2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)
October 8-12, 2022, Macau, China

2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC) | 978-1-6654-6880-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ITSC55140.2022.9921935

# Modeling and Influencing Human Attentiveness in Autonomy-to-Human Perception Hand-offs

Yash Vardhan Pant[1], Balasaravanan Thoravi Kumaravel[2], Ameesh Shah[2], Erin Kraemer[2], Marcell Vazquez-Chanlatte[2], Kshitij Kulkarni[2], Bjoern Hartmann[2], Sanjit A. Seshia[2]

*Abstract*— **It is not uncommon for autonomous systems (e.g., self-driving cars) to require the timely intervention of a human operator to ensure safe operation. It is important to design these systems such 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 Autonomous Vehicle (AV) is uncertain about the environment. We formalize the perception hand-off problem by designing a Partially Observable Markov Decision Process (POMDP) model. This model captures the latent cognitive state (attention) of the driver which can be influenced through a proposed query-based active information gathering (AIG) system for Human-Machine Interface (HMI). We design a web-based human study to identify the model parameters, and demonstrate the impact of the proposed HMI system. Results from this study show that the state of attentiveness does indeed impact the human performance, and our proposed active information gathering (AIG) actions, i.e., queries to the human driver, result in $7\%$ faster responses from the human. Simulations with the identified POMDP model show that a learnt policy for deploying the AIG actions improves the percentage of correct responses from the human in the perception hand-off by around $5.4\%$, outperforming other baselines while also using fewer of these actions.**

## I. Introduction

The safe operation of autonomous and semi-autonomous systems sometimes requir 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 [1]. 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 [2]. In such *takeovers* or *hand-offs*, the human operator needs to be attentive and have spatial awareness of the road, once the AV asks them to take control. However, they might not have the knowledge of the situation as much as the AV does, since they were not controlling the vehicle 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, *only the perception task* that the AV cannot confidently perform. This would allow the vehicle to operate under the

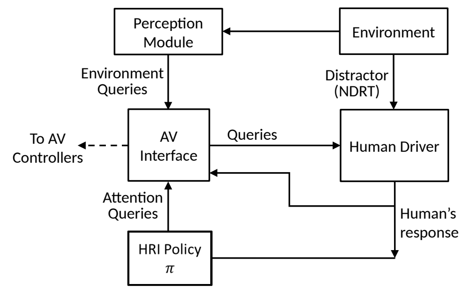


Fig. 1: An overview of the human-AV interaction in the hand-off process. The AV, through a human-machine interface (HMI), can query the human driver when its perception module requires help in decision making, or to gauge/influence the human's state of attentiveness. Further influencing the human state is a NDRT.

same (autonomous) control law and avoid sudden maneuvers unless necessary. 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.

We focus on applications in L2-L4 autonomous driving, as defined in the SAE J3016 [3], 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. The AV and the human interact via 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) 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) [4], and show the benefit of these, especially in situations where the human driver's state of attentiveness during autonomous driving is impacted by non-driving related tasks (NDRTs) or distractor tasks.

**Contributions:** In this paper, we define the *perception hand-off* problem, as a human-robot interaction in autonomous and semi-autonomous driving. The contributions of this work 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 mecha-

[1]Department of Electrical and Computer Engineering, University of Waterloo {yashpant}@uwaterloo.ca. This work was done while Yash Vardhan Pant was at UC Berkeley.
[2]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA
This work was supported in part by NSF grants CNS1545126 (VeHICaL), CCF-1837132, by an NDSEG Fellowship, the DARPA Assured Autonomy program, by Berkeley Deep Drive, and by Toyota under the iCyPhy center.

nism 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 in a setting that simulates such a perception hand-off process. This data is used to learn the proposed POMDP model, validate hypotheses on the operator behavior, and the impact of the query-based AIG mechanism; our approach achieves, on average, a $7\%$ speed up ($\approx 150ms$, see sec. IV-B) in human response times in the presence of a distraction, or NDRT[1].
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 $5.4\%$ using our approach via a simulation study (see section VI-B.1), which uses the learned model as a surrogate for the human.

## II. Related work

Here, 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:** Shia et al.[6] use measurements of the pose of the human driver of a semi-autonomous vehicle to correct the human input to the vehicle. Models with hidden latent states, usually POMDP-based, have been used to generate robot policies [7] or predict human intent [8] 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 [4] 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. A monitoring-based approach to alert the driver for a takeover is presented in [9]. 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.

**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 [1]. Dual-task experiments [10], [11], [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. Our web-based human study shows a similar trend (section IV).

**Cognitive models of humans in autonomous driving:** The human cognitive process when an AV requests the driver to take over control has been studied in [13], [14], [15]. 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).

[1]This is inline with [5], where visual stimuli to a human driving in the presence of a NDRT results in improvements of a similar order.

## 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 (see Figure 1). 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

*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.

We propose a Partially Observable Markov Decision Process with a specific structure to represent this HRI (section V). 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 [4].

### B. The human-AV interface: Querying 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 1. 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 (AIGA), or attention query:* Here, the AV does not actually require human intervention, but still queries 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

[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.

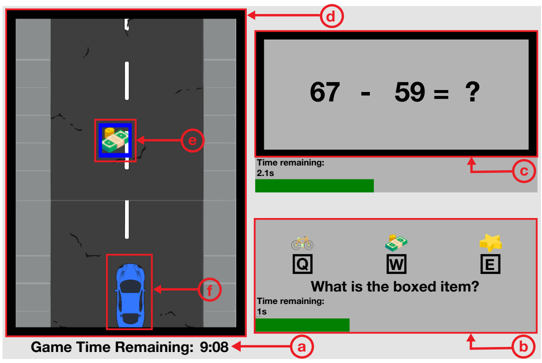


Fig. 2: The human subject experiment design for studying the hand-off process. Note: UI Text emphasized for clarity. A trial run can be seen at `https://youtu.be/LZRemqFBILA`.

there is no environment query currently on it. 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) to improve the human response for subsequent environment queries, i.e. the rate of correct responses and reduce response time.

The architecture for this hand-off HRI is shown in Figure 1. 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

We developed a web-based game that simulates the perception hand-off described above and 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 designed a *dual-task* human subject experiment where the human can interact with an AV. Dual-task refers to the the human subject performing both a driving related task and a *distractor* or non-driving related task (NDRT). The experiment, after obtaining IRB approval, was conducted on the Amazon mechanical turk platform, and we collected data from $N = 39$ users.

Figure 2 shows the UI that the subject of the experiment interacts with. The main components of this are:

1) **Driving related (primary) task:** Figure 2 shows the setup for the driving task. The user has a top-down view of an AV (Figure 2f) 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 2b) within a deadline.
2) **Distractor task:** For the NDRT (Figure 2c), the user has to solve basic arithmetic questions. They have upto 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 label the object.

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 (which appear every 10s). 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.

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. Note, the AV encounters only one object at a time.

### A. Attention query order: Interleaving of attention and environment queries

Every three objects in the driving task can be considered to form a *set*. In our study, every third object in a *set*, has an environment query asked on it. An attention query is asked on either the first or the second object of this set. A *set* can have one of three *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)
- $Order_{\varphi}$ : 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). We varied two conditions across the span of the experiment - Presence / Absence of the Distractor task - 2 levels, and the *order* in which queries are deployed - 3 levels. These conditions were varied in a within-subjects manners, meaning every subject experienced all the 2x3 combinations of the conditions.

### B. Brief summary of results from the human subject study

Through data collected from this study[3], we note two main effects on human performance:

*1. Effect of distractor:* In the presence of a distractor task, there was a statistically significant *increase* (average of $364ms$) in human response time for the primary (driving) task and a statistically significant *decrease* (by 7%) in the fraction of driving-related queries answered correctly. A

[3]Due to space constraints, we omit details about the statistical tests performed here. Details of the statistical results (and tests) from the human study can be found in the extended online report on this work [16].

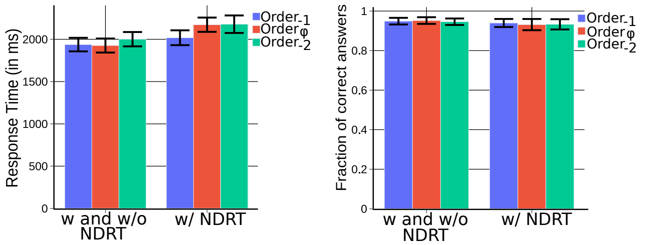


Fig. 3: Average Response Times ($RT$) and Fraction of queries correctly answered ($f$) for the different conditions

post-hoc paired-sample T-test confirmed this with $p < 0.05$ and Cohen's D$= 0.72$ (large effect size). See [16] for details.

*2. Effect of attention queries in the presence of a distractor task:* For the driving-related queries designated as environment queries, the human response time was *faster* in a statistically significant manner (**on average by $154ms$, or $7\%$**) when the environment queries immediately (i.e., on objects that appeared $10s$ ago) were preceded by an attention queries, as compared to the case when they were not. A post-hoc paired-sample T-test (details in [16]) confirmed this with $p < 0.05$ and Cohen's D$= 0.67$ (large effect size).

We thus observe that attention does indeed impact the performance of a human in the perception hand-off setting. More importantly, we also note that the *AIGA (attention queries) improves human performance*. These form the basis for the modeling in the rest of this paper. Note, in the absence of distractor task, the attention queries do not impact human performance in a statistically significant manner (details in [16]). This is expected, since with no distraction, human attention level for the driving task is expected to be high.

## 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):* We posit that relevant to the perception hand-off, the human has $L = \{l_1, \ldots, l_N\}$ 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 $T_{\text{max}} = Ddt$ 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 1). 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 $t \in T$ where, $T = \{-D, \ldots, -1, 0, 1, \ldots, D\}$. In the absence of any active queries[4], the state is 0. 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 there is no response 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 state again resets to 0. 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.

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.
- $A = \{a_\phi, a_1^{AIGA}, \ldots, a_m^{AIGA}, a_1^{PER}, \ldots, a_m^{PER}\}$ is the action space. $a_\phi$ corresponds to no action (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 (environment query) respectively. Also let the set of AIGA be $A^{AIGA}$, and the queries from the perception module be $A^{PER}$, s.t. $A = A^{AIGA} \cup A^{PER} \cup a_\phi$.
- $O = \{O_\phi, O_1, \ldots, O_P\}$ 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 $O_\phi$ corresponds to no response, and $O_1, \ldots, O_P$ 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 $k$[5].
- $\mathbb{O} : S \times A \to O$, 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.

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

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

Here, actions $a \in a_\phi \cup A^{AIGA}$ are *controllable* in the sense that they can be deployed through a policy in order to monitor or influence the human's attentiveness level $l$. The actions from the perception module $a \in A^{PER}$ are triggered whenever 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 (which formalizes assumption 3):

*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 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 $t \in T$.

*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 $s[k+1] \in S_0$. 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 $o[k+1] = o_\phi$, 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 $s[k+1] \in S_{-i-1}$, $\forall i \in \{D-1, \ldots, 0\}$.

*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 introduce structural constraints and sparsity on the state transition function $\mathbb{T}$ to capture the relevant behaviors for time-sensitive HRI.

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

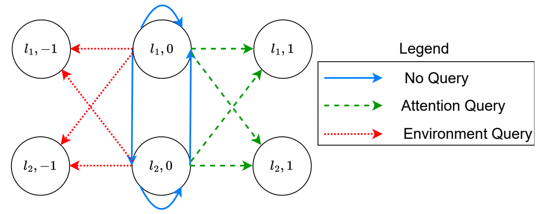

Fig. 4: The state space for the POMDP formalizing the perception hand-off experiment studied in section IV. Shown here are the possible 1-step state transitions when starting in states $l_1, 0$ or $l_2, 0$ and under the different possible actions. Also see example 1.

### B. Connecting the model to the dual-task experiment

- *Human responses when inattentive:* When the human subject is distracted, they are slower to answer queries, and also get them wrong more often (section IV). 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 5 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 $s = \{l_1, .\}$. Additionally, the probability of getting no response at the first time step in the query (corresponding to the time bin (or interval) $[0s, 1.33s]$ since query was asked) is much higher in the low attention level state.
- *Impact of AIGA/attention queries:* Section IV shows the positive impact of well timed attention queries on the human performance. 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. $\mathbb{T}(s = \{l_2, t+1\}|s = \{l_1, t\}, a \neq a_\phi, .) > \mathbb{T}(s = \{l_1, t+1\}|s = \{l_1, t\}, a \neq a_\phi, .)$, e.g., in figure 4, this implies that the probability of transitioning from $\{l_1, 0\}$ to $\{l_2, 1\}$ is higher than that of transitioning to $\{l_1, 1\}$.

To learn the POMDP 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 [17] (section V-D).

### C. Belief updates

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

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

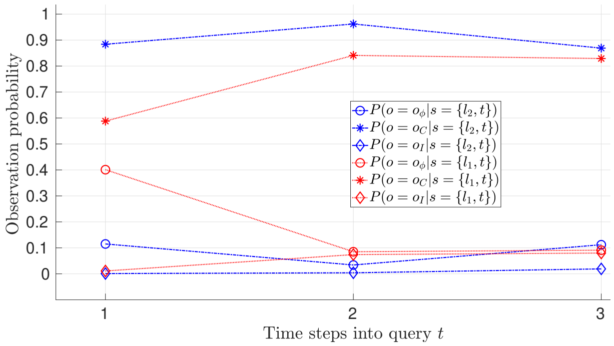


Fig. 5: Learned observation probabilities for all time steps in a query with $D = 3$. Since the attention and environment queries are displayed in an identical manner, we assume that the observation probabilities are the same for getting responses for both types of queries. These show that the human is more likely to answer queries correctly ($o = o_C$) and earlier in the high attention level state $s = \{l_2, .\}$ than in the low attention level state $s = \{l_1, .\}$.

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, $\sum_{s \in S} b_s[k] = 1 \, \forall k$.

*D. Learning the model from experimental data*

In order to learn a model similar to the one proposed above in example 1 from the dual-task experiment data, we use the Baum-Welch algorithm [17], 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 time series of observations and actions collected via the dual-task experiment. Note, unlike in a standard POMDP where 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 (definition 3).

*1) 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-D), 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:

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

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

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

This is obtained by applying the chain rule of probabilities (i.e., $P(A, B|C) = P(A|B, C)P(B|C)$) on the conditional probability $\hat{\mathbb{T}}([s', o'] \,|\, [s, o], a)$ and then using the definition of $\mathbb{T}$ and $\mathbb{O}$ (definition 3). The state transition matrix for this lifted state (given $a$, $o'$) corresponds to the Kronecker product of the 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 [17]. Following this, $\mathbb{T}$ and $\mathbb{O}$ for the model of Definition 3 can be recovered using a structured Kronecker product recovery method [18].

*E. 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 1). 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 $k$, $P(a[k] = a^{PER}) = p$.

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), as explained next.

*1) Marginalizing the actions from the perception module:* Since actions in $A^{PER}$ are not controllable, we must account for their impact before we can learn the POMDP. For simplicity, we explain this process through the POMDP for the perception hand-off as outlined in example 1. 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 an AIGA policy.

*2) Value iteration-based learned policy:* Given that solutions for POMDP policy learning are not exact and often difficult to interpret, we used the Value Iteration algorithm [19] to derive an interpretable 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 reward:

$$R = \sum_{k=0}^{\infty} \gamma^k r[k], \text{ where, } \gamma \in (0, 1] \text{ and,} \tag{3}$$

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

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 $\leq 1$ 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 reward function penalizes AIGA actions ($C_1 \geq 0$) to avoid asking too many attention queries, but 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. For the simulations in the next section, 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 reward function (3).

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

We first show the ability of the proposed model to represent the perception hand-off via experimental data gathered from 39 human subjects performing a trial of 10 minutes each (section IV). Next, we show that the model is suited for influencing the human's attentiveness levels.

**Implementation details:** The Baum-Welch algorithm of [17] for learning the model was implemented in 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 the OpenAI gym environment.

### A. Learning a model from data

We use a subset of the collected data to learn a model (see section V-D) with different sampling times $dt$ and associated number of time steps in a query before it expires, $D$. Figure 5 shows the learned observation probabilities for each time step in the query for $D = 3$.

Next, we evaluate the likelihood of the action-observation sequence (see Section V-D) 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 V-B) 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 $\geq 2$) since we compute them over sequences of hundreds of time steps.

### B. AIGA policy for perception hand-off

From the learned model (we pick the setting of $D = 3$), we learn a policy (section V-E.2) to deploy the AIGA (attention queries) to maximize a reward function (3) 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.

**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 (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 (by $\epsilon = 0.1$):

$$a[k] = \begin{cases} a^{AIGA}, & \text{if } \sum_{s'=\{l_1,.\}} b_{s'}[k] - \sum_{s'=(l_2,.)} b_{s'}[k] \geq 0.5 \\ a_{\phi}, & \text{otherwise.} \end{cases}$$

- *Random:* The policy randomly deploys attention queries, $P(a[k] = a^{AIGA}) = 0.5$.
- *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.

*1) Simulation results and summary:* In addition to modeling 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. Table II presents the averages and standard deviations over 100 runs (of 100 time steps), with random initial state $s[0] \in S_0$, of the following quantities, evaluated for the learnt and baseline policies interacting with the learnt POMDP model simulating the human driver:

**Accumulated reward:** The learned policy results in a higher accumulated reward than the baseline policies, demonstrating how AIGA in a systematic manner can show improvements over not using the AIGA or deploying it *randomly*.

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, $T_{\text{resp}}$ 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) | $T_{\text{resp}}$ | $f * 100$ | $\#a^{AIGA} : \#a^{PER}$ |
|---|---|---|---|---|
| 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 |

**Percentage of environment queries correctly responded to:** The learned policy also results in the highest percentage $f * 100$ of environment queries with correct responses, showing an improvement of $5.4\%$ over the no AIGA policy.
**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 asks the least number of attention queries per environment query. This is due to the reward function that penalizes asking attention queries, which would in practice be to avoid causing fatigue to the driver by querying them too frequently.

These simulation results show the potential for improving human responses to perception hand-off queries, demonstrating the benefit of using model-based AIGA.

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.

## VII. Discussion

**Summary:** We present a model-based formalization of the perception hand-off, or the problem of bringing the human in the decision making loop when an autonomous system is uncertain about the perceived 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.

**Limitations and future work:** Due to COVID-19, the human subject experiment 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.

Ongoing work focuses on validating the results on the impact of the policy via another human subject web experiment with the learned (and baseline) policies operating in the loop. This experiment will allow for the AIG actions to be deployed on a finer time scale than the one in this paper.

## References


[1] 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.

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

[3] "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.

[4] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan, "Information gathering actions over human internal state," in *Proceedings of the IEEE, /RSJ, International Conference on Intelligent Robots and Systems (IROS)*. IEEE, October 2016, pp. 66–73.

[5] G. Aschersleben and J. 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] V. A. Shia, Y. Gao, R. Vasudevan, K. D. Campbell, T. Lin, F. Borrelli, and R. Bajcsy, "Semiautonomous vehicular control using driver modeling," *IEEE Transactions on Intelligent Transportation Systems*, 2014.

[7] N. Gopalan and S. Tellex, "Modeling and solving human-robot collaborative tasks using pomdps," in *RSS Workshop on Model Learning for Human-Robot Communication*, vol. 32, no. 4, 2015, pp. 590–628.

[8] W. Zheng, B. Wu, and H. Lin, "Pomdp model learning for human robot collaboration," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 1156–1161.

[9] R. Calinescu, N. Alasmari, and M. Gleirscher, "Maintaining driver attentiveness in shared-control autonomous driving," *arXiv preprint arXiv:2102.03298*, 2021.

[10] S. E. Arkonac, D. P. Brumby, T. Smith, and H. V. R. 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*, ser. AutomotiveUI '19. ACM, 2019.

[11] F. Naujoks, D. Befelein, K. Wiedemann, and A. Neukum, "A review of non-driving-related tasks used in studies on automated driving," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2017, pp. 525–537.

[12] H.-I. Lee, S. Park, J. Lim, S. H. Chang, J.-H. Ji, S. Lee, J. Lee *et al.*, "Influence of driver's career and secondary cognitive task on visual search behavior in driving: a dual-task paradigm," *Advances in Physical Education*, vol. 5, no. 04, p. 245, 2015.

[13] L. Scatturin, R. Erbach, and M. 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*, ser. AutomotiveUI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 215–220.

[14] M. Scharfe and N. 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*, 2019, pp. 2734–2740.

[15] D. D. Salvucci, M. Zuber, E. Beregovaia, and D. Markley, "Distract-r: Rapid prototyping and evaluation of in-vehicle interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '05. New York, NY, USA: Association for Computing Machinery, 2005, p. 581–589. [Online]. Available: https://doi.org/10.1145/1054972.1055052

[16] Y. V. Pant, B. T. Kumaravel, A. Shah, E. Kraemer, M. Vazquez-Chanlatte, K. Kulkarni, B. Hartmann, and S. A. Seshia, "Tech report: Modeling and influencing human attentiveness in autonomy-to-human perception hand-offs," https://yashpant.github.io/assets/pdf/root_techrep2.pdf, 2022, accessed: 03-12-2021.

[17] S. Koenig and R. G. Simmons, "Unsupervised learning of probabilistic models for robot navigation," in *Proceedings of IEEE International Conference on Robotics and Automation*, 1996.

[18] S. Rambhatla, X. Li, and J. Haupt, "Provable online cp/parafac decomposition of a structured tensor via dictionary learning," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[19] R. Bellman, "A markovian decision process," *Indiana Univ. Math. J.*, vol. 6, pp. 679–684, 1957.