2-8-2019

Improving Human-Machine Collaboration Through Transparency-based Feedback – Part I: Human Trust and Workload Model

Kumar Akash

Katelyn Polson

Neera Jain

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-34 (2019) 315–321

Improving Human-Machine Collaboration Through Transparency-based Feedback – Part I: Human Trust and Workload Model ⋆

Kumar Akash * **Katelyn Polson** * **Tahira Reid** * **Neera Jain** *

* *School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: kakash@purdue.edu, polsonk@purdue.edu, tahira@purdue.edu, neerajain@purdue.edu).*

Abstract: In this paper, we establish a partially observable Markov decision process (POMDP) model framework that captures dynamic changes in human trust and workload for contexts that involve interactions between humans and intelligent decision-aid systems. We use a reconnaissance mission study to elicit a dynamic change in human trust and workload with respect to the system's reliability and user interface transparency as well as the presence or absence of danger. We use human subject data to estimate transition and observation probabilities of the POMDP model and analyze the trust-workload behavior of humans. Our results indicate that higher transparency is more likely to increase human trust when the existing trust is low but also is more likely to decrease trust when it is already high. Furthermore, we show that by using high transparency, the workload of the human is always likely to increase. In our companion paper, we use this estimated model to develop an optimal control policy that varies system transparency to affect human trust-workload behavior towards improving human-machine collaboration.

Keywords: trust in automation, human-machine interface, intelligent machines, Markov decision processes, stochastic modeling, parameter estimation, dynamic behavior

1. INTRODUCTION

Given the ubiquity of autonomous and intelligent systems, humans are increasingly interacting and collaborating with such systems in both complex situations (e.g., warfare and healthcare) and daily life (e.g., robotic vacuums). Published studies have shown that human trust in automation is an important factor that affects the outcome of the aforementioned interactions and that it can be improved by increasing the transparency of an intelligent system's decisions (Helldin, 2014; Mercado et al., 2016). Chen et al. (2014) defines transparency as "the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process." Therefore, greater transparency allows humans to make informed judgments and accordingly make better choices.

Nonetheless, high levels of trust are not always desirable and can lead to humans trusting an error-prone system. Instead, trust should be appropriately calibrated according to the system's capability (Lee and See, 2004). Moreover, high transparency involves communicating more information to the human and thus can increase the workload of the human (Lyu et al., 2017). In turn, high levels of workload can lead to fatigue, which can reduce the human's performance. Therefore, we aim to design intelligent systems that can respond to changes in human trust and workload in real-time to achieve optimal or near-optimal performance. For intelligent systems, a user interface (UI) is generally the means through which communication with the human is achieved. Therefore, the system must understand how the *transparency* of its communication through the UI affects the human's cognitive state.

Although researchers have developed various models of human trust behavior (Moe et al., 2008; Malik et al., 2009) and established the effect of transparency on trust (Helldin, 2014; Mercado et al., 2016; Wang et al., 2016a), there does not exist a quantitative model that captures the *dynamic* effect of transparency on human trust. Furthermore, published studies considering the effects of transparency on workload do not model its dynamics. Therefore, a fundamental gap remains in capturing the dynamic effect of machine transparency on human trust-workload behavior so that it can be used for improving human-machine collaboration.

In this paper, we present a partially observable Markov decision process (POMDP) model framework for capturing *dynamics of human trust and workload* for contexts that involve interaction between a human and an intelligent decision-aid system. We specifically consider a reconnaissance mission study adapted from the literature in which human subjects are aided by a virtual robotic assistant in completing a series of reconnaissance missions. We use the collected human subject data to train the POMDP

⋆ This material is based upon work supported by the National Science Foundation under Award No. 1548616. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

10.1016/j.ifacol.2019.01.028

model. We further study the effects of transparency and experience on human trust and workload using the estimated parameters. In a companion paper (Akash et al., 2018), the trained model is used to estimate human trust and workload and to develop a near-optimal control policy that varies machine transparency to improve outcomes of the human-machine collaboration.

This paper is organized as follows. Section 2 provides background on existing models of trust and workload particularly as they relate to how they are affected by transparency. The proposed POMDP framework for trust and workload is described in Section 3. Section 4 describes the reconnaissance mission study used to collect human subject data. The parameter estimation algorithm is presented in Section 3. Results and discussion are presented in Section 6, followed by concluding statements in Section 7.

2. BACKGROUND

At best we can only estimate or infer the cognitive state of a human through observations of the human. Hidden Markov models are popular for modeling human behavior (Li and Okamura, 2003; Pineau et al., 2003; Wang et al., 2009; Liu and Datta, 2012) because they provide a probabilistic framework for intent inference and incorporate uncertainty related to observations. However, a significant limitation of HMMs is that they do not include the effects of inputs or actions from intelligent systems that affect human behavior.

2.1 Markov Models for Human Trust

Several researchers have modeled human trust behavior using Markov models, particularly HMMs (Moe et al., 2008; Malik et al., 2009; ElSalamouny et al., 2009). Since human trust is not directly measurable, HMMs can instinctively be applied to infer the probability distribution of trust states. Nonetheless, in human-machine collaborations, the machine not only needs to infer a human's hidden mental state, but also needs to make decisions and take actions based on this inference. Furthermore, these actions would affect human trust behavior and should be incorporated in the model. A POMDP provides a framework that incorporates all of the modeling characteristics of HMMs and also accounts for the machine's actions. It also facilitates a framework for calculating the optimal series of actions for desired performance. POMDPs have been used in HMI contexts including automatically generating robot explanations to improve teaming performance (Wang et al., 2016b) and estimating trust in agent-agent interactions (Seymour and Peterson, 2009).

In prior work (Akash et al., 2017; Hu et al., 2018) we showed that past experience related to a machine's reliability affects human trust. Moreover, the type of error (i.e., miss or false alarm) has different effects on human trust dynamics. Misses and false alarms can only occur when the machine recommends absence of stimuli and presence of stimuli, respectively; therefore, the recommendation also has an effect on human trust behavior. However, machines cannot explicitly control the recommendation nor reliability as they depend on the environment and the true situation. Therefore, although we also model the effects of the machine's reliability and its affect on human trust-workload behavior dynamics, we only propose to use the machine's transparency of communication as a feedback control variable to improve human-machine collaboration.

2.2 Effects of Transparency on Trust and Workload

Several studies have been conducted to investigate the effect of transparency on trust. Early work conducted by Helldin (2014) suggests that increased system transparency increases trust in the system but also causes workload to increase. Mercado et al. (2016) conducted multiple studies based on Helldin's findings and confirmed that increased transparency yields higher trust but did not find that workload increases with transparency. Wang et al. have also conducted several experiments in this field (Wang et al., 2016a,b, 2015). Their studies showed that only the robot's ability to report correctly influenced trust; however, the studies also highlighted the limitation of the use of self-reported trust data.

Although higher transparency can increase human trust, it also can increase human workload. Lyu et al. (2017) showed that information volume significantly influenced driving speed and lane deviation, which indicates that 1) driving workload has an effect on driving performance and 2) high workload could cause driving performance impairment. Bohua et al. (2011) showed that cognitive difficulty increases as the amount of information increases, which shows that workload increases with more information. Therefore, we propose to model human workload along with human trust in the same framework.

2.3 Observing Human Trust and Workload

Trust and workload have previously been recorded using self-reported survey results in which questions customized to an experiment are on a Likert scale such that participants can report how much they trusted the system and understood the scenario. Workload is commonly assessed using the NASA TLX survey (Proctor and Van Zandt, 2008). However, it is not practical to collect human self-reported behavior for use with real-time feedback algorithms. Alternatively, trust can be inferred implicitly via compliance (Freedy et al., 2007; Wang et al., 2016b). Moreover, other studies have shown a correlation between workload and response time, which offers the ability for this metric to be measured implicitly as well (Helldin, 2014). Therefore, we propose to use compliance and response time as observations corresponding to trust and workload, respectively. It should be noted that other behavioral metrics like reliance, eye-tracking data, etc. can also be used but are outside the scope of this work.

3. MODELING TRUST-WORKLOAD BEHAVIOR

In this section, we describe a POMDP model of human trust and workload. We consider only contexts that involve human interaction with an intelligent decision-aid system that gives recommendations based on the presence or absence of a stimulus. Such autonomous systems only provide suggestions to the human; the final decision and/or action is taken by the human. These systems are prevalent both in safety-critical situations (e.g., assistive search

robots detecting dangers in warfare, health recommender systems detecting diseases in health-care) and in daily life (e.g., car blind-spot detectors in transport sector). Unfortunately the benefits of such systems can be lost due to a fundamental lack of human trust in the system or due to high workload.

While interacting with intelligent decision-aid systems, humans can either comply with the system's recommendation or reject it. Furthermore, there is a response time (RT) associated with the human's decision. We assume that these characteristics of human decision (compliance and response time) are dependent on human trust and workload. Moreover, we assume that human trust and workload are influenced by characteristics of the decision-aid system's recommendations. These characteristics include recommendation type (stimulus absent or present), transparency (amount of information), and past experience (faulty or reliable recommendations). We propose that increasing the transparency of the recommendation will help the human make a more informed decision, maintain trust, and thereby improve human-machine collaboration. However, this increased level of transparency requires the human to process more information, which can lead to increased workload. Therefore, there exists a trust-workload trade-off that needs to be optimized by maintaining appropriate transparency levels. With the additional assumption that the dynamics of human trust and workload follow the Markov property (Puterman, 2014), we consider the use of a POMDP for modeling the human trust-workload behavior.

A POMDP is a 7-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{O}, O, \mathcal{R}, \gamma)$ with a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, with transition probabilities $\mathcal{T}(s'|s,a)$ that govern the transition from state s to s' given the action a, and an additional set of observations $\mathcal{O}$, with observation probabilities $O(o|s)$ that govern the likelihood of observing o given the process is in state s. We will not consider the reward function $\mathcal{R}$ and discount factor γ as they are used for finding an associated optimal control policy, which will be considered in our companion paper (Akash et al., 2018). We define human trust-workload behavior as a process that we model using a POMDP.

We define the finite set of states $\mathcal{S} = [\mathit{Trust}, \mathit{Workload}]^T$ where both *trust* T and *workload* W can be either low ($\bullet_\downarrow$) or high ($\bullet_\uparrow$), i.e., Trust $\in \{T_\downarrow, T_\uparrow\}$ and Workload $\in \{W_\downarrow, W_\uparrow\}$. Human trust-workload behavior is influenced by characteristics of the system recommendations that define the finite set of actions $\mathcal{A} = [\mathit{Recommendation}, \mathit{Experience}, \mathit{Transparency}]^T$. Here, *recommendation* S_A can be either Stimulus Absent S_A^- or Stimulus Present S_A^+, *experience* E depends on the reliability of the last recommendation which can be either Faulty E^- or Reliable E^+, and *transparency* τ can be either Low Transparency τ_L, Medium Transparency τ_M, or High Transparency τ_H. The state transition probability function $\mathcal{T}(s'|s,a)$ can be represented as a $4 \times 4 \times 12$ matrix, such that $\mathcal{T}(i,j,k)$ represents the transition probability from the i^{th} state to j^{th} state given an action k. The human decision characteristics define the set of observations $\mathcal{O} = [\mathit{Compliance}, \mathit{Response\ Time}]^T$. Here, *compliance* C can be either Disagree C^- or Agree C^+ and *response time* RT can be segregated into three bins, namely, fast response time RT_F, medium response time RT_M, and slow response time RT_S. The observation probability function $O(o|s)$ can be represented as a 4×6 matrix, such that $O(i,j)$ represents the observation probability of the j^{th} observation given the state i.

We assume that trust and workload behavior are independent such that trust only affects compliance and workload only affects response time. Therefore, we identify these models independently. While it is possible for trust and workload to be coupled, a combined trust-workload model would require twice as many parameters to be trained, in turn requiring significantly more human subject data. Therefore, we proceed here with the independent model assumption, and investigation of a combined trust-workload model will be addressed in future work.

4. HUMAN SUBJECT STUDY

The focus of the experiment design, which is adapted from (Wang et al., 2015), was to capture how different levels of system transparency influence trust in autonomous systems as well as human-robot teaming performance.

Stimuli and Procedure:

A within-subjects study was performed in which participants were told they would interact with assistive robots to perform reconnaissance missions in three different locations. In each location, the participant searched 14 buildings and classified them as safe/unsafe based on the presence of either chemical or physical danger, with the goal of successfully searching all buildings as fast as possible. Prior to entering each building, the participant needed to decide if they would wear protective gear or not. They were informed that searching a building with protective gear would take approximately 15 seconds but would ensure that they would not be injured if some form of danger was present. Conversely, searching without gear would only take 5 seconds, but if danger was present, the participant would be injured and require 2 minutes to recover. In order to aid in their decision, a robotic companion surveyed each building first and provided a recommendation on whether or not protective gear was advised. Each robot was equipped with a camera to detect the presence of gunmen and a chemical sensor to detect chemicals.

In each mission, a different robot with a different transparency level provided the recommendation for each building. The low transparency robot reported if the building was safe/unsafe and thus if the gear was or was not advised. The medium transparency robot additionally included details regarding which type of danger had been detected, or if both sensors had not detected any danger. The high transparency robot included all of the information provided by the low and medium transparency robots in addition to a percent confidence in the report. These transparency levels are consistent with those proposed by Chen et al. (2014). Examples of the robot reports for danger related to the presence of a chemical are as follows:

- Low Transparency: "I have finished surveying the [building name]. I think the place is dangerous. Protective gear is needed."
- Medium Transparency: "I have finished surveying the [building name]. My sensors have detected traces of

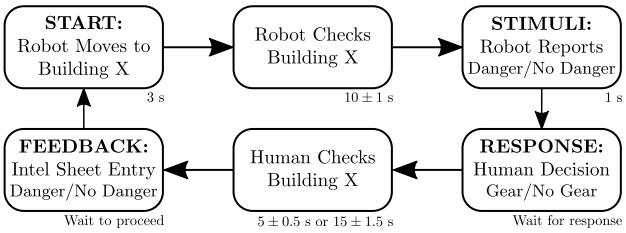


Fig. 1. The sequence of events in a single trial. The time length marked on the bottom right corner of each event indicates the time interval for which the information appeared on the computer screen.

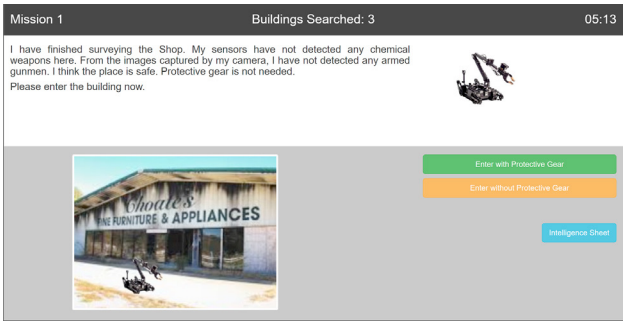


Fig. 2. Example screenshot of the interface of the experimental study.

dangerous chemicals. From the images captured by my camera, I have not detected any armed gunmen. Protective gear is needed."

- High Transparency: "I have finished surveying the [building name]. My sensors have detected traces of dangerous chemicals. From the images captured by my camera, I have not detected any armed gunmen. I am 95% confident in this assessment. I think it will be dangerous for you to enter the building without protective gear. The protective gear will slow you down a little. "

For each trial, the presence or absence of danger was equally probable. The robot reported the building status with 80% accuracy and was capable of giving both false alarms and misses. Furthermore, in high transparency, the robot reported its confidence about the assessment between 93% and 97% during reliable trials and between 80% and 90% during faulty trials. The sequence of events in each trial is shown in Fig. 1 along with a screenshot of the study interface in Fig. 2.

Participants: Seventy-nine participants (34 males, 45 females), ranging in age from 20-68 (mean 34.70 and standard deviation 9.57) from the United States participated in the study. They were recruited using Amazon Mechanical Turk (Amazon, 2005) and completed the study online. The compensation was $1.50 for their participation, and each participant electronically provided their consent. The Institutional Review Board at Purdue University approved the study.

5. MODEL PARAMETER ESTIMATION

Using the aggregated data of the 79 participants, we estimate the transition probability function, observation probability function, and the prior probabilities of states for the trust and workload models. We assume that the same models of trust and workload are representative of general human behavior. For this context, the recommendation that indicates *no danger* is defined as Stimulus Absent S_A^- and the recommendation that indicates *danger* is defined as Stimulus Present S_A^+. Furthermore, fast, medium, and slow response times are categorized for each participant based on the individual's first, second, and third tertiles of response time distribution, respectively. We consider the interaction between human and robot in each mission for each participant as a sequence of actions and observations.

We use an extended version of the Baum-Welch algorithm that is used to estimate the state transition and observation probability functions for a hidden Markov model (HMM) (see Rabiner and Juang (1986) for details). It is trivial to extend the Baum-Welch algorithm from learning hidden Markov models to learning POMDPs by taking into account the actions in every state during the estimation step (Cassandra et al., 1994); therefore, an explicit proof is not provided.

In order to prevent the Baum-Welch algorithm from overfitting the set of sequences, we split them randomly into two equal sets: a training set of sequences and a testing set of sequences. It is ensured that each of the three missions is uniformly distributed across these sets. The testing set is then used for cross-validation, stopping the Baum-Welch algorithm when the fit of generated POMDPs starts to decrease (or converge) on the testing sequence.

Finally, it should be noted that the quality of any data-based parameter estimation is only as good as the data itself. In the context of human subject data, no number of samples can fully represent the human population. In order to calculate the possible error in parameter estimation caused by the variation in sample selection, we iterated the estimation 10,000 times, with each iteration using a new randomly selected set of training and testing data. Errors caused by variation in the sample selection for a 95% confidence interval (CI) were less than 2% for all of the parameters. Thus, the parameter estimates are robust to variations in the sample selection.

6. RESULTS AND DISCUSSIONS

Here we present and analyze the resulting POMDP models of trust and workload.

6.1 Trust Model

The initial probability of states for Low Trust $T_\downarrow$ and High Trust $T_\uparrow$ are estimated as:

$$p_0(T_\downarrow) = 0.1288, \quad p_0(T_\uparrow) = 0.8712 \ . \tag{1}$$

This indicates that there is approximately an 87.12% probability that participants began the experiment with a state of high trust. This is consistent with the fact that given widespread use of automation, humans tend to trust a system even when they have no initial experience with it (Merritt and Ilgen, 2008).

The observation probability function $O_T(o|s)$ is represented in Fig. 3 and shows the probability of participants' compliance with the system's recommendations based on their trust level. Though there is more than 90% proba-

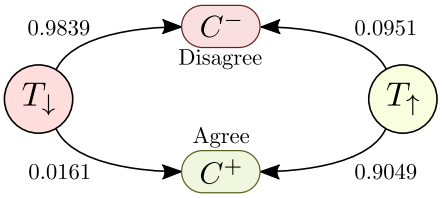


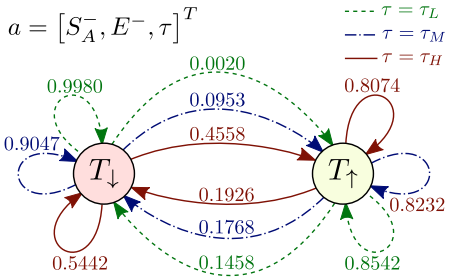
Fig. 3. Observation probability function $O_T(o|s)$ for the trust model. Probabilities of observation are shown beside the arrows.

bility that High Trust will result in a participant *agreeing* with the recommendation, there still exists an approximately 10% probability that the participant will disagree. Moreover, being in a state of Low Trust will result in the participant *disagreeing* with the recommendation with a probability of 98.3%, which is close, but not equal to, 100%.

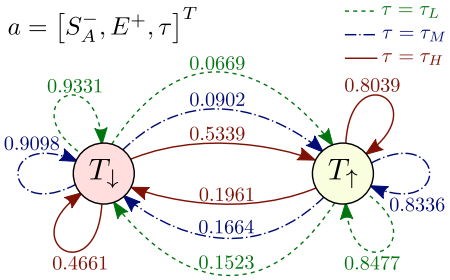
The transition probability function $\mathcal{T}_T(s'|s,a)$ is represented in Fig. 4 and shows the probability of participants transitioning from the state s to s' (where $s, s' \in \{T_\downarrow, T_\uparrow\}$) based on the action $a \in \mathcal{A}$. We first consider the case when the recommendation indicates no danger S_A^-. This is a high-risk situation in our context because incorrectly trusting the recommendation – in other words, complying with an erroneous recommendation *not to wear gear* – can lead to injury and a penalty of 2 minutes. We observe that in this case (see Fig. 4(a) and 4(b)), if the participant is in a state of Low Trust $T_\downarrow$, the probability of transitioning to a state of High Trust $T_\uparrow$ increases with an increase in transparency ($< 7\%$ for τ_L, $\approx 9\%$ for τ_M, and $> 45\%$ for τ_H). Therefore, increasing transparency when the participants' trust is low is more likely to increase their trust level in this high-risk situation. On the other hand, if the participant is in a state of High Trust $T_\uparrow$, the probability of transitioning to a state of Low Trust $T_\downarrow$ also increases with increasing transparency ($\approx 15\%$ for τ_L, $\approx 17\%$ for τ_M, and $\approx 19\%$ for τ_H). This is because the participant can make a more informed decision when the UI is more transparent and avoid errors that would result from trusting the recommendation when it is actually a poor recommendation.

Cases in which the recommendation indicates danger is present S_A^+ involve less risk for participants because if they choose to comply with the recommendation and wear protective gear despite the recommendation being incorrect, they are only delayed by 15 seconds. In this low-risk case (see Fig. 4(c) and 4(d)), we observe that the probability of transitioning to High Trust $T_\uparrow$ from any state of trust is typically higher for Low Transparency τ_L as compared to higher levels of transparencies. Therefore, in this low-risk case, the robot providing its recommendation with Low Transparency has the highest probability of increasing the trust level of the human. It is important to note that higher transparencies can lower the human's trust because higher transparencies can provide 'too much information', causing a participant to 'overthink' and subsequently leading to analysis-paralysis (Langley, 1995).

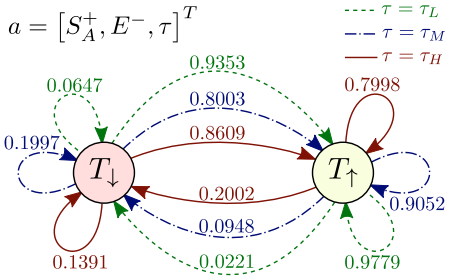
Finally, we observe that the transition probabilities to High Trust $T_\uparrow$ are typically higher in cases where the last experience was Reliable E^+ (see Fig. 4(b) and 4(d))



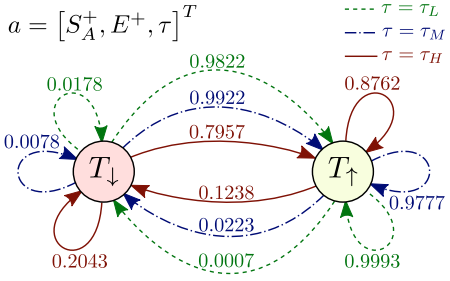
(a) Recommendation indicating no danger and Faulty last experience



(b) Recommendation indicating no danger and Reliable last experience



(c) Recommendation indicating danger and Faulty last experience



(d) Recommendation indicating danger and Reliable last experience

Fig. 4. Transition probability function $\mathcal{T}_T(s'|s,a)$ for Trust model. Probabilities of transition are shown beside the arrows.

as compared to when the last experience was Faulty E^- (Fig. 4(a) and 4(c)), as long as the recommendation and transparency remain the same. However, there are some exceptions to this observation. These include a lower transition probability from Low Trust $T_\downarrow$ to High Trust $T_\uparrow$ in the case of High Transparency τ_H and S_A^+ (see Fig. 4(c) and 4(d)), and for Medium Transparency τ_M and S_A^- (see Fig. 4(a) and 4(b)).

Therefore, higher transparencies do not always increase trust with the highest probability but instead can help

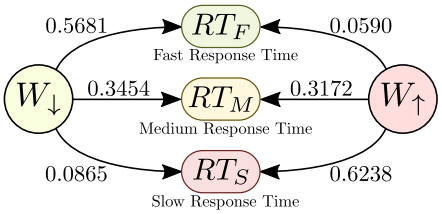


Fig. 5. Observation probability function $O_W(o|s)$ for workload model. Probabilities of observation are shown beside the arrows.

the human to make informed decisions, especially when the stakes are high. Moreover, these results suggest that the choice of transparency should not only depend on the human's current trust level, but should also consider the type of recommendation being provided by the system as well as the human's past experiences.

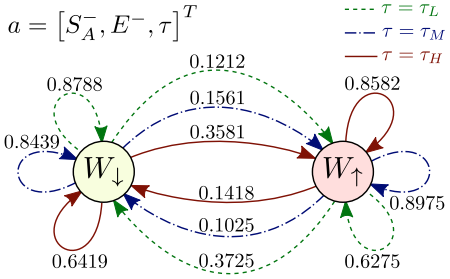
6.2 Workload Model

The initial probability of states for Low Workload $W_\downarrow$ and High Workload $W_\uparrow$ are estimated as:

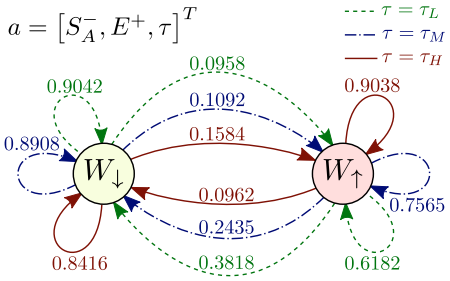
$$p_0(W_\downarrow) = 0.1970, \quad p_0(W_\uparrow) = 0.8030 \ . \tag{2}$$

This indicates that there is approximately an 80.30% probability that participants began the experiment with a state of High Workload. This is expected given that initially, participants needed to learn about the system which in turn increases their workload. The observation probability function $O_W(o|s)$ (see Fig. 5) shows the probability of participants' response times based on their workload level. In general, we observe that fast response times RT_F are more probable when the workload is low and slow response times RT_S are more probable when the workload is high. Medium response times RT_M are approximately equally probable from both states of workload.

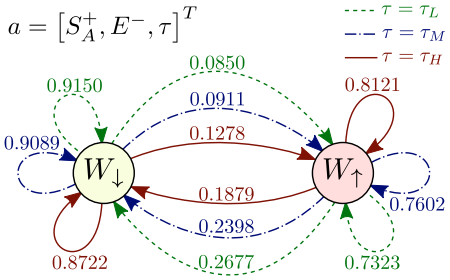
The transition probability function $\mathcal{T}_W(s'|s,a)$ is pictured in Fig. 6 and shows the probability of participants transitioning from the state s to s' based on the action $a \in \mathcal{A}$, where $s, s' \in \{W_\downarrow, W_\uparrow\}$. We observe that the probability of transitioning to a state of High Workload $W_\uparrow$ from any workload state increases with an increase in transparency for fixed recommendation and experience. Therefore, higher transparencies are more likely to increase participants' workload because participants have to process more information prior to decision-making. Moreover, in most cases, the probability of transitioning to a High Workload $W_\uparrow$ from any workload state is higher when the last experience was Faulty E^- (see Fig. 6(a) and 6(c)) as compared to when it was Reliable E^+ (see Fig. 6(b) and 6(d)) for a given recommendation and transparency. This conforms to the findings of Koehn et al. (2008) that error processing is associated with higher cognitive demands than processing UI feedback that denotes a correct response. In other words, individuals respond with faster response times in correct trials than in error trials. Also, in most cases, the probability of transitioning to High Workload $W_\uparrow$ from any workload state is higher when the recommendation indicates no danger S_A^- (see Fig. 6(a) and 6(b)) as compared to when it indicates danger S_A^+ (see Fig. 6(c) and 6(d)) for fixed experience and transparency. This is because a recommendation indicating no danger



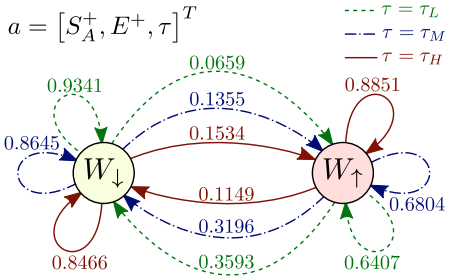
(a) Recommendation indicating no danger and Faulty last experience



(b) Recommendation indicating no danger and Reliable last experience



(c) Recommendation indicating danger and Faulty last experience



(d) Recommendation indicating danger and Reliable last experience

Fig. 6. Transition probability function $\mathcal{T}_W(s'|s,a)$ for Workload model. Probabilities of transition are shown beside the arrows.

S_A^- has a higher risk, as discussed in Section 6.1, and thus, humans consider their decision more carefully in order to avoid errors.

In summary, we have established models for human trust and workload for a decision-aid context. We observe that a higher transparency is not always more likely to increase trust in humans than lower transparencies but always is more likely to increase workload. Therefore, higher transparency is not always beneficial, and instead, system transparency should be updated based, in part, on the

state of human trust and workload. It is possible that each transparency level could be redesigned to improve the clarity of information presented and in turn, reducing their effect on workload. This, and improvements to the ecological validity of the results will be the subject of future work.

7. CONCLUSION

To attain improved human-machine collaboration, it is necessary for autonomous systems to estimate human trust and workload and respond accordingly. In turn, this requires dynamic models that capture these human states. In this paper, we used a reconnaissance mission study to elicit the dynamic change in human trust and workload with respect to the system's reliability and transparency as well as the presence or absence of danger. We established a partially observable Markov decision process (POMDP) model framework that captured dynamic changes in human trust and workload for contexts that involve the interaction of humans with an intelligent decision-aid system. We used the collected human subject data to estimate probabilities of the POMDP model and analyze the trust-workload behavior of humans. Our results indicate that higher transparency is more likely to increase human trust when the existing trust is low but is also more likely to decrease trust when it is already high. However, higher transparency always has a higher probability of increasing workload. In a companion paper, we use this estimated model to develop an optimal control policy that varies system transparency to affect human trust-workload behavior towards improving human-machine collaboration. In future work, we will examine interaction effects that may exist between trust and workload by estimating a combined model of these dynamics.

REFERENCES

Akash, K., Hu, W.L., Reid, T., and Jain, N. (2017). Dynamic modeling of trust in human-machine interactions. In *American Control Conference (ACC), 2017*, 1542–1548. IEEE.

Akash, K., Reid, T., and Jain, N. (2018). Improving Human-Machine Collaboration Through Transparency-based Feedback – Part II: Control Design and Synthesis. In *2nd IFAC Conference on Cyber-Physical & Human-Systems*. Miami, FL.

Amazon (2005). Amazon mechanical turk. [ONLINE] Available at: https://www.mturk.com/. [Accessed 20 February 2018].

Bohua, L., Lishan, S., and Jian, R. (2011). Driver's visual cognition behaviors of traffic signs based on eye movement parameters. *Journal of Transportation Systems Engineering and Information Technology*, 11(4), 22–27.

Cassandra, A.R., Kaelbling, L.P., and Littman, M.L. (1994). Acting optimally in partially observable stochastic domains. In *AAAI*, volume 94, 1023–1028.

Chen, J.Y., Procci, K., Boyce, M., Wright, J., Garcia, A., and Barnes, M. (2014). Situation awareness-based agent transparency. Technical report, Army Research Lab Aberdeen Proving Ground MD Human Research and Engineering Directorate.

ElSalamouny, E., Sassone, V., and Nielsen, M. (2009). HMM-based trust model. In *International Workshop on Formal Aspects in Security and Trust*, 21–35. Springer, Berlin, Heidelberg.

Freedy, A., DeVisser, E., Weltman, G., and Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In *2007 International Symposium on Collaborative Technologies and Systems*, 106–114.

Helldin, T. (2014). *Transparency for Future Semi-Automated Systems: Effects of transparency on operator performance, workload and trust*. Ph.D. thesis, Örebro Universitet.

Hu, W.L., Akash, K., Reid, T., and Jain, N. (2018). Computational modeling of the dynamics of human trust during human–machine interactions. *IEEE Transactions on Human-Machine Systems*. (In Press).

Koehn, J., Dickinson, J., and Goodman, D. (2008). Cognitive demands of error processing. *Psychological Reports*, 102(2), 532–538.

Langley, A. (1995). Between "paralysis by analysis" and "extinction by instinct". *Sloan Management Review*, 36(3), 63.

Lee, J.D. and See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.

Li, M. and Okamura, A.M. (2003). Recognition of operator motions for real-time assistance using virtual fixtures. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings. 11th Symposium on*, 125–131. IEEE.

Liu, X. and Datta, A. (2012). Modeling context aware dynamic trust using hidden Markov model. In *AAAI*, 1938–1944.

Lyu, N., Xie, L., Wu, C., Fu, Q., and Deng, C. (2017). Drivers cognitive workload and driving performance under traffic sign information exposure in complex environments: a case study of the highways in China. *International journal of environmental research and public health*, 14(2), 203.

Malik, Z., Akbar, I., and Bouguettaya, A. (2009). Web services reputation assessment using a hidden Markov model. In *Service-Oriented Computing*, 576–591. Springer, Berlin, Heidelberg.

Mercado, J.E., Rupp, M.A., Chen, J.Y.C., Barnes, M.J., Barber, D., and Procci, K. (2016). Intelligent agent transparency in human-agent teaming for multi-UxV management. *Human Factors*, 58(3), 401–415.

Merritt, S.M. and Ilgen, D.R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, 50(2), 194–210.

Moe, M.E.G., Tavakolifard, M., and Knapskog, S.J. (2008). Learning trust in dynamic multiagent environments using HMMs. In *Proceedings of the 13th Nordic Workshop on Secure IT Systems (NordSec 2008)*.

Pineau, J., Gordon, G., Thrun, S., et al. (2003). Point-based value iteration: An anytime algorithm for POMDPs. In *IJCAI*, volume 3, 1025–1032.

Proctor, R.W. and Van Zandt, T. (2008). *Human factors in simple and complex systems*. CRC Press.

Puterman, M.L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Rabiner, L. and Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1), 4–16.

Seymour, R. and Peterson, G.L. (2009). A trust-based multiagent system. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 3, 109–116. IEEE.

Wang, N., Pynadath, D.V., and Hill, S.G. (2016a). The impact of POMDP-generated explanations on trust and performance in human-robot teams. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 997–1005. International Foundation for Autonomous Agents and Multiagent Systems.

Wang, N., Pynadath, D.V., and Hill, S.G. (2016b). Trust calibration within a human-robot team: Comparing automatically generated explanations. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, 109–116. IEEE.

Wang, N., Pynadath, D.V., Unnikrishnan, K., Shankar, S., and Merchant, C. (2015). Intelligent agents for virtual simulation of human-robot interaction. volume 9179 of *Lecture Notes in Computer Science*. Springer International Publishing.

Wang, Z., Peer, A., and Buss, M. (2009). An HMM approach to realistic haptic human-robot interaction. In *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, 374–379. IEEE.