TripAware: Emotional and Informational Approaches to Encourage Sustainable Transportation via Mobile Applications

Jesse Zhang
John Sullivan
Vasudev Venkatesh P. B.
Kyle Tse
Andy Yan
John Leyden
Kalyanaraman Shankari
Randy H. Katz

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2019-2
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-2.html

January 11, 2019

Copyright © 2019, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Acknowledgement

We thank Prof. David Culler, for providing routine feedback and guidance with analyzing our results. We also thank Prof. Ani Adhikari, for her statistical guidance in our analysis, Prof. Eric Paulos for guidance in writing this paper, and Simon Behar for helping us design the experiment and set up our servers. Finally, much thanks to the URAP program and Stephanie for making this collaboration possible!

TripAware: Emotional and Informational Approaches to Encourage Sustainable Transportation via Mobile Applications

Jesse Zhang
UC Berkeley
jessezhang@berkeley.edu

Jack Sullivan
UC Berkeley
jsullivan27@berkeley.edu

Vasudev Venkatesh P.B.
UC Berkeley
vas.venk@berkeley.edu

Kyle Tse
UC Berkeley
tse.kylekobie@berkeley.edu

Andy Yan
UC Berkeley
yan.andy4@berkeley.edu

John Leyden
UC Berkeley
jleyden@berkeley.edu

Kalyanaraman Shankari
UC Berkeley
shankari@eecs.berkeley.edu

Randy Katz
UC Berkeley
randy@cs.berkeley.edu

ABSTRACT

This exploratory study investigates how mobile applications assist commuters with sustainable transportation choices. Our goal is to determine persuasiveness of two broad categories of features: emotional or informational. A controlled trial randomly assigned 41 users to three mobile applications: *Emotion*, *Information*, *Control*. During the ten week study, we recorded user interactions and changes in transportation habits.

Several features distinguish this study from prior work, the most salient of which are: (1) automatic trip recording, segmentation, and classification; (2) statistical assessment of metrics that reflect a user's interactions and behaviors; (3) larger and more diverse samples.

Using hypothesis testing, we found that *Emotion* resulted in greater engagement with the application ($p = 0.006, 0.035, 0.031, 0.040$) while *Information* improved the sustainability of travel behavior ($p = 0.043$). This suggests a combination of both approaches is required in order to both maintain user engagement and have an effect on carbon emissions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2019, May 2019, Glasgow, UK

© 2018 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Jesse Zhang, Jack Sullivan, Vasudev Venkatesh P.B., Kyle Tse, Andy Yan, John Leyden, Kalyanaraman Shankari, and Randy Katz. 2018. TripAware: Emotional and Informational Approaches to Encourage Sustainable Transportation via Mobile Applications. In *Proceedings of ACM CHI conference (CHI 2019).* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Among all sectors of the United States economy, transportation is the largest contributor of CO_2 emissions, responsible for 28.5% of total emissions [1]. One approach to reduce transportation emissions is through traveler behavior change, and the majority of Americans are considering this – 66% of people in the United States agree that major lifestyle changes are required to combat climate change [6]. Despite this inclination, psychological barriers, biases, misinformation, and lack of substantial incentives often reduce the likelihood of people taking less carbon-intensive commutes.

As of 2018 in the United States, 77% of people own smartphones, up from the 35% in Pew Research Center's first survey of smartphone ownership in 2011 [6]). The ubiquity of smartphones and their increasing set of features have introduced an unobtrusive way of collecting mobility data. In this paper, we explore two approaches for changing transportation habits through the E-mission platform: a mobile application created by K. Shankari that collects human travel data [17]. Our application, *TripAware*, was built upon the E-mission platform, allowing it to seamlessly track and classify user trips in the background on a secure server.

An important factor that we considered when analyzing existing emissions-reduction literature was the variety of experimental approaches. Current literature on changing transportation behavior lack a strong connection to a larger model of human behavior. This often leads to features with conflicting psychological mechanisms for behavioral change, such as "rational" vs "moral." Thus, we choose to compare two broad behavioral approaches: Informational and Emotional. This is premised upon the behavior change wheel [12] and the CALO-RE [9] taxonomies, which categorize various sources of certain behaviors and their corresponding methods of change. After considering the different types of behavior change strategies, we sought to focus on four overarching policies: relatable effort, social pressure, habit modification, and historical insight.

In this paper, we outline our two test applications, *Information* and *Emotion*, which calculate carbon emissions from transportation and utilize behavior-changing features in a randomized controlled experiment. The paper begins with discussing our motivation for the groupings and related literature. This is followed by the challenges and limitations within our experiment, our design decisions, and finally our results.

After hypothesis testing, we find that emotional persuasion–as measured by overall interaction and retention duration on *Emotion* compared to control–resulted in improved engagement with the app. We also find that the informational approach improved the sustainability of travel behavior as measured by the change in carbon emissions for users *Information* compared to control.

2 INFLUENCING HUMAN BEHAVIOR

During the design stage of our experiment, we focused heavily on rooting our design decisions in a coherent model of human behavior. The behavior change wheel (BCW) addressed our needs for such behavioral models, matching types of behaviors with types of interventions suited for them [12]. As a result, we constructed our experiment around the theory of the BCW, choosing to contrast two fundamentally different approaches to changing transportation habits.

Within the BCW framework, mobile applications can be best categorized under the communication/marketing policy, defined as “using print, electronic, telephonic or broadcast media." We link this policy to the BCW’s sources of behavior. Each source outlines the specific psychological mechanism through which a certain behavior is created and reinforced.

We focus our study upon two main sources of behavior, namely *automatic motivation* and *reflective motivation*. Automatic motivation is derived from learning to associate certain behaviors with positive or negative feelings, thereby creating impulses or counter-impulses towards a specific target behavior (in our case, switching the user to environmentally friendly modes of transportation). Reflective motivation is motivation derived from a greater understanding of the impact of one’s behavior and knowledge of alternatives available to change said behavior.

Our primary goal is to explore the effectiveness of reflective motivation versus automatic motivation to alter users’ transportation habits towards more sustainable modes of transportation. This comparison arises naturally within the context of emissions reduction: how much will people change their transportation habits if they realize there exist sustainable alternatives, as compared to if they are taught to feel negatively about unsustainable modes of transportation and positively about sustainable ones? The two questions are mutually exclusive. The former—the informational approach—presumes that the people generally want to change their transportation behavior, and just need to be shown how. The latter—the emotional approach— instead attempts to make people emotionally driven towards changing their transportation habits.

3 RELATED WORK

There has been ample prior work on various forms of persuasive sustainability. Here we outline similar approaches taken to influence and track carbon-emitting transportation behaviors. Examples of prior persuasion techniques include tracking individual progress, visual feedback, social pressure, and gamification. We use the BCW to select a subset of these approaches that are based on different psychological mechanisms. We then design an experiment to elicit the empirical evidence that is lacking from prior qualitative studies on small sample sizes.

General Applications for Emissions Reduction

From5To4 [5] is a web tool that aims to reduce energy consumption of commuter and business trips. Through gamification and an informational dashboard, it nudges individuals and employees to modify their travel behaviors and use sustainable modes of transportation for a fraction of their trips. This community based model maps regions in terms of aggregate emissions and transportation information, but requires manually recording trips. TripAware is able to automatically track trips to reduce the burden on the user.

Smartphone Applications for Emissions Reduction

In the past decade, a few research projects have explored fostering sustainable behavior by providing feedback through smartphone mobility tracking applications. The basis of this domain of study was Ubigreen [8] which explored how visual feedback of a user’s transportation behaviors would help to support greener transportation choices. As the first study of this type, Ubigreen had limitations such as a small sample size ($N = 13$), a lack of interactions between social groups, and not displaying carbon savings information.

Many studies that focus on sustainable transportation persuasion in mobile applications do not necessarily target individual carbon emissions. A large sample size was collected by Bucher et al. through an already existing fitness tracking app called *Moves* to identify fitness activities with high accuracy and generic daily transport routes and activities [3]. However, the main focus was on transportation mode and trip tracking instead of empirically assessing behavior change.

Individualized carbon data was presented through Zapico et. al's experiment with changing behavior towards low-carbon lifestyles [21], but this effort still lacked quantitative results such as change in emissions or transportation. MatkaHupi [11] utilizes gamification to engage and motivate users to change their behaviors sustainably and included quantitative results for interaction but still not for emissions change. Ecoisland [20] focuses on interventions based on persuasion and social pressure, but still lacks quantitative emissions changes.

Evaluating Persuasive Technologies

Sustainably Unpersuaded [2] and *Landscape of Sustainable HCI* [7] map out larger trends within HCI, identifying common approaches and flaws in research methodology. *Sustainability Unpersuaded* questions the premise of this study, arguing that "[proposing] technical solutions to social problems" is in fact, "not producing solutions." However this study indicates that we may not want to rule out technical solutions so quickly. With smartphones now using a significant portion of our total attention, there exists greater potential for mobile applications to sculpt our daily habits by becoming more 'social' and integrated within human existence. In other words, it becomes increasingly difficult to distinguish between technological and social solutions.

Advancing Current Literature

Sample Size. An issue with some papers that used technology to study emission reduction was their sample sizes. Takayama et al. (2009) had a sample size of 20 participants. Jylhä et al. (2013) had a sample size of just 7. While we would have preferred to have a larger sample size, 41 users strengthens our statistical tests.

Lack of Emissions Change Results. None of the papers that implemented an application similar to ours measured the actual change in users' emissions. We measure this and offer analysis in the *Analysis* section.

Application Design Considerations

Reviewing related studies gave us insights into what types of persuasion features have been tried and current limitations of sustainability persuasion studies. With the flexibility of the E-mission platform and modern smartphones, we reduced user friction significantly while simultaneously further improving upon successful features.

The features we chose to implement within our application were inspired largely from related studies. We looked at different features and chose ones that implemented the specific interventions we desired. For example, in response to Ubigreen's future work, we aimed to implement user and community interaction and individualized user carbon information through our implemented tiered system and information dashboard presented in section 4. But we were also inspired by persuasion efforts such as Zapico et. al's polar bears and utilized them for our central feature for our *Emotion* application.

Ultimately, our study is focused on comparing sources of behavioral change, not inventing radically innovative ideas. Consequently, we decided that we should build on existing platforms and ideas, like the polar bear. *TripAware* uses the open source E-mission platform which gives us transportation classification and flexibility to alter different parts of the application–from server-side data processing algorithms to front-end UI changes. We also used components from the default version of E-mission for our information application such as the last week statistics on the dashboard (Figure 3).

4 IMPLEMENTATION

The primary goal of this paper is to explore different types of mobile features and their effects on user travel behavior. From previous literature and the BCW, we identified two main types of persuasive feature groups: *Information* and *Emotion*. *Information* features primarily focused on providing direct feedback on user behavior while *Emotion* features leveraged people's emotions as a motivating factor. We categorized the study into three groups (*Control*, *Information*, and *Emotion*) and developed different versions of the existing E-mission application—one for each experiment group. The following section will discuss the various features that were implemented, why they were included, problems we encountered during implementation, and appropriate solutions.

Trip Diary

One of the main features of the original E-mission application is the trip diary, which organizes and provides an interface for a user's daily trips. In addition to displaying the locations that the user traveled to and from, the diary also informs the user of what modes of transportation were used during a trip, duration, speed, and total carbon emissions for that trip. This trip log is present all experiment groups and serves as the only feature for the control group.

Carbon Intensity

Trips are importantly coupled with carbon data. We define *carbon intensity* as the carbon metric used in the application

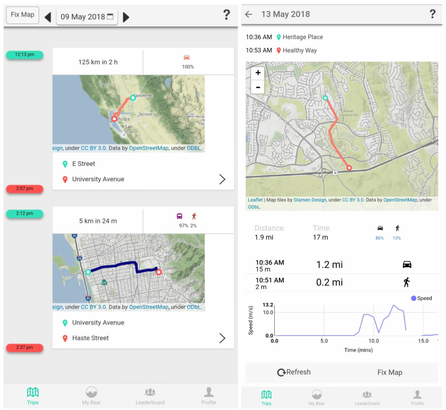

Figure 1: The diary and detailed trip information screen.

for all users, calculated with the corresponding formula:

$$ci_u = \frac{\sum_{i=1}^{n} Rew(d_i) f_i}{\sum_{i=1}^{n} d_i} \tag{1}$$

$$\text{with } Rew(d_i) = \begin{cases} d_i & d_i < t_i \\ d_i * max(-\frac{1}{2t_i} d_i + \frac{3}{2}, \frac{1}{2}) & d_i \geq t_i \end{cases} \tag{2}$$

where ci_u denotes user u's carbon intensity for the past week, n being the number of trips in the past week, f_i is the carbon footprint per mile for trip i's mode of transportation, d_i is the distance of the ith trip, t_i is the threshold value for trip i's mode of transportation, and $Rew(d_i)$ is the adjusted (rewarded) transformation of d_i. We came up with this metric to serve as a more equitable representation of a user's carbon emissions with details and notations elaborated in "The Walking College Student Problem" section.

Tiered System

Many research projects have already explored the extent to which group pressure changes one's behavior. Schultz et al. [16] discovered that people started to care less about their energy consumption when told that they were consuming less than average (the "boomerang effect"). In 2008, Oinas-Kukkonen and Harjumaa [15] postulated offering public recognition can increase the likelihood of adoption of new behaviors.

These ideas inspired us to utilize a tier system for this study. *TripAware* attempts to reduce the boomerang effect by grouping together people with similar performances. Since top performers will be compared to other top performers, they will not feel as "ahead" as they would when compared to the average user. On the other hand, users who rank low on the tier system will not be as discouraged since they will be compared to similar users (rather than "perfect" users). Fundamentally, we hypothesize that displaying users' progress towards reducing their carbon intensity would encourage them to adopt new behaviors that would help them achieve a higher rank.

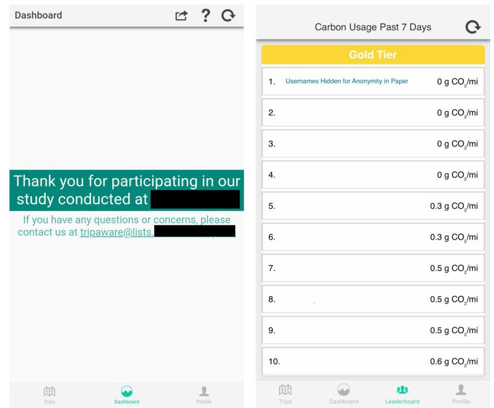

Figure 2: *Left:* The control dashboard. *Right:* The tier system (usernames hidden for anonymity).

Users were divided into three tiers: gold, silver and bronze. People were ranked by their carbon intensity in grams per mile, with the top third put into gold, the middle third into silver, and the last third into bronze. Usernames were displayed so that people can track their change relative to others while maintaining anonymity in the experiment. The tier system was updated hourly.

Control Group

The control group of the experiment served as a point of reference to measure the impact of the other versions of the application. With this in mind, we sought to minimize the application's influence on users' travel patterns. As a result, subjects in the control group were given a bare version of the original E-mission application. This version only consisted of the trip diary; the main dashboard and Habitica game platform were removed and did not include the aforementioned tier system. Lastly, in order to incentivize users to keep the application installed, the dashboard was replaced with a reminder that the experiment was still in process.

Information Group

The primary goal of the *Information* features is to make people aware of sustainable choices available to them and to make their current transportation habits and travel data transparent. In order to achieve this, we implemented a dashboard with various forms of information. These features test the reflective motivation behavior change paradigm defined in the BCW. As the information version of *TripAware* presents users with information about their performance and how to further reduce their carbon emissions, it approaches emissions reduction under the assumption that people are already willing to reduce their emissions and that they just need more information about how to do so.

TripZoom, a study that similarly sought to explore sustainable transportation via mobility data, operated under the assumption that “direct feedback can positively influence the mobility behavior of individuals" [10]. This concept inspired us to create our dashboard components. In order to determine what the components would be, we decided that the following were vital issues to address:

(1) Users should receive concrete suggestions on *how* to reduce their footprint. Even if they are aware about how poorly they are performing, they may never realize *how* they could improve.
(2) Users should know their carbon footprint and receive direct feedback if they have improved. This is necessary for instilling a sense of urgency.
(3) Users should be able to reflect on past trips and see the impact on their carbon footprint.

The first section of the dashboard addresses the first problem by displaying a suggested trip for the user. When tapped, the application will direct the user to Google/Apple Maps, where the mode of transportation and route is already entered. We decide which suggestion to display by scanning forty of the user’s most recent trips. There are three conditions that can be satisfied for a trip to generate a suggestion:

- If the trip was taken via car and covered 5 to 15 miles, then the application suggests that the user rides a bus.
- If it was taken with a car, bus, or train and covered 1 to 5 miles, the application suggests that the user bikes.
- If it was taken with a car, bus, or train and covered less than one mile, the the application suggests that the user walks or bikes.

When a suggestion is generated, the user receives an appropriate notification. New trips are checked every hour and a new suggestion is sent if possible immediately after checking trips. This system does not take into account other factors like users’ desired arrival times.

Below the suggestion is the “Weekly Stats" section. It gives users a quick summary of their recent performance, addressing problem two. As shown in Figure 3, users can see their carbon footprint, net change in carbon emissions from the last week, and distance traveled, and most common mode of transportation in the past week.

Finally, the “Recent Trips” section addresses the third issue. Each trip has a visualization for the trip transportation mode, time, distance, and carbon footprint. Tapping a trip in this section redirects the view to a more detailed summary of the trip within the diary page.

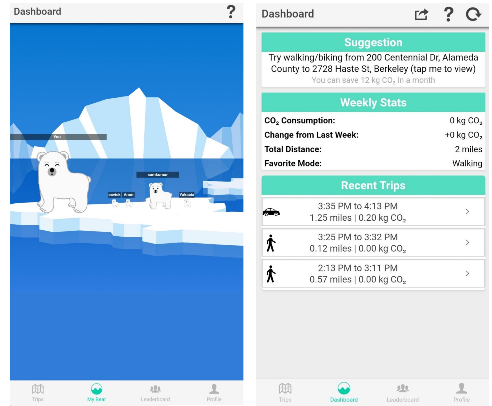

Figure 3: Dashboards: *Left: Emotion Right:* information

Emotion Group

The centerpiece of the *Emotion* application was the polar bear on the dashboard. At a high level, the bear was meant to provide Emotional incentive and to display customized feedback based on the user’s transportation habits. We want to display feedback based on carbon intensity levels over time and their tier list position through the bear and its environment. This polar bear approach tests the automatic motivation behavior change paradigm of the BCW. The polar bear’s size and emotions change to give positive or negative feedback to the user about their behaviors in order to persuade to reduce their emissions.

Utilizing relatable characters to nudge people towards sustainable behaviors has also been a popular behavioral approach in current literature. Takayama et al. (2009) created Eco-Island [20], an application that created virtual avatars for families. The avatars lived on an island where water would rise above the island as their carbon emissions rose. Notably, participants cited their motivation for reducing emissions as wanting to save the in-game island, rather than using environmental reasoning. This led us to categorize this intervention as incentivization, which is defined as “using an external stimulus, such as a condition or an object, that enhances or serves as a motive for behaviour". We thus saw this type of feature as very much in line with our mission to construct an automatic motivation source of behavioral change. The polar bear drew inspiration from these ideas, as the polar bear game built into the app was designed to be passive, attracting users to revisit the app occasionally. The polar bear was designed to be simple, having only two features, growth rate and emotion, to avoid overwhelming the user.

We picked a polar bear because polar bears are typically viewed as a symbol of environmental sustainability. In order

Table 1: Carbon Thresholds for Reward System

Transportation Mode	Penalty Threshold (miles)
Car	50
Bus	25
Train	37.5

to inform the user of their change in emissions, we used a combination of size and mood.

Polar bear size was an indication of performance consistency, a more long term metric to incentivize user behavior. The polar bear became larger to reflect long term emissions reduction, and smaller to reflect an increase in long term emissions. Its size resets if a user crosses a threshold in change in carbon emissions between the previous week and the current week.

Mood was an indication of short-term performance and changed depending on the previous day's carbon emissions. If a user had lesser carbon emissions the previous day than the previous week's average, the polar bear would be happy. Mood was inspired by Carrus et al. [4], which explained that feeling guilty about failing to protect the environment can lead to changes in behaviors that people think affect the environment. In the context of both Carrus et al. and automatic motivation from the BCW, we hypothesized that the polar bear's ability to display sadness would make people feel guilty about increasing their emissions, leading them to want to change their behavior; meanwhile, the polar bear's ability to display happiness would also bring about a feeling of happiness in the user.

Users were made aware of the underlying logic behind mood changes and growth rates via the app FAQ.

In order to curb the leaderboard problem mentioned in section 4.3, we grouped polar bears on different islands where each island represented the tier that the user was placed in. A user could always see their bear and four other randomly chosen bears from their tier. This would not only group users with people with similar performance but also prevent the island from being crowded.

5 DESIGN DECISIONS

Before implementing the aforementioned features, we encountered problems and made key design decisions for our applications. The following section will describe these challenges and outline appropriate solutions and our decisions.

Walking College Student Problem

Certain lifestyles will lead to lower carbon intensity values and will be advantageous for their leaderboard rankings. In our case, determining a user's leaderboard position in their tier based on only their carbon footprint can potentially rank users with shorter commutes higher. For instance, a typical college student in an urban setting can dominate the Tiered System as their classes and responsibilities are all within walking distance. We formalized this idea as the walking college student problem (WCSP). Other studies tracking and comparing carbon emissions of users with varying commutes will have the same problem.

Options in approaching an unbiased carbon footprint ranking system include:

(1) Ranking with raw carbon emissions.
(2) Tiers based on transportation modes such as car, bus, and train tiers.
(3) Reward and distance adjustment to carbon intensity value to balance users with differing commutes.

We decided the best solution to WCSP was a reward system that decreases a trip's carbon intensity if that trip distance is above a certain threshold. This approach aims to both mediate the differences between individuals with shorter and longer commutes and persuade users to use more sustainable modes of transportation when traveling small distances. Rewards are calculated as shown in equation 2 defined in Section 4.2 and intuitively returns a modified distance that is then multiplied by carbon footprint over actual trip distance, leading to a decreased carbon intensity value for trips farther than a mode-specific threshold. In detail, if the trip distance is below the certain trip's transportation threshold, the $Rew()$ function leaves distance unchanged, leaving the carbon intensity value untampered with. Furthermore, if the trip has no carbon emissions (they walked or biked), then the reward function just leaves the distance unchanged as the trip has no carbon footprint regardless. Otherwise, if the input distance is greater than or equal to the transportation threshold, we decrease the distance ratio as distance increases (with the distance ratio equation $-\frac{1}{2t_i}d_i + \frac{3}{2}$ coming from the line between $(t_i, 1)$ and $(2t_i, \frac{1}{2})$, the ratio being capped at one half, where the threshold t_i depends on the mode of transportation (Table 1)). Thus the output distance is restricted to the interval $[\frac{1}{2}d_i, d_i]$, showing that the reward system only benefits long commuters as the goal is to penalize long commuters less if they likely have no other accessible, cleaner mode of transportation for some of their longer commutes. We cap the discounted distance at $\frac{1}{2}$ of their real distance because we don't want to discount long commutes' carbon intensities by too much, as they are still contributing carbon emissions through these commutes.

The threshold values were chosen to reflect "acceptable" commute distances for each mode of transportation. For example, a person that commutes via bus for 30 miles is less likely to be able to find a more sustainable route such as biking for that particular commute. With the threshold value being equal to 25, that trip's carbon intensity value is multiplied by a ratio of $\frac{9}{10}$.

Mode Inference

The E-mission platform [17] utilizes a two-level classifier with GIS integration to classify trips into various transportation mode categories. Bucher et al. [3] suggested cross-referencing public transportation networks to improve transportation mode inference, something that the E-mission platform implemented as we were writing and testing *TripAware* (GIS integration). This improved the accuracy of our mode inference greatly, as our applications would often classify trips taken via public transportation as car trips beforehand.

Determining Tiers

As mentioned in Section 4.3, a leaderboard system could be implemented to encourage users to do better relative to their peers. However, seeing the exceptional performance of people at the top discourages users from trying to climb the leaderboard. There are several ways to approach this issue, such as not having a leaderboard at all, limiting the problem by having friend-group specific leaderboards, or grouping together users on the leaderboard into categorical tiers. For example, users could be grouped together by their most common transportation mode. This would somewhat limit the variance of performance within a tier because users who use similar modes of transportation would have similar carbon intensity values.

TripAware addresses this problem by grouping users into 3 tiers determined by carbon intensity. When users view the leaderboard, they focus more on their tier. By comparing their performance with other users in the same tier, users are able to outperform the people around them, and be encouraged to keep climbing tiers.

Focus on Transportation Emissions

In our project, we focused solely on tracking transportation emissions, intentionally limiting our focus to reductions made as a direct result of changes in transportation habits. We did not take into consideration the impact any other form of behaviour change (eg. eating less red meat) had to a user's overall carbon footprint. This choice is in line with the primary objective of project, that is to guage the impact of different behavioral approaches in modifying transportation behavior. Furthermore, we did not include measures that would further refine our carbon estimate, such as taking into account the car model used (and its fuel efficiency), and whether the user was carpooling. We did not include these features so as to simplify user experience as much as possible. These features would have required users to input details after each trip, which would have made our application far more obtrusive and further reduced user engagement.

6 RECRUITING

With IRB approval and appropriate ethics clearance, we launched the *TripAware* study with a website specifically designed to convince users to download our applications.

TripAware's recruiting website was designed to be modern-looking and brief. We chose this approach because we had received feedback that readers would lose interest with text heavy or old-looking websites. We ultimately ended up not recruiting as many people as we had initially planned to, likely due to the website's brevity and perceived lack of credibility without our institution's logo (see Figure 4). We initially believed that emphasizing a monetary incentive would interest people to join the study; however, this may have equally detracted from the site's legitimacy. What strategy for recruitment in these types of studies is still an open question.

Social media promotion and advertising was meant to maximize the variety of users in the study. *TripAware* social media accounts were made on Twitter and Facebook, along with a logo and informative blog post. Additionally, influencers such as Meredith Lee from the West Big Data Innovation Hub promoted *TripAware* on social media. Lastly, \$148.35 was spent on Facebook advertisements (including the one in figure 4, which targeted users from New York, Boston, DC, Seattle, and the Bay Area. 18292 unique people saw the advertisement and 132 people clicked on the link directing to the *TripAware* website. Once on the website, one of the three versions of the app would be randomly assigned to a user and the download links and graphics would be modified so that the user can only download the app assigned to them.

Once downloaded, the app displayed a consent form that asked participants for permission to track their location and acknowledge that no personally identifiable data would ever be shared to anyone outside of the research team. Participants then had to fill out an initial survey. Overall, 16 users were recruited via presentations, 19 from family and friends, 1 from social media, and 5 declined to state. The majority of the participants were 18-30 years old. The median income among those who reported it was between \$100,000 - \$200,000. Respondents also answered questions regarding how much they were willing to walk or bike per day, resulting in a median of 31 - 60 minutes for both (out of respondents who had bicycles/could walk). 56.1% of respondents stated that they believed that public transportation was easily accessible to them, with 39% stating that it wasn't and 4.9% declining to state. Finally, 73.2% of participants believed that

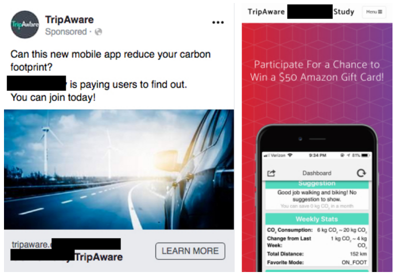

Figure 4: *Left:* This specific Facebook ad generated 7587 views, 52 clicks, and cost $45.39. *Right:* The recruiting website links to and displays one of the three apps randomly.

Table 2: Average Frequency of Actions per User

	Control	Information	Emotion
Diary Check	16.667	33.25	12.727
Expanded Trip	1.000	1.875	3.273
Opened App	11.556	37.000	68.000*

alternative modes of transportation like walking and biking were easily accessible to them, with 24.4% responding no and 2.4% declining to state.

7 RESULTS

Our goal is to determine if the applications focusing on *Emotion* and *Information* were more effective in certain areas than the control skeleton application, and if both are more effective, compare *Emotion* and *Information* based on the same metrics. More specifically, due to the broad nature of our experiment, we measure multiple metrics (in *section 7.2*) that would provide us with many perspectives of each experiment group.

Before analyzing our results, we must mention three issues that we ran into while deciding how to present our data. We had to slightly reduce our sample size for much of our analyses by removing users who participated for less than a week (most of these people never took a single trip) where applicable. Depending on the metric, our sample size hovered between 6-12 for each group. Furthermore, as with many other studies of this nature, our participants could join and leave at different times throughout the study, creating a lot of variance in participation lengths and limiting the number of recognized statistical tests we could use to compare group differences. Finally, due to the low sample size, outliers could affect hypothesis test results. All three problems are addressed by permutation testing, and our entire hypothesis testing process is described in the next section.

Table 3: Notification Opening Frequency per User

	Information	Emotion
Mean	2.308	5.25
Median	1	1.5
Standard Deviation	3.0225	9.22

Hypothesis Testing

Authors suggest using nonparametric exact testing if the sample size is small [13], [18], [19] For all relevant analyses, we performed a permutation test to test the null hypothesis that the observed difference in means of the statistic being measured across the two groups being compared was due to random chance, using a p-value of 0.05.

We performed our permutation tests with 100,000 iterations of permuting. Exact permutation tests require that we observe all possible permutations of the data, so instead we approximated it with 100,000 permutations, which is even more than what many others suggest [14].

Analysis

We decided to look at eight different metrics at the end of the experiment that are split into two categories: app interaction and behavior change. Our code can be viewed on Github at https://tinyurl.com/y7jjt63o.

App Interaction.

(1) **Opened App**
We recorded the number of times users opened the application and computed the average per person per group. We compared differences in averages across groups when performing the tests. Averages are shown in Table 2. *Emotion* and *Information* received a statistically significant level of interaction as compared to control ($p = 0.006, 0.045$ respectively). However *Emotion-Information* was insignificant ($p = 0.189$).
(2) **Diary Check**
We recorded how many times users looked at the diary and computed the average number of views per person per group. We permuted the difference in averages across groups. Results are shown in Table 2. All results showed to be statistically insignificant (*Emotion-Control*: $p = 0.273$, *Information-Control*: $p = 0.230$, *Emotion-Information*: $p = 0.101$).
(3) **Expanded Trips**
We kept track of how many trips users viewed and the average number of views per person per group. We permuted the difference in average number of views per person across groups. Results are shown in Table

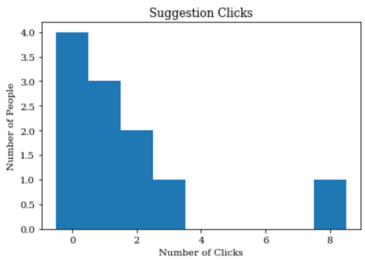

Figure 5: Suggestion taps histogram.

2. We find that *Emotion* vs *Control* resulted in a significant difference ($p = 0.035$). Both other comparisons are insignificant (*Information-Control*: $p = 0.133$ and *Emotion-Information*: $p = 0.183$).

(4) **Retention Duration**
We calculated retention durations by finding the difference between the latest time and the earliest time in which the user opened the app. We permuted the average of these times across groups. On average, *Emotion*users used the app for 53.067 days, information users 30.644 days, and control users 27.606 days. After permutation tests, we conclude statistical significance between *Emotion* and *Control* and *Emotion* and *Information* ($p = 0.031, 0.040$ respectively). *Information-control* was not significant ($p = 0.398$).

(5) **Notification Taps**
We kept track of the number of times *Emotion* and *Information* users tapped on notifications. Unfortunately, we did not keep track of the total number of notifications sent so we can not present proportions. Results are displayed in Table 3. *Control* is not displayed because it did not send users notifications.

(6) **Suggestion Taps**
We recorded the number of times users tapped on a suggestion (taking them to Google/Apple Maps). In Figure 5, we can see that most users did not heavily interact with the suggestion system.

Behavior Change.

(1) **Change in Carbon Intensity**
Intuitively, change in carbon intensity for each user is represented as a normalized value of how much their carbon intensity changed over time. Specifically, a user's change in intensity was calculated by summing the difference between the carbon intensities of week i and week $i + 1$ for all weeks, as a percentage of their overall average carbon intensity. For some user u, this is calculated as:

$$\frac{\sum_{i=1}^{n-1} cu_{i+1} - cu_i}{\frac{1}{n}\sum_{i=1}^{n} cu_i} * 100 \tag{3}$$

Table 4: Average Total Carbon Change per User

	Carbon Change (%)
Control	-25.53
Emotion	-74.04
Information	-120.28*

where cu_i represents the carbon intensity for user u at week i, and they participated for n weeks. Results are shown in Table 4. We conclude that there is no significant difference between *Emotion* and *Information* ($p = 0.139$) and *Emotion* and *Control* ($p = 0.116$), but there is a significant difference between *Information* and *Control* ($p = 0.043$).

(2) **Change in Transportation Modes**
Intuitively, change in transportation mode for each user is represented as a normalized value of how each transportation mode changed over time. Specifically, a user's change in each transportation mode was calculated by summing the difference between the proportion of trips of week i and week $i + 1$ for all weeks, as a percentage of their overall average transportation mode proportion. For some user u and transportation mode m, this is calculated as:

$$\frac{\sum_{i=1}^{n-1} ratio_{m,i+1} - ratio_{m,i}}{\frac{1}{n}\sum_{i=1}^{n} ratio_{m,i}} * 100 \tag{4}$$

where $ratio_m, i$ represents the proportion of week i's user trips matching the tm transportation mode over all trips for week i, and they participated for n weeks. To determine the change in transportation modes from week to week, we calculated the proportion of trips for three modes—walking/biking, car, and bus—for each week over the span of ten weeks. We also recorded trips taken on train but decided to omit the information due to the lack of data points. Sample sizes for certain modes of transportation have smaller samples sizes (e.g. bus having $n = 20$ across all three groups from 41) due to some users never using that transportation mode. We calculated the weekly change in proportion of trips spent in each mode over the span of ten weeks. Change in proportions were calculated by summing week-to-week proportion differences as a percentage of that user's overall average carbon value. All combinations of group and transportation modes had p values far above a 0.05 threshold (the range was $0.12 - 0.48$).

Limitations of the Study

This section describes some limitations of the *TripAware* study and what future studies can do to mitigate these effects.

Small Sample Size with Study Duration. TripAware had 41 participants. Although this study had more participants than those cited in other papers, the number was not enough as some of those users did not interact with the app as much as we had expected or did not keep the app for a sufficient period of time. Furthermore the study lasted for ten weeks. Although this time span is sufficient, there was still difficulty finding consistency within the results due to our decision to accept subjects on a rolling basis. This study could have benefited from a longer testing time period, combined with a very large sample size that allows us to have enough people at the start rather than have to accept on a rolling basis.

Website Appeal. Despite generating 132 website visits via Facebook ads, at most 6 people joined the study through Facebook. In hindsight, focusing on monetary incentives did not work for our recruitment process.

Onboarding Process. To join the study, participants were asked to download the app and then click on a link on the landing page to officially participate. This was necessary as we only had one base app to download, users had to download the different versions via the aforementioned link. Some users also accidentally disabled location tracking for the app, preventing any trips from being recorded. These issues could have been circumvented by having distinct applications on the App Store and simplifying the onboarding process.

User Bias. Due to the recruiting process containing targeted presentations and advertisements, there was the potential for bias towards recruiting people that care more about reducing their carbon emissions. We hoped to offset this with the promise of a chance to win ten $50 Amazon gift cards, although we can not be sure exactly why people joined the study. This question should have been in our initial survey.

Additional Limitations. Trip mode thresholds in our reward function (see section 5.1) could have been lower as users that we aimed to assist rarely had their carbon values reduced. Also the range $[t_i, 2t_i]$ of the *Rew*() function was computed incorrectly as it was quadratic instead of linear. It would have been better to not include that range function and only have two linear lines (d_i and $\frac{1}{2}d_i + \frac{1}{2}t_i$). This quadratic component results in slightly underestimating carbon intensity, but only impacted around 0.25% of total trip sections. In addition, keeping track of more data would've been helpful to better analyze our results. For example, we did not keep track of the number of notifications sent for each group (only number opened). Also, our results are specific to our set of features, so more studies are needed to further analyze how emotional or informational approaches can help change people's behavior.

8 CONCLUSION

Both *Emotion* and *Information* have statistically significant effects on users' behaviors when compared to the control application. Emotional features kept the users more engaged (app opening frequency, expanded trips, retention duration), while informational features caused a significant week-to-week reduction in carbon change and improvement in app opening frequency. Perhaps participants in the *Emotion* group became engaged with the polar bear, but the link between the polar bears' happiness levels and a user's carbon emissions was not strong enough to cause a significant change. Meanwhile, informational features such as the dashboard could have kept people more aware of their emissions, but failed to keep them engaged enough to stay in the study significantly longer than the control group; for example, participants did not interact heavily with the recommendation, something we originally believed people would check frequently.

We found one significant difference between *Emotion* and *Information*. *Emotion* demonstrated a longer retention duration. This reinforces our belief that the emotional features kept users more engaged, even compared to *Information*, which presents more features than *Control* does. Overall, we can conclude that informational and emotional features are better than having no features at all; our results illuminate possible behavioral influence from each set of features but are too limited to fully explain the effectiveness of the two sets when compared to each other. Tying in the Behavior Change Wheel and our original question, we observe that—in the context of our experiment— reflective motivation led to changes in our target behavior (reducing carbon emissions) while automatic motivation led to changes in behaviors not directly linked to our target behavior.

ACKNOWLEDGMENTS

We thank *Prof. David Culler*, for providing routine feedback and guidance with analyzing our results. We also thank *Prof. Ani Adhikari*, for her statistical guidance in our analysis, *Prof. Eric Paulos* for guidance in writing this paper, and *Simon Behar* for helping us design the experiment and set up our servers.

REFERENCES

[1] *Inventory of U.S. greenhouse gas emissions and sinks 1990-2016.* U.S. Environmental Protection Agency.

[2] H. Brynjarsdottir, M. HÃěkansson, J. Pierce, E. Baumer, C. Disalvo, and P. Sengers. Sustainably unpersuaded. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI 12*, 2012.

[3] D. Bucher, F. Cellina, F. Mangili, R. Martin, R. Rudel, A.-E. Rizzoli, and O. Elabed. Exploiting fitness apps for sustainable mobility - challenges deploying the goeco! app. 09 2016.

[4] G. Carrus, P. Passafaro, and M. Bonnes. Emotions, habits and rational choices in ecological behaviours: The case of recycling and use of

public transportation. 28:51–62, 03 2008.

[5] H. K. M. B. A. N. J. C. E. G. R. Castell, N.; Liu. Mobile technologies and personalized environmental information for supporting sustainable mobility in oslo: The citi-sense-mob approach. *In Proceedings of the 28th International Conference on Informatics for Environmental Protection, Oldenburg, Germany, 10âĂŞ12 September 2014.*

[6] P. R. Center. Mobile fact sheet. 02 2018.

[7] C. Disalvo, P. Sengers, and H. BrynjarsdÃşttir. Mapping the landscape of sustainable hci. *Proceedings of the 28th international conference on Human factors in computing systems - CHI 10*, 2010.

[8] J. Froehlich, T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J. A. Landay. Ubigreen. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 2009.

[9] M. Hagger, D. Keatley, and D. K. C. Chan. Calo-re taxonomy of behavior change techniques. pages 100–105, 01 2014.

[10] P. Holleis, M. Luther, G. Broll, H. Cao, J. Koolwaaij, A. Peddemors, P. Ebben, M. Wibbels, K. Jacobs, and S. Raaphorst. Tripzoom: A system to motivate sustainable urban mobility. 05 2018.

[11] A. Jylhä, P. Nurmi, M. Sirén, S. Hemminki, and G. Jacucci. Matkahupi: A persuasive mobile application for sustainable mobility. pages 227–230, 2013.

[12] S. Michie, M. M. van Stralen, and R. West. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1):42, Apr 2011.

[13] F. J. Mundry R. Use of statistical programs for nonparametric tests of small samples often leads to incorrect p values:examples from animal behavior. pages 256–259, 1998.

[14] T. Nichols and A. Holmes. Nonparametric permutation tests for functional neuroimaging: A primer with examples. page 9, 2001.

[15] H. Oinas-Kukkonen and M. Harjumaa. A systematic framework for designing and evaluating persuasive systems. pages 164–176, 2008.

[16] P. W. Schultz, J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434, 2007. PMID: 17576283.

[17] K. Shankari, M. A. Bouzaghrane, S. M. Maurer, P. Waddell, D. E. Culler, and R. H. Katz. E-mission: An open-source smartphone platform for collecting human travel data. *Transportation Research Record: Journal of the Transportation Research Board*.

[18] C. N. Siegal S. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 1988.

[19] R. F. Sokal RR. *Introduction to Biostatistics*. W.H. Freeman, 1987.

[20] C. Takayama, V. Lehdonvirta, M. Shiraishi, Y. Washio, H. Kimura, and T. Nakajima. Ecoisland: A system for persuading users to reduce co2 emissions. pages 59–63, 03 2009.

[21] J. L. Zapico, M. Turpeinen, and N. Brandt. Climate persuasive services: Changing behavior towards low-carbon lifestyles. pages 14:1–14:8, 2009.