Model-based Formalization of the Autonomy-to-Human Perception Hand-off

Yash Vardhan Pant
Balasaravanan Thoravi Kumaravel
Ameesh Shah
Erin Kraemer
Marcell Vazquez-Chanlatte
K Kulkarni
Björn Hartmann
Sanjit A. Seshia

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2021-8
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2021/EECS-2021-8.html

March 15, 2021

Acknowledgement

This work was supported in part by NSF CPS Frontier project VeHICaL (CNS-1545126), by Toyota under the iCyPhy center, and by Berkeley Deep Drive.

Model-based Formalization of the Autonomy-to-Human Perception Hand-off

Yash Vardhan Pant, Balasaravanan Thoravi Kumaravel, Ameesh Shah, Erin Kraemer,
Marcell Vazquez-Chanlatte, Kshitij Kulkarni, Bjoern Hartmann, Sanjit A. Seshia
University of California, Berkeley, CA, USA

***Abstract*—Autonomous and semi-autonomous systems can encounter situations where timely attention of a human operator is required to take over some aspect of decision making or control. For certain human robot interaction (HRI) applications, like Autonomous Vehicle (AV) operations, these decisions could be both time-critical and safety-critical. Given this, it is important to ensure that the human is brought into the decision making loop in a manner that enables them to make a timely and correct decision. In this paper, we consider one such application, which we refer to as the *perception hand-off problem*, which brings the driver into the loop when the perception module of an AV is uncertain about the environment. We formalize the *perception hand-off* problem using a Partially Observable Markov Decision Process (POMDP) model with a problem specific structure. This model captures the latent cognitive state of the driver which can be influenced through a query-based Human-Machine Interface (HMI). Through a human-study experiment on the perception hand-off problem for object recognition, we learn such a model and validate our hypotheses about the hand-off problem and the impact of our query-based HMI. The results from this study show that the state of attentiveness does indeed impact the human performance, and our proposed active information gathering (AIG) actions, or queries, result in 7% faster responses from the human. We also use this experimental data to learn the proposed POMDP model parameters. Simulations with this identified model show that a policy for deploying the AIG actions improves the percentage of correct responses from the human in the perception hand-off by around 5.5%, outperforming other baselines while also using fewer of these actions.**

I. Introduction

The safe operation of autonomous and semi-autonomous systems sometimes requires intervention from a human operator. However, the human operator may not always be in a state to make a correct and timely decision, leading to safety violations with potentially fatal consequences [13]. In recent years, issues with the perception module of autonomous vehicles (AVs) have been a dominant cause for a human to take over control of the vehicle [3]. In such *takeovers* or *hand-offs*, the human operator needs to be attentive and have spatial awareness once the AV asks them to take control, but might not have as complete a picture as the AV does since they are were not controlling the vehicle up until the hand-off was initiated. We posit that in such scenarios, continued semi-autonomous operation could be possible by handing off to the human just the perception task that the AV cannot confidently perform. This would allow the vehicle to operate under the same (autonomous) control law and avoid sudden maneuvers unless necessary. It would also enable gathering data on-the-fly, which could be useful for improving the robustness of the perception algorithms. We refer to this human-robot interaction as the *perception hand-off*. This hand-off requires a system to effectively alert the human and bring them into the decision-making loop in a manner that ensures overall system safety. In this paper, we present *an approach to formally model the Human-Robot interaction in the perception hand-off problem and develop an active information gathering scheme* that enables us to leverage a *query-based human machine interface (HMI) to both estimate and influence the human state* to improve response time and correctness. Our framework can be viewed as a formalization and automation of techniques such as *pointing-and-calling* [1], a method used by train operators in Japan wherein they points at signs and verbalizes the command that they will execute, and which has been shown to reduce operator error and improve response times [24].

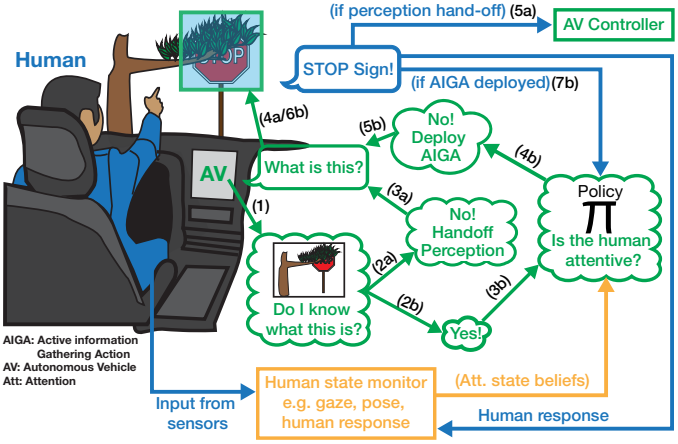

Fig. 1: Illustration of our approach for the perception hand-off.

We focus on applications in L3/L4 autonomous driving, as defined in the SAE J3016 standard [2] for AVs, and develop a framework for bringing the human in the decision-making loop at a high attention level to safely execute a perception hand-off. Here, the AV and the human interact through a HMI, through which the AV can query the human for two purposes: a) the AV is unsure of the environment and requires human input in decision making, or b) a Human-Robot Interaction (HRI) policy (designed for maximizing safety in a perception hand-

off) wants to either infer the state of the human's attentiveness or influence it in preparation for an upcoming (potential) perception hand-off. We refer to the latter as an *active information gathering action* (AIGA) [18], and show the benefit of these, especially in situations where the human driver's state of attentiveness during autonomous driving is also impacted by non-driving related tasks (NDRTs) or distractor tasks.

Showing the potential applicability of our approach, in an online survey that we hosted on Amazon mechanical turk (mturk), 45 of 56 users preferred that their AV ask non-critical questions if it can help them respond faster and more accurately in a critical scenario. A majority (49 out of 56) also preferred to be kept in the loop if the AV is uncertain about the environment during autonomous operation. The following example illustrates the perception hand-off.

Example 1 (Perception hand-off in autonomous driving)**.** *To demonstrate how our system can be of use in a perception hand-off scenario, consider the following example. The AV is operating in an area that predominantly has stop signs at intersections. As show in Figure 1, the AV approaches a partially occluded stop sign. In such a setup, based on a system that monitors the human (e.g. gaze tracking, pose detection, or location of driver's hands), the AV maintains a belief about the human's level of attentiveness and uses the responses from the human in the cases below to update it:*

Case 1: *The AV has correctly identified the stop sign with high confidence (Fig.1-2b). Subsequently, a human-robot interaction policy determines (Fig.1-3b) if it needs to deploy an AIGA (Fig.1-5b) that both alters the human's attention state and allows to further update the belief (Fig.1-7b).*

Case 2: *The AV is unable to identify the stop sign with high confidence(Fig.1-2a), the human operator is asked to identify it. Their response is then used by the AV to perform an appropriate behavior, e.g. stop at the sign and then proceed. We refer to this as a "perception hand-off" (Fig.1-3a).*

A. Overview of our approach and outline of the paper

In section II, we cover some of the existing research that is relevant in the context of this work. Section III states the problem statements we aim to solve, and gives an outline of the key components of our framework to do so. As a first step, we study the perception hand-off through a human subject study (section IV), deployed on Amazon's Mechanical Turk (mturk) platform[1], where the human subject helps identify objects on the road that the AV is unsure of, while also performing a non-driving related task (NDRT) or *distractor* task. Data from this experiment shows that the NDRT increases the human's response time and lowers accuracy of their responses. It also shows that the proposed active information gathering (AIG) mechanism results in faster responses from the human even in the presence of NDRTs. We propose a Partially Observable Markov Decision Process (POMDP) model to represent this HRI (section V) to capture such behavior, and learn the model parameters through the data gathered via the mturk study. The structure of the model makes it well suited for the perception hand-off, and also makes it amenable to learn policies for influencing human behavior in this HRI (section V-C). Simulation studies (Section VI) show the benefits of our approach and its potential to improve the driver's response time and rate of making correct decisions.

[1]Since in-lab studies were not possible during the pandemic, we designed an experiment that could be deployed online to reach a wide user base

B. Contributions

The main contributions of this paper are:

1) A model-based formalization of the perception hand-off process for time-critical human operator decision-making in autonomous/semi-autonomous systems;
2) A query-based active information gathering mechanism to use the HMI to gauge and influence the attention of the human operator in a closed-loop manner;
3) A human subject study to gather data on human operator performance in a setting that simulates such a perception hand-off process. This data is used to learn the proposed POMDP model and validate hypotheses on the operator behavior and the impact of the query-based AIG mechanism, in particular showing on average a 7% speed up in response times when the human is distracted, and
4) A model-based policy that uses the query-based AIG mechanism to influence the operator attention in order improve their performance on these hand-off tasks.

We demonstrate that the rate of making correct decisions improves by 5.5% using our approach via a simulation study, which uses the learned model as a surrogate for the human.

II. Related work

In this paper, we study the problem of safe interaction between a human operator and an autonomous/semi-autonomous vehicle. In this section, we cover some of the relevant work in this context from across different research areas.

Model-based Human-Robot Interaction: In [23], measurements of the pose of the human driver of a semi-autonomous vehicle are used to correct the human input to the vehicle. Models with hidden latent states, usually POMDP-based, have been used to generate robot policies [9] or predict human intent [26] in collaborative human-robot tasks. These works however do not consider the case where the robot can *actively* gather information, i.e. take actions to estimate or influence the latent (human) state. The work in [18] takes a step in this direction, where an autonomous vehicle takes actions to actively estimate whether the driver of a nearby human operated vehicle is *attentive* or *inattentive*. However, the human latent state is assumed to be time invariant. In this work, we consider the problem of active information gathering to both estimate and influence the mental state of the human operator, which is time

varying. A monitoring-based approach to alert the driver for a takeover is presented in [7]. The space of states and actions there is similar to ours, but unlike our approach, they assume full state observability. They also assume *a priori* knowledge of a transition model, while one of our main contributions is designing an experiment to gather data to learn such a model. Finally, in the context of HRI in autonomous driving, there has been extensive research into the problem of AV control in the presence of human driven vehicles [11, 19, 25].

Dual-task driving studies: In situations where the safe operation of an AV requires the assistance of a human operator, the human's behavior is not guaranteed to be timely, or even correct. This can mostly be attributed to human operators of vehicles performing non-driving related tasks [13]. Dual-task experiments [4, 14, 5, 12] have been designed to study driver behavior in the presence of non-driving, or *distractor* tasks. The findings in these state that the presence of a distractor task impairs the driver's performance on driving-related tasks and increases their response times. As will be seen later in the paper, similar to these driving simulator-based studies, the web-based experiment we use for data gathering and model learning also exhibits these trends.

Cognitive models of humans in autonomous driving: Seminal work by Card et. al [8] proposed the Model Human Processor (MHP), that explicitly modeled the human as a set of computer memories (such as long-term memory, working memory) and processors bound by certain "principles of operation". It characterized attributes such as decay time and access time of information stored in them. Using these, one can empirically estimate time taken by a user to complete a task, as the time for human brain to process and act on different types of input stimuli. Similar techniques have been used in other cognitive models such as GOMS and KLM [8]. A major drawback is that, these require a prior knowledge of the specific task(s) to estimate human performance on them. But, for a human in a self-driving car, the distractor task that they perform varies and, is unknown ahead of time. This necessitates a data-driven and real-time adaptive approach.

The human cognitive process when an AV requests the driver to take over control has been studied in [21, 22, 20]. Unlike these works that aim to model the underlying cognitive processes step-by-step, we aim to develop a computational latent state model that can be influenced by an external process, i.e. the Human-Machine Interface (HMI). Also, in our version of the takeover or hand-off process, the human does not take over full control of the vehicle, but is tasked with recognizing an object on the road when the AV cannot. In other applications, time constrained decision making of humans has been studied in puzzle solving [16] and gambling [15].

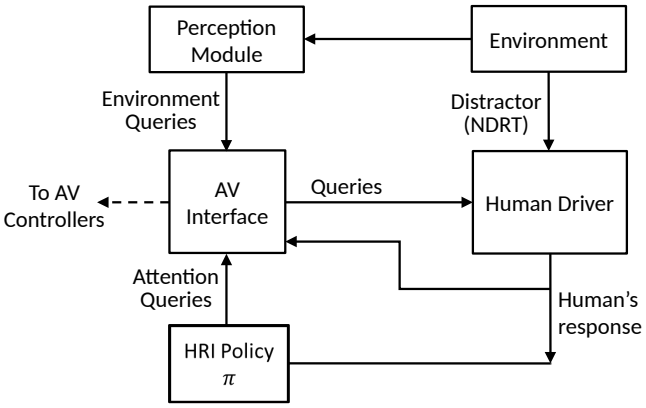

Fig. 2: An overview of the human-AV interaction in the hand-off process. The AV, through a human-machine interface (HMI), can query the human driver when its perception module requires help in decision making, or to gauge/influence the human's state of attentiveness. Further influencing the human state is a NDRT.

III. Formalizing the Autonomy-to-Human Perception Hand-off

We develop a model-based framework to represent and influence the human behavior during the perception hand-off HRI process, an overview of which is shown in Figure 2. First, we address the need for developing a latent variable model of the hand-off HRI that is suited for closed loop control and can be interpreted for online monitoring of the human operator's attentiveness. We also propose the use of query-based active information gathering actions that enable us to do so.

A. Modeling the human response

The first step in our framework is to develop a model for the hand-off HRI that can be used for closed-loop and online interaction between the AV and the human operator.

Problem 1 (Modeling)**.** *Develop a model for the human operator's response (timing and correctness) to perception-hand-off queries from the autonomous system, that can account for the (latent) human attentiveness levels, transitions between them, and the impact of queries on them.*

In this work, we propose a Partially Observable Markov Decision Process with a specific structure to represent this HRI. Section V covers the details regarding the states, actions, observations and the transition structure in this model. Our model allows for the latent state of the human to change over time, and be influenced by the AIGAs, distinguishing our approach from other works like [18].

B. The human-AV interface: Querying the driver for hand-off

Next, we also discuss the interface between the human operator and the AV. In the version of the perception hand-off problem considered here, the AV occasionally requires human intervention in decision making, e.g. identifying an object on the road. In our framework, this is posed to the human as

queries, which must be answered within a given deadline. The queries are displayed to the human via a HMI[2], which also registers the response from the human, as shown in figure 2. This response would be used by the AV to decide which behavioral action (e.g. lane change or emergency braking) to execute, that however is beyond the scope of the current work.

A query displayed via the HMI could be of two types:

1) *From the perception module, or environment query:* These are asked when the AV's perception module is unsure how to interpret the environment and requires the human to make a decision. We refer to these as *environment queries* as they are triggered by factors external to the AV.
2) *Active information gathering action, or attention query:* Here, the AV does not actually require human intervention, but nevertheless relies on a *policy* to query the driver to either influence or better estimate a latent state.

Assumption 1 (Precedence of Environment queries)**.** *An attention query can be only be displayed via the HMI if there is no environment query actively displayed. An attention query can also be preempted by an environment query.*

This assumption is formalized in section V. Note that, an environment query is due to factors external to the AV (which cannot be directly controlled). This can be interpreted as a second player's actions (environment) in a two player game, where the first player (the AV's HRI policy) takes actions in the form of attention queries.

Next, we consider the problem of developing a policy for scheduling attention queries to increase the human's attentiveness towards driving related tasks.

Problem 2 (Policy for the hand-off HRI)**.** *Develop a policy for deploying the AIGAs (attention queries) in order to improve the human response to subsequent environment queries, i.e. the rate of correct responses and reduce response time.*

The overall architecture for this hand-off HRI is shown in Figure 2. In order to study human behavior to this setup, and to gather data to learn the proposed model, we developed a proof-of-concept user study that simulates this hand-off process.

IV. Dual-task web experiment for the perception hand-off process

For the study, we developed a web-based game that simulated the perception hand-off process as in the previous section. We used it to study the impact of human attention levels and the HMI on the timing and accuracy of decision making in hand-off situations, we developed a *dual-task* human subject experiment in the form of a game where the human can interact with an AV. The dual-task here refers to the fact that the human subject performs both a driving related task and a *distractor* or non-driving related task (NDRT). The experiment was conducted on the Amazon mechanical turk platform, and we had 40 users who took part in it.

[2]The formal design of such an interface is beyond the scope of this paper, however we consider a graphical interface (see Section IV) that allows us to study and collect data for the perception hand-off HRI.

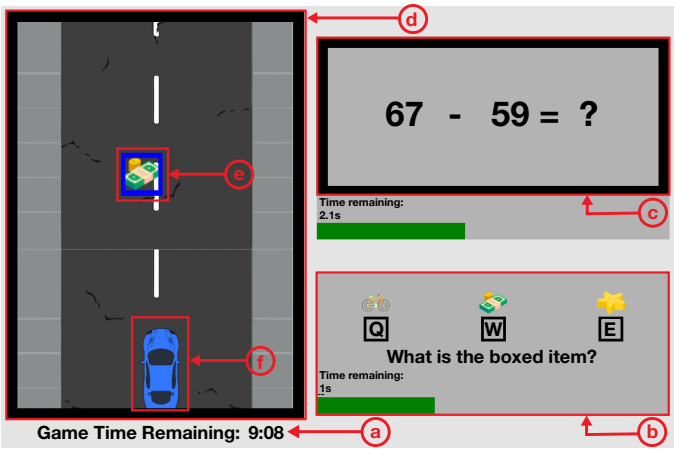

Fig. 3: The human subject experiment design for studying the hand-off process. Note: UI Text emphasized for clarity. An example of an experimental trial can be seen at https://youtu.be/LZRemqFBILA.

Figure 3 shows the UI that the subject of the experiment interacts with. The main components of this, as shown clockwise from the left, are as follows. See Appendix A for details.

1) **Driving related (primary) task:** Figure 3d shows the setup for the driving task. The user has a top-down view of an AV (Figure 3f) driving on a straight one-way road with objects on it. Only one object is present on the road at a time; a new one is spawned every 10s. For some of these objects, the AV requires user input (via keyboard), within a deadline of $4s$, to label them correctly (using legend in 3b) within a fixed amount of time.
2) **Distractor task:** For the NDRT (Figure 3c), the user has to solve basic arithmetic questions within a given time. They have exactly 5 seconds to answer each question, and the task display is toggled on/off every 150s.
3) **Human-Machine Interface:** This displays: a) The queries to the human to identify an object on the road, and b) The information regarding which key corresponds to the different classes among which the human must choose to associate the object with.

Note that in order to simulate an AV and to deal with the constraints of designing and deploying the experiment, the subject cannot directly control the car in the driving related task. For objects that the AV can identify on its own, it performs an appropriate behavior to either avoid or collect the objects. In cases where it requires the user to identify an object, the car takes an action only after receiving user input. The full experiment takes 10 minutes.

A. Interleaving of attention and environment queries

While each of attention queries and environment queries by themselves are identical, the primary difference between

the two types is the *order* in which they are deployed. AV encounters only one object at a time, and the time difference between the objects in 10s. Every *set* of three objects forms a condition for Hypotheses 2A-5B, and takes 30s. More details on *orders* can be found in Appendix C.

B. Summary of data and hypotheses on human performance

1. Are the distractor tasks effective? First, we ensure that the deployed distractor task was effective in distracting the users from the primary task of identifying objects. To do that, we tested the performance of the users on two key metrics: a) Average Response time of answering driving-related queries, RT; b) Fraction of primary queries answered correctly, f, and formulated the following hypothesis.

Hyp. 1A. *The average response time (RT) of driving-related queries in the presence of a distractor task will be greater than when distractor task is absent*

Hyp. 1B. *The fraction correctly answered (f) of driving-related queries in the presence of a distractor task will be lower than when distractor task is absent*

We carried out paired-sample t-tests and observed a statistically significant difference in both average response times and fraction of queries answered correctly. In the presence of distractor tasks, RT was significantly higher (M = 2171ms, SD=520ms) than in its absence (M = 1807ms, SD = 449ms); $t(39) = 7.82$, $p < 0.05$, Cohen's D = 0.72. In case of f, it was significantly lower in the presence of distractor task (M = 0.894, SD = 0.13) when compared to its absence (M = 0.966, SD = 0.063); $t(39) = 3.907$, $p < 0.05$, Cohen's D = 0.67.

Discussion: These results suggest that the distractor task indeed distracts the user from the driving task, both by increasing the time for task completion, and reducing their correctness. This ensures that the users, in the presence of distractor tasks, are operating in dual-task setting.

2. Effect of attention queries: We proceed to study the effect of attention queries on the user's performance on environment queries. We expect that the presence of attention queries on objects increases user awareness of the driving task. Hence, they might perform better on a succeeding query(s).

To test this, we formulated hypotheses (Hyp. 4A, 4B, 5A, 5B), and tested them. However, we were unable to reject the null hypotheses at a significance level of $p < 0.05$. Overall, no significant effect of attention queries were found on the environment queries across the entire experiment. These are reported in the Appendix D for reference.

3. Effect of attention queries in a dual-task setting: Subsequently, we continued to study the effect of attention queries on the user's performance on environment queries, but now specifically in a dual-task setting, in which the distractor task (NDRT) was present. We formulated the following hypotheses:

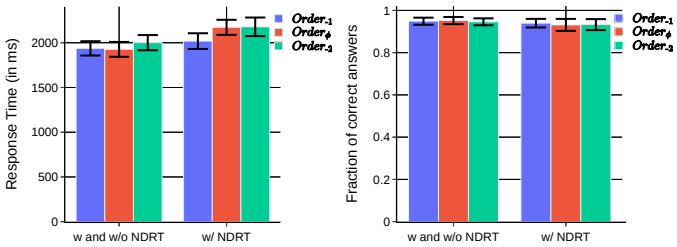

Fig. 4: Average Response Times (RT) and Fraction of queries correctly answered (f) for different conditions employed in Hyp. 2A, 2B, 3A and 3B

Hyp. 2A. *In a dual-task setting, the RT of environment queries such that there is an attention query on the preceding object (Experimental Condition - $Order_{\text{-}1}$) will be less than the RT of env. queries where there is no att. query on at least two objects before it (Control Condition - $Order_{\phi}$).*

Hyp. 2B. *In a dual-task setting, the f of environment queries such that there is an attention query on the preceding object (Experimental Condition - $Order_{\text{-}1}$) will be greater than the f of env. queries where there is no att. query on at least two objects before it (Control Condition - $Order_{\phi}$).*

Hyp. 3A. *In a dual-task setting, the RT of env. queries such that there is an att. query on the second object preceding it (Experimental Condition - $Order_{\text{-}2}$), will be less than the RT of env. queries where there is no att. query on at least two objects before it (Control Condition - $Order_{\phi}$).*

Hyp. 3B. *In a dual-task setting, the f of env. queries such that there is an att. query on the second object preceding it (Experimental Condition - $Order_{\text{-}2}$), will be greater than the f of env. queries where there is no att. query on at least two objects before it (Control Condition - $Order_{\phi}$).*

We carried out paired-sample t-tests for these and observed a statistically significant difference in results of Hyp. 2A. The average response times (in the presence of distractor task) of Control Condition ($Order_{\phi}$): M=2172ms and SD=533ms; For the Experimental condition ($Order_{\text{-}1}$): M=2018ms and SD=555ms; t(39) = 3.24, $p < 0.05$, Cohen's D = 0.27 (Small effect size). Hence we accept Hypothesis 2A.

We did not find statistically significant differences when testing for Hypotheses 2B, 3A and 3B. These are visualized in Figure 4. and reported in the Appendix D for reference.

Discussion: In Hyp. 3A, we do not see any statistically significant difference in performance of an environment query when an attention query was asked on the second object preceding it (20s before). However, when the time difference between attention query and the environment query was reduced (to 10s), as in Hyp. 2A, we observed that the reduction in response time (of around $150ms$, or by 7%) caused by the presence of attention query was statistically significant.

The presence of an attention query on an object had a

minimal effect on the correctness (f) of response to a query on a subsequent object. This could be because the primary task was easy enough that it was answered correctly in most cases, or due to the high time difference between attention and environment queries. This needs to be explored in a future study with a harder primary task, and where we can deploy the queries on a finer time scale. However, attention queries can increase the user's general awareness of the primary task, thereby reducing their average response time.

V. Finite state POMDP model for the human-AV interaction in the hand-off process

In this section, we develop a Partially Observable Markov Decision Process (POMDP) to represent the HRI for the perception hand-off and model the impact of the active information gathering actions (problem 1). The partial observability is over the internal level of human attention, which we allow to be time varying and which has a direct impact on the human's behavior in a perception hand-off, e.g. due to the NDRT as seen in section IV. Outside of a controlled environment, such external factors cannot be measured directly; therefore, we assume probabilistic transitions between attention levels. This allows the AV to maintain a belief over the human's attention.

Definition 1 (Human attentiveness level)**.** *First, we hypothesize that relevant to the perception hand-off, the human has $L = \{l_1, \dots, l_N\}$ levels of attention. At a discrete time step k, the human attentiveness state can take a single value in L. The attention levels are ordered $l_{i+1} \succ l_i$, with $\succ$ denoting a total order, such that higher levels imply higher attention.*

We are interesting in developing a discrete time model, where time step k corresponds to time kdt. Here, dt is the sampling time. The queries to the human have an associated deadline of $T_{\max} = Ddt$ seconds, or D time steps. The queries from the HMI act as *actions*, or inputs to the human, and the response to those queries is the *output*, or observation from the human (Figure 2). Associated with whether the actions are active information gathering queries or from the perception module, there is a counter that keeps track of how many time steps have elapsed since the query was asked.

Definition 2 (Query model)**.** *A query has states $T \in \{-D, \dots, -1, 0, 1, \dots, D\}$. For an active information gathering query, the state of the query increments from 1 to D in steps of 1 at each discrete time step if the human does not respond to the query. If the human does not respond by the query deadline, or D^{th} state, then the query* times out *and the state resets to 0. If there is a response at the t^{th} query state ($1 \leq t \leq D$), the query state again resets to 0. In the case when the query is from the perception module, e.g. an environment query as in the experiment of section IV, the query state decrements from -1 to $-D$ and resets based on whether the human responds within the query deadline or not. In the absence of any active queries*[3]*, the query state is 0.*

We now define the POMDP obtained by considering a probabilistic model for transitions of the human attentiveness states and combining this with the query model.

Definition 3 (Perception hand-off model)**.** *The perception hand-off process is then modeled by a POMDP, which is a tuple $(S, A, O, R, \mathbb{T}, \mathbb{O}, \gamma)$, where:*

- *$S = L \times T$ is the state space. Here, each state $s = \{l_i, t\} \in S$ represents the internal attention level of the human and the (time) state of the query model.*
- *$A = \{a_\phi, a_1^{AIGA}, \dots, a_m^{AIGA}, a_1^{PER}, \dots, a_m^{PER}\}$ is the action space. a_ϕ corresponds to no action, or no query displayed on the HMI. a_j^{AIGA} or a_j^{PER} refer to the i^{th} type of active information gathering actions (e.g. attention query) or i^{th} type of query from the perception module (e.g. environment query) respectively. Also let the set of AIGA be A^{AIGA}, and the queries from the perception module be A^{PER}, s.t. $A = A^{AIGA} \cup A^{PER} \cup a_\phi$.*
- *$O = \{O_\phi, O_1, \dots, O_P\}$ is the observation space, which consists of responses from the human to the query or other auxiliary measurements on the human, e.g. from driver gaze tracking or pose detection. Here O_ϕ corresponds to no response, and $O_1, \dots, O_P$ are the possible responses to the displayed query.*
- *$R : S \times A \times O \to \mathbb{R}$ is a reward function that captures the utility of the human's response to a query.*
- *$\mathbb{T} : S \times A \times O \to S$ is the state transition function which contains conditional probabilities of the form $\mathbb{T}(s'|s, a, o)$. Here s' refers to the state of the model at a time step $k+1$, and s, a, o refer to the state ,action and observation (respectively) at time step k*[4]*.*
- *$\mathbb{O} : S \times A \to O$, the observation function $\mathbb{O}$ contains conditional probabilities of the form $\mathbb{O}(o|s, a)$ and represents the probability of the human giving a particular response to a query based on the attentiveness level and time steps elapsed in the query.*
- *$\gamma \in (0, 1)$ is a discount factor.*

Here, actions $a \in a_\phi \cup A^{AIGA}$ are *controllable* in the sense that they can be deployed through a policy (see figure 2) in order to monitor or influence the human's attentiveness level l. The actions from the perception module $a \in A^{PER}$ are triggered when the perception module needs to actually perform a perception hand-off. In order to ensure that the HMI is not displaying an AIGA when a perception hand-off needs to happen, we impose the following assumption on the structure:

Assumption 2 (Precedence of $a \in A^{PER}$ over $a \in A^{AIGA}$)**.** *If a policy wants to deploy an AIGA at the same time that the*

[3]Queries that have not timed out and for which the HMI has not yet received a response from the human.

[4]Unlike a standard POMDP, the state transitions are conditioned on the output due to the counters of the query state as in definition 2.

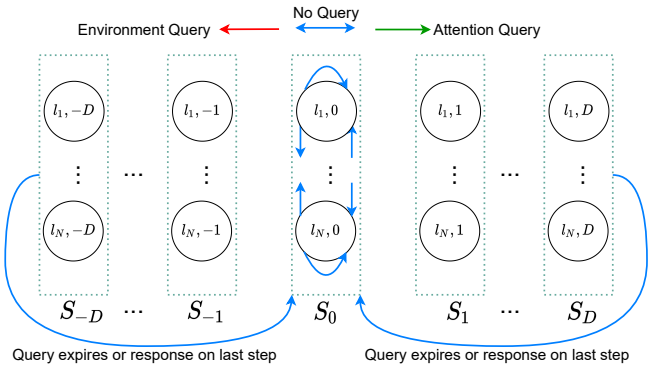

Fig. 5: The state-space for the proposed POMDP model for the AV-human decision making hand-off.

perception module requires a perception hand-off, the HMI will override the policy and perform the perception hand-off, i.e. $a \in A^{PER}$ *has precedence over* $a \in A^{AIGA}$.

A. The problem specific structure of the model

Here, we define some elements of structure of the POMDP developed above that make it specifically suited for modeling the perception hand-off process. Let $S_t \subset S$ represent the set of states $s = \{., t\}$ where the query state is $t \in T$. The state space S and these subsets are shown in figure 5.

1) Absence of an active query: At a time step k, when there is no active query on the HMI, the state takes a value $s[k] \in S_0$. If there is no new query at time step k, i.e. $a[k] = a_\phi$, then $s[k+1] \in S_0$. Note, only states in S_0 can self transition in the absence of an active query.

2) Active information gathering action: Assume that $s[k] \in S_0$. If $a[k] \in A^{AIGA}$, then $s[k+1] \in S_1$, the further evolution of states is covered in the cases below:

Case 1: No response at time step $k+1$. If $o[k+1] = o_\phi$, then the query remains active and the next state $s[k+2] \in S_2$ and so on until either a response is reached or the query times out.

Case 2: Response from human at time step $k+1$. If $o[k+1] \neq o_\phi$, then the query is now inactive and the query state resets s.t. $s[k+2] \in S_0$.

Case 3: Query time out. If there is no human response until the point $s[k+D] \in S_D$, and the human does not respond on the last time step of the query, i.e. $o[k+D] = o_\phi$, then the query times out and is inactive, and the state resets s.t. $s[k+D+1] \in S_0$.

3) Actions from the perception module: For actions from the perception module $a[k] \in A^{PER}$, the state transitions and observations have a similar structure as for the AIGAs. We use a notation here that the counter for states when $a[k] \in A^{PER}$ decrements (see the query model, definition 2) s.t. for no response from the human at a state $s[k] \in S_{-i}$, the next state is $s[k+1] \in S_{-i-1}$, $\forall i \in \{D-1, \ldots, 0\}$.

The above structural constraints are visualized in Figure 5.

4) Precedence of actions from the perception module: Finally, another structural constraint is imposed by assumption 2. This implies that if $s[k] \in S_i$, $i \in \{0, \ldots, D\}$ and $a[k] \in A^{PER}$, then $s[k+1] \in S_{-1}$.

The modeling choices highlighted above introduces structural constraints, and sparsity on the state transition function $\mathbb{T}$ and capture the relevant behaviors for the time-sensitive HRI that is the perception hand-off. The following example covers an instance of the model that is specific to this work.

Example 2 (A POMDP for the Handoff Experiment)**.** *We model the perception hand-off experiment in section IV through the following modeling choices: 1a) The human has two attentiveness levels* $L = \{l_1, l_2\}$*, where the two levels* l_1 *and* l_2 *correspond the human being inattentive or attentive respectively, 1b) The* $4s$ *deadline for answering queries is discretized into* $D \geq 1$ *time buckets, 2) The action space is* $A = \{a_\phi, a^{AIGA}, a^{PER}\}$*, where* a^{AIGA} *is the* attention *query and the environment query is* a^{PER}*, 3) The observation space is* $O = \{o_\phi, o_C, o_I\}$ *where* o_ϕ *is no response to a query,* o_C *is a correct and* o_I *is an incorrect response.*

Figure 8 in the Appendix shows the structure of such a model, and we discuss how insights from the hypothesis tests in section IV inform the model's state transition and observation probabilities in Appendix E.

B. Learning the Model from experimental data

In order to learn a model similar to the one proposed above in example 2 from the dual-task experiment data, we use the Baum-Welch algorithm [10], that aims to find the POMDP state transition ($\mathbb{T}$) and observation ($\mathbb{O}$) parameters that maximize the likelihood $\max_{\mathbb{T},\mathbb{O}} P(\mathbf{o}|\mathbf{a}; \mathbb{T}, \mathbb{O})$ via Expectation Maximization (EM). Here, $\mathbf{o} = o[1], \ldots, o[k_{\max}]$ and $\mathbf{a} = a[1], \ldots, a[k_{\max}]$ are the discretized time series of observations and actions collected via the dual-task experiment. Note, unlike in a standard POMDP where the state transition probabilities are conditioned only on the current state and action, our model has state transition probabilities that are additionally also conditioned on the current observation (see definition 3). Appendix F briefs how we adapt our model to be able to use the Baum-Welch algorithm to learn it from data.

C. Learning a policy for the HMI

With a model of the perception hand-off HRI and a method to learn it from data, we next want to exploit the model's suitability for control by developing a policy to use the AIGAs and influence the human to make better decisions in perception hand-offs (Problem 1). Given the model structure and assumptions in Section V, this policy is dependent on the perception module's behavior (Figure 2). To take this into account, we make the following simplifying assumption.

Assumption 3 (Probability of Environment queries)**.** *Environment queries are deployed at random with a constant probability* p*, i.e. at any time step* k*,* $P(a[k] = a^{PER}) = p$.

TABLE I: Ratio of likelihoods (over data for 250 time steps) of the learned POMDP model versus POMDPs with the same structure but randomly generated parameters. The evaluation is done over models with varying number of time steps per query D, i.e. discretizing the $4s$ until the query deadline into bins with different sampling times dt. We compare the likelihoods to the average likelihood from 10 random models ($Ratio_{avg}$) and to the random model with the highest likelihood ($Ratio_{best}$). A ratio ≤ 1 implies the random model fits the data as well or better than the learned model, while ratios > 1 imply that the learned model better represents the data.

Query deadline	$D = 1$	$D = 2$	$D = 3$	$D = 4$
$Ratio_{avg}$	**52.6**	10.0	11.32	17.9
$Ratio_{best}$	**34.4**	4.1	4.6	6.5

This assumption allows us to develop a model where we can marginalize out the impact of the environment queries on the dynamics and have a transition function dependent only on the AIGA (the attention query). Appendix G1 covers the details of this. Given a learned POMDP and assumption 3, we use an off-the-shelf approach (Appendix G) to obtain a policy for using the attention queries to maximize a reward (3) that encourages correct and faster responses from states s s.t.$s = \{., t\}$, $t \in \{-D, \ldots, -1\}$.

VI. Case Study: Learning a model and a policy for the human-AV perception hand-off process

In this section, we first show the ability of the proposed model to represent the perception hand-off data gathered via the experiment in section IV, which gives us time series data for 40 human subjects performing a trial of 10 minutes each. Next, we show that the model is suited to control, or influence, the human's attentiveness levels.

A. Learning a model from data

As outlined in Section V-B, we can learn the state transition and observation probabilities of our model from collected data. We use a subset of the collected data (over multiple human subjects in the experiment) to learn a model with different sampling times dt and associated number of time steps in a query before it expires, D. Section E1 discusses some of the choices and insights from the dual-task experiment used in learning the model. Figure 9 shows the learned observation probabilities for states with low and high attention levels for each time step in the query for $D = 3$.

Next, we evaluate the likelihood of the action-observation sequence (see Section V-B) over a smaller subset of the data, and compare it to the likelihood of obtaining this sequence from models with a similar structure but randomly generated state and observation transition probabilities. Table I shows how the learned models (for different values of D) are a much better fit than the random models. In addition to the specific structure of our model, this can also be partly attributed to the modeling insights (Section E) used to initialize the Baum-Welch algorithm. The model with $D = 1$ has the highest (absolute and relative) likelihood over the experimental data. This model lumps all responses and response times into a single time step in the query, and this results in a simpler model working on a coarse time scale ($dt = 4s$ sampling time) that can represent the aggregate data better. Due to the coarse timescale however, this model is not well suited to closed loop applications. Models with $D > 1$ also fit the data well, while allowing for more fine grained (in time) HRI. The likelihoods are in general small (of the order of 10^{-2} for $D = 1$, and 10^{-3} for D ≥ 2) since we compute them over sequences of hundreds of time steps.

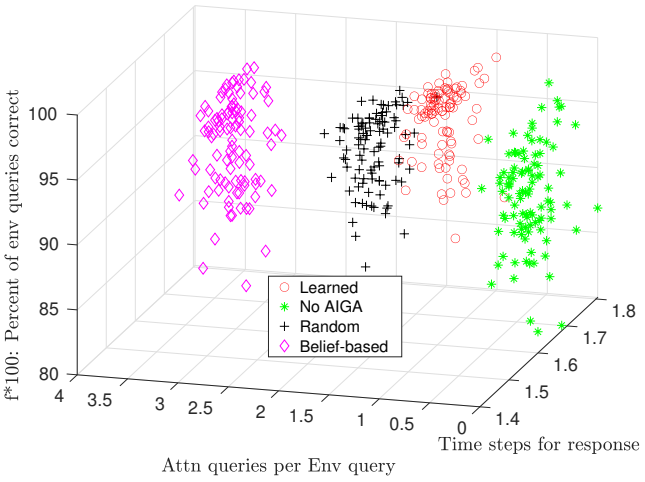

Fig. 6: Averages for 100 simulations runs (of 100 time steps, for each policy) of the percentage of environment queries answered correctly, time steps taken for a response, and number of attention queries asked per environment query. The learned policy results in a higher average of environment queries with correct responses, as well as uses less attention queries per environment query asked. Also see table II.

B. AIGA policy for perception hand-off

From the learned model (we pick the setting of $D = 3$), we can also learn a policy as in Section V-C to deploy the AIGA (attention queries) to maximize a reward function that corresponds to valuing correct and early responses from the human to actions from the perception module, or the actual perception hand-off (environment) queries. We also compare this learned policy to three baselines: a) a *random* policy that asks an attention query with probability of 0.5 at any time step, b) a *no AIGA* policy that does not ask any attention queries, and c) a *belief-based* heuristic that deploys attention queries when the belief over states with a low attention level is sufficiently high, see section G3 for details. Table II shows how the learned policy results in a higher reward, and also shows the other quantities relevant to the perception hand-off.

C. Summary

In addition to being able to model the experimental data gathered for the perception hand-off as gathered by the web experiment, the proposed POMDP is also suitable for learning a policy to interact with the human during such hand-offs as shown in Table II. The table presents the averages and standard deviations, over 100 runs, with random initial state $s[0] \in S_0$, of 100 time steps each, of the following quantities:

Accumulated reward: The learned policy (see Section V-C) results in a higher accumulated reward than the baseline policies. Notably, this also shows how deploying the AIGA in a systematic manner can show improvements over not using the AIGA or deploying it *randomly*.

Percentage of environment queries correctly responded to: For the chosen parameters for the reward function (3) (see section H2), the learned policy also results in the highest percentage $f*100$ of environment queries with correct responses.

Time to respond to environment queries: In terms of number of time steps for a response, the belief-based policy results in fastest responses on average. This is possibly due to the higher number attention queries deployed by this policy, as opposed to the learned policy or even the random policy.

Number of AIG actions taken per perception hand-off (environment) query: The learned policy again results in the least number of attention queries per environment query asked. This due to the reward function that penalizes asking attention queries, which would in practice be to avoid causing a fatigue to the driver by querying them too frequently.

Figure 6 shows the averages of these quantities for each of the 100 runs for all the policies. Note, even though the belief-based policy asks the most number of attention queries per environment queries, the learned policy results in a higher f, showing the benefit of using the AIGA in a systematic manner. Additional details about the implementation and simulation results are in Appendix H.

VII. Discussion

Summary: In this paper we present a model-based formalization of the perception hand-off, or the problem of bringing the human in the decision making loop when the perception model of an autonomous system is uncertain about the environment. We collect data on such a Human-Robot Interaction via a web-based human study, and use it to learn parameters for the proposed model and to also explore the use of an active information gathering (AIG) mechanism (attention queries) to influence the human attentiveness level. We also learn a policy for leveraging the AIG mechanism and show the benefit of our approach through the experimental data and simulations.

TABLE II: Performance of policies on learned model with $D = 3$. The table shows the means $\pm$ standard deviations across 100 simulation runs of 100 time steps each. Here, $f * 100$ represents the percentage of environment queries (a^{PER}) that were correctly responsed too, T_{resp} is the number of time steps taken on average for a response, $\#a^{AIGA} : \#a^{PER}$ is the averaged ratio of attention queries asked for one environment query. Finally, R is the average accumulated reward for each policy.

Policy	Reward (R)	T_{resp}	$f * 100$	$\#a^{AIGA} : \#a^{PER}$
Learned	**15.52** $\pm$ 5.27	1.56 $\pm$ 0.05	**98.2**$\pm$ 2.8	**0.83** $\pm$ 0.12
No AIGA	11.29 $\pm$ 5.55	1.57 $\pm$ 0.07	92.8$\pm$ 3.9	0
Random	11.78$\pm$ 6.42	1.55 $\pm$ 0.04	95.4 $\pm$ 3.1	1.48 $\pm$ 0.14
Belief	13.83 $\pm$ 4.39	**1.54** $\pm$ 0.03	95.9 $\pm$ 2.8	2.9 $\pm$ 0.11

Limitations and future work: Our work is limited in many ways. First, the human subject experiment for the perception hand-off was conducted via a web experiment in a game-like environment. Here, we lacked many of the signals that could otherwise be collected in an in person study simulating autonomous driving, e.g. gaze tracking, pose detection etc. The experiment design in its current form also restricted the use of AIG actions to once every 10s. While this still resulted in a statistically significant speed up in the response time of the human, the lack of fine grained control resulted in an insignificant increase in the correctness of the human responses when the AIG mechanism was used. Apart from the limitations of the experimental setup, the proposed model for this perception hand-off interaction lacks the ability to capture the *alarm fatigue* that would be created by too many attention queries. This both makes the problem of reward engineering to penalize these queries to limit too many queries at the cost of overall performance hard, and also would also imply that policies that simply deploys attention queries as often as possible result in the faster responses, as seen in section VI-C. Finally, the policy learning was done with a simplifying assumption on the environment queries, which would not necessarily hold in a realistic setting.

Future work will focus on validating the results regarding the impact of the policy via another human subject web experiment with the learned (and baseline) policies operating in the loop. This experiment would allow for the AIG actions to be deployed on a finer time scale than the one in this paper. Next, we will study the perception hand-off via an in-lab virtual reality-based study. This would necessitate the use of more expressive human machine interfaces (HMIs) with AIG actions that are not just visual. Additional continuous-time signals collected from this experiment (e.g. gaze tracking, galvanic skin resistance) would require us to develop a hybrid model that can account for both continuous and discrete observations. The policy design problem will also be done with more realistic assumptions on the perception hand-offs.

Conclusion: This work represents a first effort at model-based use a HMI for influencing human behavior during a perception hand-off. The experimental and simulation results are encouraging and show the potential of further developing such an approach for a real world autonomous driving setting.

ACKNOWLEDGMENT

The authors thank Steven Le, the creator of the driving game *getaway* (https://github.com/le-s/getaway), for allowing us to use the source code of the game as a starting point for our experiment. We would also like to thank everyone who participated in the experiment on mturk. This work was supported in part by NSF CPS Frontier project VeHICaL (CNS-1545126), by Toyota under the iCyPhy center, and by Berkeley Deep Drive.

REFERENCES

[1] Why Japans Rail Workers Cant Stop Pointing at Things. https://www.atlasobscura.com/articles/pointing-and-calling-japan-trains, 2017. Accessed: 02-08-2021.

[2] SAE Standards News: J3016 automated-driving graphic update. https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic, 2019. Accessed: 02-08-2021.

[3] STOP THE TROLLEY! California Autonomous Driving Test Statistics 2019. http://keerthanapg.com/stop-the-trolley/, 2020. Accessed: 02-25-2021.

[4] Sinan E. Arkonac, Duncan P. Brumby, Tim Smith, and Harsha Vardhan Ramesh Babu. In-car distractions and automated driving: A preliminary simulator study. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, AutomotiveUI '19, page 346351, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450369206. doi: 10.1145/3349263.3351505. URL https://doi.org/10.1145/3349263.3351505.

[5] Gisa Aschersleben and Jochen Müsseler. Dual-task performance while driving a car: Age-related differences in critical situations. In *Proceedings of the 8th annual conference of the cognitive science society of Germany. Saarbrücken*, 2008.

[6] Richard Bellman. A markovian decision process. *Indiana Univ. Math. J.*, 6:679–684, 1957. ISSN 0022-2518.

[7] Radu Calinescu, Naif Alasmari, and Mario Gleirscher. Maintaining driver attentiveness in shared-control autonomous driving. *arXiv preprint arXiv:2102.03298*, 2021.

[8] Stuart K. Card, Thomas P. Moran, and Allen Newell. The model human processor- an engineering model of human performance. *Handbook of perception and human performance.*, 2(45–1), 1986.

[9] Nakul Gopalan and Stefanie Tellex. Modeling and solving human-robot collaborative tasks using pomdps. In *RSS Workshop on Model Learning for Human-Robot Communication*, volume 32, pages 590–628, 2015.

[10] S. Koenig and R. G. Simmons. Unsupervised learning of probabilistic models for robot navigation. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 3, pages 2301–2308 vol.3, 1996. doi: 10.1109/ROBOT.1996.506507.

[11] Minae Kwon, Erdem Biyik, Aditi Talati, Karan Bhasin, Dylan P. Losey, and Dorsa Sadigh. When humans aren't optimal: Robots that collaborate with risk-aware humans. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2020. doi: 10.1145/3319502.3374832.

[12] Hye-In Lee, Seungha Park, Jongil Lim, Seung Ho Chang, Jung-Hyun Ji, Seungmin Lee, Jihye Lee, et al. Influence of drivers career and secondary cognitive task on visual search behavior in driving: a dual-task paradigm. *Advances in Physical Education*, 5(04):245, 2015.

[13] National Transportation Safety Board (NTSB). Highway Accident Report: Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian. https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf, 2019. Accessed: 02-08-2021.

[14] Frederik Naujoks, Dennis Befelein, Katharina Wiedemann, and Alexandra Neukum. A review of non-driving-related tasks used in studies on automated driving. In *International Conference on Applied Human Factors and Ergonomics*, pages 525–537. Springer, 2017.

[15] Lisa Ordez and Lehman Benson. Decisions under time pressure: How time constraint affects risky decision making. *Organizational Behavior and Human Decision Processes*, 71(2):121–140, 1997. ISSN 0749-5978. doi: https://doi.org/10.1006/obhd.1997.2717. URL https://www.sciencedirect.com/science/article/pii/S0749597897927175.

[16] Pedro Ortega and Alan Stocker. Human decision-making under limited time. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

[17] Sirisha Rambhatla, Xingguo Li, and Jarvis Haupt. Provable online cp/parafac decomposition of a structured tensor via dictionary learning. *Advances in Neural Information Processing Systems*, 33, 2020.

[18] Dorsa Sadigh, S. Shankar Sastry, Sanjit A. Seshia, and Anca Dragan. Information gathering actions over human internal state. In *Proceedings of the IEEE, /RSJ, International Conference on Intelligent Robots and Systems (IROS)*, pages 66–73. IEEE, October 2016. doi: 10.1109/IROS.2016.7759036.

[19] Dorsa Sadigh, S. Sankar Sastry, and Sanjit A. Seshia. Verifying robustness of human-aware autonomous cars. In *Proceedings of the 2nd IFAC, Conference on Cyber-Physical and Human Systems*, December 2018. doi: 10.1016/j.ifacol.2019.01.055.

[20] Dario D. Salvucci, Mark Zuber, Ekaterina Beregovaia, and Daniel Markley. Distract-r: Rapid prototyping and evaluation of in-vehicle interfaces. In *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems, CHI '05, page 581589, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 1581139985. doi: 10.1145/1054972.1055052. URL https://doi.org/10.1145/1054972.1055052.

[21] Lara Scatturin, Rainer Erbach, and Martin Baumann. Cognitive psychological approach for unraveling the take-over process during automated driving. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, AutomotiveUI '19, page 215220, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450369206. doi: 10.1145/3349263.3351501.

[22] Marlene Scharfe and Nele Russwinkel. A cognitive model for understanding the takeover in highly automated driving depending on the objective complexity of non-driving related tasks and the traffic environment. In *CogSci*, pages 2734–2740, 2019.

[23] Victor A Shia, Yiqi Gao, Ramanarayan Vasudevan, Katherine Driggs Campbell, Theresa Lin, Francesco Borrelli, and Ruzena Bajcsy. Semiautonomous vehicular control using driver modeling. *IEEE Transactions on Intelligent Transportation Systems*, 15(6):2696–2709, 2014.

[24] Kazumitsu Shinohara, Hiroshi Naito, Yuko Matsui, and Masaru Hikono. The effects of finger pointing and calling on cognitive control processes in the task-switching paradigm. *International Journal of Industrial Ergonomics*, 43(2):129–136, 2013. ISSN 0169-8141. doi: https://doi.org/10.1016/j.ergon.2012.08.004. URL https://www.sciencedirect.com/science/article/pii/S0169814112000728.

[25] Weilong Song, Guangming Xiong, and Huiyan Chen. Intention-aware autonomous driving decision-making in an uncontrolled intersection. *Mathematical Problems in Engineering*, 2016, 2016.

[26] W. Zheng, B. Wu, and H. Lin. Pomdp model learning for human robot collaboration. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 1156–1161, 2018. doi: 10.1109/CDC.2018.8618904.

APPENDIX

A. Details on each component of the web experiment

1) Driving related task: The AV encounters objects from 3 different classes, at any instant in time, there is no more than one object on the road. For some instances of these objects, the AV is unsure about which class the object is from, and requires input from the user. For these objects, a flashing blue box appears on top of them, and the HMI displays a query that the user should respond to, before it is too late for AV to perform an appropriate behavior ($4s$ in our experiment). A correct (incorrect) response results in the blue box changing its color to flashing green (red).

2) Distractor task: The distractor task is a series of randomly generated arithmetic questions involving subtraction of (upto) 2 digit numbers. All answers are a single digit. The user has exactly 5 seconds to answer and is then informed whether or not their submitted answer is correct or incorrect. The presence of the distractor task toggles every 150s.

3) HMI for querying the user: The primary task of the experiment is object identification. A new object appears on the road every 10 seconds. Some objects will require a response from the user, others do not. When prompted for a response, users have four seconds to identify the object from three options. They are notified whether or not they answered a query correctly (through flashing colored borders around the full panel as well as the object).

B. Crowd-sourcing deployment

The game was hosted on Heroku, utilizing the psiturk library for ease of implementation in mturk. Before beginning the experiment, the users watched a four-minute instructional video on youtube. Data is logged throughout gameplay, including timestamps and sent to a secure database following the conclusion of the game. Following the completion of the game, users took a survey embedded in the web app. Following that, they were given the option of completing a bonus questionnaire. The data was cleaned and parsed using the pandas python library. We then performed data validation manually removing participant data where the data showed that they were not focusing. We ensured this by checking the user's browser behavior throughout the experiment, seeing when they navigated away from the page, minimized it, or switched tabs.

C. Interleaving of attention and environment queries

Every three objects can be considered to form a *set*. The AV came across these objects as sets, one after the other. Every third object in a *set*, has an environment query asked over it. An attention query is asked on either the first or the second object of this set. A *set* can have one of three following *orders*:

- $Order_{\text{-}1}$: Attention query is asked on the object, immediately before the one with environment query. (also, the 2^{nd} object in the set)
- $Order_{\text{-}2}$: Attention query is asked on the object, that is two before the one with environment query. (also, the 1^{st} object in the set)
- $Order_{\phi}$: No attention query is asked on the set

The *order* in these *sets* is varied across the span of the experiment. At the end of every *set*, the *order* for the next *set* is determined with an equal probability (0.33).

D. Hypotheses testing from the experimental data

In this subsection, we report the results of experiments to study the effect of attention queries on the users performance on environment queries. We expected that the presence of

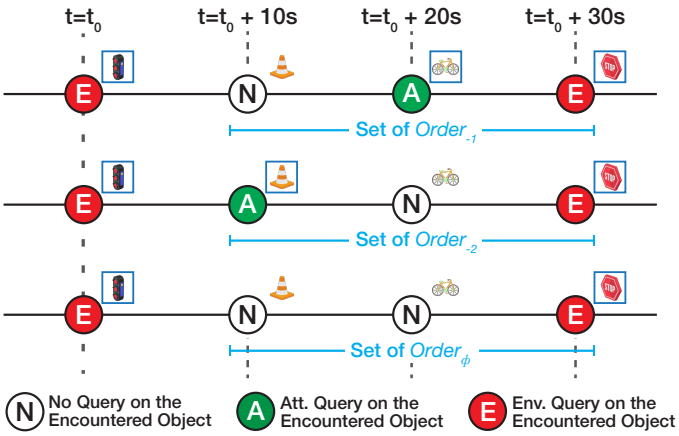

Fig. 7: Different Types *Orders* possible in deployment of attention and Environment queries on objects encountered by the AV

TABLE III: List of all hypotheses tested in the dual-task web experiment

Hypothesis	Brief Description
1A	RT with distractor greater than without
1B	f with distractor lower than without
2A	RT of $Order_{\text{-}1}$ less than $Order_{\phi}$ (distractor present)
2B	f of $Order_{\text{-}1}$ greater than $Order_{\phi}$ (distractor present)
3A	RT of $Order_{\text{-}2}$ less than $Order_{\phi}$ (distractor present)
3B	f of $Order_{\text{-}2}$ greater than $Order_{\phi}$ (distractor present)
4A	RT of $Order_{\text{-}1}$ less than $Order_{\phi}$
4B	f of $Order_{\text{-}1}$ greater than $Order_{\phi}$
5A	RT of $Order_{\text{-}2}$ less than $Order_{\phi}$
5B	f of $Order_{\text{-}2}$ greater than $Order_{\phi}$

attention queries on objects would increase user awareness of the driving task. This might lead them to perform better on a succeeding query(s). We formulated the following Hypotheses and tested them using a paired-sample t-test.

Hyp. 4A. *Across conditions, the RT of an environment query such that there is an attention query on the preceding object (Experimental Condition - $Order_{\text{-}1}$) will be less than ones where there is no such attention query on at least two objects before it (Control Condition - $Order_{\phi}$).*

Hyp. 4B. *Across conditions, the f of an environment query such that there is an attention query on the preceding object (Experimental Condition - $Order_{\text{-}1}$) will be greater than ones where there is no such attention query on at least two objects before it (Control Condition - $Order_{\phi}$).*

Hyp. 5A. *Across conditions, the RT of an environment query such that there is an attention query on the second object preceding it (Experimental Condition - $Order_{\text{-}2}$), will be less than ones where there is no such attention query on at least two objects before it (Control Condition - $Order_{\phi}$).*

Hyp. 5B. *Across conditions, the f of an environment query such that there is an attention query on the second object preceding it (Experimental Condition - $Order_{\text{-}2}$), will be greater than ones where there is no such attention query on at least two objects before it (Control Condition - $Order_{\phi}$).*

We were unable to reject the null hypotheses at a significance level of $p < 0.05$. No significant effect of attention queries were found on the environment queries across the conditions (distractor tasks present or absent). We report the hypothesis-wise results below, for reference.

Hyp. 4A: For Experimental Condition ($Order_{\text{-}1}$), Mean RT = 1937ms, SD = 506ms; For Control Condition ($Order_{\phi}$), Mean RT = 1926ms, SD = 525ms

Hyp. 4B: For Experimental Condition ($Order_{\text{-}1}$), Mean f = 0.949, SD = 0.106; For Control Condition ($Order_{\phi}$), Mean f = 0.952, SD = 0.107

Hyp. 5A: For Experimental Condition ($Order_{\text{-}2}$), Mean RT = 2001ms, SD = 529ms; For Control Condition ($Order_{\phi}$), Mean RT = 1922ms, SD = 531ms

Hyp. 5B: For Experimental Condition ($Order_{\text{-}2}$), Mean f = 0.946, SD = 0.103; For Control Condition ($Order_{\phi}$), Mean f = 0.951, SD = 0.108

We did not find a statistically significant effect of attention queries on the performance of environment queries.

However, when this is studied specifically across a dual-task setting, where the distractor task is present, we observed a statistically significant reduction in response time, in case of $Order_{\text{-}1}$ when compared to $Order_{\phi}$. We report the hypotheses-wise results of Hyp.2A-3B below, for reference.

Hyp. 2A: For Experimental Condition ($Order_{\text{-}1}$), Mean RT = 2018ms, SD = 555ms; For Control Condition ($Order_{\phi}$), Mean RT = 2172ms, SD = 533ms; Statistically significant with $t(39) = 3.24$, $p < 0.05$, Cohen's D = 0.27

Hyp. 2B: For Experimental Condition ($Order_{\text{-}1}$), Mean f = 0.940, SD = 0.124; For Control Condition ($Order_{\phi}$), Mean f = 0.932, SD = 0.173

Note, this slight average increase in response accuracy due to attention queries was not statistically significant.

Hyp. 3A: For Experimental Condition ($Order_{\text{-}2}$), Mean RT = 2178ms, SD = 647ms; For Control Condition ($Order_{\phi}$), Mean RT = 2169ms, SD = 539ms

Hyp. 3B: For Experimental Condition ($Order_{\text{-}2}$), Mean f = 0.933, SD = 0.156; For Control Condition ($Order_{\phi}$), Mean f = 0.930, SD = 0.175

These are discussed in section IV-B. Additionally, Table III provides a summary of all the hypotheses that we tested.

E. Model structure for the hand-off experiment

As described in example 2, we can model the perception hand-off studied in the dual-task web experiment (section IV) as a Partially Observable Markov Decision Process of the form in Definition 3. Figure 8 shows a simplified version of this model, with $D = 1$ or the entire $4s$ duration to respond to a query discretized into a single time step

1) Connecting the model structure to the dual-task experiment: Some notable features connecting the model to insights from the dual-task experiment data are:

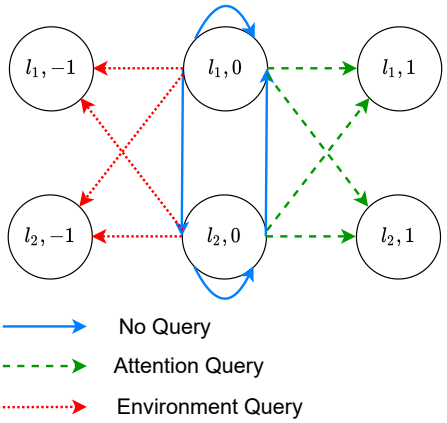

Fig. 8: The state space for the POMDP model formalizing the perception hand-off experiment studied in section IV. Shown here are the possible 1-step state transitions when starting in states $l_1, 0$ or $l_2, 0$ and under the different possible actions. Also see example 2.

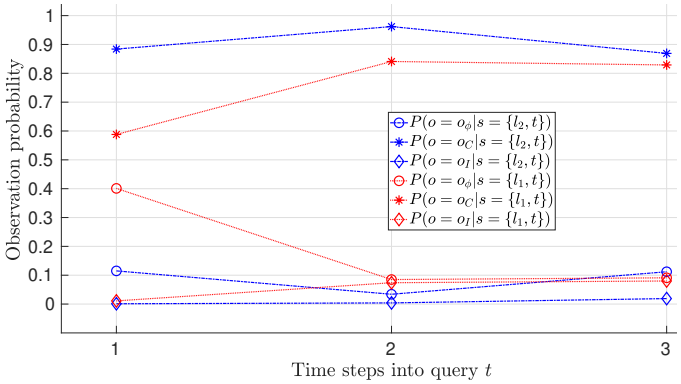

Fig. 9: Learned observation probabilities for all time steps in a query with $D = 3$. In our model since the attention and environment queries are displayed in an identical manner, we assume that the observation probabilities are the same for getting responses for both types of queries. The learned probabilities suggest that the human is more likely to answer queries correctly and earlier in the high attention level state $s = \{l_2, .\}$ than in the low attention level state $s = \{l_1, .\}$.

- *Human responses when inattentive:* As seen in hypotheses 1A and 1B, when the human subject is distracted, they are slower to answer queries, and also get them wrong more often. This is captured in the model as $\mathbb{O}(o_I | s = \{l_1, .\} > \mathbb{O}(o_I | s = \{l_2, .\}$, i.e. probability of incorrect response is higher when the attention level is low. Also, $\mathbb{O}(o_\phi | s = \{l_1, t\} > \mathbb{O}(o_\phi | s = \{l_1, t\}, t \in T$ (see definition 2), i.e. the probability of the human not responding at a time step in the query is higher if they are at a lower attention level. Figure 9 shows the learned observation probabilities for a model with $D = 3$ time steps in a query. Note how the probability of getting a correct response at a high attention level state $s = \{l_2, .\}$ is higher than that at low attention level states $s = \{l_1, .\}$. Also notable from the figure is the probability of getting no response at the first time step in the query (corresponding to the time interval $[0s, 1.33s]$ since query was asked) is much higher in the low attention level state.
- *Impact of AIGA/attention queries:* Hypothesis 2A shows the impact of well timed attention queries. The model captures this behavior by increasing the probability of switching to a state with a higher attention level once a query has been asked, i.e. $\mathbb{T}(s = \{l_2, t+1\} | s = \{l_1, t\}, a \neq a_\phi, .) > \mathbb{T}(s = \{l_1, t+1\} | s = \{l_1, t\}, a \neq a_\phi, .)$. As show in Figure 8, this implies that the probability of transitioning from $\{l_1, 0\}$ to $\{l_2, 1\}$ (or $\{l_2, -1\}$) is higher than that of transitioning to $\{l_1, 1\}$ (or $\{l_1, -1\}$).

To learn such a POMDP model from data, we use these insights in creating the initial POMDP transition and observation functions which are then iterated upon by the Baum-Welch algorithm [10] (also see section V-B).

2) Belief updates: From the model in Definition 3, we can monitor the human attention levels by using the Bayesian belief update below:

$$b_{s'}[k+1] = \eta^{-1} \mathbb{O}(o | s', a) \sum_{s \in S} \mathbb{T}(s' | s, a, o) b_s[k], \text{where,}$$
$$\eta = \sum_{s' \in S} \mathbb{O}(o' | s', a) \sum_{s \in S} \mathbb{T}(s' | s, a, o) b_s[k] \quad (1)$$

Here $b_s[k]$ represents the belief (or probability) that the actual state is s at time step k. Also, o' represents the observation at time $k+1$, a the action at time step k and o the observation at time k. Note, $\sum_{s \in S} b_s[k] = 1 \, \forall k$.

F. Transforming the POMDP for model learning

As explained in Section V, the state transition probabilities of the perception hand-off POMDP are conditioned on the previous state, action and observation. To use the Baum-Welch algorithm to learn this POMDP from data (section V-B), we need to transform it into a standard POMDP where state transitions are conditioned only on the state and action. To do so, we *lift* the state, and define a new state:

$$\hat{s}[k] = \begin{bmatrix} s[k] \\ o[k-1] \end{bmatrix} \in \hat{S} = S \times O$$

The state transition function for the next lifted state $\hat{s}'$ is now given by conditional probabilities of the form, $\hat{\mathbb{T}}(\hat{s}' | \hat{s}, a)$ This is now dependent only on the previous lifted state $\hat{s}$ and the action a. The corresponding observation function is simply: $\hat{\mathbb{O}}(o | \hat{s}, a) = \mathbb{O}(o | s, a)$. Note, the lifted state transition function is related to $\mathbb{T}$ and $\mathbb{O}$ as:

$$\hat{\mathbb{T}}([s', o'] \,|\, [s, o], a) = \mathbb{T}(s' \,|\, s, a, o') \mathbb{O}(o' | s, a)$$

The state transition matrix for this lifted state for a given a, o' can be computed by the Kronecker product of the associated matrices $\mathbb{T}(. | ., a, o')$ and $\mathbb{O}(o' | ., a)$. The resultant transition

probability matrices ($\hat{\mathbb{T}}$) and observation probability matrices ($\hat{\mathbb{O}}$) for the lifted state can be learned from the experimental data via the Baum-Welch algorithm [10]. The parameters $\mathbb{T}$ and $\mathbb{O}$ for the original model of Definition 3 can be recovered using a structured Kronecker product recovery method [17].

G. Learning a policy for the POMDP model

1) Marginalizing the actions from the perception module: The perception hand-off POMDP in Definition 3 has two types of actions, the set of controllable active information gathering actions (AIGAs) A^{AIGA} (and the no query action a_ϕ) and the set of actions from the perception module A^{PER}. In section V-C, we aim to learn a policy to use the AIGA to maximize a reward (3) over human responses to the queries from the perception module. Since the actions in A^{PER} are not controllable, we need to account for their impact before we can learn such a model. For the sake of simplicity, we explain this process through the POMDP for the perception hand-off as outlined in example 2. Assumption 3 states that the environment queries (a^{PER}) are deployed at any time step with a constant probability p. Using this, we can now *marginalize* out this action from the POMDP (a^{PER}), and obtain a transition function $\mathbb{T}_p$ over only the AIGA/attention queries (a^{AIGA}) as follows:

$$\begin{aligned}&\mathbb{T}_p(s'|s, a \in \{a^{AIGA}, a_\phi\}, o) = \\ &p\mathbb{T}(s'|s, a^{PER}, o) + (1-p)\mathbb{T}(s'|s, a \in \{a^{AIGA}, a_\phi\}, o)\end{aligned} \tag{2}$$

The resulting POMDP is then used to learn a policy to deploy the AIGA, as is explained next.

2) Value iteration-based learned policy: Given that solutions for POMDP policy learning are not exact and are often difficult to interpret, we used the Value Iteration algorithm [6] to derive an interpretable exact policy with respect to the POMDP in (2), and then compute the optimal action to take based on our belief of the POMDP at a given point in time. Results of policy performance were generated by deploying each policy in an environment based on the learned model, and recording the actions and rewards earned by the policy (which did not have access to the underlying state of the environment). The policy aims to maximize the following reward:

$$R = \sum_{k=0}^{\infty} \gamma^k r[k], \tag{3}$$

where, $\gamma \in (0, 1]$ and

$$r[k] = \begin{cases} -C_1, & \text{if } a[k] = a^{AIGA} \\ C_2, & \text{if } o[k] = o_C,\ s[k] \in S_i,\ i \in \{1, \dots, D\} \\ C_3\lambda^{|-i|}, & \text{if } o[k] = o_C,\ s[k] \in S_i,\ i \in \{-1, \dots, -D\} \\ -C_4, & \text{if } o[k] = o_I,\ s[k] \in S_i,\ i \in \{-1, \dots, -D\} \\ 0, & \text{otherwise.} \end{cases}$$

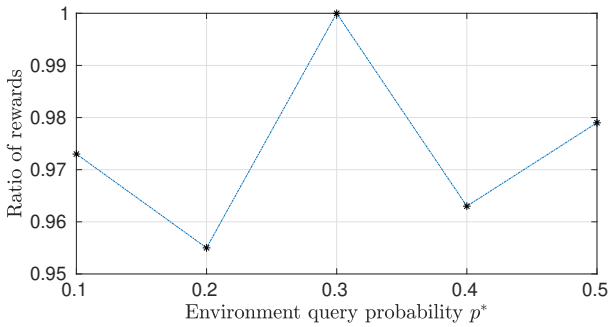

Fig. 10: Ratio of accumulated rewards for a policy learned assuming $p = 0.3$ versus a policy learned with the actual environment query probability p^*, averaged over 100 runs of 100 time steps each.

This reward function penalizes ($C_1 \geq 0$) AIGA actions to avoid asking too many attention queries to the human, but also rewards correct responses to attention queries $C_2 \geq 0$. For environment queries, it rewards correct responses earlier in the query $\lambda^k C_3 \geq 0$, where $\lambda \in (0, 1]$ is a factor that lowers the reward for responses later on. Finally, we also penalize incorrect responses to environment queries $C_4 \geq 0$, i.e. the human making a wrong decision in the perception hand-off.

3) Baselines for comparison: We compare the learned policy that maximizes the reward defined above to three baselines:

- *Belief-based:* This policy uses the belief over the states of the model at each time (1) to deploy attention queries if the sum of belief over states with the lower attention level i.e. over s' s.t., $s' = \{l_1, .\}$ is greater than the sum of belief over states with the higher attention level:

$$a[k] = \begin{cases} a^{AIGA}, & \text{if } \sum_{s'=\{l_1,.\}} b_{s'}[k] - \sum_{s'=(l_2,.)} b_{s'}[k] \geq \epsilon \\ a_\phi, & \text{otherwise.} \end{cases}$$

- *Random:* This policy randomly deploys attention queries s.t $P(a[k] = a^{AIGA}) = 0.5$. Note, these actions are still subject to assumption 2.
- *No AIGA:* Here, we don't use the AIGA. This baselines corresponds to the perception hand-off happening without any human monitoring or HMI policy in place.

H. Additional simulation results

1) Implementation details: The Baum-Welch algorithm of [10] for learning the model was implemented Python 3.7, as was the value-iteration algorithm for learning a policy. Simulation evaluations of the policy interacting with the learned model were done via an implementation using the openAI gym environment. For the reward function (3), we use the following parameters $\gamma = 0.99$, $\lambda = 0.95$, $C_1 = C_2 = 0.01$, $C_3 = 1, C_4 = -2$. For the belief-based baseline, we used $\epsilon = 0.1$.

2) Robustness of learned policy: We also evaluate the robustness of the learned policy to perturbations in the environment queries probability action p (see assumption 3). Figure 10 shows the impact of learning a policy assuming

a probability of $= 0.3p$, but evaluating it for a model with probability $p^* \in [0.1, 0.5]$. As seen in the figure, there is a graceful degradation in the accumulated rewards when there is a mismatch in p and p^*.