3^d(2\mathfrak{C}+3\sqrt{\log T})^2}{L_z\varepsilon m_0}2^{(N+2)(d+1)}\right).$$

Putting together the two cases we obtain a total regret bound of

$$\mathcal{O}\left(\frac{3^d(2\mathfrak{C}+3\sqrt{\log T})^2}{L_z\varepsilon m_0}2^{(N+2)(d+1)}+T2^{-N}L_z\varepsilon\right).$$

We also can upper bound the regret by $8TL_z\varepsilon$, since all models active after phase $\lfloor\log_2(1/(L_z\varepsilon))\rfloor$ have $\Delta(\theta)\leq 8L_z\varepsilon$ and there are at most T time steps in total. This means that we can bound the regret by:

$$\mathcal{O}\left(\min\left\{TL_z\varepsilon,\frac{3^d(2\mathfrak{C}+3\sqrt{\log T})^2}{L_z\varepsilon m_0}2^{(N+2)(d+1)}+T2^{-N}L_z\varepsilon\right\}\right).$$

We now choose N to minimize this bound. We let $\eta=2^{-N}$ and choose some $\eta\in(0,1)$. The error from optimizing over $\eta\in(0,1)$ instead of an integral value of N contributes at most constant factors. This means that we can upper bound the regret by:

$$\mathcal{O}\left(\min\left\{TL_z\varepsilon,\frac{12^d(2\mathfrak{C}+3\sqrt{\log T})^2}{L_z\varepsilon m_0}\eta^{-(d+1)}+T\eta L_z\varepsilon\right\}\right),$$

for any $\eta\in(0,1)$. Note that, if $\eta\geq 1$, the second term in the bound is at least as large as the first term, hence we can choose any $\eta>0$. In particular, we can further upper bound the regret by

$$\mathcal{O}\left(\min_{\eta>0}\left(\frac{12^d(2\mathfrak{C}+3\sqrt{\log T})^2}{L_z\varepsilon m_0}\eta^{-(d+1)}+T\eta L_z\varepsilon\right)\right).$$

Now, we set

$$\eta=\left(\frac{12^d\left(3\sqrt{\log T}+2\mathfrak{C}\right)^2}{TL_z^2\varepsilon^2 m_0}\right)^{\frac{1}{d+2}}.$$

Thus, we finally get a regret bound of

$$\mathcal{O}\left(T^{\frac{d+1}{d+2}}(L_z\varepsilon)^{\frac{d}{d+2}}\left(\frac{\left(\sqrt{\log T}+\mathfrak{C}\right)^2}{m_0}\right)^{\frac{1}{d+2}}\right),$$

as desired. □

I.2.5 Proof of Theorem 113

Now, we are ready to prove Theorem 113.

First, we handle the case where the clean event defined in (I.1) does not hold and the concentration bound is violated. By Lemma 258, this happens with probability at most T^{-2}. The regret incurred in each deployment is at most 1 and there are T deployments, so these events contribute a negligible factor T^{-1} to the expected regret.

For the case where the clean event holds we can build on Lemma 261, Lemma 262, and Lemma 263. From Lemma 262, we obtain a bound for the total regret incurred in phases up to $\lfloor \log_2(1/(L_z\varepsilon)) \rfloor$. By Lemma 261 we can set the parameter d in Lemma 263 to be the $(L_z\varepsilon)$-sequential zooming dimension, and thus from Lemma 263 we obtain a regret bound for all later phases.

Putting all this together yields the desired bound.

I.3 Regret analysis of Algorithm 7

The proof of Theorem 115 relies on two key lemmas. One proves that $\mathcal{C}_t$ are valid confidence sets for μ_* at every step, and the other one proves a regret bound assuming that $\mathcal{C}_t$ are valid confidence sets.

Throughout we denote by $\mathcal{B}_m$ the unit ball in $\mathbb{R}^m$. For a vector x and matrix M, we will use the notation $\|x\|_M = \sqrt{x^\top M x}$.

An important object in the proofs will be $S_t := \sum_{i=1}^t \theta_i \bar{z}_{0,i}^\top$, where $\bar{z}_{0,i} = \frac{1}{m_0}\sum_{j=1}^{m_0} z_i^{(j)} - \mu_*^\top \theta_i$. Essentially $\bar{z}_{0,i}$ is the average over m_0 samples from $\mathcal{D}_0$, collected at step i. We will also denote $V_t(\lambda) = (\lambda I + \sum_{i=1}^t \theta_i \theta_i^\top)$, for an arbitrary offset $\lambda > 0$, and $V_t \equiv V_t(0)$. Note that in the algorithm statement we use $\Sigma_t = V_t\left(\frac{1}{m_0}\right)$.

I.3.1 Clean event

As for Algorithm 6, we introduce a clean event. In this case, the clean event will be defined as

$$E_{\text{clean}} = \{\forall t \in \mathbb{N} : \mu_* \in \mathcal{C}_t\}, \tag{I.3}$$

where $\mathcal{C}_t$ are the confidence sets constructed in Algorithm 7.

The technical subtlety lies in the fact that the points θ_t are chosen adaptively, hence one cannot simply apply standard least-squares confidence intervals to argue that the sets $\mathcal{C}_t$ are valid. The same difficulty is resolved in the analysis of the LinUCB algorithm and our proof builds on the proof technique of that analysis.

Before stating the main technical lemma, we start with an auxiliary result that we will use in the proof.

Lemma 264. *Suppose that $\mathcal{D}_0$ is 1-subgaussian. Then, for all $x \in \mathcal{B}_m$ and $y \in \mathbb{R}^{d_\Theta}$, the process*

$$M_t(x, y) = \exp\left(y^\top S_t x - \frac{1}{2m_0}\|y\|_{V_t}^2\right)$$

is a supermartingale with respect to the natural filtration, with $M_0(x, y) = 1$.

Proof. Since $\bar{z}_{0,i}$ are $\frac{1}{\sqrt{m_0}}$-subgaussian, we know that all one-dimensional projections are also $\frac{1}{\sqrt{m_0}}$-subgaussian, hence $\bar{z}_{0,i}^\top x$ are independent $\frac{1}{\sqrt{m_0}}$-subgaussian as well. Using this, we know

$$\mathbb{E}\left[\exp(y^\top \theta_t z_{0,t}^\top x) \;\middle|\; \mathcal{F}_{t-1}\right] \le \exp\left(\frac{(y^\top \theta_t)^2}{2m_0}\right) = \exp\left(\frac{\|y\|_{\theta_t\theta_t^\top}^2}{2m_0}\right)$$

almost surely. Hence,

$$\begin{aligned}
\mathbb{E}[M_t(x, y) \mid \mathcal{F}_{t-1}] &= \mathbb{E}\left[\exp\left(y^\top S_t x - \frac{1}{2m_0}\|y\|_{V_t}^2\right) \;\middle|\; \mathcal{F}_{t-1}\right] \\
&= M_{t-1}(x, y)\mathbb{E}\left[\exp\left(y^\top \theta_t z_{0,t}^\top x - \frac{1}{2m_0}\|y\|_{\theta_t\theta_t^\top}^2\right) \;\middle|\; \mathcal{F}_{t-1}\right] \\
&\le M_{t-1}(x, y)
\end{aligned}$$

almost surely. Furthermore, $M_0(x, y) = 1$ is trivially true. □

Now we are ready to state the main technical lemma about the validity of $\mathcal{C}_t$.

Lemma 265. *We have that*

$$\mathbb{P}\left\{E_{\text{clean}}\right\} \ge 1 - T^{-2}.$$

Proof. First we will show that for any $\delta \in (0, 1)$,

$$\mathbb{P}\left\{\exists t \in \mathbb{N} : \|V_t(\lambda)^{-1/2} S_t\|^2 \ge \frac{1}{m_0}\left(8m + 4\log\left(\frac{1}{\delta}\right) + 2\log\left(\frac{\det(V_t(\lambda))}{\lambda^{d_\Theta}}\right)\right)\right\} \le \delta, \tag{I.4}$$

for all $\lambda > 0$.

Let $\Sigma = \frac{m_0}{\lambda} I \in \mathbb{R}^{d_\Theta \times d_\Theta}$ and let h be the density of $\mathcal{N}(0, \Sigma)$. Then, for any fixed $x \in \mathcal{B}_m$ and $M_t(x, y)$ as in Lemma 264, define

$$\bar{M}_t(x) = \int_{\mathbb{R}^{d_\Theta}} M_t(x, y) h(y) = \frac{1}{\sqrt{(2\pi)^{d_\Theta}\det(\Sigma)}} \int_{\mathbb{R}^{d_\Theta}} \exp\left(y^\top S_t x - \frac{1}{2m_0}\|y\|_{V_t}^2 - \frac{1}{2}\|y\|_{\Sigma^{-1}}^2\right) dy.$$

Notice that we can write

$$y^\top S_t x - \frac{1}{2m_0}\|y\|_{V_t}^2 - \frac{1}{2}\|y\|_{\Sigma^{-1}}^2 = \frac{1}{2}\|S_t x\|_{(\Sigma^{-1} + \frac{V_t}{m_0})^{-1}}^2 - \frac{1}{2}\left\|y - \left(\Sigma^{-1} + \frac{V_t}{m_0}\right)^{-1} S_t x\right\|_{\Sigma^{-1} + \frac{V_t}{m_0}}^2.$$

Thus, by integrating out the Gaussian density, we get

$$
\begin{aligned}
&\bar{M}_t(x)\\
&= \exp\left(\frac{1}{2}\|S_t x\|^2_{(\Sigma^{-1}+\frac{V_t}{m_0})^{-1}}\right) \frac{1}{\sqrt{(2\pi)^{d_\Theta}\det(\Sigma)}} \int_{\mathbb{R}^{d_\Theta}} \exp\left(-\frac{1}{2}\left\|y - \left(\Sigma^{-1}+\frac{V_t}{m_0}\right)^{-1} S_t x\right\|^2_{\Sigma^{-1}+\frac{V_t}{m_0}}\right) dy\\
&= \exp\left(\frac{1}{2}\|S_t x\|^2_{(\Sigma^{-1}+\frac{V_t}{m_0})^{-1}}\right) \left(\frac{\det((\Sigma^{-1}+\frac{V_t}{m_0})^{-1})}{\det(\Sigma)}\right)^{1/2}\\
&= \exp\left(\frac{m_0}{2}\|V_t^{-1/2}(\lambda) S_t x\|^2\right)\left(\frac{\lambda^{d_\Theta}}{\det(V_t(\lambda))}\right)^{1/2}.
\end{aligned}
$$

Now, by Lemma 20.3 in (Lattimore and Szepesvári, 2020), since $M_t(x, y)$ is a supermartingale then $\bar{M}_t(x)$ is a non-negative supermartingale with $\bar{M}_0(x) = 1$. Thus, we can apply the maximal inequality to get

$$
\begin{aligned}
&\mathbb{P}\left\{\exists t \in \mathbb{N} : \log \bar{M}_t(x) \geq \log(1/\delta)\right\}\\
&\quad = \mathbb{P}\left\{\exists t \in \mathbb{N} : \frac{m_0}{2}\|V_t(\lambda)^{-1/2} S_t x\|^2 - \frac{1}{2}\log\left(\frac{\det(V_t(\lambda))}{\lambda^{d_\Theta}}\right) \geq \log(1/\delta)\right\} \leq \delta.
\end{aligned} \tag{I.5}
$$

Inequality (I.5) is valid for all fixed $x \in \mathcal{B}_m$; to prove inequality (I.4), we use a covering argument. Let $N_{\frac{1}{2},m}$ denote a $\frac{1}{2}$-net of $\mathcal{B}_m$, and note that we can make $|N_{\frac{1}{2},m}| \leq 5^m$. Then,

$$
\|V_t(\lambda)^{-1/2} S_t\| = \max_{x \in \mathcal{B}_m} \|V_t(\lambda)^{-1/2} S_t x\| \leq 2 \max_{x \in N_{\frac{1}{2},m}} \|V_t(\lambda)^{-1/2} S_t x\|.
$$

Therefore, we can apply a union bound to conclude that for all $s > 0$,

$$
\begin{aligned}
\mathbb{P}\left\{\exists t \in \mathbb{N} : \|V_t(\lambda)^{-1/2} S_t\|^2 \geq s\right\} &\leq \mathbb{P}\left\{\exists t \in \mathbb{N} : \max_{x \in N_{1/2,m}} \|V_t(\lambda)^{-1/2} S_t x\|_2^2 \geq \frac{s}{4}\right\}\\
&\leq \sum_{x \in N_{1/2,m}} \mathbb{P}\left\{\exists t \in \mathbb{N} : \|V_t(\lambda)^{-1/2} S_t x\|_2^2 \geq \frac{s}{4}\right\}.
\end{aligned}
$$

By picking $s = \frac{1}{m_0}(8m + 4\log\frac{1}{\delta} + 2\log(\frac{\det(V_t(\lambda))}{\lambda^{d_\Theta}})) \geq \frac{1}{m_0}(4\log\frac{5^m}{\delta} + 2\log(\frac{\det(V_t(\lambda))}{\lambda^{d_\Theta}}))$ and applying Equation (I.5), we get

$$
\mathbb{P}\left\{\exists t \in \mathbb{N} : \|V_t(\lambda)^{-1/2} S_t\|^2 \geq \frac{1}{m_0}\left(8m + 4\log\left(\frac{1}{\delta}\right) + 2\log\left(\frac{\det(V_t(\lambda))}{\lambda^{d_\Theta}}\right)\right)\right\} \leq \sum_{x \in N_{1/2,m}} \frac{\delta}{5^m} \leq \delta.
$$

This completes the proof of inequality (I.4).

It remains to relate this bound to the definition of $\mathcal{C}_t$. We can write

$$
\hat{\mu}_t - \mu_* = V_t(\lambda)^{-1} S_t + V_t(\lambda)^{-1} V_t \mu_* - \mu_*,
$$

and therefore

$$\begin{aligned}\|V_t(\lambda)^{1/2}(\hat{\mu}_t - \mu_*)\| &= \|V_t(\lambda)^{-1/2}S_t + V_t(\lambda)^{1/2}(V_t(\lambda)^{-1}V_t - I)\mu_*\| \\ &\leq \|V_t(\lambda)^{-1/2}S_t\| + \sqrt{\|\mu_*^\top (V_t(\lambda)^{-1}V_t - I)V_t(\lambda)(V_t(\lambda)^{-1}V_t - I)\mu_*\|} \\ &= \|V_t(\lambda)^{-1/2}S_t\| + \sqrt{\lambda}\sqrt{\|\mu_*^\top (I - V_t(\lambda)^{-1}V_t)\mu_*\|} \\ &= \|V_t(\lambda)^{-1/2}S_t\| + \sqrt{\lambda}\|\mu_*\|,\end{aligned}$$

where the second equality follows by writing $V_t = V_t(\lambda) - \lambda I$. Note additionally that by $\max\{\|\theta\| : \theta \in \Theta\} \leq 1$ and the AM-GM inequality,

$$\det(V_t(\lambda)) \leq \left(\frac{1}{d_\Theta}\mathrm{trace} V_t(\lambda)\right)^{d_\Theta} \leq \left(\frac{d_\Theta \lambda + t}{d_\Theta}\right)^{d_\Theta}.$$

Applying Equation (I.4), setting $\delta = \frac{1}{T^2}$ and $\lambda = \frac{1}{m_0}$ completes the proof. □

I.3.2 Regret bound on the clean event

The place where the structure of the performative risk comes into play is the following lemma, where we relate the suboptimality of the deployed model θ_t to properties of the confidence set $\mathcal{C}_t$.

Lemma 266. *Suppose that the clean event (I.3) holds. Then, we can bound the suboptimality of θ_t by*

$$\Delta(\theta_t) \leq \min\left\{1, L_z \sup_{\mu,\mu' \in \mathcal{C}_t} \|(\mu - \mu')^\top \theta_t\|\right\}.$$

Proof. In what follows, all expectations are taken only over a sample $z_0 \sim \mathcal{D}_0$ independent of everything else (i.e., all other random quantities are conditioned on).

Since the loss is bounded, we know $\Delta(\theta_t) \leq 1$. For the other bound, notice that

$$\Delta(\theta_t) = \mathbb{E}\ell(z_0 + \mu_*^\top \theta_t; \theta_t) - \mathbb{E}\ell(z_0 + \mu_*^\top \theta_{\mathrm{PO}}; \theta_{\mathrm{PO}}).$$

By the definition of the algorithm and the clean event, we can lower bound the second term $\mathbb{E}\ell(z_0 + \mu_*^\top \theta_{\mathrm{PO}}; \theta_{\mathrm{PO}})$ as follows:

$$\mathbb{E}\ell(z_0 + \mu_*^\top \theta_{\mathrm{PO}}; \theta_{\mathrm{PO}}) \geq \mathrm{PR}_{\mathrm{LB}}(\theta_{\mathrm{PO}}) \geq \mathrm{PR}_{\mathrm{LB}}(\theta_t) = \mathbb{E}\ell(z_0 + \tilde{\mu}_t^\top \theta_t; \theta_t),$$

for some $\tilde{\mu}_t \in \mathcal{C}_t$. This means that:

$$\Delta(\theta_t) \leq \mathbb{E}\ell(z_0 + \mu_*^\top \theta_t; \theta_t) - \mathbb{E}\ell(z_0 + \tilde{\mu}_t^\top \theta_t; \theta_t).$$

To finish, we use Lipschitzness of the loss to upper bound this by $L_z\|(\mu_* - \tilde{\mu}_t)^\top \theta_t\|$. Using the clean event, we can further upper bound this by $L_z \sup_{\mu,\mu' \in \mathcal{C}_t} \|(\mu - \mu')^\top \theta_t\|$ as desired. □

We now use this bound on the suboptimality of deployed models, along with the structure of the confidence sets, to bound the regret on the clean event.

Lemma 267. *Let $1 \leq \beta_1 \leq \beta_2 \leq \ldots \beta_T$ and assume that the loss $\ell(z;\theta)$ is L_z-Lipschitz in z. Assume that the event*

$$\mu_* \in \mathcal{C}_t \subseteq \left\{ \mu \in \mathbb{R}^{d_\Theta \times m} : \left\| V_{t-1}^{1/2}\left(\frac{1}{m_0}\right)(\mu - \hat{\mu}_{t-1}) \right\|^2 \leq \beta_t \right\}$$

holds true, for all $2 \leq t \leq T$. Then, on this event, Algorithm 7 satisfies:

$$\sum_{t=1}^{T} \Delta(\theta_t) = \tilde{\mathcal{O}}\left(1 + \sqrt{d_\Theta T \beta_T \log\left(\frac{d_\Theta + T m_0}{d_\Theta}\right)} \max\{L_z, 1\}\right).$$

Proof. As in the proof of Lemma 266, all expectations are taken only over a sample $z_0 \sim \mathcal{D}_0$ independent of everything else (i.e., all other random quantities are conditioned on).

First, we separately bound the regret of the first step as $\mathcal{O}(1)$, using the fact that the loss is bounded in $[0, 1]$.

For the remainder of the steps, we apply Lemma 266 to upper bound $\Delta(\theta_t)$. Using this, coupled with structure of $\mathcal{C}_t$, we can obtain the following upper bound, for any $\lambda > 0$:

$$\begin{aligned}
\Delta(\theta_t) &\leq \min\left\{1, L_z \sup_{\mu,\mu' \in \mathcal{C}_t} \|(\mu - \mu')^\top \theta_t\|\right\} \\
&\leq \min\left\{1, L_z \sup_{\mu,\mu' \in \mathcal{C}_t} \|(\mu - \mu')^\top V_{t-1}^{1/2}(\lambda)\| \cdot \|V_{t-1}^{-1/2}(\lambda)\theta_t\|\right\} \\
&\leq \min\left\{1, 2L_z\sqrt{\beta_t}\|V_{t-1}^{-1/2}(\lambda)\theta_t\|\right\} \\
&\leq 2\sqrt{\beta_T}\min\left\{1, L_z\|V_{t-1}^{-1/2}(\lambda)\theta_t\|\right\},
\end{aligned}$$

where the last line uses the fact that $\beta_T \geq \max\{1, \beta_t\}$.

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{t=2}^{T} \Delta(\theta_t) &\leq \sqrt{T \sum_{t=2}^{T} \Delta(\theta_t)^2} \\
&\leq 2\sqrt{T\beta_T \sum_{t=2}^{T} \min\left\{1, L_z^2 \|V_{t-1}^{-1/2}(\lambda)\theta_t\|^2\right\}} \\
&\leq 2\sqrt{T\beta_T \sum_{t=2}^{T} \min\left\{1, \max\{1, L_z^2\} \|V_{t-1}^{-1/2}(\lambda)\theta_t\|^2\right\}} \\
&\leq 2\sqrt{T \max\{1, L_z^2\}\beta_T \sum_{t=2}^{T} \min\left\{1, \|V_{t-1}^{-1/2}(\lambda)\theta_t\|^2\right\}} \\
&= 2\max\{1, L_z\}\sqrt{T\beta_T \sum_{t=2}^{T} \min\left\{1, \|V_{t-1}^{-1/2}(\lambda)\theta_t\|^2\right\}}.
\end{aligned}$$

Finally, we use Lemma 19.4 in (Lattimore and Szepesvári, 2020) that says

$$\sum_{t=2}^{T} \min\left\{1, \|V_{t-1}^{-1/2}(\lambda)\theta_t\|^2\right\} \leq 2d_\Theta \log\left(\frac{\mathrm{trace} V_0(\lambda) + T}{d_\Theta \det(V_0(\lambda))^{1/d_\Theta}}\right) = 2d_\Theta \log\left(\frac{d_\Theta\lambda + T}{d_\Theta\lambda}\right).$$

Using this expression in the equation above and setting $\lambda = \frac{1}{m_0}$ yields the final result. $\square$

I.3.3 Proof of Theorem 115

We take $\sqrt{\beta_t} = \max\left\{1, \sqrt{\frac{1}{m_0}}M_* + \sqrt{\frac{8m + 8\log T + 2d_\Theta \log\left(\frac{d_\Theta + tm_0}{d_\Theta}\right)}{m_0}}\right\}$. By the constraint that $m_0 = o(\log T)$, we see that second branch dominates over the first one and so, for large enough T, $\sqrt{\beta_t} = \sqrt{\frac{1}{m_0}}M_* + \sqrt{\frac{8m + 8\log T + 2d_\Theta \log\left(\frac{d_\Theta + tm_0}{d_\Theta}\right)}{m_0}}$. Lemma 265 shows that:

$$\mu_* \in \mathcal{C}_t \subseteq \left\{\mu \in \mathbb{R}^{d_\Theta \times m} : \left\| V_{t-1}^{1/2}\left(\frac{1}{m_0}\right)(\mu - \hat{\mu}_{t-1})\right\|^2 \leq \beta_t\right\}.$$

Moreover, the contribution of the complement of the clean event to the overall regret is negligible. Plugging this choice of β_t into the bound of Lemma 267 completes the proof of Theorem 115.

I.4 Further details on zooming dimension

I.4.1 Discussion of zooming dimension definitions

We note that Definition 12 slightly differs from the definition presented in (Kleinberg et al., 2008). The statement of Definition 12 eases the comparison of the zooming algorithm of Kleinberg et al. to our new algorithm.

First, we introduce a multiplier α to emphasize that the zooming dimension implicitly depends on the Lipschitz constant of the problem (assumed to be fixed and equal to 1 by Kleinberg et al.), which can be smaller when we make full use of performative feedback.

Second, Definition 12 is slightly more conservative in two ways. One is that we intersect a cover of any subset of $\{\theta : \Delta(\theta) \leq 16\alpha s\}$ with $\{\theta : 16\alpha r < \Delta(\theta) \leq 32\alpha r\}$, rather than directly take a cover of the latter set. The other one is that we take a supremum over all covers with radius coarser than r, i.e. $s \in [r, 1]$, instead of only r. These differences are minor technicalities that we do not expect to alter the zooming dimension in a meaningful way, neither formally nor conceptually.

Lastly, rather than requiring the size of the relevant set of points to be at most of order $(1/s)^d$, we require the size to be at most of order $(3/s)^d$. In this regard, Definition 12 is less conservative than the zooming dimension in (Kleinberg et al., 2008). We make this modification so that for the Euclidean ball of dimension d_Θ of radius 1, which contains Θ, the zooming dimension is guaranteed to be at most d_Θ. This would not be true without the factor of 3. We note that the analysis of adaptive zooming in (Kleinberg et al., 2008) can be modified in a straightforward way to allow for this change, only altering constant factors in the regret bounds.

I.4.2 Gains of sequential zooming dimension

We provide an example where the sequential zooming dimension is strictly smaller than the zooming dimension.

Example 15. *Suppose that model parameters are 2-dimensional, $L_z\varepsilon = 1/32$, and the distribution map is a fixed distribution:* $\mathrm{PR}(\theta) = \mathrm{R}(\theta', \theta)$ *for all* θ, θ'. *Let* $\theta_0 = 0, \theta_1 = [1/2, 0], \theta_2 = [1/4, \sqrt{3}/4]$. *Suppose that* $\mathrm{PR}(\theta_0) = 0$, $\mathrm{PR}(\theta_1) = 1/8$, $\mathrm{PR}(\theta_2) = 15/64$, *and* $\mathrm{PR}(\theta) = 1$ *otherwise.*

Lemma 268. *In Example 15, the $(L_z\varepsilon)$-zooming dimension is at least $d \geq 0.39$, and the $(L_z\varepsilon)$-sequential zooming dimension is at most $d \leq \log_6(1.5) \approx 0.23$.*

Proof. We begin by observing that for all $0 < s \leq 1$, it holds that:

$$\{\theta \mid \Delta(\theta) \leq 16L_z\varepsilon s\} \subseteq \{\theta \mid \mathrm{PR}(\theta) \leq 16L_z\varepsilon\} = \{\theta \mid \mathrm{PR}(\theta) \leq 1/2\} = \{\theta_0, \theta_1, \theta_2\}.$$

Note that θ_0 achieves the optimal performative risk and thus does not appear in any suboptimality band $\{\theta \mid 16L_z\varepsilon r \leq \mathrm{PR}(\theta) < 32L_z\varepsilon r\}$.

First, we show that the zooming dimension is at least $\log_6(2) \approx 0.39$. Let $s = 1/2 - \varepsilon$ for ε sufficiently small. Consider the set $\{\theta \mid \Delta(\theta) \leq 16L_z\varepsilon s\} = \{\theta_0, \theta_1, \theta_2\}$. A minimal covering of the set will necessarily consist of all three points $\{\theta_0, \theta_1, \theta_2\}$. We see that the suboptimality band $\{\theta \mid 16L_z\varepsilon r \leq \mathrm{PR}(\theta) < 32L_z\varepsilon r\}$ for $r = 1/4$ contains $\{\theta_1, \theta_2\}$. Taking $\varepsilon \to 0$, we see that the zooming dimension is at least $\log_6(2)$.

Next, we show that the sequential zooming dimension is $d \leq \log_6(1.5) \approx 0.23$. Let $\mathcal{C}$ be a minimal covering of a subset of $\{\theta \mid \Delta(\theta) \leq 16L_z\varepsilon s\}$. For $s \geq 1/2$, we see that $\{\theta \mid \Delta(\theta) \leq 16L_z\varepsilon s\} = \{\theta_0, \theta_1, \theta_2\}$, and $\mathcal{C}$ contains at most 1 point. If $s < 1/2$, then $\mathcal{C}$ might contain up to 3 points. If $\mathcal{C}$ does not contain both θ_1 and θ_2, then any suboptimality band $\{\theta \mid 16L_z\varepsilon r \leq \mathrm{PR}(\theta) < 32L_z\varepsilon r\}$ contains at most 1 point from $\mathcal{C}$. If $\mathcal{C}$ contains both θ_1 and θ_2, then we leverage the sequential properties of the sequential zooming dimension. We claim that pulling θ_1 first results in θ_2 being eliminated. Notice that if θ_1 is pulled first, then $\mathrm{PR}_{\min}$ will be equal to $1/8$. $\mathrm{PR}_{\mathrm{LB}}$ will be equal to $\mathrm{R}(\theta_1, \theta_2) + L_z\varepsilon||\theta_1 - \theta_2|| = \mathrm{PR}(\theta_2) + L_z\varepsilon||\theta_1 - \theta_2|| = 15/64 - 1/64 = 14/64 = 7/32$. We see that $\mathrm{PR}_{\min} + 4L_z\varepsilon s \leq 1/8 + 1/16 = 3/16$. Since $7/32 > 3/16$, we see that θ_1 will be eliminated. This means that in expectation, at most 1.5 arms are pulled. This yields the desired bound.

□

I.5 Details of numerical illustrations

For the purpose of the illustrations in Figure 13.1, Figure 13.2, and Figure 13.3 we use a one-dimensional example where $\theta \in \mathbb{R}$. The performative effects are modeled by a linear shift, i.e.,

$$\mathrm{R}(\phi, \theta) = f(\theta) + \alpha\phi,$$

where f is a multi-modal function illustrated in the respective figures and specified as

$$f(\theta) = c_0 \cos(c_1\theta) + c_2 \sin(c_3(\theta - c_4)).$$

The shaded gray area in the figures illustrates the confidence sets computed as

$$\mathrm{PR}_{\mathrm{LB}}(\theta) = \max_{\theta' \in \mathcal{S}} \mathrm{PR}(\theta') - L_{\mathrm{PR}}\|\theta - \theta'\|, \qquad \mathrm{PR}_{\mathrm{UB}}(\theta) = \min_{\theta' \in \mathcal{S}} \mathrm{PR}(\theta') + L_{\mathrm{PR}}\|\theta - \theta'\|$$

for the baseline approach, and as

$$\mathrm{PR}_{\mathrm{LB}}(\theta) = \max_{\theta' \in \mathcal{S}} \mathrm{R}(\theta', \theta) - L_\phi\|\theta - \theta'\|, \qquad \mathrm{PR}_{\mathrm{UB}}(\theta) = \min_{\theta' \in \mathcal{S}} \mathrm{R}(\theta', \theta) + L_\phi\|\theta - \theta'\|$$

for the performative confidence bounds. We use $\mathcal{S} := \{\theta_1, \theta_2\}$ as shown in the figures. The Lipschitz constant L_{PR} of the performative risk $\mathrm{PR}(\theta) = \mathrm{R}(\theta, \theta)$ is evaluated numerically for each figure.

For Figure 13.1 and Figure 13.2 we use the following parameters: $c_0 = -1$, $c_1 = 0.7$, $c_2 = 0.3$, $c_3 = 3$, $c_4 = 0.5$, $\alpha = 1$, and a conservative Lipschitz bound $L_\phi = 1.6$ for the performative confidence bounds and $L_{\mathrm{PR}} = 3.8$ for the performative risk.

For Figure 13.3 we use $c_0 = -3$, $c_1 = 1$, $c_2 = 0.9$, $c_3 = 3$, $c_4 = 0.5$, $\alpha = 0.5$, and a conservative Lipschitz bound $L_\phi = 1.3$ for illustrating the performative confidence bounds. If exact knowledge of the shifts were available these bounds could be made even tighter.

Appendix J

Appendix for Chapter 14

J.1 Worked-out examples, auxiliary notation, and auxiliary lemmas

J.1.1 Worked out version of Example 12 for γ-tolerant benchmark

We work out the γ-tolerant benchmark for Example 12 in more detail. Consider instance $\mathcal{I}$ (leftmost table) in Table 14.2 (with $0.4 > \gamma \geq 4\delta$), which we will use to illustrate our benchmark. We show that $\beta_1^{\text{tol}} = 0.5 + \delta$ and $\beta_2^{\text{tol}} = 4\delta$. To calculate the benchmarks, we compute the sum of the ε-relaxed Stackelberg value and ε-regularizer for different values of ε and then take a minimum. We will show that the minimum turns out to be achieved at $\varepsilon = \delta$.

First, for $\varepsilon = 0$ this benchmark is equal to the Stackelberg equilibrium, which gives values $0.5 + \delta, 0.4$ for the leader and follower respectively. For $\varepsilon \in (0, \delta)$, the ε-relaxed

	b_1	b_2	b_3
a_1	$(1, 0.5 + 2\delta)$	$(0.7, 0.5 + \delta)$	$(1.1, 0)$
a_2	$(0.8, 3.5 \cdot \delta)$	$(1.2, 3 \cdot \delta)$	$(0.9, 4 \cdot \delta)$
a_3	$(0.5, 0.5)$	$(0.7, 0)$	$(2, 0.1)$

Table J.1: Calculating the δ-tolerant benchmark: Note that (a_1, b_1) is the Stackelberg equilibrium, which by Theorem 116 cannot in general be learned with sublinear regret. For each row, cells shaded in blue if they are within the δ best response for the follower ($\mathcal{B}_\delta(a_i)$). Entry (a_2, b_1) (with purple text) gives the leader's δ-relaxed Stackelberg utility - the leader's best action, assuming the follower picks the worst item within the δ-response ball. Rows a_1, a_2 (shaded in red) are in $\mathcal{A}_\delta$, the set of actions where the leader has a chance of doing at least as well as the δ-relaxed Stackelberg utility $((a_2, b_1))$. Finally, (a_2, b_3) (in green) gives the follower's best response, assuming the leader picks the worst action for it within $\mathcal{A}_\delta$.

	b_1	b_2	b_3
a_1	$(1, 0.5 + 2\delta)$	$(0.7, 0.5 + \delta)$	$(1.1, 0)$
a_2	$(0.8, 3.5 \cdot \delta)$	$(1.2, 3 \cdot \delta)$	$(0.9, 4 \cdot \delta)$
a_3	$(0.5, 0.5)$	$(0.7, 0)$	$(2, 0.1)$

Table J.2: Calculating the **self-δ**-tolerant benchmark: Note that $\mathcal{B}_\delta, \mathcal{A}_\delta$ are defined the same as in the γ-tolerant benchmark in Table J.1, so the only difference is the location of the δ-relaxed Stackelberg utility values for the leader and the follower, which are calculated by finding the *worst* expected reward for each within the $\mathcal{B}_\delta, \mathcal{A}_\delta$ sets. Here, they occur for the leader in (a_1, b_2) (in purple) and for the follower in (a_2, b_2) (in green).

Stackelberg value stays the same while the regularizer increases. For $\varepsilon = \delta$, the behavior of the ε-Stackelberg utility becomes more complicated.

- **Follower ε-best-response set:** In this instance, $\mathcal{B}_\delta(a_1) = \{a_1\}$: for arm a_1, because $0.4 > \delta$, only $\{b_1\}$ is in the best-response set. However, $\mathcal{B}_\delta(a_2) = \{b_1, b_2\}$: both arms for the follower are within δ of optimal.

- **Leader ε-relaxed Stackelberg utility:** This term captures the best utility that the leader can achieve if the follower worst-case ε-best-responds according to $\arg\min_{b \in \mathcal{B}_\delta}(a)$. Since $\mathcal{B}_\delta(a_1) = \{b_1\}$, we see that $\min_{b \in \mathcal{B}_\delta}(a_1) = 0.5 + \delta$. However, for a_2, $\min_{b \in \mathcal{B}_\delta}(a) = v_1(a_2, b_2) = 0.4$. The leader's best action is to pick a_1, so the δ-relaxed Stackelberg utility is equal to $0.5 + \delta$.

- **Leader ε-best-response sets:** We construct the $\mathcal{A}_\delta$ set by considering all actions a where the *best-case* outcome within the $\mathcal{B}_\delta(a)$ gives reward at least within δ of our benchmark value of $0.5 + \delta$. We can see $\mathcal{A}_\delta = \{a_1, a_2\}$ because they both contain an item within δ of the benchmark value ((a_1, b_1) or (a_2, b_1) respectively).

- **Follower's ε-relaxed Stackelberg utility:** This term considers the worst-case action within $\mathcal{A}_\delta$ for the follower. If the leader picks a_1, the only response is b_1 which gives value 0.4, while if the leader picks a_2, the best response is b_2 which gives value $3 \cdot \delta$. The minimum of these, plus a regularizer term, gives a benchmark of $4 \cdot \delta$.

The above analysis shows that for $\varepsilon = \delta$, the ε-relaxed Stackelberg utility plus the ε-regularizer are equal to $(0.5 + 2\delta, 4\delta)$ for the leader and follower, respectively. For $\varepsilon \in (\delta, \gamma)$, the best response sets will not change, but the penalty for ε will increase, so these will not affect the infimum. Taking the minimum over the calculated benchmarks for $\varepsilon \in \{0, \gamma\}$ gives $0.5 + \delta, 4\delta$ for the leader and follower respectively.

J.1.2 Worked out version of Example 12 for self-γ-tolerant benchmark

We work out the self-γ-tolerant benchmark for Example 12 in more detail. Again, consider $\mathcal{I}$ in Table 14.2, which we also used to illustrate the γ-tolerant benchmark in Example 12. Recall that for $\varepsilon = 0$, we recover the Stackelberg equilibrium benchmark of $(0.5 + \delta, 0.1)$ for the leader and follower, respectively. For $\varepsilon \in (0, \delta)$ the $\mathcal{B}_\varepsilon(a), \mathcal{A}_\varepsilon$ sets don't change, but the penalty increases, so this is irrelevant for the infimum. Recall that from that analysis, we found that $\mathcal{B}_\delta(a_1) = \{b_1\}, \mathcal{B}_\delta(a_2) = \{a_1, a_2\}$, and $\mathcal{A}_\delta = \{a_1, a_2\}$. The self$-\gamma$-tolerance benchmark only requires each agent to compete with the *worst* element within the product set $\mathcal{A}_\delta \times \mathcal{B}_\delta(a)$ (if we consider the instance where $\varepsilon = \delta$).

For the given instance, this gives the benchmarks for the leader and follower of $0.4 + \delta$ and $2 \cdot \delta + \delta$, where we have added a δ regularizer penalty to both. Finally, we note that for $\varepsilon \in (\delta, 0.1)$, again the $\mathcal{B}_\varepsilon(a), \mathcal{A}_\varepsilon$ sets do not change but the penalty increases, so these are again irrelevant for the infimum. Taking the minimum of the benchmarks over $\varepsilon \in \{0, \delta\}$ gives $0.4 + \delta, 3\delta$ for the leader and follower respectively. Note that this differs from the γ-tolerant benchmark for the follower only by δ, but differs by 0.1 (a constant) for the leader.

J.1.3 Additional worked out example for the benchmark

Tables J.1 and J.2 contain worked examples of how the benchmarks are calculated for more complex examples.

J.1.4 Additional Notation and Auxiliary Lemmas

We introduce the following notation and auxiliary lemmas which will be convenient in our proofs.

Notation for Player Histories. First, we introduce the following notation for the player histories that will be convenient to use in algorithmic specifications and proofs.

In a *weakly decentralized Stackelberg game (WeakDSG)*, let the leader's history up to time step t be the set of arms that were pulled, as well as the reward for the leader at each time step:

$$H_{1,t} := \{(t', a_{t'}, b_{t'}, r_{1,t'}(a_{t'}, b_{t'})) \mid 1 \le t' < t\}.$$

In a *strongly decentralized Stackelberg game (StrongDSG)*, the leader cannot even observe the action chosen by the follower, but the follower's information remains unchanged. That is $H_{1,t} := \{(t', a_{t'}, r_{1,t'}(a_{t'}, b_{t'})) \mid 1 \le t' < t\}$.

Let the follower's history be

$$H_{2,t} := \{(t', a_{t'}, b_{t'}, r_{2,t'}(a_{t'}, b_{t'})) \mid 1 \le t' < t\}.$$

When the follower runs a separate algorithm on each choice of $a \in \mathcal{A}$ and does not share information across arms (e.g., in Proposition 117, Theorem 118, Chapter 128, and Chapter 126), then the follower's history for the arm $a \in \mathcal{A}$ is given by:

$$H_{2,t,a} := \{(n_{t'+1}(a), b_{t'}, r_{2,t'}(a_{t'}, b_{t'})) \mid 1 \le t' < t, a_{t'} = a\},$$

where $n_{t'+1}(a)$ is the number of times that arm a is pulled prior to the $(t'+1)$th time step.

Auxiliary lemma for regret analysis. Next, we introduce the following auxiliary lemma which will be useful in the regret analysis.

Lemma 269. *Let $\mathcal{C}$ be a finite set of arms and let $T \ge 1$ be a time horizon. Let $(c_1, \ldots, c_T) \in \mathcal{C}^T$ denote any history of arm pulls. Let $n_t(c) = \sum_{t=1}^{t-1} \mathbb{1}[c_{t'} = c]$ denote the number of times that c is pulled prior to time step t. Then it holds that:*

$$\sum_{c \in \mathcal{C}} \frac{1}{\sqrt{n_t(c)}} \le O\left(\sqrt{T \cdot |\mathcal{C}|}\right)$$

Proof. We observe that

$$\sum_{c \in \mathcal{C}} \frac{1}{\sqrt{n_t(c)}} = \sum_{c \in \mathcal{C}} \sum_{n=1}^{n_t(c)} \frac{1}{\sqrt{n}} \le_{(A)} \sum_{c \in \mathcal{C}} O\left(\sqrt{n_t(c)+1}\right) \le_{(B)} O\left(\sqrt{T \cdot |\mathcal{C}|}\right) \le,$$

where (A) follows from an integral bound and (B) follows from Jensen's inequality. □

J.2 Proofs of regret lower bounds

Our regret bounds analyze a centralized setting (Chapter J.2.1) and build on standard tools (Lattimore and Szepesvári, 2020) for regret lower bounds (Chapter J.2.2). We prove Proposition 123 in Chapter J.2.3, Theorem 116 in Chapter J.2.4, and Theorem 120 in Chapter J.2.5.

J.2.1 Centralized environment

When analyzing regret lower bounds, it is also convenient to consider a *centralized environment* where a single player controls the actions of both players and observes all past actions. While the centralized environment is not our primary focus, it can (informally speaking) be viewed as a limiting case of the decentralized setting with extremely sophisticated players who could communicate their strategies to each other. We define the history for the centralized environment to be:

$$H_t^{\mathrm{C}} = \{(t', a_{t'}, b_{t'}, r_{1,t'}(a_{t'}, b_{t'}), r_{2,t'}(a_{t'}, b_{t'})) \mid 1 \le t' \le t\}.$$

The centralized player chooses an algorithm `ALG` mapping a history to a joint distribution over pairs of actions.

We show that centralized algorithms are strictly more general than decentralized environments, in that any rewards realized in a decentralized environment can also be realized in a centralized environment.

Lemma 270. *Consider a StrongDSG or WeakDSG. Fix an instance $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ and time horizon T. For any pair of algorithms $\texttt{ALG}_1$ and $\texttt{ALG}_2$, there exists a centralized algorithm $\texttt{ALG}$ such that the leader rewards $(r_{1,1}(a_1, b_1), \ldots, r_{1,T}(a_T, b_T))$ are identically distributed for $\texttt{ALG}$ and $(\texttt{ALG}_1, \texttt{ALG}_2)$ and the follower rewards $(r_{2,1}(a_1, b_1), \ldots, r_{2,T}(a_T, b_T))$ are also identically distributed for $\texttt{ALG}$ and $(\texttt{ALG}_1, \texttt{ALG}_2)$.*

Lemma 270 follows immediately from designing $\texttt{ALG}$ to "simulate" histories for the leader and the follower (by projecting away the information unavailable to each player) and then to choose arms by applying $\texttt{ALG}_1$ and $\texttt{ALG}_2$ on these histories.

J.2.2 Useful lemmas

Our regret bounds leverage the following standard tools (Lattimore and Szepesvári, 2020) which we restate for completeness. Like in Lattimore and Szepesvári (2020), we will use the Bretagnolle–Huber inequality.

Theorem 271 (paraphrased from Theorem 14.2 in (Lattimore and Szepesvári, 2020))**.** *Let P and Q be probability measures on the same measurable space $(\Omega, \mathcal{F})$, and let $E \in \mathcal{F}$ be an arbitrary event. Then it holds that:*

$$P(G) + Q(G^c) \geq \frac{1}{2} e^{-KL(P,Q)}$$

where $G^c = \Omega \setminus G$ is the complement of G and $KL(P, Q)$ is the KL divergence between P and Q.

We similarly work with the canonical bandit model (Chapter 4.6 in Lattimore and Szepesvári (2020)) but with some modifications because there are two observed rewards (for the leader and the follower) in our setup. We call the analogous setup in our setting the *canonical centralized bandit model.* Note that the sample space of the probability space is now $(((\mathcal{A} \times \mathcal{B}) \times \mathbb{R} \times \mathbb{R})^T$ (instead of $([k] \times \mathbb{R})^T$, like in the typical canonical bandit model).

We show an analogous *divergence decomposition* (Lemma 15.1 in Lattimore and Szepesvári (2020)) applies to our setting. For this result, fix $\mathcal{A}$ and $\mathcal{B}$, and let v and $\tilde{v}$ be two different specifications of utilities. For $i \in \{1, 2\}$, let $r_i(a, b)$ denote the reward distribution $N(v_i(a, b), 1)$ and let $\tilde{r}_i(a, b)$ denote the reward distribution $N(\tilde{v}_i(a, b), 1)$.

Theorem 272 (adapted from Lemma 15.1 in (Lattimore and Szepesvári, 2020))**.** *Fix an algorithm $\texttt{ALG}$ for the centralized environment. Let P (resp. $\tilde{P}$) denote the probability measure corresponding to the canonical centralized bandit model for $\texttt{ALG}$ applied to $(\mathcal{A}, \mathcal{B}, v)$ (resp. $(\mathcal{A}, \mathcal{B}, \tilde{v})$). Let $n_T(a, b) = \sum_{t=1}^{T} \mathbb{1}[a_t = a, b_t = b]$ denote the number of times that arm (a, b) is pulled. Then it holds that:*

$$D(P, \tilde{P}) = \sum_{(a,b) \in \mathcal{A} \times \mathcal{B}} \mathbb{E}_P[n_T(a, b)] \cdot (D(r_1(a, b), \tilde{r}_1(a, b)) + D(R_2(a, b), \tilde{r}_2(a, b)).$$

where $D(\cdot,\cdot)$ denotes the KL divergence, where $r_i(a,b)$ denotes the reward distribution $N(v_i(a,b),1)$ and $\tilde{r}_i(a,b)$ denotes the reward distribution $N(\tilde{v}_i(a,b),1)$ for $i=1,2$.

Proof. This follows from the exact same argument as the proof in Lattimore and Szepesvári (2020), where X_t is interpreted as the *pair* of rewards $(r_{1,t}(a_t,b_t), r_{2,t}(a_t,b_t))$ (or $(\tilde{r}_{1,t}(a_t,b_t), \tilde{r}_{2,t}(a_t,b_t))$ observed at time step t. Let $r(a,b)$ be the product distribution $r_1(a,b)\times r_2(a,b)$, and let $\tilde{r}(a,b)$ be the product distribution $\tilde{r}_1(a,b)\times \tilde{r}_2(a,b)$. This yields:

$$D(P,\tilde{P}) = \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} \mathbb{E}_P[n_T(a,b)]\cdot D(r(a,b),\tilde{r}(a,b)).$$

The result follows from applying the "chain rule" which implies that the KL divergence of a product distribution is the sum of KL divergences of the individual distributions:

$$D(r(a,b),\tilde{r}(a,b)) = D(r_1(a,b),\tilde{r}_1(a,b)) + D(r_2(a,b),\tilde{r}_2(a,b).$$

□

Recall that we assume Gaussian noise, which further simplifies Theorem 272. By applying standard KL divergence bounds for univariate Gaussians, we obtain the following corollary of Theorem 272.

Corollary 273. *Fix an algorithm* $\texttt{ALG}$ *for the centralized environment. Let P (resp. $\tilde{P}$) denote the probability measure corresponding to the canonical centralized bandit model for* $\texttt{ALG}$ *applied to $(\mathcal{A},\mathcal{B},v)$ (resp. $(\mathcal{A},\mathcal{B},\tilde{v})$). Let $n_T(a,b)=\sum_{t=1}^T \mathbb{1}[a_t=a,b_t=b]$ denote the number of times that arm (a,b) is pulled. Then it holds that:*

$$D(P,\tilde{P}) = \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} \mathbb{E}_P[n_T(a,b)]\cdot \frac{(v_1(a,b)-\tilde{v}_1(a,b))^2+(v_2(a,b)-\tilde{v}_2(a,b))^2}{2}.$$

J.2.3 Proof for Proposition 123

We prove Proposition 123, restated below.

Proposition 123. *Consider StrongDSGs or WeakDSGs with actions sets $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}|\geq 2$ and $|\mathcal{B}|\geq 2$. For any algorithms* $\texttt{ALG}_1$ *and* $\texttt{ALG}_2$*, there exists an instance $\mathcal{I}^*=(\mathcal{A},\mathcal{B},v_1,v_2)$ such that at least one of the players incurs $\Omega(\sqrt{T\cdot(|\mathcal{A}|-1)\cdot|\mathcal{B}|})$ regret with respect to the self-γ-tolerant benchmarks $\beta_1^{\text{self-tol}}$ and $\beta_2^{\text{self-tol}}$, that is:* $\max(R_1(T;\mathcal{I}^*),R_2(T;\mathcal{I}^*)) = \Omega(\sqrt{T\cdot(|\mathcal{A}|-1)\cdot|\mathcal{B}|})$.

Proof of Proposition 123. Fix $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}|\geq 2$ and $|\mathcal{B}|\geq 1$.

It suffices to prove this lower bound in a *centralized* environment where a single learner can choose action pairs (a,b) and observes rewards for both players (Lemma 270). We define

	b_1	$\ldots$	b'	$\ldots$
a_1	(δ,δ)	(δ,δ)	(δ,δ)	(δ,δ)
$\vdots$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
a'	$(0,0)$	$(0,0)$	$*$	$(0,0)$
$\vdots$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$

Table J.3: Hard instance for Proposition 123, where $*$ is equal to $(0,0)$ for instance $\mathcal{I}_{a_1,b_1}$, and $(2\delta,2\delta)$ otherwise.

a family of instances in the centralized game and evaluate the self-tolerant benchmarks on this family of instances. Arbitrarily pick some $a_1 \in \mathcal{A}$ to be the "base" action. Let $\mathcal{F}_{\delta,\mathcal{A},\mathcal{B}}$ be the family of $(|\mathcal{A}|-1)\cdot|\mathcal{B}|+1$ instances of the form $(\mathcal{A},\mathcal{B},v_1,v_2)$ for varying settings of v_1 and v_2, where we index the instances by $(a',b') \in ((\mathcal{A}\setminus\{a_1\})\times\mathcal{B})\cup\{(a_1,b_1)\}$. The utility functions for the instance $\mathcal{I}_{(a',b')}$ are equal to the terms below (illustrated in Table J.3):

$$v_1(a,b) = v_2(a,b) = \begin{cases} \delta & \text{if } a = a_1 \\ 0 & \text{if } (a',b') \neq (a,b), a \neq a_1 \\ 2\delta & \text{if } (a',b') = (a,b), a \neq a_1 \end{cases}$$

We claim that the $\beta_1^{\text{self-tol}} = \beta_2^{\text{self-tol}} = \delta$ for the instance $\mathcal{I}_{(a_1,b_1)}$ and $\beta_1^{\text{self-tol}} = \beta_2^{\text{self-tol}} = 2\delta$ for the instances $\mathcal{I}_{(a',b')}$ where $(a',b') \neq (a_1,b_1)$. To see this, observe that on the instance $\mathcal{I}_{(a_1,b_1)}$, it holds that $\mathcal{B}_\varepsilon(a_1) = \mathcal{B}$ and $\mathcal{A}_\varepsilon = \{a_1\}$ if $\varepsilon < \delta$. Thus, it holds that $\min_{a\in\mathcal{A}_\varepsilon}\min_{b\in\mathcal{B}_\varepsilon(a)} v_1(a,b)+\varepsilon \geq \delta$ for all ε, so the benchmark is equal to

$$\beta_1^{\text{self-tol}} = \beta_2^{\text{self-tol}} = \delta,$$

as desired. On instances $\mathcal{I}_{(a',b')}$ where $(a',b') \neq (a_1,b_1)$, it holds that $\mathcal{B}_\varepsilon(a') = \{b'\}$ if $\varepsilon < 2\delta$ and $\mathcal{A}_\varepsilon = \{a'\}$ if $\varepsilon < \delta$. If $\varepsilon < \delta$ or $\varepsilon \geq 2\delta$, then $\min_{a\in\mathcal{A}_\varepsilon}\min_{b\in\mathcal{B}_\varepsilon(a)} v_1(a,b)+\varepsilon \geq 2\delta$. If $\delta \leq \varepsilon < 2\delta$, then $\mathcal{A}_\varepsilon = \{a',a_1\}$ and it also holds that $\min_{a\in\mathcal{A}_\varepsilon}\min_{b\in\mathcal{B}_\varepsilon(a)} v_1(a,b)+\varepsilon \geq 2\delta$. This means that the self-tolerant benchmarks are equal to:

$$\beta_1^{\text{self-tol}} = \beta_2^{\text{self-tol}} = 2\delta,$$

as desired.

Because the utilities in $\mathcal{F}_{\delta,\mathcal{A},\mathcal{B}}$ and the benchmarks are the same for the leader and follower, we see that the regret is also the same for both players. Thus, for the remainder of the analysis, we do not need to distinguish between the regret of the leader and the regret of the follower. Let $R(T;\mathcal{I})$ denote the regret incurred on instance $\mathcal{I}$. Since the benchmarks are equal to the maximum reward across all pairs of arms, the expected regret is always nonnegative.

Fix any ALG for the centralized environment. For each $(a,b) \in ((\mathcal{A} \setminus \{a_1\}) \times \mathcal{B}) \cup \{(a_1,b_1)\}$, let $P_{a,b}$ denote the probability measure over canonical centralized bandit model when ALG is applied to the instance $\mathcal{I}_{a,b}$ (see Chapter J.2.2). Let $n_T(a,b) = \sum_{t=1}^{T} \mathbb{1}[a_t = a, b_t = b]$ be the random variable denoting the number of times that (a,b) is pulled. We define:

$$(a_m, b_m) := \underset{(a,b)\in\mathcal{A}\times\mathcal{B}|a\neq a_1}{\arg\min} \mathbb{E}_{P_{(a_1,b_1)}}[n_T(a,b)]$$

to be the arm pulled the minimum number of times in expectation over $P_{(a_1,b_1)}$ (i.e., the expectation when ALG is applied to the instance $\mathcal{I}_{a_1,b_1}$). This means that

$$\mathbb{E}_{P_{(a_1,b_1)}}[n_T(a_m,b_m)] \leq \frac{T}{(|\mathcal{A}|-1)\cdot|\mathcal{B}|}.$$

We will construct δ such that the regret is high on at least one of the instances $\mathcal{I}_{(a_1,b_1)}$ and $\mathcal{I}_{(a_m,b_m)}$.

Now, let G denote the event that $\sum_{b\in\mathcal{B}} n_T(a_1,b) \leq T/2$ (i.e., the arm a_1 is pulled less than $T/2$ times). It is easy to see that the regret satisfies:

$$R(T;\mathcal{I}_{a_1,b_1}) \geq \frac{\delta\cdot T}{2}\cdot P_{a_1,b_1}[G]$$

$$R(T;\mathcal{I}_{a_m,b_m}) \geq \frac{\delta\cdot T}{2}\cdot P_{a_m,b_m}[G^c]$$

where G^c is the complement of G. We apply Theorem 271 to see that:

$$\begin{aligned}
R(T;\mathcal{I}_{a_1,b_1}) + R(T;\mathcal{I}_{a_m,b_m}) &= \frac{\delta\cdot T}{2}\left(P_{a_1,b_1}[G] + P_{a_m,b_m}[G^c]\right) \\
&\geq_{(1)} \frac{\delta\cdot T}{2}\cdot\frac{1}{2}\exp\left(-KL(P_{a_1,b_1},P_{a_m,b_m})\right) \\
&\geq_{(2)} \frac{\delta\cdot T}{2}\cdot\frac{1}{2}\exp\left(-\mathbb{E}_{P_{a_1,b_1}}[n_T(a_m,b_m)]\cdot(2\delta)^2\right) \\
&\geq_{(3)} \frac{\delta\cdot T}{4}\cdot\exp\left(-\frac{4\cdot\delta^2\cdot T}{(|\mathcal{A}-1)|\mathcal{B}|}\right).
\end{aligned}$$

where (1) applies Theorem 271 and (2) applies Corollary 273, and (3) applies the fact that $n_T(a_m,b_m) \leq \frac{T}{(|\mathcal{A}|-1)\cdot|\mathcal{B}|}$. If we set $\delta = \Theta(\sqrt{\frac{|\mathcal{A}-1)|\mathcal{B}|}{T}})$, then we obtain a bound of $\Theta(\sqrt{T\cdot(|\mathcal{A}|-1)\cdot|\mathcal{B}|})$. Since expected regret is nonnegative for these instances (see discussion above), this implies that either $R(T;\mathcal{I}_{a_1,b_1}) = \Omega(\sqrt{T\cdot(|\mathcal{A}|-1)\cdot|\mathcal{B}|})$ or $R(T;\mathcal{I}_{a_m,b_m}) = \Omega(\sqrt{T\cdot(|\mathcal{A}|-1)\cdot|\mathcal{B}|})$ as desired.

□

J.2.4 Proof of Theorem 116

Theorem 116. *Consider StrongDSGs or WeakDSGs. For any algorithms* $\mathtt{ALG}_1$ *and* $\mathtt{ALG}_2$*, there exists an instance* $\mathcal{I}^*$ *with* $|\mathcal{A}| = |\mathcal{B}| = 2$ *where at least one of the players incurs linear regret with respect to the Stackleberg benchmarks* β_1^{orig} *and* β_2^{orig}*. That is, it holds that* $\max(R_1(T;\mathcal{I}^*), R_2(T;\mathcal{I}^*)) = \Omega(T)$.

Proof. It suffices to prove this lower bound in a *centralized* environment where a single learner can choose action pairs (a, b) and observes rewards for both players (Lemma 270). We construct a pair of instances $\mathcal{I}$ and $\tilde{\mathcal{I}}$ such that at least one of the players incurs linear regret on at least one of the instances. In particular, we take $\mathcal{I}$ and $\tilde{\mathcal{I}}$ to be the instances depicted in Table 14.1 with $\delta = O(1/\sqrt{T})$ (reproduced here for convenience).

	b_1	b_2
a_1	$(0.6, \delta)$	$(0.2, \mathbf{0})$
a_2	$(0.5, 0.6)$	$(0.4, 0.4)$

(a) Mean rewards $(v_1(a, b), v_2(a, b))$ for $\mathcal{I}$

	b_1	b_2
a_1	$(0.6, \delta)$	$(0.2,\boldsymbol{2\delta})$
a_2	$(0.5, 0.6)$	$(0.4, 0.4)$

(b) Mean rewards $(\tilde{v}_1(a, b), \tilde{v}_2(a, b))$ for $\tilde{\mathcal{I}}$

We first compute the benchmarks on these two instances. On instance $\mathcal{I}$, it holds that $(a^*, b^*) = (a_1, b_1)$, $\beta_1^{\text{orig}} = 0.6$ and $\beta_2^{\text{orig}} = \delta \geq 0$. On the other hand, on instance $\tilde{\mathcal{I}}$, it holds that $(a^*, b^*) = (a_2, b_1)$, $\beta_1^{\text{orig}} = 0.5$, and $\beta_2^{\text{orig}} = 0.6$. It is easy to see that $R_1(T;\mathcal{I})$ and $R_2(T;\tilde{\mathcal{I}})$ are always *nonnegative*.

Fix any $\mathtt{ALG}$ for the centralized environment. Let P (resp. $\tilde{P}$) denote the probability measure over canonical centralized bandit model when $\mathtt{ALG}$ is applied to the instance $\mathcal{I}$ (resp. $\tilde{\mathcal{I}}$) (see Chapter J.2.2). We will show that the regret is high on at least one of the instances $\mathcal{I}$ and $\tilde{\mathcal{I}}$.

Now let $n_T(a, b) = \sum_{t=1}^{T} \mathbb{1}[a_t = a, b_t = b]$ be the random variable denoting the number of times that (a, b) is pulled, and let G denote the event that $n_T(a_1, b_1) \leq T/2$ (i.e., the arm (a_1, b_1) is pulled less than $T/2$ times). It is easy to see that the regret satisfies:

$$R_1(T;\mathcal{I}) \geq \frac{0.1 \cdot T}{2} \cdot P[G]$$

$$R_2(T;\tilde{\mathcal{I}}) \geq \frac{(0.6 - \delta) \cdot T}{2} \cdot \tilde{P}[G^c]$$

	b_1	$\ldots$	b'	$\ldots$
a_1	$(0.5, 3 \cdot \delta)$	$(0.5, 3 \cdot \delta)$	$(0.5, 3 \cdot \delta)$	$(0.5, 3 \cdot \delta)$
$\vdots$	$(0.5 + \delta, \delta)$	$(0, 0)$	$*$	$(0, 0)$
$\vdots$	$(0.5 + \delta, \delta)$	$(0, 0)$	$*$	$(0, 0)$
$\vdots$	$(0.5 + \delta, \delta)$	$(0, 0)$	$*$	$(0, 0)$

Table J.5: Hard instance for Theorem 120, where $*$ is equal to $(0, 0)$ for instance $\mathcal{I}_{a_1,b_1}$, and $(0, 2\delta)$ otherwise. Note that this example is structurally similar to the illustrative example in Table 14.3, but with $|\mathcal{A}|, |\mathcal{B}| \geq 2$.

where G^c is the complement of G. We apply Theorem 271 to see that:

$$\begin{aligned}
R_1(T;\mathcal{I}) + R_2(T;\tilde{\mathcal{I}}) &= \frac{0.1 \cdot T}{2} \cdot P[G] + \frac{(0.6-\delta) \cdot T}{2} \cdot \tilde{P}[G^c] \\
&\geq \frac{0.1 \cdot T}{2} \cdot \left(P[G] + \tilde{P}[G^c]\right) \\
&\geq_{(1)} \frac{0.1 \cdot T}{2} \cdot \frac{1}{2}\exp\left(-KL(P, \tilde{P})\right) \\
&\geq_{(2)} \frac{0.1 \cdot T}{2} \cdot \frac{1}{2}\exp\left(-\mathbb{E}_P[n_T(a_1, b_2)] \cdot \frac{(2 \cdot \delta)^2}{2}\right) \\
&\geq_{(3)} \frac{0.1 \cdot T}{4} \cdot \exp\left(-2 \cdot \delta^2 \cdot T\right).
\end{aligned}$$

where (1) applies Theorem 271 and (2) applies Corollary 273, and (3) uses the fact that $n_T(a_1, b_2) \leq T$. If we take $\delta = O(T^{-1/2})$, then we obtain a bound of $\Omega(T)$. Since these expected regrets are always nonnegative (see discussion above), this implies that either $R_1(T;\mathcal{I}) = \Omega(T)$ or $R_2(T;\tilde{\mathcal{I}}) = \Omega(T)$ as desired. □

J.2.5 Proof of Theorem 120

Theorem 120. *Consider StrongDSGs or WeakDSGs with action sets $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}| \geq 2$ and $|\mathcal{B}| \geq 2$. For any algorithms $\texttt{ALG}_1$ and $\texttt{ALG}_2$, there exists an instance $\mathcal{I}^* = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ such that at least one of the players incurs $\Omega(T^{2/3} \cdot (|\mathcal{B}|)^{1/3})$ regret with respect to the γ-tolerant benchmarks β_1^{tol} and β_2^{tol}:*

$$\max(R_1(T;\mathcal{I}^*), R_2(T;\mathcal{I}^*)) = \Omega(T^{2/3} \cdot (|\mathcal{B}|)^{1/3}).$$

Proof. Fix $\mathcal{A}$ and $\mathcal{B}$ such that $|\mathcal{A}| \geq 2$ and $|\mathcal{B}| \geq 2$.

It suffices to prove this lower bound in a *centralized* environment where a single learner can choose action pairs (a, b) and observes rewards for both players (Lemma 270). We define a family of instances in the centralized game and evaluate the self-tolerant benchmarks on this family of instances. Arbitrarily pick some $(a_1, b_1) \in \mathcal{A} \times \mathcal{B}$ to be the "base" action. Let $\mathcal{F}_{\delta,\mathcal{A},\mathcal{B}}$ be the family of $|\mathcal{B}|$ instances of the form $(\mathcal{A}, \mathcal{B}, v_1, v_2)$ for varying settings of v, where we index the instances by $\mathcal{B}$. The utility functions for the instance $\mathcal{I}_{b'}$ are equal to terms below (illustrated in Table J.5):

$$v_1(a,b) = \begin{cases} 0.5 & \text{if } a = a_1 \\ 0.5 + \delta & \text{if } a \neq a_1, b = b_1 \\ 0 & \text{if } a \neq a_1, b \neq b_1. \end{cases}$$

$$v_2(a,b) = \begin{cases} 3\delta & \text{if } a = a_1 \\ \delta & \text{if } a \neq a_1, b = b_1 \\ 2\delta & \text{if } b = b', a \neq a_1, b \neq b_1 \\ 0 & \text{if } b \neq b', a \neq a_1, b \neq b_1 \end{cases}$$

We claim that the $\beta_1^{\text{tol}} = 0.5 + \delta$ and $\beta_2^{\text{tol}} = \delta$ for the instance $\mathcal{I}_{(a_1,b_1)}$ and $\beta_1^{\text{tol}} = 0.5$ and $\beta_2^{\text{tol}} = 3\delta$ for the instances $\mathcal{I}_{(a',b')}$ where $(a', b') \neq (a_1, b_1)$.

- *Instance $\mathcal{I}_{b_1}$:* If $\varepsilon < \delta$, it holds that $\mathcal{B}_\varepsilon(a) = \{b_1\}$ for $a \neq a_1$ and $\mathcal{A}_\varepsilon = \mathcal{A} \setminus \{a_1\}$. If $\varepsilon \geq \delta$, then it holds that $\mathcal{B}_\varepsilon(a) = \mathcal{B}$ and $\mathcal{A}_\varepsilon = \mathcal{A}$. Altogether, this means that $\beta_1^{\text{tol}} = 0.5 + \delta$ and $\beta_2^{\text{tol}} = \delta$.

- *Instances $\mathcal{I}_{b'}$ where $b' \neq b_1$:* If $\varepsilon < \delta$, it holds that $\mathcal{B}_\varepsilon(a) = \{b'\}$ for $a \neq a_1$ and $\mathcal{A}_\varepsilon = \{a_1\}$. If $\delta \leq \varepsilon < 2\delta$, then it holds that $\mathcal{B}_\varepsilon(a_1) = \mathcal{B}$ and $\mathcal{B}_\varepsilon(a) = \{b', b_1\}$ for $a \neq a_1$, and $\mathcal{A}_\varepsilon = \mathcal{A}$. If $\varepsilon \geq 2\delta$, then it holds that $\mathcal{B}_\varepsilon(a) = \mathcal{B}$ and $\mathcal{A}_\varepsilon = \mathcal{A}$. Altogether, this means that $\beta_1^{\text{tol}} = 0.5$ and $\beta_2^{\text{tol}} = 3\delta$.

It is easy to see that the regret $R_1(T; \mathcal{I}_{b_1})$ and the regret $R_2(T; \mathcal{I}_b)$ for $b \neq b_1$ are always nonnegative.

Fix an `ALG` be an algorithm for the centralized environment. For each $b \in \mathcal{B}$, let P_b denote the probability measure over canonical centralized bandit model when `ALG` is applied to the instance $\mathcal{I}_b$ (see Chapter J.2.2). Let $n_T(a,b) = \sum_{t=1}^{T} \mathbb{1}[a_t = a, b_t = b]$ be the random variable denoting the number of times that (a, b) is pulled. We define:

$$b_m := \operatorname*{arg\,min}_{b \in \mathcal{B} | b \neq b_1} \mathbb{E}_{P_{b_1}} \left[\sum_{a \neq a_1} n_T(a,b) \right]$$

to be the arm b such that the set of arms (a', b) for $a' \neq a_1$ is pulled the minimum number of times in expectation over P_{b_1} (i.e., the expectation when `ALG` is applied to the instance $\mathcal{I}_{b_1}$). This means that

$$\sum_{b \neq b_1} \sum_{a \neq a_1} \mathbb{E}_{P_{b_1}}[n_T(a,b)] \geq (|\mathcal{B}| - 1) \sum_{a \neq a_1} \mathbb{E}_{P_{b_1}}[n_T(a, b_m)].$$

We will construct δ such that the regret is high on at least one of the instances $\mathcal{I}_{b_1}$ and $\mathcal{I}_{b_m}$.

Now, let G denote the event that $\sum_{a \neq a_1} n_T(a, b_1) \leq T/2$ (i.e., arms of the form (a', b_1) for $a' \neq a$ are pulled less than $T/2$ times). It is easy to see that the regret satisfies:

$$R_1(T; \mathcal{I}_{b_1}) \geq \frac{\delta \cdot T}{2} \cdot P_{b_1}[E]$$

$$R_2(T; \mathcal{I}_{b_m}) \geq \frac{2 \cdot \delta \cdot T}{2} \cdot P_{b_m}[E^c]$$

$$R_1(T; \mathcal{I}_{b_1}) \geq (0.5 + \delta) \cdot \mathbb{E}\left[\sum_{a \neq a_1, b \neq b_1} n_T(a, b)\right] \geq 0.5 \cdot \mathbb{E}\left[\sum_{a \neq a_1, b \neq b_1} n_T(a, b)\right].$$

where G^c is the complement of G. We apply Theorem 271 to see that:

$$\begin{aligned}
&2 \cdot R_1(T; \mathcal{I}_{b_1}) + R_2(T; \mathcal{I}_{b_m}) \\
&= \frac{\delta \cdot T}{2} \cdot P_{b_1}[E] + \frac{2 \cdot \delta \cdot T}{2} \cdot P_{b_m}[E^c] + 0.5 \cdot \mathbb{E}\left[\sum_{a \neq a_1, b \neq b_1} n_T(a, b)\right] \\
&\geq \frac{\delta \cdot T}{2} \cdot (P_{b_1}[E] + P_{b_m}[E^c]) + 0.5 \cdot \mathbb{E}\left[\sum_{a \neq a_1, b \neq b_1} n_T(a, b)\right] \\
&\geq_{(1)} \frac{\delta \cdot T}{2} \exp\left(-KL(P_{b_1}, P_{b_m})\right) + +0.5 \cdot \mathbb{E}\left[\sum_{a \neq a_1, b \neq b_1} n_T(a, b)\right] \\
&\geq_{(2)} \frac{\delta \cdot T}{2} \cdot \frac{1}{2} \exp\left(-\mathbb{E}_{P_{b_1}}\left[\sum_{a \neq a_1} n_T(a, b_m)\right] \cdot \frac{(2\delta)^2}{2}\right) + 0.5 \cdot \mathbb{E}\left[\sum_{a \neq a_1, b \neq b_1} n_T(a, b)\right] \\
&\geq_{(3)} \frac{\delta \cdot T}{2} \cdot \frac{1}{2} \exp\left(-\mathbb{E}_{P_{b_1}}\left[\sum_{a \neq a_1} n_T(a, b_m)\right] \cdot \frac{(2\delta)^2}{2}\right) + 0.5(|\mathcal{B}| - 1) \cdot \mathbb{E}\left[\sum_{b \neq b_1} n_T(a, b_m)\right]
\end{aligned}$$

where (1) applies Theorem 271 and (2) applies Corollary 273 and where (3) uses the fact that $\sum_{b \neq b_1} \sum_{a \neq a_1} \mathbb{E}_P[n_T(a, b)] \geq (|\mathcal{B}| - 1) \sum_{a \neq a_1} \mathbb{E}_P[n_T(a, b_m)]$.

We claim that the expression is $\Omega(T^{2/3}(|\mathcal{B}| - 1)^{1/3})$. We split into two cases based on the value of $\mathbb{E}\left[\sum_{a \neq a_1} n_T(a, b_m)\right]$: $\mathbb{E}\left[\sum_{a \neq a_1} n_T(a, b_m)\right] \geq \Theta(T^{2/3}(|\mathcal{B}| - 1)^{-2/3})$ and $\mathbb{E}\left[\sum_{a \neq a_1} n_T(a, b_m)\right] \leq \Theta(T^{2/3}(|\mathcal{B}| - 1)^{-2/3})$.

1. *Case 1:* $\mathbb{E}\left[\sum_{a \neq a_1} n_T(a, b_m)\right] \geq \Theta(T^{2/3}(|\mathcal{B}| - 1)^{-2/3})$. In this case, we see that $0.5(|\mathcal{B}| - 1) \cdot \mathbb{E}\left[\sum_{b \neq b_1} n_T(a, b_m)\right] = \Omega(T^{2/3}(|\mathcal{B}| - 1)^{1/3})$.

2. *Case 2:* $\mathbb{E}\left[\sum_{a \neq a_1} n_T(a, b_m)\right] \leq \Theta(T^{2/3}(|\mathcal{B}|-1)^{-2/3})$. In this case, we can write:

$$\frac{\delta \cdot T}{2} \cdot \frac{1}{2} \exp\left(-\mathbb{E}_{P_{b_1}}\left[\sum_{a \neq a_1} n_T(a, b_m)\right] \cdot \frac{(2\delta)^2}{2}\right) \geq \frac{\delta \cdot T}{2} \cdot \frac{1}{2} \exp\left(-\Theta\left(T^{2/3}(|\mathcal{B}|-1)^{-2/3} \cdot \delta^2\right)\right).$$

In this case, we set $\delta = \Theta(T^{-1/3}(|\mathcal{B}|-1)^{1/3})$ and the expression becomes $\Omega(T^{2/3}(|\mathcal{B}|-1)^{1/3})$.

This proves that $2 \cdot R_1(T; \mathcal{I}_{b_1}) + R_2(T; \mathcal{I}_{b_m}) = \Omega(T^{2/3}(|\mathcal{B}|-1)^{1/3})$.

Since expected regret is nonnegative for these instances (see discussion above), this implies that either $R_1(T; \mathcal{I}_{b_1}) = \Omega(T^{2/3}(|\mathcal{B}|-1)^{1/3})$ or $R_2(T; \mathcal{I}_{b_m}) = \Omega(T^{2/3}(|\mathcal{B}|-1)^{1/3})$ as desired. □

J.3 Proofs for Chapter 14.4

J.3.1 Proof of Proposition 117

We prove Proposition 117.

Proposition 117. *Consider StrongDSGs where the follower runs a separate instantiation of* $\texttt{ExploreThenCommit}(E, \mathcal{B})$ *for every* $a \in \mathcal{A}$*. Moreover, suppose that the leader runs* $\texttt{ExploreThenCommit}(E' \cdot |\mathcal{B}|, \mathcal{A})$ *for any* $E' \leq E$ *(i.e., the leader's exploration phase ends before the follower's exploration phase). Then, there exists an instance* $\mathcal{I}^*$ *such that both players incur linear regret with respect to the* γ*-tolerant benchmarks* β_1^{tol} *and* β_2^{tol}*: that is,* $\min(R_1(T; \mathcal{I}^*), R_2(T; \mathcal{I}^*)) = \Omega(T)$.

This proof holds for $\gamma < 0.1$ (the construction can be generalized to other constant γ by adjusting the values of the mean rewards; we present this construction which builds on Table 14.2).

	b_1	b_2
a_1	$(0.6, 0.4)$	$(0.2, 0)$
a_2	$(0.5, 0.3)$	$(0.4, 0.2)$

Table J.6: A single instance, illustrating the γ-tolerant benchmark - variant of Table 14.2 with $\delta = 0.1$

Proof. We take $\mathcal{I}^*$ to be the instance $\mathcal{I}$ in Table J.6 (equivalent to Table 14.2 with $\delta = 0.1$).

The fact that $E' < E$ means that the leader's exploration phase takes place entirely during the follower's exploration phase. Moreover, since the leader's exploration parameter $E' \cdot |\mathcal{B}|$ is divisible by $|\mathcal{B}|$, for every arm $a \in \mathcal{A}$, the follower pulls every arm $b \in \mathcal{B}$ an equal number of times. Given that follower explores evenly between the two arms b_1 and b_2, the leader's expected average reward $\mathbb{E}[\hat{v}_1(a_1)]$ from a_1 during the first $E' \cdot |\mathcal{B}|$ rounds is given by $(0.6 + 0.2)/2 = 0.4$ and the leader's expected average reward $\mathbb{E}[\hat{v}_1(a_2)]$ average reward from a_2 is given by $(0.5 + 0.4)/2 = 0.45$.

The proofs boils down to analyzing the relationship between the distributions $\hat{v}_1(a_1)$ and $\hat{v}_1(a_2)$. Note that we allow E, E' to be arbitrary, so we cannot use standard concentration bounds. Instead, we leverage the symmetry of the distribution of the empirical mean $\hat{v}_1(a_1)$ (this follows from the fact that $\hat{v}_1(a_1) - \mathbb{E}[\hat{v}_1(a_1)]$ is distributed as a Gaussian). This means that:

$$\mathbb{P}[\hat{v}_1(a_1) > 0.4] = \mathbb{P}[\hat{v}_1(a_1) < 0.4] = 0.5.$$

(The probability $\mathbb{P}[\hat{v}_1(a_1) = \mathbb{E}[\hat{v}_1(a_1)] = 0.4]$ is equal to 0.) Similarly, we see that:

$$\mathbb{P}[\hat{v}_1(a_2) > 0.45] = \mathbb{P}[\hat{v}_1(a_2) < 0.45] = 0.5.$$

Because the stochastic rewards have independent randomness, we know that with probability at least 0.25 we have $\hat{v}_1(a_1) < 0.4$ and $\hat{v}_1(a_2) > 0.45$. When this occurs, the leader commits to pulling arm a_2.

Regardless of the follower's choice of action (b_1 or b_2) in the commit phase, this means that the follower obtains reward at most 0.3 and the leader obtains reward at most 0.5. However, recall that we found that the γ-tolerant benchmark (for $\gamma = 0.1$) are $\beta_1^{\text{tol}} = 0.6$ and $\beta_2^{\text{tol}} = 0.4$. This leads to linear regret (at least $0.25 \cdot 0.1 \cdot T$) for both players, even with respect to the γ-tolerant benchmark. $\square$

J.3.2 Proof of Theorem 118

Theorem 118. *Consider a StrongDSG where the follower runs a separate instantiation of* `ExploreThenCommit`$(E_2, \mathcal{B})$ *for every $a \in \mathcal{A}$, and where the leader runs* `ExploreThenCommitThrowOut`$(E_1, E_2 \cdot |\mathcal{B}|, \mathcal{A})$. *If $E_2 = \Theta(|\mathcal{A}|^{-2/3}|\mathcal{B}|^{-2/3} \cdot (\log T)^{1/3} T^{2/3})$, and $E_1 = \Theta(|\mathcal{A}|^{-2/3} \cdot (\log T)^{1/3} T^{2/3})$, then, the regret with respect to the γ-tolerant benchmarks is bounded as:*

$$\max(R_1(T), R_2(T)) = O\left(|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3} T^{2/3}\right).$$

In this theorem, we will assume $\gamma = \omega\left(T^{-1/3}\,|\mathcal{A}|^{1/3}\,|\mathcal{B}|^{1/3} \cdot (\log(T)^{1/3})\right)$ (see Chapter 14.6.1 for a discussion of γ) .

Notation. We will use the following notation in the proof. For $a \in \mathcal{A}$ and $b \in \mathcal{B}$, let $\hat{v}_2(a, b)$ denote the empirical mean of observations that the follower has seen for arm a during the first $E_2 \cdot |\mathcal{A}| \cdot |\mathcal{B}|$ time steps. For $a \in \mathcal{A}$, let $\hat{v}_1(a)$ denote the empirical mean of observations that the leader has seen for arm a during the first time steps $t \in [E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + 1, E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|]$. We denote by $\tilde{b}(a) = \arg\max_{b \in \mathcal{B}} \hat{v}_2(a, b)$ the arm that follower has committed to for rounds $t > E_2 \cdot |\mathcal{A}| \cdot |\mathcal{B}|$ onwards. We denote by $\tilde{a} = \arg\max_{a \in \mathcal{A}} \hat{v}_1(a)$ the arm that the leader has committed to for rounds $t > E_1 \cdot |\mathcal{A}|$.

Clean event. We define the clean event $G := G_L \cap G_F$ to be the intersection of a clean event G_L for the leader and a clean event G_F for the follower. Informally speaking, the clean event for the leader is the event that for all arms, the empirical mean reward $\hat{v}_1(a)$ is close to the true reward $v_1(a, \tilde{b}(a))$. The event G_L is formalized as follows:

$$\forall a \in \mathcal{A} : |\hat{v}_1(a) - v_1(a, \tilde{b}(a))| \leq \frac{10\sqrt{\log T}}{\sqrt{E_1}}.$$

Similarly, informally speaking, the clean event for the follower is the event that for all arms, the empirical mean reward $\hat{v}_2(a, b)$ is close to the true reward $v_2(a, b)$. The event G_F is formalized as follows:

$$\forall a \in \mathcal{A}, b \in \mathcal{B} : |\hat{v}_2(a, b) - v_2(a, b)| \leq \frac{10\sqrt{\log T}}{\sqrt{E_2}}.$$

We prove that the clean event occurs with high probability.

Lemma 274. *Assume the notation above. Let the follower run a separate instantiation of* `ExploreThenCommit`$(E_2, \mathcal{B})$ *for every $a \in \mathcal{A}$, and let the leader run* `ExploreThenCommitThrowOut`$(E_1, E_2 \cdot |\mathcal{B}|, \mathcal{A})$*. Then the clean event occurs with probability* $\mathbb{P}[G] \geq 1 - (|\mathcal{A}| \cdot |\mathcal{B}| + |\mathcal{A}|)T^{-3}$.

Proof. First, we consider the follower's clean event G_F. For each $a \in \mathcal{A}, b \in \mathcal{B}$, the follower has seen E_2 samples, so by a Chernoff bound, we have that

$$P\left[|\hat{v}_2(a, b) - v_2(a, b)| \geq \frac{10\sqrt{\log T}}{\sqrt{E_2}}\right] \leq T^{-3}.$$

We union bound over $a \in \mathcal{A}$ and $b \in \mathcal{B}$.

Next, we consider the leader's clean event G_L. Note that $\hat{v}_1(a)$ estimate is derived from rewards only after the follower has committed to a best response, so it is drawn from a distribution centered at $v_1(a, \tilde{b}(a))$, with E_1 samples. Again by applying a Chernoff bound, we see that

$$P\left[|\hat{v}_1(a) - v_1(a, \tilde{b}(a))| \geq \frac{10\sqrt{\log T}}{\sqrt{E_1}}\right] \leq T^{-3}.$$

We union bound over $a \in \mathcal{A}$.

Finally, we apply another union bound which leads $\mathbb{P}[G] \geq 1 - (|\mathcal{A}| \cdot |\mathcal{B}| + |\mathcal{A}|) \cdot T^{-3})$. $\square$

We also prove the following lower bounds on the leader's utility and follower's utility from the actions $\tilde{a}$ and $\tilde{b}(\tilde{a})$ that they commit to.

Lemma 275. *Assume the notation above. Let the follower run a separate instantiation of* `ExploreThenCommit`$(E_2, \mathcal{B})$ *for every $a \in \mathcal{A}$, and let the leader run*

`ExploreThenCommitThrowOut`$(E_1, E_2 \cdot |\mathcal{B}|, \mathcal{A})$. *Suppose that the clean event G holds. Then, for some $\varepsilon^* = \Theta\left(\max\left(\frac{\sqrt{\log T}}{\sqrt{E_1}}, \frac{\sqrt{\log T}}{\sqrt{E_2}}\right)\right)$, it holds that:*

$$v_1(\tilde{a}, \tilde{b}(\tilde{a})) \geq \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) - \varepsilon^*$$

and that

$$v_2(\tilde{a}, \tilde{b}(\tilde{a})) \geq \min_{a \in \mathcal{A}_{\varepsilon^*}} \max_{b \in \mathcal{B}} v_2(a, b) - \varepsilon^*.$$

Proof of Lemma 275. We assume that the clean event G holds. We take $\varepsilon^* = \Theta\left((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1/3} \cdot T^{-1/3}\right)$ with sufficiently high implicit constant.

First, we show that the follower chooses a near-optimal action for every $a \in \mathcal{A}$: that is, $v_2(a, \tilde{b}(a)) \geq \max_{b \in \mathcal{B}} v_2(a, b) - \varepsilon^*$. Since G_F holds, for every $a \in \mathcal{A}$ and $b \in \mathcal{B}$, we know that $|\hat{v}_2(a, b) - v_2(a, b)| \leq \frac{10\sqrt{\log T}}{\sqrt{E_2}}$. Based on our setting of E_2 and because $\tilde{b}(a) = \arg\max_{b \in \mathcal{B}} \hat{v}^2_{a,b}$, it holds that:

$$v_2(a, \tilde{b}(a)) \geq \left(\max_{b \in \mathcal{B}} v_2(a, b)\right) - \frac{20\sqrt{\log T}}{\sqrt{E_2}} \geq \left(\max_{b \in \mathcal{B}} v_2(a, b)\right) - \varepsilon^*,$$

as desired.

Next, we show that the leader chooses a near-optimal action: that is, $v_1(\tilde{a}, \tilde{b}(\tilde{a})) \geq \max_{a \in \mathcal{A}} v_1(a, \tilde{b}(a)) - \varepsilon^*$. Since G_L holds, we know that $|\hat{v}_1(a) - v_1(a, \tilde{b}(a))| \leq \frac{10\sqrt{\log T}}{\sqrt{E_1}}$. Based on our setting of E_2 and because $\tilde{a} = \arg\max_{a \in \mathcal{A}} \hat{v}_1(a)$, it holds that:

$$v_1(\tilde{a}, \tilde{b}(\tilde{a})) \geq \left(\max_{a \in \mathcal{A}} v_1(a, \tilde{b}(a))\right) - \frac{20\sqrt{\log T}}{\sqrt{E_1}} \geq \left(\max_{a \in \mathcal{A}} v_1(a, \tilde{b}(a))\right) - \varepsilon^*.$$

as desired.

To bound the leader's utility, observe that $v_2(a, \tilde{b}(a)) \geq \max_{b \in \mathcal{B}} v_2(a, b) - \varepsilon^*$ implies that $b \in \mathcal{B}_{\varepsilon^*}(a)$. This, coupled with the other bound, means that:

$$v_1(\tilde{a}, \tilde{b}(\tilde{a})) \geq \max_{a \in \mathcal{A}} v_1(a, \tilde{b}(a)) - \varepsilon^* \geq \left(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a, b)\right) - \varepsilon^*.$$

To bound the follower's utility, observe that $v_1(\tilde{a}, \tilde{b}(\tilde{a})) \geq \max_{a \in \mathcal{A}} v_1(a, \tilde{b}(a)) - \varepsilon^*$ and $v_2(a, \tilde{b}(a)) \geq \max_{b \in \mathcal{B}} v_2(a, b) - \varepsilon^*$ together imply that

$$\max_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(\tilde{a}, b) \geq v_1(\tilde{a}, \tilde{b}(\tilde{a})) \geq \max_{a \in \mathcal{A}} v_1(a, \tilde{b}(a)) - \varepsilon^* \geq \left(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a, b)\right) - \varepsilon^*,$$

which implies that $a \in \mathcal{A}_{\varepsilon^*}$. This means that

$$v_2(\tilde{a}, \tilde{b}(\tilde{a})) \geq \left(\max_{b \in \mathcal{B}} v_2(\tilde{a}, b)\right) - \varepsilon^* \geq \min_{a \in \mathcal{A}_{\varepsilon^*}} \max_{b \in \mathcal{B}} v_2(a, b) - \varepsilon^*.$$

$\square$

We now prove Theorem 118.

Proof of Theorem 118. Assume that the clean event G holds. This occurs with probability at least $1 - (|\mathcal{A}| \cdot |\mathcal{B}| + |\mathcal{A}|)T^{-3}$ (Lemma 274), so the clean event not occuring counts negligibly towards regret.

First, we consider the first $E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|$ time steps. Each time step results in $O(1)$ regret for both players. Based on the settings of E_1 and E_2, these phases contribute a regret of:

$$E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}| = O\left(|\mathcal{A}|^{1/3} \cdot |\mathcal{B}|^{1/3} \cdot (\log T)^{1/3} \cdot T^{2/3}\right).$$

We focus on $t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|$ for the remainder of the analysis. Our main ingredient is Lemma 275. Note that $\varepsilon^* = \Theta\left(\max\left(\frac{\sqrt{\log T}}{\sqrt{E_1}}, \frac{\sqrt{\log T}}{\sqrt{E_2}}\right)\right) = \Theta\left((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1/3} \cdot T^{-1/3}\right)$ based on the settings of E_1 and E_2. The regret of the leader can be bounded as:

$$\begin{aligned}
&\beta_1^{\text{tol}} \cdot (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_1(a_t, b_t) \\
&\leq (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot \left(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) + \varepsilon^*\right) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_1(\tilde{a}, \tilde{b}(\tilde{a})) \\
&= (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot \varepsilon^* + (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \left(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) - v_1(\tilde{a}, \tilde{b}(\tilde{a}))\right) \\
&\leq_{(A)} 2 \cdot T \cdot \varepsilon^* \\
&\leq O\left(T^{2/3} (\log T)^{1/3} |\mathcal{A}|^{1/3} |\mathcal{B}|^{1/3}\right).
\end{aligned}$$

where (A) follows from Lemma 275. The regret of the follower can similarly be bounded as:

$$\begin{aligned}
&\beta_1^{\text{tol}} \cdot (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_2(a_t, b_t) \\
&\leq (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot \left(\min_{a \in \mathcal{A}_{\varepsilon^*}} \max_{b \in \mathcal{B}} v_2(a, b) + \varepsilon^*\right) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_2(\tilde{a}, \tilde{b}(\tilde{a})) \\
&= (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot \varepsilon^* + (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \left(\min_{a \in \mathcal{A}_{\varepsilon^*}} \max_{b \in \mathcal{B}} v_2(a, b) - v_2(\tilde{a}, \tilde{b}(\tilde{a}))\right) \\
&\leq_{(B)} 2 \cdot T \cdot \varepsilon^* \\
&\leq O\left(T^{2/3} (\log T)^{1/3} |\mathcal{A}|^{1/3} |\mathcal{B}|^{1/3}\right).
\end{aligned}$$

where (B) follows from Lemma 275. This proves the desired result.

□

J.3.3 Proof of Theorem 119

Theorem 119. *Let $E = \Theta(|\mathcal{A}|^{-2/3}(|\mathcal{B}| \log T)^{1/3} T^{2/3})$. Consider a StrongDSG, where $\texttt{ALG}_2$ is any algorithm with high-probability instantaneous regret $g(t, T, \mathcal{B}) = O\left((|\mathcal{A}||\mathcal{B}| \log T)^{1/3} T^{-1/3}\right)$ for $t > E$ and $g(t, T, \mathcal{B}) = 1$ for $t \leq E$, and where $\texttt{ALG}_1 = \texttt{ExploreThenUCB}(E)$. Then, it holds that the regret with respect to the γ-tolerant benchmarks β_1^{tol} and β_2^{tol} is bounded as:*

$$\max(R_1(T), R_2(T)) = O\left(|\mathcal{A}|^{1/3} |\mathcal{B}|^{1/3} (\log T)^{1/3} T^{2/3}\right).$$

We assume $\gamma = \omega\left(\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3}T^{-1/3}\right)$.

Notation. We will use the following notation in the proof. Let $\varepsilon^* = \max_{t>E} g(t, T, \mathcal{B})$. Let $\tilde{a} = \arg\max_{a\in\mathcal{A}} \min_{\mathcal{B}_{\varepsilon^*}(a)} v_1(a, b)$ be the optimal action for the leader if the follower can worst-case ε^*-best-respond to any action. Let $\hat{v}_{1,t}(a)$ be the empirical mean specified in `ExplorethenUCB` at the beginning of time step t: this is the empirical mean of all observations that the leader has seen for arm a prior to time step t during the UCB phase (i.e., after time step $E \cdot |\mathcal{A}| + 1$ and prior to time step t). Moreover, for each arm $a \in \mathcal{A}$, let $S(a) = \{t > E \cdot |\mathcal{A}| \mid a_t = a\}$ be the set of time steps where arm a is pulled during the UCB phase, and let $n_{E\cdot|\mathcal{A}|,t}(a) = |\{E \cdot |\mathcal{A}| < t' < t \mid a_{t'} = a\}|$ be the number of times that a is pulled during the UCB phase prior to time step t'.

Clean event. We define the clean event $G := G_L \cap G_F$ to be the intersection of a clean event G_L for the leader and a clean event G_F for the follower. Informally speaking, the clean event for the leader is the event that for all arms $a \in \mathcal{A}$ and for all time steps t, the empirical mean $\hat{v}_{1,t}(a)$ is close to the true average of the mean rewards across actions taken by b when the leader has chosen action a. The event G_L is formalized as follows:

$$\forall a \in \mathcal{A}, t \leq T : \left| \frac{1}{n_{E\cdot|\mathcal{A}|,t}(a)} \sum_{E\cdot|\mathcal{A}|<t'<t|a_{t'}=a} v_1(a_{t'}, b_{t'}) - \hat{v}_{1,t}(a) \right| \leq \frac{10\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a)}}.$$

The clean event G_F for the follower is the event that the follower picks an item within the ε^* best response set: $\forall t > E \cdot |\mathcal{A}| : b_t \in \mathcal{B}_{\varepsilon^*}(a_t)$.

We first prove that the clean event G occurs with high probability.

Lemma 276. *Assume the notation above. Let $\mathtt{ALG}_2$ be any algorithm with high-probability instantaneous regret g where $g(t, T, \mathcal{B}) = O(E^{-1/2}|\mathcal{B}|^{1/2}(\log T)^{1/2})$ for $t > E$ and $g(t, T, \mathcal{B}) = 1$ for $t \leq E$, and let $\mathtt{ALG}_1 = \mathtt{ExploreThenUCB}(E)$. Then, the event G occurs with high probability: $\mathbb{P}[G] \geq 1 - T^{-3}(|\mathcal{A}| + 1)$.*

Proof. We first show that $\mathbb{P}[G_F] \geq 1 - |\mathcal{A}| \cdot T^{-3}$. A sufficient condition for this event to hold is that:

$$\forall t > E \cdot |\mathcal{A}| : v_2(a_t, b_t) \geq \max_{b\in\mathcal{B}} v_2(a_t, b) - \max_{t>E} g(t, T, \mathcal{B}).$$

Since the exploration phases pulls every arm $a \in \mathcal{A}$ a total of E times, the high-probability instantaneous regret assumption guarantees that this event holds with probability at least $1 - |\mathcal{A}| \cdot T^{-3}$, as desired.

We next show that $\mathbb{P}[G_L] \geq 1 - T^{-3}$. This follows from a a Chernoff bound (and using the analogue of one of the canonical bandit models in Lattimore and Szepesvári (2020)) combined with a union bound.

The lemma follows from another union bound over G_L and G_F. □

Our main lemma provides, an upper bound on $\frac{1}{n_{E\cdot|\mathcal{A}|,T}(a')-1} \sum_{t\in S(a')\setminus\{\max(S(a'))\}} v_1(a_t, b_t)$, which is the average of the mean rewards obtained on a' across all time steps t where

a' is pulled (except for the last round), for each arm $a' \in \mathcal{A}$. In particular, we upper bound this quantity by the worst-case optimal reward under ε-best-responses by the follower $(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a,b))$ minus the twice the size of the confidence set of a.

Lemma 277. *Assume the notation above. Suppose that the clean event G holds. Then it holds that:*

$$\frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a')-1} \sum_{t \in S(a') \setminus \{\max(S(a'))\}} v_1(a_t,b_t) \geq \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a')-1}}.$$

Proof. We assume that the clean event $G = G_L \cap G_F$ holds. Note that $t^* = \max(S(a'))$ denotes the last time step during which a' is chosen. Recalling that $\tilde{a} = \arg\max_{a \in \mathcal{A}} \min_{\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b)$, let $S = S(\tilde{a}) \cap [E \cdot |\mathcal{A}| + 1, t^* - 1]$ be the set of time steps during the UCB phase prior to time step t^* where arm $\tilde{a}$ is pulled. We see that at the beginning of time step t^*, it holds that:

$$\begin{aligned}
\frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a')-1} \sum_{t \in S(a') \setminus \{t^*\}} v_1(a_t,b_t) &\geq_{(1)} \hat{v}_{1,t^*}(a') - \frac{10\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a')-1}} \\
&\geq v_{1,t^*}^{\text{UCB}}(a') - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a')-1}} \\
&\geq v_{1,t^*}^{\text{UCB}}(\tilde{a}) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a')-1}} \\
&= \hat{v}_{1,t^*}(\tilde{a}) + \frac{10\sqrt{\log T}}{\sqrt{|S|}} - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a')-1}} \\
&\geq_{(2)} \frac{1}{|S|} \sum_{t \in S} v_1(a_t,b_t) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a')-1}} \\
&\geq_{(3)} \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a')-1}}
\end{aligned}$$

where (1) and (2) use the clean event G_L. Step (3) uses the clean event G_F which guarantees that $b_t \in \mathcal{B}_{\varepsilon^*}(a_t)$ for all t, which means that for any $t \in S$, it holds that:

$$v_1(a_t,b_t) = v_1(\tilde{a},b_t) \geq \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(\tilde{a},b) = \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a,b)$$

as desired. □

Now we are ready to prove Theorem 119.

Proof of Theorem 119. Assume that the clean event G holds. This occurs with probability at least $1-(1+|\mathcal{A}|)T^{-3}$ (Lemma 276), so the clean event not occurring counts negligibly towards regret.

The regret in the explore phase is bounded by $O(1)$ in each round, the total regret from that phase is $O(T^{2/3}|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3})$ for either player.

The remainder of the analysis boils down to bounding the regret in the UCB phase. We separately analyze the regret of the leader and the follower. Observe that $\varepsilon^* = \max_{t>E} g(t,T,\mathcal{B}) = O\left(\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3}T^{-1/3}\right)$ based on the assumption on the follower's algorithm.

Regret for the leader. We bound the regret as:

$$\begin{aligned}
&\beta_1^{\mathrm{tol}} \cdot (T - E\cdot|\mathcal{A}|) - \sum_{t=E\cdot|\mathcal{A}|+1}^{T} v_1(a_t,b_t) \\
&\leq \sum_{t=E\cdot|\mathcal{A}|+1}^{T} \left(\varepsilon^* + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t,b_t)\right) \\
&= \sum_{a\in\mathcal{A}} \sum_{t\in S(a)} \left(\varepsilon^* + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t,b_t)\right) \\
&\leq |\mathcal{A}| + \sum_{a\in\mathcal{A}} \sum_{t\in S(a)\setminus\{\max(S(a))\}} \left(\varepsilon^* + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t,b_t)\right) \\
&\leq |\mathcal{A}| + \underbrace{\varepsilon^*\cdot T}_{(1)} \\
&+ \underbrace{\sum_{a\in\mathcal{A}} (n_{E\cdot|\mathcal{A}|,T+1}(a) - 1)\cdot \left(\max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a)-1} \sum_{t\in S(a)\setminus\{\max(S(a))\}} (v_1(a_t,b_t))\right)}_{(2)}
\end{aligned}$$

The term $|\mathcal{A}|$ computes negligibly, term (1) is equal to $\Theta(\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}(\log T)^{1/3}T^{2/3})$, and term (2) can be bounded by:

$$\begin{aligned}
&\sum_{a\in\mathcal{A}} (n_{E\cdot|\mathcal{A}|,T+1}(a) - 1)\cdot \left(\max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a)-1} \sum_{t\in S(a)\setminus\{\max(S(a))\}} (v_1(a_t,b_t))\right) \\
&\leq \sum_{a\in\mathcal{A}} (n_{E\cdot|\mathcal{A}|,T+1}(a) - 1)\cdot \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a)} - 1} \\
&\leq O\left(\sqrt{\log T}\cdot \sum_{a\in\mathcal{A}} \sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a) - 1}\right) \\
&\leq O\left(\sqrt{|\mathcal{A}|T\log T}\right),
\end{aligned}$$

where the first inequality uses Lemma 277 and the last inequality uses Jensen's inequality.

Regret for the follower. Note that $\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)$ denotes the set of time steps where an action in $\mathcal{A}_{\varepsilon^*}$ is chosen. We bound the regret as:

$$\beta_2^{\text{tol}} \cdot (T - E \cdot |\mathcal{A}|) - \sum_{t=E\cdot|\mathcal{A}|+1}^{T} v_2(a_t, b_t)$$
$$\leq \underbrace{\left(\sum_{t=E\cdot|\mathcal{A}|+1}^{T} \mathbb{1}[t \notin \cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)]\right)}_{(1)} + \underbrace{\sum_{t\in\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)} \left(\min_{a\in\mathcal{A}_{\varepsilon^*}} \max_{b\in\mathcal{B}} v_2(a,b) - v_2(a_t,b_t)\right)}_{(2)} + \underbrace{\varepsilon^* \cdot |\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)|}_{(3)}$$

We first bound term (1), which can be rewritten as $\sum_{t=E\cdot|\mathcal{A}|+1}^{T} \mathbb{1}[t \notin \cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)] = \sum_{a\notin\mathcal{A}_{\varepsilon^*}} n_{E\cdot|\mathcal{A}|,T}(a)$. This counts the number of times that arms outside of $\mathcal{A}_{\varepsilon^*}$ are pulled during the UCB phase. The key intuition is when an arm $a_t \notin \mathcal{A}_{\varepsilon^*}$, it holds that:

$$v_1(a_t, b_t) \leq \max_{b\in\mathcal{B}_{\varepsilon^*}(a')} v_1(a_t, b) < \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \varepsilon^*,$$

where the first inequality uses the fact that $b_t \in \mathcal{B}_{\varepsilon^*}(a_t)$ (which follows from the clean event G_F) and the second inequality uses the fact that $a_t \notin \mathcal{A}_{\varepsilon^*}$. This implies that for any $a' \notin \mathcal{A}_{\varepsilon^*}$, the average reward across all time steps (except for the last time step) where a' is pulled satisfies:

$$\frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a') - 1} \sum_{t\in S(a')\setminus\{\max(S(a'))\}} v_1(a_{t'}, b_{t'}) < \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \varepsilon^*.$$

However, by Lemma 277, we can also lower bound the average reward across all time steps (except for the last time step) where a' is pulled in terms of $n_{E\cdot|\mathcal{A}|,T+1}(a')$ as follows:

$$\frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a') - 1} \sum_{t\in S(a')\setminus\{\max(S(a'))\}} v_1(a_{t'}, b_{t'}) \geq \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{10\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a') - 1}}.$$

Putting these two inequalities together, we see that:

$$\frac{10\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a') - 1}} \geq \varepsilon^*,$$

which bounds the number of times that a' is pulled during the UCB phase as follows:

$$n_{E\cdot|\mathcal{A}|,T+1}(a') \leq \Theta\left(\frac{\log T}{(\varepsilon^*)^2}\right) = \Theta\left((\log T)^{1/3} T^{2/3} |\mathcal{A}|^{-2/3} |\mathcal{B}|^{-2/3}\right).$$

This means that:

$$\sum_{t=E\cdot|\mathcal{A}|+1}^{T} \mathbb{1}[t \notin \cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)] = \sum_{a\notin\mathcal{A}_{\varepsilon}} n_{E\cdot|\mathcal{A}|,T+1}(a) \leq \Theta\left((\log T)^{1/3} T^{2/3} |\mathcal{A}|^{1/3} |\mathcal{B}|^{-2/3}\right)$$

	b_1	b_2
a_1	$(1,0)$	$(1-x,y)$
a_2	$(1-2x,2y)$	$(1-3x,3y)$

Table J.7: Set $x, y \in (0, 1/3)$ to obtain an example where both players have completely inverted ordered preferences over outcomes, but for $x, y > \mathcal{O}(1/T)$ have bounded continuity.

Next, we bound term (2):

$$\begin{aligned}\min_{a\in\mathcal{A}_{\varepsilon^*}}\max_{b\in\mathcal{B}} v_2(a,b) - \mathbb{E}[v_2(a_t,b_t)] &\leq \sum_{t\in\cup_{a\in\mathcal{A}_{\varepsilon^*}}S(a)}\left(\max_{b\in\mathcal{B}} v_2(a_t,b) - \mathbb{E}[v_2(a_t,b_t)]\right)\\ &\leq |\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)|\cdot\varepsilon^*\\ &\leq T\cdot\varepsilon^*\\ &= \Theta\left((\log T)^{1/3}T^{2/3}|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}\right).\end{aligned}$$

Finally, we bound term (3) as $\varepsilon^*\cdot|S| \leq T\cdot\varepsilon^* = \Theta\left((\log T)^{1/3}T^{2/3}|\mathcal{A}|^{1/3}|\mathcal{B}|^{1/3}\right)$. Putting this all together yields the desired bound. □

J.4 Proofs for Chapter 14.5

J.4.1 Alignment and continuity discussion

We note that constant L^* still allows for a rich space of disagreement on values. We will formalize our discussion on the distinction between requiring that the leader and the follower have the same relative ordering on every pair of (a, b) outcomes (*ordered alignment*) and that they agree on which pairs of outcomes are *sufficiently different* (*continuity*). In particular, our Lipschitz condition requires continuity, but still allows for arbitrarily misordered alignment. As an example, Table J.7 gives an example where the leader and the follower have completely inverted preferences over every outcome, but have utility that is $\max\left(\frac{x}{y},\frac{y}{x}\right)$ Lipschitz continuous.

J.4.2 Proofs and examples for Chapter 14.5.1

In this section, we prove Theorem 121, restated below for convenience.

Theorem 121. *Consider a StrongDSG where $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ has Lipschitz constant L^*. Let $\texttt{ALG}_2$ be any algorithm satisfying high-probability anytime regret $h(t, T, \mathcal{B}) = C'\sqrt{|\mathcal{B}|t\log T}$ where C' is a constant, and let $\texttt{ALG}_1 = \texttt{LipschitzUCB}(L, C'\sqrt{|\mathcal{B}|})$ for any $L \geq L^*$. Then both players achieve the following regret bounds with respect to the original Stackelberg* *benchmarks* β_1^{orig} *and* β_2^{orig}*: that is,* $R_1(T;\mathcal{I}) = O\left(L\sqrt{T|\mathcal{A}||\mathcal{B}|\log T}\right)$ *and* $R_2(T;\mathcal{I}) = O\left(L^2\sqrt{T|\mathcal{A}|\cdot|\mathcal{B}|\log T}\right)$.

Notation. Let $\hat{v}_{1,t}(a)$ be the empirical mean specified in `LipschitzUCB` at the beginning of time step t, which is the mean of the leader's stochastic rewards $\{r_{1,t'}(a_{t'}, b_{t'}) \mid a_{t'} = a, 1 \leq t' < t\}$. We also define $\hat{v}_{1,t}(a,b)$ to be the mean of the leader's stochastic rewards for the arm (a,b) up through time step $t-1$ (the set given by $\{r_{1,t'}(a_{t'}, b_{t'}) \mid a_{t'} = a, b_{t'} = b, 1 \leq t' < t\}$). Note that this quantity is not computable by the leader in a StrongDSG, but we nonetheless find it convenient to consider in the analysis. Let $n_t(a) = |1 \leq t' < t \mid a_t = a|$ be the number of times that a has been chosen prior to time step t. Let $n_t(a,b) = |1 \leq t' < t \mid a_t = a, b_t = b|$ be the number of times that (a,b) has been chosen prior to time step t. For each arm $a \in \mathcal{A}$, let $b^*(a) = \arg\max_{b\in\mathcal{B}} v_2(a,b)$ be the follower's best response.

Clean event. We define the clean event $G = G_L \cap G_F$ to be the intersection of a clean event G_L for the leader and a clean event G_F for the follower. Informally speaking, the clean event for the leader is the event that for all pairs of arms, the empirical mean reward $\hat{v}_{1,t}(a,b)$ is close to the true reward $v_1(a,b)$. The event G_L is formalized as follows:

$$\forall a \in \mathcal{A}, t \leq T : |\hat{v}_{1,t}(a,b) - v_1(a,b)| \leq \frac{10\sqrt{\log T}}{\sqrt{n_t(a)}}.$$

Informally speaking, the clean event for the follower is the event that the follower satisfies high-probability anytime regret bounds. The event G_F is formalized as follows:

$$\forall a \in \mathcal{A}, t \leq T : \sum_{1\leq t'<t \mid a_t = a} (v_2(a, b^*(a)) - v_2(a_t, b_t)) \leq C'\sqrt{|\mathcal{B}|n_t(a)\log T}$$

We first prove that the clean event G occurs with high probability.

Lemma 278. *Assume the setup of Theorem 121 and the notation above. Then the clean event occurs with high probability:* $\mathbb{P}[G] \geq 1 - T^{-3}(|\mathcal{A}|+1)$.

Proof. We union bound over G_L and G_F. The analysis for G_F follows from the high-probability anytime regret bound assumption. The analysis for G_L follows from a Chernoff bound (and using the analogue of one of the canonical bandit models in Lattimore and Szepesvári (2020)) combined with a union bound. □

The following lemma guarantees, for each arm $a \in \mathcal{A}$, that the empirical mean $\hat{v}_{1,t}(a)$ is close to the mean reward if the follower were to best-respond $\max_{b\in\mathcal{B}} v_1(a,b)$. Conceptually speaking, this lemma guarantees that the confidence sets for the leader are always "correct".

Lemma 279. *Assume the setup of Theorem 121 and the notation above. Suppose that the clean event G holds. Then for any $t \leq T$ and $a \in \mathcal{A}$, it holds that:*

$$|\hat{v}_{1,t}(a) - v_1(a, b^*(a))| \leq \frac{10\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a)}} + C' \cdot L \cdot \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a)}}.$$

Proof. We observe that:

$$\begin{aligned}
&|\hat{v}_{1,t}(a) - v_1(a, b^*(a))| \\
&= \left| \left(\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot \hat{v}_1(a,b) \right) - v_1(a, b^*(a)) \right| \\
&= \left| \left(\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot \hat{v}_1(a,b) \right) - \frac{1}{n_t(a)} \left(\sum_{b \in \mathcal{B}} n_t(a,b) \cdot v_1(a, b^*(a)) \right) \right| \\
&\leq \frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |\hat{v}_1(a,b) - v_1(a, b^*(a))| \\
&\leq \underbrace{\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |\hat{v}_1(a,b) - v_1(a,b)|}_{(A)} + \underbrace{\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |v_1(a,b) - v_1(a, b^*(a))|}_{(B)}.
\end{aligned}$$

First, we will bound term (A), which relates the error of the estimate of $v_1(a,b)$. We see that:

$$\begin{aligned}
\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |\hat{v}_1(a,b) - v_1(a,b)| &\leq_{(1)} \frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot \frac{10\sqrt{\log T}}{\sqrt{n_t(a,b)}} \\
&= \frac{10\sqrt{\log T}}{n_t(a)} \sum_{\sum_{b \in \mathcal{B}}} \sqrt{n_t(a,b)} \\
&\leq_{(2)} \frac{10\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a)}}.
\end{aligned}$$

where (1) uses the clean event G_L and (2) uses Jensen's inequality.

Term (B) represents represents the difference in the leader's utility between the arm chosen by the follower and the follower's best-response. We can bound this as:

$$\begin{aligned}
\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |v_1(a,b) - v_1(a, b^*(a))| &\leq_{(1)} \frac{L^*}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |v_2(a,b) - v_2(a, b^*(a))| \\
&=_{(2)} \frac{L^*}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot (v_2(a, b^*(a)) - v_2(a,b)),
\end{aligned}$$

where (1) uses the Lipschitz property and (2) uses the fact that $b^*(a)$ is the best arm for the follower, given that the leader pulls arm a. Using the clean event G_F and that $L \geq L^*$, we see that:

$$\frac{L^*}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot (v_2(a, b^*(a)) - v_2(a,b)) = \frac{L^*}{n_t(a)} \sum_{1 \leq t' < t | a_t = a} (v_2(a, b^*(a)) - v_2(a_t, b_t))$$
$$\leq C' \cdot L \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a)}}.$$

Taken together, these terms give the desired bound. □

It will also be convenient to bound the following two quantities which surface in our regret analysis. At a conceptual level, B_1 captures the sum of the sizes of the confidence sets of the arms pulled by the leader, and the term B_2 captures the cumulative suboptimality of the follower relative to the action a that they are provided in each time step.

Lemma 280. *Assume the setup of Theorem 121 and the notation above. Suppose that the clean event G holds. Then it holds that:*

$$B_1 := \sum_{t=1}^{T} \left(\frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} \right) \leq O\left(L\sqrt{T|\mathcal{A}||\mathcal{B}| \log T} \right)$$
$$B_2 := \sum_{t=1}^{T} (v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)) \leq O\left(\sqrt{T|\mathcal{A}||\mathcal{B}| \log T} \right)$$

Proof. To bound B_2, we see that:

$$\begin{aligned}
B_2 &= \sum_{t=1}^{T} (v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)) \\
&= \sum_{a \in \mathcal{A}} \sum_{t \in T | a_t = a} (v_2(a, b^*(a)) - v_2(a, b_t)) \\
&\leq_{(A)} \sum_{a \in \mathcal{A}} C' \cdot \sqrt{|\mathcal{B}| \cdot n_T(a) \log T} \\
&= C' \cdot \sqrt{|\mathcal{B}| \log T} \cdot \sum_{a \in \mathcal{A}} \sqrt{n_T(a)} \\
&\leq_{(B)} C' \cdot \sqrt{T|\mathcal{A}||\mathcal{B}| \log T},
\end{aligned}$$

where (A) uses the event G_F and (B) uses Jensen's inequality.

To bound B_1, we note that we must upper bound this both with a) the gap of the confidence interval, as well as b) the error on the leader's estimates of their value for arm a.

Taken together, this yields;

$$
\begin{aligned}
B_1 &= \sum_{t=1}^{T} \left(\frac{10\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} \right) \\
&= \sum_{t=1}^{T} \frac{10\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} + \sum_{t=1}^{T} C' \cdot L \cdot \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} \\
&\leq (10\sqrt{|\mathcal{B}| \log T} + C' \cdot L\sqrt{|\mathcal{B}| \log T}) \sum_{t=1}^{T} \frac{1}{\sqrt{n_t(a_t)}} \\
&\leq_{(A)} (10\sqrt{|\mathcal{B} \log T} + C' \cdot L\sqrt{|\mathcal{B}| \log T}) \cdot (2 \cdot \sqrt{T|\mathcal{A}|} + |\mathcal{A}|) \\
&= O\left(L\sqrt{T|\mathcal{A}||\mathcal{B}| \log T} \right).
\end{aligned}
$$

where (A) follows from Lemma 269

□

We now prove Theorem 121.

Proof of Theorem 121. Assume that clean event G holds. This occurs with probability at least $1-(|\mathcal{A}+1)T^{-3}$ (Lemma 278), so the clean event not occurring counts negligibly towards regret.

Moreover, let $(a^*, b^*(a^*))$ be the Stackelberg equilibrium. Let $\alpha_t(a) = \frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a)}} + C \cdot L \cdot \frac{\sqrt{\log T}}{\sqrt{n_t(a)}}$ be the confidence bound size at time step t and let $v_{1,t}^{\mathrm{UCB}}(a) = \hat{v}_{1,t}(a) + \alpha_t(a)$ denote the UCB estimate in $\texttt{LipschitzUCB}(L, C)$ computed during time step t prior to reward at time step t being observed.

We can bound the leader's regret as:

$$
\begin{aligned}
R_1(T) &= \sum_{t=1}^{T}(v_1(a^*, b^*(a^*)) - v_1(a_t, b_t)) \\
&= \sum_{t=1}^{T} (v_1(a^*, b^*(a^*)) - v_1(a_t, b^*(a_t))) + \sum_{t=1}^{T} (v_1(a_t, b^*(a_t)) - v_1(a_t, b_t)) \\
&\leq_{(A)} \sum_{t=1}^{T} (\hat{v}_1(a^*) + \alpha_t(a^*) - \hat{v}_1(a_t) + \alpha_t(a_t)) + \sum_{t=1}^{T} |v_1(a_t, b^*(a_t)) - v_1(a_t, b_t)| \\
&\leq \sum_{t=1}^{T} \left(v_{1,t}^{\mathrm{UCB}}(a^*) - v_1^{\mathrm{UCB}}(a_t) + 2 \cdot \alpha_t(a_t)\right) + L \cdot \sum_{t=1}^{T} |v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)| \\
&\leq 2 \cdot \sum_{t=1}^{T} \alpha_t(a_t) + L \cdot \sum_{t=1}^{T} (v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)) \\
&= 2 \cdot \sum_{t=1}^{T} \left(\frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} \right) + L \cdot B_2 \\
&= 2 \cdot B_1 + L \cdot B_2 \\
&\leq_{(B)} O\left(L\sqrt{T|\mathcal{A}||\mathcal{B}| \log T}\right)
\end{aligned}
$$

where (A) uses Lemma 279 and (B) uses Lemma 280.

We also bound the follower's regret as:

$$
\begin{aligned}
R_2(T) &= \sum_{t=1}^{T} \left(v_2(a^*, b^*(a^*)) - v_2(a_t, b_t)\right) \\
&= \sum_{t=1}^{T} \left(v_2(a^*, b^*(a^*)) - v_2(a^*, b^*(a_t))\right) + \sum_{t=1}^{T} \left(v_2(a^*, b^*(a_t)) - v_2(a_t, b_t)\right) \\
&= \sum_{t=1}^{T} L \cdot |v_1(a^*, b^*(a^*)) - v_1(a_t, b^*(a_t))| + B_2 \\
&=_{(A)} \sum_{t=1}^{T} L \cdot \left(v_1(a^*, b^*(a^*)) - v_1(a_t, b^*(a_t))\right) + B_2 \\
&\leq_{(B)} \sum_{t=1}^{T} L \cdot \left(\hat{v}_{1,t}(a^*) + \alpha_t(a^*) - \hat{v}_{1,t}(a_t) + \alpha_t(a^*)\right) + B_2 \\
&= \sum_{t=1}^{T} L \cdot \left(v_{1,t}^{\text{UCB}}(a^*) - v_{1,t}^{\text{UCB}}(a_t) + 2 \cdot \alpha_t(a_t)\right) + B_2 \\
&\leq \sum_{t=1}^{T} L \cdot \left(2 \cdot \alpha_t(a_t)\right) + B_2 \\
&= 2L \cdot \sum_{t=1}^{T} \left(\frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}}\right) + B_2 \\
&= 2L \cdot B_1 + B_2 \\
&\leq_{(C)} O\left(L^2 \sqrt{T|\mathcal{A}||\mathcal{B}| \log T}\right)
\end{aligned}
$$

where (A) uses the fact that a^* is the action chosen by the leader at the Stackelberg equilibrium where (B) uses Lemma 279 and (C) uses Lemma 280.

□

J.4.3 Proof of Theorem 122

We prove Theorem 122, restated below.

Theorem 122. *Consider a WeakDSG, where for each $a \in \mathcal{A}$, the algorithm $\texttt{ALG}_2$ runs a separate instantiation of* `ActiveArmElimination` *with parameters $M_1, \ldots, M_P$ (where $M_i = \Theta(\log T \cdot 2^{2i})$ denotes the number of times that each arm is pulled in phase i). Let $\texttt{ALG}_1 =$* `PhasedUCB`*$(M_1, \ldots, M_P)$. Then it holds that the regret with respect to the self-γ-tolerant benchmarks $\beta_1^{\text{self-tol}}$ and $\beta_2^{\text{self-tol}}$ is bounded as:*

$$\max(R_1(T), R_2(T)) = O\left(\sqrt{|\mathcal{A}| \cdot |\mathcal{B}| \cdot T \cdot \log T}\right).$$

This theorem assumes that $\gamma = \Omega\left(T^{-1/4}\sqrt{|\mathcal{A}||\mathcal{B}| \cdot \log T}\right)$.

Notation. Let $\hat{v}_{1,t}(a,b)$ denote the empirical mean of the leader's observed rewards

$$\{r_{1,t'}(a,b) \mid 1 \leq t' < t, a_t = a, b_t = b\}$$

for (a,b) up to time step t. (The leader can observe this information in a WeakDSG.) Let $v_{1,t}^{\mathrm{UCB}}(a,b)$ denote the UCB estimate in `PhasedUCB` during time step t. Let $n_t(a) = |\{1 \leq t' < t \mid a_t = a\}|$ be the number of times that arm a is pulled before time step t. Let $n_t(a,b) = |\{1 \leq t' < t \mid a_t = a, b_t = b\}|$ be the number of times that arms (a,b) are pulled before time step t. Let C be a constant such that `ActiveArmElimination` has high-probability instantaneous regret $g(t,T,\mathcal{B}) = C \cdot \sqrt{|\mathcal{B}| \log T/t)}$ (such a constant C exists by Chapter 126). Let $\mathcal{B}_t(a)$ be the computation of the active set at line 3 of `PhasedUCB` during time step t. Let $s_t(a)$ be the value of the variable $s'(a)$ at the end of the `ComputeActiveArms` algorithm, when it is called at the beginning of time step t in `PhasedUCB`. Let (a^*, b^*) be the Stackelberg equilibrium.

Clean event. We define the clean event $G := G_L \cap G_F \cap G_{L,F}$ to be the intersection of a clean event G_L for the leader, a clean event G_F for the follower, and a clean event $G_{L,F}$ for the follower (using the leader's assessment of the follower). Informally speaking, the clean event G_L for the leader is the event that the empirical mean $\hat{v}_1(a,b)$ is always sufficiently close to the true mean reward $v_1(a,b)$. We formalize the clean event G_L as follows:

$$\forall t \in T, a \in \mathcal{A}, b \in \mathcal{B} : |\hat{v}_{1,t}(a,b) - v_1(a,b)| \leq \frac{10\sqrt{\log T}}{\sqrt{n_t(a,b)}}.$$

The clean event G_F for the follower is the event that the follower satisfies the high-probability instantaneous regret guarantee:

$$\forall t \leq T : \left| v_2(a_t, b_t) - \max_{b \in \mathcal{B}} v_2(a_t, b) \right| \leq C \cdot \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a)}}.$$

The final clean event $G_{L,F}$ is the event that the active arm set $\mathcal{B}_t(a^*)$ for the Stackelberg action always contains the follower's best-response:

$$\forall t \in T, b \in \mathcal{B} : \arg\max_{b \in \mathcal{B}} v_2(a^*, b) \in \mathcal{B}_t(a^*).$$

Lemma 281. *Assume the setup of Theorem 122 and notation above. Then the clean event G occurs with high probability:* $\mathbb{P}[G] \geq 1 - (2 \cdot |\mathcal{A}| + 1) \cdot T^{-3}$.

Proof. We union bound for G_F, G_L, and $G_{L,F}$. The analysis for G_L follows from a Chernoff bound (and using the analogue of one of the canonical bandit models in Lattimore and Szepesvári (2020)) combined with a union bound. The analysis for G_F follows from Chapter 126. The analysis for $G_{L,F}$ follows from standard properties of `ActiveArmElimination` (e.g., see Lattimore and Szepesvári (2020)) combined with a union bound over $\mathcal{A}$. $\square$

The first lemma shows that if the follower runs ActiveArmElimination, for every $a \in \mathcal{A}$ and $b \in \mathcal{B}_t(a)$, we can upper and lower bound the number of pulls $n_t(a,b)$ in terms of the last phase that the follower has completed (as assessed by the leader).

Lemma 282. *Assume the setup of Theorem 122 and notation above. Then for every time step t, and every $a \in \mathcal{A}$ and $b \in \mathcal{B}_t(a)$, it holds that:*

$$n_t(a,b) \in \left[\sum_{i=1}^{s_t(a)} M_i, \sum_{i=1}^{s_t(a)+1} M_i + 1\right]$$

Proof. This follows from the implementation of `ComputeActiveArms` combined with the specification of `ActiveArmElimination`, which guarantees that the follower has finished phase $s_t(a)$ by the end of round $t-2$ and is at most one step into phase $s_t(a)+2$. □

The next lemma guarantees that at every time step t, the chosen pair of actions (a_t, b_t) are in the ε_t-best-response sets for each player, where ε_t depends on the number of times $n_t(a_t)$ that arm a_t has been chosen so far.

Lemma 283. *Assume the setup of Theorem 122 and notation above. Suppose that the clean event G holds. Then for every time step t, it holds that for*

$$v_1(a_t, b_t) \geq \min_{a \in \mathcal{A}^1_{\varepsilon_t}} \min_{b \in \mathcal{B}_{\varepsilon_t}(a)} v_1(a,b)$$

$$v_2(a_t, b_t) \geq \min_{a \in \mathcal{A}^1_{\varepsilon_t}} \min_{b \in \mathcal{B}_{\varepsilon_t}(a)} v_2(a,b).$$

for $\varepsilon_t = \Theta(\sqrt{|\mathcal{B}| \cdot \log T / n_t(a_t)})$.

Proof. It suffices to show that $a_t \in \mathcal{A}_{\varepsilon_t}$ and $b_t \in \mathcal{B}_{\varepsilon_t}(a_t)$.

By the clean event G_F, it immediately follows that $b_t \in \mathcal{B}_{\varepsilon_t}(a_t)$.

To show that $a_t \in \mathcal{A}_{\varepsilon_t}$, it suffices to show that

$$\max_{b \in \mathcal{B}_{\varepsilon_t}(a_t)} v_1(a_t, b) \geq \max_{a' \in \mathcal{A}} \min_{b' \in \mathcal{B}_{\varepsilon}(a')} v_1(a', b') - \varepsilon_t,$$

which can be written as $\max_{a' \in \mathcal{A}} \min_{b' \in \mathcal{B}_{\varepsilon_t}(a')} v_1(a', b') \leq \max_{b \in \mathcal{B}_{\varepsilon_t}(a_t)} v_1(a,b) + \varepsilon_t$. To see this, observe that:

$$
\begin{aligned}
\max_{a' \in \mathcal{A}} \min_{b' \in \mathcal{B}_{\varepsilon_t}(a')} v_1(a', b') &\leq v_1(a^*, b^*) \\
&\leq_{(A)} \max_{b \in \mathcal{B}'(a^*)} v_{1,t}^{\mathrm{UCB}}(a^*, b) \\
&\leq \max_{b \in \mathcal{B}'_t(a_t)} v_{1,t}^{\mathrm{UCB}}(a_t, b) \\
&\leq_{(B)} \max_{b \in \mathcal{B}'_t(a_t)} \left(v_1(a_t, b) + 20 \cdot \sqrt{\frac{\log T}{n_t(a_t, b)}} \right) \\
&\leq_{(C)} \max_{b \in \mathcal{B}'_t(a_t)} \left(v_1(a_t, b) + 20 \cdot \sqrt{\frac{\log T}{\sum_{i=1}^{s_t(a)} M_i}} \right) \\
&\leq_{(D)} \max_{b \in \mathcal{B}'_t(a_t)} (v_1(a_t, b)) + \Theta\left(\sqrt{\frac{|\mathcal{B}| \log T}{n_t(a_t)}} \right) \\
&\leq \max_{b \in \mathcal{B}'_t(a_t)} v_1(a_t, b) + \varepsilon_t \\
&\leq_{(E)} \max_{b \in \mathcal{B}_{\varepsilon_t}(a_t)} v_{1,t}^{\mathrm{UCB}}(a_t, b) + \varepsilon_t.
\end{aligned}
$$

where (A) uses the event $G_{L,F}$, (B) uses the event G_L, (C) applies the lower bound in Lemma 282, (D) uses the upper bound in Lemma 282 to see that:

$$
n_t(a_t) \leq \sum_{b \in \mathcal{B}} n_t(a_t, b) \leq \sum_{b \in \mathcal{B}} \left(\left(\sum_{i=1}^{s_t(a)+1} M_i \right) + 1 \right) \leq \Theta\left(|\mathcal{B}| \cdot \sum_{i=1}^{s_t(a)} M_i \right)
$$

since every arm is pulled and (E) uses the clean event G_F. □

Now, we prove Theorem 122.

Proof of Theorem 122. Assume that the clean event G occurs. This occurs with probability at least $1 - (2 \cdot |\mathcal{A}| + 1) \cdot T^{-3}$ (Lemma 281), so the clean event not occurring counts negligibly towards regret.

We apply Lemma 283 to see that at time step t, it holds that for $\varepsilon_t = \Theta(\sqrt{|\mathcal{B}| \cdot \log T / n_t(a_t)})$, it holds that

$$
v_1(a_t, b_t) \geq \min_{a \in \mathcal{A}_{\varepsilon_t}} \min_{b \in \mathcal{B}_{\varepsilon_t}(a)} v_1(a, b)
$$

$$
v_2(a_t, b_t) \geq \min_{a \in \mathcal{A}_{\varepsilon_t}} \min_{b \in \mathcal{B}_{\varepsilon_t}(a)} v_2(a, b).
$$

For the leader, this implies that:

$$\begin{aligned}
R_1(T) &= \beta_1^{\text{self-tol}} \cdot T - \sum_{t=1}^{T} v_1(a_t, b_t) \\
&\leq \sum_{t=1}^{T} \left(\varepsilon_t + \min_{a \in \mathcal{A}_{\varepsilon_t}} \min_{b \in \mathcal{B}_{\varepsilon_t}(a)} v_1(a,b) - \sum_{t=1}^{T} v_1(a_t, b_t) \right) + \sum_{t=1}^{T} \mathbb{1}[\varepsilon_t > \gamma] \\
&\leq \left(\sum_{t=1}^{T} \varepsilon_t \right) + \sum_{t=1}^{T} \mathbb{1}[\varepsilon_t > \gamma].
\end{aligned}$$

For the follower, this similarly implies that:

$$\begin{aligned}
R_2(T) &= \beta_2^{\text{self-tol}} \cdot T - \sum_{t=1}^{T} v_2(a_t, b_t) \\
&\leq \sum_{t=1}^{T} \left(\varepsilon_t + \min_{a \in \mathcal{A}_{\varepsilon_t}} \min_{b \in \mathcal{B}_{\varepsilon_t}(a)} v_2(a,b) - \sum_{t=1}^{T} v_2(a_t, b_t) \right) \\
&\leq \left(\sum_{t=1}^{T} \varepsilon_t \right) + \sum_{t=1}^{T} \mathbb{1}[\varepsilon_t > \gamma].
\end{aligned}$$

To bound $\sum_{t=1}^{T} \varepsilon_t$, we observe that:

$$\begin{aligned}
\sum_{t=1}^{T} \varepsilon_t &= \sum_{t=1}^{T} \Theta\left(\sqrt{\frac{|\mathcal{B}| \cdot \log T}{n_t(a_t)}} \right) \\
&= \Theta\left(\sqrt{|\mathcal{B}| \cdot \log T} \cdot \sum_{t=1}^{T} \frac{1}{\sqrt{n_t(a_t)}} \right) \\
&\leq_{(A)} O\left(\sqrt{|\mathcal{B}| \cdot \log T} \cdot \sqrt{|\mathcal{A}| \cdot T} \right)
\end{aligned}$$

where (A) follows from Lemma 269. This gives the desired upper bound.

To bound $\sum_{t=1}^{T} \mathbb{1}[\varepsilon_t > \gamma]$, based on the setting of ε_t, we observe that $\varepsilon_t \leq \gamma$ when $n_{a_t} = O\left(\frac{|\mathcal{B}| \cdot (\log T)}{\varepsilon_t^2}\right)$. This means that $\mathbb{1}[\varepsilon_t > \gamma]$ occurs in at most $\Theta\left(\frac{|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T)}{\gamma^2}\right)$ time steps. As long as $\gamma = \Omega\left(T^{-1/4}\sqrt{|\mathcal{A}||\mathcal{B}| \cdot \log T}\right)$, this term contributes $O\left(\sqrt{|\mathcal{B}| \cdot \log T} \cdot \sqrt{|\mathcal{A}| \cdot T}\right)$ to regret.

$\square$

J.5 Proofs for Chapter 14.6

J.5.1 Proof of Theorem 124

Theorem 124. *Suppose that $c \geq 1$ and $d \leq 1$, and let $\eta := 2/(2+d)$. Consider a StrongDSG, where the follower runs a separate instantiation of* ***ExploreThenCommit***$(E_2, \mathcal{B})$ *for every $a \in \mathcal{A}$, and the leader runs* ***ExploreThenCommitThrowOut***$(E_1, E_2 \cdot |\mathcal{B}|, \mathcal{A})$*. If $E_2 = \Theta(|\mathcal{A}|^{-\eta}|\mathcal{B}|^{-\eta} \cdot (\log T)^{1-\eta}(c \cdot T)^{\eta})$, and $E_1 = \Theta(|\mathcal{A}|^{-\eta} \cdot (\log T)^{1-\eta}(c \cdot T)^{\eta})$, then the leader and follower regret with respect to the generalized (c, d, γ)-tolerant benchmarks are both at most:*

$$\max(R_1(T), R_2(T)) = O\big((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta} \cdot (c \cdot T)^{\eta}\big).$$

The proof follows a similar argument to the proof of Theorem 118 and borrows some lemmas from Appendix J.3.2

Proof of Theorem 124. Assume that the clean event G holds. This occurs with probability at least $1 - (|\mathcal{A}| \cdot |\mathcal{B}| + |\mathcal{A}|)T^{-3}$ (Lemma 274), so the clean event not occuring counts negligibly towards regret.

First, we consider the first $E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|$ time steps. Each time step results in $O(1)$ regret for both players. Based on the settings of E_1 and E_2, these phases contribute a regret of:

$$E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}| = O\left(|\mathcal{A}|^{1-\eta} \cdot |\mathcal{B}|^{1-\eta} \cdot (\log T)^{1-\eta}(c \cdot T)^{\eta}\right).$$

We focus on $t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|$ for the remainder of the analysis. Our main ingredient is Lemma 275. Note that $\varepsilon^* = \Theta\left(\max\left(\frac{\sqrt{\log T}}{\sqrt{E_1}}, \frac{\sqrt{\log T}}{\sqrt{E_2}}\right)\right) = \Theta\left((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{\eta/2} \cdot (c \cdot T)^{-\eta/2}\right)$ based on the settings of E_1 and E_2. The regret of the leader can be bounded as:

$$\begin{aligned}
&\beta_1^{\text{tol}} \cdot (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_1(a_t, b_t) \\
&\leq (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot \left(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) + c \cdot (\varepsilon^*)^d\right) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_1(\tilde{a}, \tilde{b}(\tilde{a})) \\
&= (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot c \cdot (\varepsilon^*)^d + (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|)\left(\max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(\tilde{a}, \tilde{b}(\tilde{a}))\right) \\
&\leq_{(A)} T \cdot c \cdot (\varepsilon^*)^d + T \cdot \varepsilon^* \\
&\leq_{(B)} T \cdot c \cdot (\varepsilon^*)^d \\
&\leq \Theta\left((c \cdot T)^{1-(\eta \cdot d/2)} \cdot (|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{\eta \cdot d/2}\right) \\
&= \Theta\left((c \cdot T)^{\eta} \cdot (|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta}\right).
\end{aligned}$$

where (A) follows from Lemma 275 and (B) uses the fact that $c \geq 1$ and $d \leq 1$. The regret of the follower can similarly be bounded as:

$$\begin{aligned}
&\beta_1^{\text{tol}} \cdot (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_2(a_t, b_t) \\
&\leq (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot \left(\min_{a \in \mathcal{A}_{\varepsilon^*}} \max_{b \in \mathcal{B}} v_2(a, b) + c \cdot (\varepsilon^*)^d \right) - \sum_{t > E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| + E_1 \cdot |\mathcal{A}|} v_2(\tilde{a}, \tilde{b}(\tilde{a})) \\
&= (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \cdot c \cdot (\varepsilon^*)^d + (T - E_2 \cdot |\mathcal{B}| \cdot |\mathcal{A}| - E_1 \cdot |\mathcal{A}|) \left(\min_{a \in \mathcal{A}_{\varepsilon^*}} \max_{b \in \mathcal{B}} v_2(a, b) - v_2(\tilde{a}, \tilde{b}(\tilde{a})) \right) \\
&\leq_{(B)} T \cdot c \cdot (\varepsilon^*)^d + T \cdot \varepsilon^* \\
&\leq O\left(T^{2/3} (\log T)^{1/3} |\mathcal{A}|^{1/3} |\mathcal{B}|^{1/3} \right).
\end{aligned}$$

where (B) follows from Lemma 275. This proves the desired result. □

J.5.2 Proof of Theorem 125

Theorem 125. *Suppose that $c \geq 1$ and $d \leq 1$, and let $\eta := 2/(2 + d)$. Let $E = \Theta(|\mathcal{A}|^{-\eta} (|\mathcal{B}| \log T)^{1-\eta} (c \cdot T)^{\eta})$. Consider a StrongDSG where `ALG`$_2$ is any algorithm with high-probability instantaneous regret $g(t, T, \mathcal{B}) = O\left((|\mathcal{A}| \cdot |\mathcal{B}| \cdot \log T)^{\eta/2} \cdot (c \cdot T)^{-\eta/2} \right)$ for $t > E$ and $g(t, T, \mathcal{B}) = 1$ for $t \leq E$, and where `ALG`$_1$ = `ExploreThenUCB`(E). Then, then the leader and follower regret with respect to the generalized (c, d, γ)-tolerant benchmarks are both bounded as:*

$$\max(R_1(T), R_2(T)) = O((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta} \cdot (c \cdot T)^{\eta}).$$

The proof follows a similar argument to the proof of Theorem 118 and borrows some lemmas from Appendix J.3.3

Proof of Theorem 125. Assume that the clean event G holds. This occurs with probability at least $1 - (1 + |\mathcal{A}|)T^{-3}$ (Lemma 276), so the clean event not occurring counts negligibly towards regret.

The regret in the explore phase is bounded by $O(1)$ in each round, the total regret from that phase is $E \cdot |\mathcal{A}| = O((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta} \cdot (c \cdot T)^{\eta})$ for either player.

The remainder of the analysis boils down to bounding the regret in the UCB phase. We separately analyze the regret of the leader and the follower. Observe that $\varepsilon^* = \max_{t > E} g(t, T, \mathcal{B}) = O\left((|\mathcal{A}| \cdot |\mathcal{B}| \log T)^{\eta/2} \cdot (c \cdot T)^{-\eta/2} \right)$ based on the assumption on the follower's algorithm.

Regret for the leader. We bound the regret as:

$$
\begin{aligned}
&\beta_1^{\mathrm{tol}} \cdot (T - E \cdot |\mathcal{A}|) - \sum_{t=E\cdot|\mathcal{A}|}^{T} v_1(a_t, b_t) \\
&\leq \sum_{t=E\cdot|\mathcal{A}|+1}^{T} \left(c \cdot (\varepsilon^*)^d + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t, b_t) \right) \\
&= \sum_{a\in\mathcal{A}} \sum_{t\in S(a)} \left(c \cdot (\varepsilon^*)^d + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t, b_t) \right) \\
&\leq |\mathcal{A}| + \sum_{a\in\mathcal{A}} \sum_{t\in S(a)\setminus\{\max(S(a))\}} \left(c \cdot (\varepsilon^*)^d + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t, b_t) \right) \\
&\leq |\mathcal{A}| + \underbrace{c \cdot (\varepsilon^*)^d \cdot T}_{(1)} \\
&\quad + \underbrace{\sum_{a\in\mathcal{A}} (n_{E\cdot|\mathcal{A}|,T+1}(a) - 1) \cdot \left(\max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a) - 1} \sum_{t\in S(a)\setminus\{\max(S(a))\}} (v_1(a_t, b_t)) \right)}_{(2)}
\end{aligned}
$$

The term $|\mathcal{A}|$ computes negligibly and term (1) is equal to $O((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{\eta \cdot d/2} \cdot (c \cdot T)^{1-\eta \cdot d/2}) = O((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta} \cdot (c \cdot T)^{\eta})$. Term (2) can be bounded by by the same argument as Theorem 119, which we repeat for completeness:

$$
\begin{aligned}
&\sum_{a\in\mathcal{A}} (n_{E\cdot|\mathcal{A}|,T+1}(a) - 1) \cdot \left(\max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a) - 1} \sum_{t\in S(a)\setminus\{\max(S(a))\}} (v_1(a_t, b_t)) \right) \\
&\leq \sum_{a\in\mathcal{A}} (n_{E\cdot|\mathcal{A}|,T+1}(a) - 1) \cdot \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a)} - 1} \\
&\leq O\left(\sqrt{\log T} \cdot \sum_{a\in\mathcal{A}} \sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a) - 1} \right) \\
&\leq O\left(\sqrt{|\mathcal{A}| T \log T} \right),
\end{aligned}
$$

where the first inequality uses Lemma 277 and the last inequality uses Jensen's inequality.

Regret for the follower. Note that $\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)$ denotes the set of time steps where an action in $\mathcal{A}_{\varepsilon^*}$ is chosen. We bound the regret as:

$$
\begin{aligned}
&\beta_2^{\mathrm{tol}} \cdot (T - E \cdot |\mathcal{A}|) - \sum_{t=E\cdot|\mathcal{A}|}^{T} v_2(a_t, b_t) \\
&\leq \underbrace{\left(\sum_{t=E\cdot|\mathcal{A}|}^{T} \mathbb{1}[t \notin \cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)] \right)}_{(1)} + \underbrace{\sum_{t\in\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)} \left(\min_{a\in\mathcal{A}_{\varepsilon^*}} \max_{b\in\mathcal{B}} v_2(a,b) - v_2(a_t, b_t) \right)}_{(2)} + \underbrace{c \cdot (\varepsilon^*)^d \cdot |\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)|}_{(3)}
\end{aligned}
$$

Term (1) can be bounded by a similar argument to Theorem 119, which we repeat for completeness. This term can be rewritten as $\sum_{t=E\cdot|\mathcal{A}|}^{T} \mathbb{1}[t \notin \cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)] = \sum_{a\notin\mathcal{A}_{\varepsilon^*}} n_{E\cdot|\mathcal{A}|,T}(a)$. This counts the number of times that arms outside of $\mathcal{A}_{\varepsilon^*}$ are pulled during the UCB phase. The key intuition is when an arm $a_t \notin \mathcal{A}_{\varepsilon^*}$, it holds that:

$$v_1(a_t, b_t) \leq \max_{b\in\mathcal{B}_{\varepsilon^*}(a')} v_1(a_t, b) < \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) - \varepsilon^*,$$

where the first inequality uses the fact that $b_t \in \mathcal{B}_{\varepsilon^*}(a_t)$ (which follows from the clean event G_F) and the second inequality uses the fact that $a_t \notin \mathcal{A}_{\varepsilon^*}$. This implies that for any $a' \notin \mathcal{A}_{\varepsilon^*}$, the average reward across all time steps (except for the last time step) where a' is pulled satisfies:

$$\frac{1}{n_{E\cdot|\mathcal{A}|,T}(a') - 1} \sum_{t\in S(a')\setminus\{\max(S(a'))\}} v_1(a_{t'}, b_{t'}) < \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) - \varepsilon^*.$$

However, by Lemma 277, we can also lower bound the average reward across all time steps (except for the last time step) where a' is pulled in terms of $n_{E\cdot|\mathcal{A}|,T}(a')$ as follows:

$$\frac{1}{n_{E\cdot|\mathcal{A}|,T}(a') - 1} \sum_{t\in S(a')\setminus\{\max(S(a'))\}} v_1(a_{t'}, b_{t'}) \geq \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) - \frac{10\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T}(a') - 1}}.$$

Putting these two inequalities together, we see that:

$$\frac{10\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T}(a') - 1}} \geq \varepsilon^*,$$

which bounds the number of times that a' is pulled during the UCB phase as follows:

$$n_{E\cdot|\mathcal{A}|,T}(a') \leq \Theta\left(\frac{\log T}{(\varepsilon^*)^2}\right) = \Theta\left((|\mathcal{A}|\cdot|\mathcal{B}|)^{-\eta}\cdot(\log T)^{1-\eta}\cdot(c\cdot T)^{\eta}\right).$$

This means that:

$$\sum_{t=E\cdot|\mathcal{A}|}^{T} \mathbb{1}[t \notin \cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)] = \sum_{a\notin\mathcal{A}_{\varepsilon}} n_{E\cdot|\mathcal{A}|,T}(a) \leq \Theta\left((|\mathcal{A}|\cdot\log T)^{1-\eta}\cdot(|\mathcal{B}|)^{-\eta}\cdot(c\cdot T)^{\eta}\right)$$

Next, we bound term (2):

$$\begin{aligned}
\min_{a\in\mathcal{A}_{\varepsilon^*}} \max_{b\in\mathcal{B}} v_2(a,b) - \mathbb{E}[v_2(a_t,b_t)] &\leq \sum_{t\in\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)} \left(\max_{b\in\mathcal{B}} v_2(a_t,b) - \mathbb{E}[v_2(a_t,b_t)]\right) \leq |\cup_{a\in\mathcal{A}_{\varepsilon^*}} S(a)|\cdot\varepsilon^* \\
&\leq T\cdot\varepsilon^* \\
&\leq T\cdot c\cdot(\varepsilon^*)^d \\
&= O((|\mathcal{A}|\cdot|\mathcal{B}|\cdot(\log T))^{\eta\cdot d/2}\cdot(c\cdot T)^{1-\eta\cdot d/2}) \\
&\leq O((|\mathcal{A}|\cdot|\mathcal{B}|\cdot(\log T))^{1-\eta}\cdot(c\cdot T)^{\eta}).
\end{aligned}$$

Finally, we bound term (3) as

$$\begin{aligned}\varepsilon^* \cdot |S| \le T \cdot \varepsilon^* \le T \cdot c \cdot (\varepsilon^*)^d &\le O((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{\eta \cdot d/2} \cdot (c \cdot T)^{1-\eta \cdot d/2}) \\ &= O((|\mathcal{A}| \cdot |\mathcal{B}| \cdot (\log T))^{1-\eta} \cdot (c \cdot T)^{\eta}).\end{aligned}$$

□

J.6 Proofs for Chapter 14.7

J.6.1 Proofs for Chapter 14.7.1

The follower algorithms $\texttt{ALG}_2$ that we analyze in this section run a separate instantation of a standard bandit algorithm for every $a \in \mathcal{A}$. We show that if $\texttt{ALG}$ satisfies a high-probability instantaneous (resp. anytime) regret bound, the same high-probability instantaneous (resp. anytime) regret bound is inherited for $\texttt{ALG}_2$ (recall that in Chapter 14.2.4 we defined high-probability instantaneous regret and high-probability anytime regret for both single-bandit learners which act in isolation and follower algorithms).

Lemma 284. *Suppose that the follower algorithm $\texttt{ALG}_2$ runs a separate instantation, for every $a \in \mathcal{A}$, of an single-bandit learning algorithm $\texttt{ALG}$ operating on the arms $\mathcal{B}$. If $\texttt{ALG}$ satisfies high-probability instantaneous regret g, then $\texttt{ALG}_2$ satisfies high-probability instantaneous regret g. Similarly, if $\texttt{ALG}$ also satisfies high-probability anytime regret h, then $\texttt{ALG}_2$ also satisfies high-probability anytime regret h.*

Proof. We use the following notation in the proof. Let $n_t(a)$ be the number of times that arm a has been pulled up prior to time step t. Following Chapter J.1.4, the follower's history can be represented as:

$$H_{2,t} := \{(t', a_{t'}, b_{t'}, r_{2,t'}(a_{t'}, b_{t'})) \mid 1 \le t' < t, a_{t'} = a\},$$

and the follower's history on the arm $a \in \mathcal{A}$ can be represented as:

$$H_{2,t,a} := \{(n_{t'+1}(a), b_{t'}, r_{2,t'}(a_{t'}, b_{t'})) \mid 1 \le t' < t, a_{t'} = a\}.$$

Using this notation and by the definition of $\texttt{ALG}_2$, we see that $\texttt{ALG}_2(a_t, H_{2,t}) = \texttt{ALG}(H_{2,t,a_t})$. We use this relationship to analyze the high-probability instantaneous regret and high-probability anytime regret of $\texttt{ALG}_2$.

High-probability instantaneous regret. Let the time horizon be T, and suppose that $\texttt{ALG}$ satisfies high-probability instantaneous regret $g(t, T, \mathcal{B})$ for every $1 \le t \le T$. Using this combined with the fact that $\texttt{ALG}_2(a_t, H_{2,t}) = \texttt{ALG}(H_{2,t,a_t})$, we see that for each $a \in \mathcal{A}$:

$$\mathbb{P}\left[\forall t \in [T] \mid v_2(a_t, b_t) \ge \max_{b \in \mathcal{B}} v_2(a_t, b) - g(n_{t+1}(a), T)\right] \ge 1 - T^{-3}.$$

Taking a union bound over $a \in \mathcal{A}$ demonstrates that:

$$\mathbb{P}\left[\forall t \in [T], a \in \mathcal{A} \mid v_2(a_t, b_t) \geq \max_{b \in \mathcal{B}} v_2(a_t, b) - g(n_{t+1}(a), T)\right] \geq 1 - |\mathcal{A}| \cdot T^{-3},$$

so $\texttt{ALG}_2$ satisfies high-probability instantaneous regret g.

High-probability anytime regret. Let the time horizon be T, and suppose that $\texttt{ALG}$ satisfies high-probability anytime regret $h(t, T, \mathcal{B})$ for every $1 \leq t \leq T$. Using this combined with the fact that $\texttt{ALG}_2(a_t, H_{2,t}) = \texttt{ALG}(H_{2,t,a_t})$, we see that for each $a \in \mathcal{A}$:

$$\mathbb{P}\left[\forall t \in [T] \mid \sum_{t' \leq t \mid a_{t'} = a} \max_{b \in \mathcal{B}} v_2(a, b) - \sum_{t' \leq t \mid a_{t'} = a} v_2(a, b_{t'}) \leq h(n_{t+1}(a), T)\right] \geq 1 - T^{-3}.$$

Taking a union bound over $a \in \mathcal{A}$ demonstrates that:

$$\mathbb{P}\left[\forall t \in [T], a \in \mathcal{A} \mid \sum_{t' \leq t \mid a_{t'} = a} \max_{b \in \mathcal{B}} v_2(a, b) - \sum_{t' \leq t \mid a_{t'} = a} v_2(a, b_{t'}) \leq h(n_{t+1}(a), T)\right] \geq 1 - |\mathcal{A}| \cdot T^{-3},$$

so $\texttt{ALG}_2$ satisfies high-probability anytime regret h. □

Using Lemma 284, it suffices to analyze the high-probability instantaneous regret and high-probability anytime regret of the following standard bandit algorithms as single-bandit learners with arms $\mathcal{B}$, mean rewards $v_2(b)$, and stochastic rewards $r_{2,t}(b)$. In the proofs, we let $n_t(b)$ denote the number of times that arm b has been pulled prior to time step t.

Proposition 126. *Suppose that for every $a \in \mathcal{A}$, the follower runs a separate instantiation of $\texttt{ActiveArmElimination}(M_1, \ldots, M_P)$ (Algorithm 14) with $M_i = \Theta(\log T \cdot 2^{2i})$. Then the follower satisfies high-probability instantaneous regret $g(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot \log(T)/t}$, which implies $g(t, T, \mathcal{B}) = O\left((|\mathcal{A}||\mathcal{B}| \log T)^{1/3} T^{-1/3}\right)$ for $t \geq \Theta(|\mathcal{A}|^{-2/3}(|\mathcal{B}| \log T)^{1/3} T^{2/3})$. Moreover, the follower satisfies high-probability anytime regret $h(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot \log(T) \cdot t})$.*

Proof of Chapter 126. We first show the high-probability instantaneous regret bound and then deduce the high-probability anytime regret bound.

High-probability instantaneous regret bound. By Lemma 284, it suffices to show the bound for $\texttt{ActiveArmElimination}$ (using phase lengths $M_i = \Theta(\log(T) \cdot 2^{2i})$) as a single-bandit learner with arms $\mathcal{B}$, mean rewards $v_2(b)$, and stochastic rewards $r_{2,t}(b)$. We let $n_t(b)$ denote the number of times that arm b has been pulled prior to time step t in the current phase. Let $\hat{v}_{2,t}(b)$ denote the empirical mean reward for arm b over the rewards observed prior to time step t in the *previous* (last completed) phase. Let $\mathcal{B}'_{t,\text{curr}}$ be the set of arms active in the *current* phase, and let $\mathcal{B}'_{t,\text{prev}}$ be the set of arms active in the *previous* (last completed) phase. For each time step t, let s'_t denote the index of the *previous* (last completed) phase at time step t.

Let the clean event G denote the event that at every time step t, it holds that:

$$\forall t \in [T], b \in \mathcal{B}'_{t,\text{prev}} : |v_2(b) - \hat{v}_{2,t}(b)| \leq \frac{10\sqrt{\log T}}{\sqrt{M_{s'_t}}}.$$

Applying a Chernoff bound and a union bound, it holds that $P[G] \geq 1 - T^{-3}$.

We condition on the clean event G for the remainder of the analysis. Let $b^* = \arg\max_{b \in \mathcal{B}} v_2(b)$. Using the elimination rule, we can bound the suboptimality of each arm $b \in \mathcal{B}'_{t,\text{curr}}$:

$$\begin{aligned}
&|v_2(b^*) - v_2(b)| \\
&\leq |\hat{v}_{2,t}(b^*) - v_2(b^*)| + |\hat{v}_{2,t}(b) - v_2(b)| + |\hat{v}_{2,t}(b^*) - \hat{v}_{2,t}(b)| \\
&\leq 40\frac{\sqrt{\log(T)}}{\sqrt{M_{s'_t}}} \\
&\leq \Theta(2^{-s'_t}).
\end{aligned}$$

It suffices to lower bound $2^{-2 \cdot s'_t}$. We observe that:

$$t \leq |\mathcal{B}| \left(M_{s'_t+1} + \sum_{s=1}^{s'_t} M_s \right) \leq \Theta(|\mathcal{B}| \cdot \log(T) \cdot 2^{2 \cdot s'_t}),$$

where the last expression uses the geometric rate of increase of $M_i = \Theta(\log(T) \cdot 2^{2i})$. This implies that

$$2^{-s'_t} = O(\sqrt{|\mathcal{B}| \cdot \log T / t}).$$

Altogether, this implies that:

$$v_2(b_t) \geq \max_{b \in \mathcal{B}} v_2(b) - O(\sqrt{|\mathcal{B}| \cdot \log T / t}),$$

as desired.

High-probability anytime regret bound. Using Observation 14.7.1, it holds that the high-probability anytime regret can be bounded as:

$$\sum_{t'=1}^{t} O\left(\sqrt{\frac{\log(T) \cdot |\mathcal{B}|}{t'}}\right) = \sqrt{\log(T) \cdot |\mathcal{B}|} \cdot O\left(\sum_{t'=1}^{t} \frac{1}{\sqrt{i}}\right) \leq_{(A)} \Theta(\sqrt{\log(T) \cdot t \cdot |\mathcal{B}|})$$

where (A) follows from an integral bound and Jensen's inequality. This proves the desired bound. $\square$

Proposition 127. *Suppose that the follower runs a separate instantiation of* `ExploreThenCommit`$(E, \mathcal{B})$ *(Algorithm 8) for every $a \in \mathcal{A}$. Then, the follower satisfies high-probability instantaneous regret $g(t, T, \mathcal{B}) = \mathcal{O}(\sqrt{\log T / E})$ for all time steps $t \geq E \cdot |\mathcal{B}|$. If $E = \Theta((|\mathcal{A} \cdot |\mathcal{B}|)^{-2/3} (\log T)^{1/3} T^{2/3})$, then $g(t, T, \mathcal{B}) = O\left((|\mathcal{A}||\mathcal{B}| \log T)^{1/3} T^{-1/3}\right)$ for $t \geq \Theta(|\mathcal{A}|^{-2/3} (|\mathcal{B}| \log T)^{1/3} T^{2/3})$.*

Proof of Chapter 127. By Lemma 284, it suffices to show the instantaneous regret bound for `ExploreThenCommit` as a single-bandit learner with arms $\mathcal{B}$, mean rewards $v_2(b)$, and stochastic rewards $r_{2,t}(b)$. We let $n_t(b)$ denote the number of times that arm b has been pulled prior to time step t. Let $\hat{v}_{2,t}(b)$ denote the empirical mean reward for arm b over the rewards observed prior to time step t.

Let the clean event G capture the event that the empirical mean of every arm is close to the true mean whenever $t > E \cdot |\mathcal{B}|$ time steps, that is:

$$\forall b \in \mathcal{B}, t > E \cdot |\mathcal{B}| : |\hat{v}_{2,t}(b) - v_2(b)| \leq 10 \cdot \frac{\sqrt{\log(T)}}{\sqrt{E}}$$

Applying a Chernoff bound (and using the analogue of one of the canonical bandit models in Lattimore and Szepesvári (2020)), it holds that $P[G] \geq 1 - T^{-3}$.

Now, conditioning on the clean event G, we see that after time step $t > E \cdot |\mathcal{B}|$, it holds that:

$$|\hat{v}_{2,t}(b) - v_2(b)| \leq 10 \frac{\sqrt{\log(T)}}{\sqrt{E}}.$$

Since the algorithm chooses the arm with highest empirical mean from the first $E \cdot |\mathcal{B}|$ time steps is selected, this means that:

$$\max_{b \in \mathcal{B}} v_2(b) - v_2(b) \leq 20 \cdot \frac{\sqrt{\log(T)}}{\sqrt{E}}$$

for any $t > E \cdot |\mathcal{B}|$. □

Proposition 128. *Suppose that the follower runs a separate instantiation of UCB for every $a \in \mathcal{A}$. Then, the follower satisfies high-probability anytime regret bound $h(t, T, \mathcal{B}) = O(\sqrt{|\mathcal{B}| \cdot t \cdot \log(T)})$.*

Proof of Chapter 128. By Lemma 284, it suffices to show the anytime regret bound for UCB as a single-bandit learner with arms $\mathcal{B}$, mean rewards $v_2(b)$, and stochastic rewards $r_{2,t}(b)$. We let $n_t(b)$ denote the number of times that arm b has been pulled prior to time step t. Let $\hat{v}_{2,t}(b)$ denote the empirical mean reward for arm b over the rewards observed prior to time step t.

We define the *clean event* G as the true mean being contained within the upper and lower confidence bounds for each arm a, that is:

$$\forall b \in \mathcal{B}, t \leq T : |\hat{v}_{2,t}(b) - v_2(b)| \leq 10 \cdot \sqrt{\frac{\log(T)}{n_t(b)}}.$$

By a Chernoff bound (and using the analogue of one of the canonical bandit models in Lattimore and Szepesvári (2020)) followed by a union bound, we have that $P[G] \geq 1 - T^{-3}$.

We condition on G for the remainder of the analysis. Since the arm with highest upper confidence bound is always chosen and since G holds, the selected arm b_t's true mean $v_2(b_t)$ falls within the $2 \cdot \sqrt{\frac{\log(T)}{n_t(b_t)}}$ bound. By Lemma 269, this means that the regret at any time step t for any arm $a \in \mathcal{A}$ is upper bounded by:

$$10 \cdot \sum_{t'=1}^{t} \sqrt{\frac{\log(T)}{n_{t'}(b_{t'})}} \leq 10 \cdot \sqrt{\log(T) \cdot |\mathcal{B}| \cdot t}$$

as desired. $\square$

J.6.2 Proofs for Chapter 14.7.2

We prove Theorem 129 in Appendix J.6.2 and we prove Theorem 130 in Appendix J.6.2.

Proof of Theorem 129

We prove Theorem 129, following a similar argument to the proof of Theorem 119.

Theorem 129. *Let $c_1 \in (0,1)$ and $c_2, c_3 > 0$. Let $E = \Theta(|\mathcal{A}|^{-1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)}(\log T)^{c_3/(1+c_1)} \cdot T^{1/(1+c_1)})$. Consider a StrongDSG where $\texttt{ALG}_2$ is any algorithm with high-probability instantaneous regret $g(t,T,\mathcal{B}) = O\left(E^{-c_1}|\mathcal{B}|^{c_2}(\log T)^{c_3}\right)$ for $t > E$ and $g(t,T,\mathcal{B}) = 1$ for $t \leq E$, and where $\texttt{ALG}_1 = \texttt{ExploreThenUCB}(E)$. Then, it holds that the regret $\max(R_1(T), R_2(T))$ with respect to the γ-tolerant benchmarks β_1^{tol} and β_2^{tol} is bounded as:*

$$O\left(T^{1/(1+c_1)} \cdot |\mathcal{A}|^{c_1/(1+c_1)} \cdot |\mathcal{B}|^{c_2/(1+c_1)} \cdot (\log T)^{c_3/(1+c_1)}\right) + \Theta\left(\sqrt{T|\mathcal{A}|\log T}\right).$$

We assume $\gamma = \omega\left(|\mathcal{A}|^{c_1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)}(\log T)^{c_3/(1+c_1)}T^{-1/(1+c_1)}\right)$.

Notation and Clean Event. We use the same notation as in the proof of Theorem 119. We also define the clean event $G := G_L \cap G_F$ to be the same as in the proof of Theorem 119.

We prove that the clean event G occurs with high probability, generalizing Lemma 276.

Lemma 285. *Assume the notation above. Let $\texttt{ALG}_2$ be any algorithm with high-probability instantaneous regret g where $g(t,T,\mathcal{B}) = O(E^{-c_1}|\mathcal{B}|^{c_2}(\log T)^{c_3})$ for $t > E$ and $g(t,T,\mathcal{B}) = 1$ for $t \leq E$, and let $\texttt{ALG}_1 = \texttt{ExploreThenUCB}(E)$. Then, the event G occurs with high probability: $\mathbb{P}[G] \geq 1 - T^{-3}(|\mathcal{A}| + 1)$.*

Proof. We first show that $\mathbb{P}[G_F] \geq 1 - |\mathcal{A}| \cdot T^{-3}$. A sufficient condition for this event to hold is that:

$$\forall t > E \cdot |\mathcal{A}| : v_2(a_t, b_t) \geq \max_{b \in \mathcal{B}} v_2(a_t, b) - \max_{t > E} g(t, T, \mathcal{B}).$$

Since the exploration phases pulls every arm $a \in \mathcal{A}$ a total of E times, the high-probability instantaneous regret assumption guarantees that this event holds with probability at least $1 - |\mathcal{A}| \cdot T^{-3}$, as desired.

We next show that $\mathbb{P}[G_L] \geq 1 - T^{-3}$. This follows from a Chernoff bound (and using the analogue of one of the canonical bandit models in Lattimore and Szepesvári (2020)) combined with a union bound.

The lemma follows from another union bound over G_L and G_F. $\square$

Now we are ready to prove Theorem 129.

Proof of Theorem 129. Assume that the clean event G holds. This occurs with probability at least $1 - (1 + |\mathcal{A}|)T^{-3}$ (Lemma 285), so the clean event not occurring counts negligibly towards regret.

The regret in the explore phase is bounded by $O(1)$ in each round, the total regret from that phase is $O(T^{1/(1+c_1)}|\mathcal{A}|^{c_1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)}(\log T)^{c_3/(1+c_1)})$ for either player.

The remainder of the analysis boils down to bounding the regret in the UCB phase. We separately analyze the regret of the leader and the follower. Observe that $\varepsilon^* = \max_{t>E} g(t, T, \mathcal{B}) = O\left(|\mathcal{B}|^{c_2}(\log T)^{c_3}E^{-c_1}\right) = O\left(|\mathcal{A}|^{c_1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)}(\log T)^{c_3/(1+c_1)}T^{-c_1/(1+c_1)}\right)$ is based on the assumption on the follower's algorithm.

Regret for the leader. We use a similar analysis as in the proof of Theorem 119, repeating the full analysis for completeness.

$$
\begin{aligned}
&\beta_1^{\mathrm{tol}} \cdot (T - E \cdot |\mathcal{A}|) - \sum_{t=E\cdot|\mathcal{A}|+1}^{T} v_1(a_t, b_t) \\
&\leq \sum_{t=E\cdot|\mathcal{A}|+1}^{T} \left(\varepsilon^* + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t,b_t)\right) \\
&= \sum_{a\in\mathcal{A}} \sum_{t\in S(a)} \left(\varepsilon^* + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t,b_t)\right) \\
&\leq |\mathcal{A}| + \sum_{a\in\mathcal{A}} \sum_{t\in S(a)\setminus\{\max(S(a))\}} \left(\varepsilon^* + \max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - v_1(a_t,b_t)\right) \\
&\leq |\mathcal{A}| + \underbrace{\varepsilon^* \cdot T}_{(1)} \\
&\quad + \underbrace{\sum_{a\in\mathcal{A}} (n_{E\cdot|\mathcal{A}|,T+1}(a) - 1) \cdot \left(\max_{a\in\mathcal{A}} \min_{b\in\mathcal{B}_{\varepsilon^*}(a)} v_1(a,b) - \frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a) - 1} \sum_{t\in S(a)\setminus\{\max(S(a))\}} (v_1(a_t,b_t))\right)}_{(2)}
\end{aligned}
$$

The term $|\mathcal{A}|$ computes negligibly, term (1) is equal to

$$\Theta(\mathcal{A}|^{c_1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)}(\log T)^{c_3/(1+c_1)}T^{1/(1+c_1)}),$$

and term (2) can be bounded by:

$$
\begin{aligned}
&\sum_{a\in\mathcal{A}}(n_{E\cdot|\mathcal{A}|,T}(a)-1)\cdot\left(\max_{a\in\mathcal{A}}\min_{b\in\mathcal{B}_{\varepsilon^*}(a)}v_1(a,b)-\frac{1}{n_{E\cdot|\mathcal{A}|,T+1}(a)-1}\sum_{t\in S(a)\setminus\{\max(S(a))\}}(v_1(a_t,b_t))\right)\\
&\leq\sum_{a\in\mathcal{A}}(n_{E\cdot|\mathcal{A}|,T+1}(a)-1)\cdot\frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a)}-1}\\
&\leq O\left(\sqrt{\log T}\cdot\sum_{a\in\mathcal{A}}\sqrt{n_{E\cdot|\mathcal{A}|,T+1}(a)-1}\right)\\
&\leq O\left(\sqrt{|\mathcal{A}|T\log T}\right),
\end{aligned}
$$

where the first inequality uses Lemma 277 and the last inequality uses Jensen's inequality.

Regret for the follower. We modify the analysis from the proof of Theorem 119. We bound the regret as:

$$
\begin{aligned}
&\beta_2^{\text{tol}}\cdot(T-E\cdot|\mathcal{A}|)-\sum_{t=E\cdot|\mathcal{A}|+1}^{T}v_2(a_t,b_t)\\
&\leq\underbrace{\sum_{t=E|\mathcal{A}|+1}^{T}\left(\min_{a\in\mathcal{A}_{\varepsilon_t}}\max_{b\in\mathcal{B}}v_2(a,b)-v_2(a_t,b_t)\right)}_{(1)}+\underbrace{\sum_{t=E|\mathcal{A}|+1}^{T}\varepsilon_t}_{(2)}
\end{aligned}
$$

where

$$
\varepsilon_t=\begin{cases}1 & \text{if } n_{E\cdot|\mathcal{A}|,t}(a_t)=1\\ \max\left(\varepsilon^*,20\frac{\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a_t)}}\right) & \text{else .}\end{cases}
$$

We bound term (1). We first show that $a_t \in \mathcal{A}_{\varepsilon_t}$:

$$\begin{aligned}
\max_{b \in \mathcal{B}_{\varepsilon_t}(a_t)} v_1(a_t, b) &\geq \max_{b \in \mathcal{B}_{\varepsilon^*}(a_t)} v_1(a_t, b) \\
&\geq_{(A)} \frac{1}{n_{E\cdot|\mathcal{A}|,t}(a_t)} \sum_{E\cdot|\mathcal{A}|<t'<t|a_{t'}=a_t} v_1(a_{t'}, b_{t'}) \\
&\geq_{(B)} \hat{v}_{1,t}(a_t) - \frac{10\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a_t)}} \\
&= v_{1,t}^{\mathrm{UCB}}(a_t) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a_t)}} \\
&= \max_{a \in \mathcal{A}} \left(v_{1,t}^{\mathrm{UCB}}(a)\right) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a_t)}} \\
&\geq_{(C)} \max_{a \in \mathcal{A}} \left(\frac{1}{n_{E\cdot|\mathcal{A}|,t}(a_t)} \sum_{E\cdot|\mathcal{A}|<t'<t|a_{t'}=a} v_1(a_{t'}, b_{t'}) \right) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a_t)}} \\
&\geq_{(D)} \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon^*}(a)} v_1(a, b) - \frac{20\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a_t)}} \\
&\geq \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}_{\varepsilon_t}(a)} v_1(a, b) - \varepsilon_t.
\end{aligned}$$

where (A) and (D) uses the event G_F, and (B) and (C) use the event G_L. Applying G_F again, this implies that:

$$\begin{aligned}
\min_{a \in \mathcal{A}_{\varepsilon_t}} \max_{b \in \mathcal{B}} v_2(a, b) - v_2(a_t, b_t) &\leq v_2(a_t, b) - v_2(a_t, b_t) \\
&\leq \varepsilon^*.
\end{aligned}$$

Putting this all together, term (1) is bounded by

$$\begin{aligned}
\sum_{t=E|\mathcal{A}|+1}^{T} \left(\min_{a \in \mathcal{A}_{\varepsilon_t}} \max_{b \in \mathcal{B}} v_2(a, b) - v_2(a_t, b_t) \right) &\leq \varepsilon^* \cdot T \\
&= \Theta(\mathcal{A}|^{c_1/(1+c_1)} |\mathcal{B}|^{c_2/(1+c_1)} (\log T)^{c_3/(1+c_1)} T^{1/(1+c_1)}).
\end{aligned}$$

We next bound term (2) as follows:

$$\begin{aligned}
\sum_{t=E|\mathcal{A}|+1}^{T} \varepsilon_t &= |\mathcal{A}| + \sum_{a\in\mathcal{A}} \sum_{t\in S(a)\setminus \min(S(a))} \max\left(\varepsilon^*, 20\frac{\sqrt{\log T}}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a)}}\right) \\
&\leq |\mathcal{A}| + \varepsilon^* \cdot T + 20\sqrt{\log T} \cdot \sum_{a\in\mathcal{A}} \sum_{t\in S(a)\setminus \min(S(a))} \frac{1}{\sqrt{n_{E\cdot|\mathcal{A}|,t}(a)}} \\
&\leq_{(A)} |\mathcal{A}| + \varepsilon^* \cdot T + O\left(\sqrt{T|\mathcal{A}|\log T}\right) \\
&\leq \Theta(\mathcal{A}|^{c_1/(1+c_1)}|\mathcal{B}|^{c_2/(1+c_1)}(\log T)^{c_3/(1+c_1)}T^{1/(1+c_1)}) + O\left(\sqrt{T|\mathcal{A}|\log T}\right),
\end{aligned}$$

where (A) uses Lemma 269.

Putting this all together yields the desired bound. □

Proof of Theorem 130

We prove Theorem 130, following a similar approach to the proof of Theorem 121.

Theorem 130. *Let $c_1 \in (0,1)$, $c_2, c_3 > 0$, and $C' > 0$ be arbitrary constants. Consider a StrongDSG where $\mathcal{I} = (\mathcal{A}, \mathcal{B}, v_1, v_2)$ has Lipschitz constant L^*. Let $\texttt{ALG}_2$ be any algorithm satisfying high-probability anytime regret $h(t,T,\mathcal{B}) = C' \cdot t^{c_1} \cdot |\mathcal{B}|^{c_2} \cdot (\log(T))^{c_3}$. Let $\texttt{ALG}_1 = \texttt{LipschitzUCBGen}(L, C'B^{c_2}, c_1, c_3)$ for any $L \geq L^*$. Then both players achieve the following regret bounds with respect to the original Stackelberg benchmarks β_1^{orig} and β_2^{orig}: that is, $R_1(T;\mathcal{I}) = O\left(\sqrt{T|\mathcal{A}||\mathcal{B}|\log T} + L|\mathcal{A}|^{1-c_1}|\mathcal{B}|^{c_2}(\log T)^{c_3}T^{c_1}\right)$ and $R_2(T;\mathcal{I}) = O\left(L\sqrt{T|\mathcal{A}||\mathcal{B}|\log T} + L^2|\mathcal{A}|^{1-c_1}|\mathcal{B}|^{c_2}T^{c_1}(\log T)^{c_3}\right)$.*

Notation. We use the same notation as in the proof of Theorem 121.

Clean event. We again define the clean event $G = G_L \cap G_F$ to be the intersection of a clean event G_L for the leader and a clean event G_F for the follower. The event G_L is the same as in the proof of Theorem 121. The event G_F is formalized as follows:

$$\forall a \in \mathcal{A}, t \leq T : \sum_{1\leq t'<t|a_t=a} (v_2(a, b^*(a)) - v_2(a_t, b_t)) \leq C'(n_t(a))^{c_1}|\mathcal{B}|^{c_2}(\log T)^{c_3}$$

We first generalize Lemma 278.

Lemma 286. *Assume the setup of Theorem 121 and the notation above. Then the clean event occurs with high probability: $\mathbb{P}[G] \geq 1 - T^{-3}(|\mathcal{A}| + 1)$.*

Proof. We union bound over G_L and G_F. The analysis for G_F follows from the high-probability anytime regret bound assumption. The analysis for G_L follows from a Chernoff bound (and using the analogue of one of the canonical bandit models in Lattimore and Szepesvári (2020)) combined with a union bound. □

The following lemma generalizes Lemma 279.

Lemma 287. *Assume the setup of Theorem 130 and the notation above. Suppose that the clean event G holds. Then for any $t \leq T$ and $a \in \mathcal{A}$, it holds that:*

$$|\hat{v}_{1,t}(a) - v_1(a, b^*(a))| \leq \frac{10\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a)}} + C' \cdot L \cdot (n_t(a))^{c_1 - 1} \cdot |\mathcal{B}|^{c_2} (\log T)^{c_3}.$$

Proof. The proof follows similarly to the proof of Lemma 279. We observe that:

$$\begin{aligned}
&|\hat{v}_{1,t}(a) - v_1(a, b^*(a))| \\
&= \left| \left(\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot \hat{v}_1(a,b) \right) - v_1(a, b^*(a)) \right| \\
&= \left| \left(\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot \hat{v}_1(a,b) \right) - \frac{1}{n_t(a)} \left(\sum_{b \in \mathcal{B}} n_t(a,b) \cdot v_1(a, b^*(a)) \right) \right| \\
&\leq \frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |\hat{v}_1(a,b) - v_1(a, b^*(a))| \\
&\leq \underbrace{\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |\hat{v}_1(a,b) - v_1(a,b)|}_{(A)} + \underbrace{\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |v_1(a,b) - v_1(a, b^*(a))|}_{(B)}.
\end{aligned}$$

The bound of term (A) proceeds the same as before, and repeat the proof for completeness:

$$\begin{aligned}
\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |\hat{v}_1(a,b) - v_1(a,b)| &\leq_{(1)} \frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot \frac{10\sqrt{\log T}}{\sqrt{n_t(a,b)}} \\
&= \frac{10\sqrt{\log T}}{n_t(a)} \sum_{\sum_{b \in \mathcal{B}}} \sqrt{n_t(a,b)} \\
&\leq_{(2)} \frac{10\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a)}}.
\end{aligned}$$

where (1) uses the clean event G_L and (2) uses Jensen's inequality.

The bound of term (B) proceeds similarly, with some minor modifications:

$$\begin{aligned}
\frac{1}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |v_1(a,b) - v_1(a, b^*(a))| &\leq_{(1)} \frac{L^*}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot |v_2(a,b) - v_2(a, b^*(a))| \\
&=_{(2)} \frac{L^*}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot (v_2(a, b^*(a)) - v_2(a,b)),
\end{aligned}$$

where (1) uses the Lipschitz property and (2) uses the fact that $b^*(a)$ is the best arm for the follower, given that the leader pulls arm a. Using the clean event G_F and that $L \geq L^*$, we see that:

$$\begin{aligned}\frac{L^*}{n_t(a)} \sum_{b \in \mathcal{B}} n_t(a,b) \cdot (v_2(a, b^*(a)) - v_2(a,b)) &= \frac{L^*}{n_t(a)} \sum_{1 \leq t' < t | a_t = a} (v_2(a, b^*(a)) - v_2(a_t, b_t)) \\ &\leq C' \cdot L \cdot (n_t(a))^{c_1 - 1} |\mathcal{B}|^{c_2} (\log T)^{c_3}.\end{aligned}$$

Taken together, these terms give the desired bound. □

We next generalize Lemma 280.

Lemma 288. *Assume the setup of Theorem 130 and the notation above. Suppose that the clean event G holds. Then it holds that:*

$$B_1 := \sum_{t=1}^{T} \left(\frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot |\mathcal{B}| \frac{\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} \right) \leq O\left(\sqrt{|\mathcal{A}||\mathcal{B}| T \log T} + L \cdot |\mathcal{A}|^{1-c_1} |\mathcal{B}|^{c_2} (\log T)^{c_3} T^{c_1} \right)$$

$$B_2 := \sum_{t=1}^{T} (v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)) \leq O\left(|\mathcal{A}|^{1-c_1} |\mathcal{B}|^{c_2} \cdot (\log T)^{c_3} T^{c_1} \right)$$

Proof. To bound B_2, we see that:

$$\begin{aligned} B_2 &= \sum_{t=1}^{T} (v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)) \\ &= \sum_{a \in \mathcal{A}} \sum_{t \in T | a_t = a} (v_2(a, b^*(a)) - v_2(a, b_t)) \\ &\leq_{(A)} \sum_{a \in \mathcal{A}} C'(n_T(a))^{c_1} |\mathcal{B}|^{c_2} \cdot (\log T)^{c_3} \\ &= C' |\mathcal{B}|^{c_2} \cdot (\log T)^{c_3} \cdot \sum_{a \in \mathcal{A}} (n_{T+1}(a))^{c_1} \\ &\leq_{(B)} O\left(|\mathcal{A}|^{1-c_1} |\mathcal{B}|^{c_2} \cdot (\log T)^{c_3} T^{c_1} \right) \end{aligned}$$

where (A) uses the event G_F and (B) uses Jensen's inequality.

To bound B_1:

$$
\begin{aligned}
B_1 &= \sum_{t=1}^{T} \left(\frac{10\sqrt{|\mathcal{B}| \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot |n_t(a_t)|^{c_1-1} |\mathcal{B}|^{c_2} (\log T)^{c_3} \right) \\
&=_{(A)} O\left(\sqrt{|\mathcal{A}||\mathcal{B}|T \log T}\right) + C' \cdot L \cdot |\mathcal{B}|^{c_2} (\log T)^{c_3} \cdot \sum_{t=1}^{T} |n_t(a_t)|^{c_1-1} \\
&=_{(A)} O\left(\sqrt{|\mathcal{A}||\mathcal{B}|T \log T}\right) + C' \cdot L \cdot |\mathcal{B}|^{c_2} (\log T)^{c_3} \cdot \sum_{a \in \mathcal{A}} \sum_{t | a_t = a} |n_t(a_t)|^{c_1-1} \\
&\leq_{(B)} O\left(\sqrt{|\mathcal{A}||\mathcal{B}|T \log T} + L \cdot |\mathcal{B}|^{c_2} (\log T)^{c_3} \cdot \sum_{t | a_t = a} |n_{T+1}(a)|^{c_1} \right) \\
&\leq_{(C)} O\left(\sqrt{|\mathcal{A}||\mathcal{B}|T \log T} + L \cdot |\mathcal{A}|^{1-c_1} |\mathcal{B}|^{c_2} (\log T)^{c_3} T^{c_1}\right).
\end{aligned}
$$

where (A) follows from Lemma 269, (B) follows from an integral bound, and (C) follows from Jensen's inequality.

□

We now prove Theorem 121.

Proof of Theorem 130. Assume that clean event G holds. This occurs with probability at least $1-(|\mathcal{A}+1)T^{-3}$ (Lemma 286), so the clean event not occurring counts negligibly towards regret.

Moreover, let $(a^*, b^*(a^*))$ be the Stackelberg equilibrium. Let

$$
\alpha_t(a) = \frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a)}} + C' \cdot L \cdot |n_t(a)|^{c_1-1} |\mathcal{B}|^{c_2} (\log T)^{c_3}
$$

be the confidence bound size at time step t and let $v_{1,t}^{\mathrm{UCB}}(a) = \hat{v}_{1,t}(a) + \alpha_t(a)$ denote the UCB estimate in `LipschitzUCBGen`(L, C) computed during time step t prior to reward at time step t being observed.

We can bound the leader's regret as:

$$
\begin{aligned}
R_1(T) &= \sum_{t=1}^{T}(v_1(a^*, b^*(a^*)) - v_1(a_t, b_t)) \\
&= \sum_{t=1}^{T} (v_1(a^*, b^*(a^*)) - v_1(a_t, b^*(a_t))) + \sum_{t=1}^{T} (v_1(a_t, b^*(a_t)) - v_1(a_t, b_t)) \\
&\leq_{(A)} \sum_{t=1}^{T} (\hat{v}_1(a^*) + \alpha_t(a^*) - \hat{v}_1(a_t) + \alpha_t(a_t)) + \sum_{t=1}^{T} |v_1(a_t, b^*(a_t)) - v_1(a_t, b_t)| \\
&\leq \sum_{t=1}^{T} \left(v_{1,t}^{\mathrm{UCB}}(a^*) - v_1^{\mathrm{UCB}}(a_t) + 2 \cdot \alpha_t(a_t)\right) + L \cdot \sum_{t=1}^{T} |v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)| \\
&\leq 2 \cdot \sum_{t=1}^{T} \alpha_t(a_t) + L \cdot \sum_{t=1}^{T} (v_2(a_t, b^*(a_t)) - v_2(a_t, b_t)) \\
&= 2 \cdot \sum_{t=1}^{T} \left(\frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot |n_t(a_t)|^{c_1 - 1} |\mathcal{B}|^{c_2} (\log T)^{c_3} \right) + L \cdot B_2 \\
&= 2 \cdot B_1 + L \cdot B_2 \\
&\leq_{(B)} O\left(\sqrt{|\mathcal{A}||\mathcal{B}|T \log T} + L \cdot |\mathcal{A}|^{1-c_1} |\mathcal{B}|^{c_2} (\log T)^{c_3} T^{c_1}\right)
\end{aligned}
$$

where (A) uses Lemma 287 and (B) uses Lemma 288.

We also bound the follower's regret as:

$$
\begin{aligned}
R_2(T) &= \sum_{t=1}^{T} \left(v_2(a^*, b^*(a^*)) - v_2(a_t, b_t)\right) \\
&= \sum_{t=1}^{T} \left(v_2(a^*, b^*(a^*)) - v_2(a^*, b^*(a_t))\right) + \sum_{t=1}^{T} \left(v_2(a^*, b^*(a_t)) - v_2(a_t, b_t)\right) \\
&= \sum_{t=1}^{T} L \cdot |v_1(a^*, b^*(a^*)) - v_1(a_t, b^*(a_t))| + B_2 \\
&=_{(A)} \sum_{t=1}^{T} L \cdot (v_1(a^*, b^*(a^*)) - v_1(a_t, b^*(a_t))) + B_2 \\
&\leq_{(B)} \sum_{t=1}^{T} L \cdot (\hat{v}_{1,t}(a^*) + \alpha_t(a^*) - \hat{v}_{1,t}(a_t) + \alpha_t(a^*)) + B_2 \\
&= \sum_{t=1}^{T} L \cdot \left(v_{1,t}^{\mathrm{UCB}}(a^*) - v_{1,t}^{\mathrm{UCB}}(a_t) + 2 \cdot \alpha_t(a_t)\right) + B_2 \\
&\leq \sum_{t=1}^{T} L \cdot (2 \cdot \alpha_t(a_t)) + B_2 \\
&= 2L \cdot \sum_{t=1}^{T} \left(\frac{10\sqrt{\mathcal{B} \log T}}{\sqrt{n_t(a_t)}} + C' \cdot L \cdot |n_t(a_t)|^{c_1 - 1} |\mathcal{B}|^{c_2} (\log T)^{c_3} \right) + B_2 \\
&= 2L \cdot B_1 + B_2 \\
&\leq_{(C)} O\left(L\sqrt{|\mathcal{A}||\mathcal{B}|T \log T} + L^2 \cdot |\mathcal{A}|^{1-c_1} |\mathcal{B}|^{c_2} (\log T)^{c_3} T^{c_1} \right)
\end{aligned}
$$

where (A) uses the fact that a^* is the action chosen by the leader at the Stackelberg equilibrium where (B) uses Lemma 287 and (C) uses Lemma 288.

$\square$