2-8-2019

Improving Human-Machine Collaboration Through Transparency-based Feedback – Part II: Control Design and Synthesis

Kumar Akash

Tahira Reid

Neera Jain

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-34 (2019) 322–328

Improving Human-Machine Collaboration Through Transparency-based Feedback – Part II: Control Design and Synthesis ⋆

Kumar Akash * **Tahira Reid** * **Neera Jain** *

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

Abstract: To attain improved human-machine collaboration, it is necessary for autonomous systems to infer human trust and workload and respond accordingly. In turn, autonomous systems require models that capture both human trust and workload dynamics. In a companion paper, we developed a trust-workload partially observable Markov decision process (POMDP) model framework that captured changes in human trust and workload for contexts that involve interaction between a human and an intelligent decision-aid system. In this paper, we define intuitive reward functions and show that these can be readily transformed for integration with the proposed POMDP model. We synthesize a near-optimal control policy using transparency as the feedback variable based on solutions for two cases: 1) increasing human trust and reducing workload, and 2) improving overall performance along with the aforementioned objectives for trust and workload. We implement these solutions in a reconnaissance mission study in which human subjects are aided by a virtual robotic assistant in completing a series of missions. We show that it is not always beneficial to aim to improve trust; instead, the control objective should be to optimize a context-specific performance objective when designing intelligent decision-aid systems that influence trust-workload behavior.

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

1. INTRODUCTION

With the increasing use of autonomous and intelligent systems, humans must interact and collaborate with these systems in both complex situations (e.g., warfare and health-care) and daily life (e.g., robotic vacuums). To maximize the benefits of these interactions, human trust in the system plays an important role (Lee and See, 2004; Sheridan and Parasuraman, 2005). More specifically, published studies have shown that human trust can be improved by increasing the transparency of intelligent systems' 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." Greater transparency allows humans to make informed judgments and accordingly make better decisions. Nonetheless, very high levels of human trust are not always desirable and can lead to humans trusting an error-prone system. Moreover, high transparency requires communicating more information to the human and can thus increase the workload level of the human (Lyu et al., 2017). In turn, high levels of workload can lead to fatigue, which can reduce the human's performance (Bohua et al., 2011). Therefore, we aim to design intelligent systems that can respond to changes in human trust and workload to achieve optimal performance.

Although researchers have developed various models of human trust (Moe et al., 2008; Malik et al., 2009; Akash et al., 2017; Hu et al., 2018) and workload (Wickens, 2008; Parasuraman, 2000), there does not exist a closed-loop framework for influencing human trust and workload to improve human-machine collaboration. Furthermore, published studies have shown that transparency affects both human trust (Helldin, 2014; Mercado et al., 2016) and workload (Lyu et al., 2017; Bohua et al., 2011) but has not been systematically used to control trust-workload behavior. Therefore, a fundamental gap remains in using machine transparency to dynamically improve human-machine collaboration.

In a companion paper titled *"Improving Human-Machine Collaboration Through Transparency-based Feedback – Part I: Human Trust and Workload Model"* (Akash et al., 2018), we developed a partially observable Markov decision process (POMDP) framework for estimating human trust and workload as it changes with machine transparency. The model captures changes in trust and workload for contexts that involve interaction between a human and an intelligent decision-aid system. In this paper, we establish

⋆ 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.

2405-8963 © 2019, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.01.026

a systematic method for shaping the reward function for the trust-workload POMDP model framework so as to close the loop between human and machine. We design and synthesize feedback control policies that vary machine transparency based on solutions for two cases: 1) reward functions designed to improve human trust and reduce workload, and 2) reward functions designed to improve a context-specific performance metric along with trust and workload. We implement these control policies in a reconnaissance mission study in which human subjects are aided by a virtual robotic assistant. Finally, we analyze the performance of these two control policies against an open-loop baseline.

This paper is organized as follows. Section 2 provides background on our POMDP model for trust-workload behavior. The proposed framework to obtain reward functions for the trust-workload model is described in Section 3. The algorithm used for determining the two sets of reward functions along with a near-optimal control policy are presented in Section 4. Section 5 describes the reconnaissance mission study used to test two feedback control policies. Results and discussion are presented in Section 6, followed by concluding statements in Section 7.

2. BACKGROUND

In the aforementioned companion paper (Akash et al., 2018), we established a POMDP model for trust-workload behavior of humans during interactions with an intelligent decision-aid system. The model consists of a finite set of states,

$$\mathcal{S} = [Trust,\, Workload]^T\,,$$

where both *trust* T and *workload* W can be either low ($\bullet_{\downarrow}$) or high ($\bullet_{\uparrow}$), that is, Trust $\in \{T_{\downarrow}, T_{\uparrow}\}$ and Workload $\in \{W_{\downarrow}, W_{\uparrow}\}$. We define a finite set of actions

$$\mathcal{A} = [Recommendation,\ Experience,\, Transparency]^T,$$

where *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. Finally, we define a finite set of observations

$$\mathcal{O} = [Compliance, Response\ Time]^T\,,$$

where *compliance* C can be either Disagree C^- or Agree C^+ and *response time* RT can be either fast response time RT_F, medium response time RT_M, or slow response time RT_S.

We collected human subject data using study adapted from the literature in which human subjects were aided by a virtual robotic assistant while completing a series of reconnaissance missions. Participants interacted with assistive robots to perform reconnaissance missions in three different locations. In each location, the participants searched 14 buildings and classified them as safe or unsafe based on the absence or presence of danger, respectively. Their goal was to successfully search all buildings as fast as possible. Prior to entering each building, the participants needed to decide if they would wear protective gear or not. They were informed that searching a building with the protective gear would take approximately 15 seconds but would ensure they would not be injured if danger was present. Conversely, searching without the gear would only take 5 seconds, but if danger was present, they would be injured and a 2-minute recovery time penalty would be applied. In order to assist the participant in their decision-making, the robotic companion surveyed each building first and provided a recommendation as to whether or not the protective gear was advised. In each mission, a different robot with a different transparency level provided the recommendation for each building. Data from 79 participants was collected and used to estimate the transition probability functions $\mathcal{T}(s'|s,a)$ and observation probability functions $O(o|s)$ for separate and independent POMDP models for trust and workload. We refer the reader to Part I of this paper (Akash et al., 2018) for the parameter values of the estimated probability functions.

Here in Part II, we consider reward functions $\mathcal{R}_T(s'|s,a)$ and $\mathcal{R}_W(s'|s,a)$ with respect to trust and workload, respectively, along with a discount factor γ to find the optimal control policy that varies machine transparency to improve the human-machine collaboration.

3. REWARDS FOR TRUST-WORKLOAD POMDP

We use a general definition of the reward function $\mathcal{R}(s'|s,a)$, defined as the reward received for transitioning from a state s to s' given an action a. A discounting factor γ is used to discount the future rewards so that immediate rewards are preferred. An optimal control policy using transparency as the feedback variable maximizes the expected total reward earned. In this section we define the rewards for the context of humans interacting with intelligent decision-aid systems.

State Rewards: During a human-machine interaction, we do not want the human to have low trust in the machine. Therefore, we assign a reward of $-\zeta$ (negative reward implies a penalty) for transitioning to the state of Low Trust $T_{\downarrow}$ from any existing state of trust given any action. Furthermore, we want the human to avoid high levels of workload; therefore we assign a reward of $-\eta$ for transitioning to the state of High Workload $W_{\uparrow}$. These rewards can be represented as

$$\mathcal{R}_T^S(s' = T_{\downarrow}|s,a) = -\zeta; \quad s \in \{T_{\downarrow}, T_{\uparrow}\} \text{ and } a \in \mathcal{A}, \quad (1)$$

$$\mathcal{R}_W^S(s' = W_{\uparrow}|s,a) = -\eta; \quad s \in \{W_{\downarrow}, W_{\uparrow}\} \text{ and } a \in \mathcal{A}. \quad (2)$$

Here, $\mathcal{R}_T^S$ and $\mathcal{R}_W^S$ are the state reward functions for the trust and workload models, respectively. The relative values of ζ and η determine the relative importance of trust and workload, respectively.

Performance Rewards: Apart from maintaining trust and workload in human-machine collaborations, it is important to achieve the goals that are specific to a given interaction or collaboration. Machines are never completely reliable and are instead prone to errors and failures. Thus, it is not always beneficial for the human to trust the machine. Instead, in the context of a human being helped by an intelligent decision-aid system, we want the human to make *correct* decisions; in other words, we want the human to comply with the system's recommendation when it is correct, and not comply when it is incorrect. In order to enforce this in our framework, we introduce a penalty when the human makes incorrect decisions.

Table 1. Reliability characteristics of the decision-aid system representing the probabilities of the assistive system's inference given the true situation.

		Decision-aid System's Inference	
		Stimulus Absent S_A^-	Stimulus Present S_A^+
True Situation	Stimulus Absent S^-	$1-\alpha$	α
	Stimulus Present S^+	β	$1-\beta$

Although we have defined the recommendation S_A of the decision-aid system based on its inference about the situation, we also need to distinguish what the human infers about the situation and what the true situation is. For example, in the context of the reconnaissance mission described earlier, it is possible for the recommendation of the decision-aid system to indicate the presence of danger but for the human to believe that the decision-aid system is unreliable. In this situation, the human may infer that there is no danger, when in fact danger is present. We denote the true (or actual) absence or presence of the stimulus as $S \in \{S^-, S^+\}$, the decision-aid system's inference or recommendation as $S_A \in \{S_A^-, S_A^+\}$, and the human's inference as $S_H \in \{S_H^-, S_H^+\}$. Here, $\bullet^-$ and $\bullet^+$ represent absence and presence of stimulus, respectively. We denote the probability of true presence of stimulus $\Pr(S^+) := d$ and, therefore, $\Pr(S^-) = 1-d$. Furthermore, an agent (decision-aid system or human) can make two types of errors, namely, *alpha*-errors or *beta*-errors.

Definition 1. An *alpha*-error is the error an agent makes by inferring a true absence of the stimulus S^- as the presence of the stimulus (S_A^+ or S_H^+).

Definition 2. A *beta*-error is the error an agent makes by inferring a true presence of the stimulus S^+ as the absence of the stimulus (S_A^- or S_H^-).

In practice, the reliability with which a decision-aid system makes correct predictions is a system characteristic and known *a priori*; therefore, we denote the probabilities of the decision-aid system making a beta-error or alpha-error as $\Pr(S_A^-|S^+) = \beta$ and $\Pr(S_A^+|S^-) = \alpha$ respectively. These reliability characteristics of the decision-aid system are summarized in Table 1.

In order to enforce penalties in our framework when the human makes incorrect decisions (i.e., makes an error), we assign a reward of $-\kappa$ when the human makes an *alpha*-error and a reward of $-\delta$ when the human makes a *beta*-error. The relative values of κ and δ determine the relative importance of alpha- and beta-errors made by the human, respectively. It should be noted that an error made by the human is dependent on whether or not the decision-aid system made an error. A human can make an error by either agreeing with the decision-aid system's erroneous recommendation or disagreeing with the decision-aid system's correct recommendation. These rewards (penalties) are summarized in Table 2.

Until this point, we have defined these performance rewards in terms of the observations of our POMDP framework (i.e. human compliance) and the true situation. In a POMDP framework, the reward function must be defined in the form $\mathcal{R}(s'|s,a)$. Therefore, we transform these performance rewards to derive the expected performance rewards for transitioning to state $s' \in \{T_\downarrow, T_\uparrow\}$ from any state $s \in \{T_\downarrow, T_\uparrow\}$ given action $a \in \mathcal{A}$, i.e., $\mathcal{R}_T^P(s'|a)$. Here, we only consider states of trust (and not workload) because compliance is only dependent on trust behavior in our independent models of human trust and workload. Since the human's decision is dependent on their trust level and the recommendation provided by the decision-aid system, we only consider the next state s' and the system's recommendation S_A for calculating the expected reward. Therefore,

Table 2. Performance rewards based on errors made by the human.

		Human's Inference	
		Stimulus Absent S_H^-	Stimulus Present S_H^+
True Situation	Stimulus Absent S^-	0	$-\kappa$
	Stimulus Present S^+	$-\delta$	0

$$\mathcal{R}_T^P(s'|s, a = [S_A, E, \tau]^T) = \mathcal{R}_T^P(s'|S_A) \ . \tag{3}$$

Proposition 1. Given a reward function $r : S \times S_H \to \mathbb{R}$ defined in terms of true absence or presence of stimulus $S \in \{S^-, S^+\}$ and the human's inference $S_H \in \{S_H^-, S_H^+\}$ as shown in Table 2, an equivalent standard reward function in the form $\mathcal{R}_T^P(s'|S_A)$ calculated as $\mathrm{E}[r|s', S_A]$ is given by

$$\begin{aligned}
\mathcal{R}_T^P(T_\downarrow|S_A^-) &= -O_T(C^-|T_\downarrow)(1-\lambda)\kappa - O_T(C^+|T_\downarrow)\lambda\delta \ , \\
\mathcal{R}_T^P(T_\uparrow|S_A^-) &= -O_T(C^-|T_\uparrow)(1-\lambda)\kappa - O_T(C^+|T_\uparrow)\lambda\delta \ , \\
\mathcal{R}_T^P(T_\downarrow|S_A^+) &= -O_T(C^-|T_\downarrow)(1-\mu)\delta - O_T(C^+|T_\downarrow)\mu\kappa \ , \\
\mathcal{R}_T^P(T_\uparrow|S_A^+) &= -O_T(C^-|T_\uparrow)(1-\mu)\delta - O_T(C^+|T_\uparrow)\mu\kappa \ ,
\end{aligned} \tag{4}$$

where $O_T(o|s)$ is the observation probability function, $\mathrm{E}[\bullet]$ is the expected value of $\bullet$, $\lambda := \Pr(S^+|S_A^-)$, and $\mu := \Pr(S^-|S_A^+)$.

We can calculate λ and μ from Table 1 using Bayes' theorem as

$$\lambda = \frac{\beta d}{\beta d + (1-\alpha)(1-d)}, \ \ \mu = \frac{\alpha(1-d)}{\alpha(1-d) + (1-\beta)d}. \tag{5}$$

Proof. We show the proof for the first reward function in (4), where $s' = T_\downarrow$ and $S_A = S_A^-$. The other three equations can be proved similarly. Using the law of total probability over a disjoint set C, we get

$$\begin{aligned}
\mathcal{R}_T^P(T_\downarrow|S_A^-) &= \mathrm{E}[r|T_\downarrow, S_A^-] \\
&= \sum_C \Pr(C|T_\downarrow, S_A^-)\mathrm{E}[r|C, T_\downarrow, S_A^-] \ .
\end{aligned} \tag{6}$$

Compliance $C \in \{C^-, C^+\}$ is only dependent on the trust state T and not on the decision-aid system's recommendation S_A. Similarly, the performance rewards we defined are only dependent on the human's decision S_H and the actual situation S (and not on the human trust state T). Therefore, (6) can be simplified as

$$\mathcal{R}_T^P(T_\downarrow|S_A^-) = \sum_C \Pr(C|T_\downarrow)\mathrm{E}[r|C, S_A^-] \ . \tag{7}$$

When $C = C^-$, the human disagreeing C^- to a recommendation of stimulus absent S_A^- is equivalent to the human inferring the situation as stimulus present S_H^+; therefore, we can write

$$\begin{aligned}\mathrm{E}[r|C^-, S_A^-] &= \mathrm{E}[r|S_H^+, S_A^-] \\ &= \sum_S \Pr(S|S_H^+, S_A^-)\mathrm{E}[r|S, S_H^+, S_A^-] \ .\end{aligned} \tag{8}$$

Since the true situation S is independent of the human's inference S_H, and the reward r is only dependent on the true situation S and the human's inference S_H,

$$\begin{aligned}\mathrm{E}[r|C^-, S_A^-] &= \sum_S \Pr(S|S_A^-)\mathrm{E}[r|S, S_H^+] \\ &= -(1-\lambda)\kappa \ .\end{aligned} \tag{9}$$

Similarly, we derive

$$\mathrm{E}[r|C^+, S_A^-] = -\lambda\delta \ . \tag{10}$$

Using (7),(9), and (10), we get

$\mathcal{R}_T^P(T_\downarrow|S_A^-) = -O_T(C^-|T_\downarrow)(1-\lambda)\kappa - O_T(C^+|T_\downarrow)\lambda\delta$. ■

Remark 1. This result can be extended to the case when the rewards for making correct decisions, i.e., $r(S^-, S_H^-)$ and $r(S^+, S_H^+)$ are non-zero.

Therefore, using (1) and (4), the cumulative reward for trust is $\mathcal{R}_T = \mathcal{R}_T^S + \mathcal{R}_T^P$ and using (2), the reward for workload is $\mathcal{R}_W = \mathcal{R}_W^S$. Finally, to select an appropriate discount factor γ we consider the number of trials per mission in our study, i.e., $N = 14$. We select the discount factor γ such that the reward of the 14^{th} trial has a weight of e^{-1}; such a value of γ can be approximated as

$$\gamma = \frac{N}{N+1} = 0.933 \tag{11}$$

With this reward function and discount factor, we calculate the solution for the POMDP model in the next section.

4. POMDP SOLUTION

To determine the optimal transparency for a given human-machine interaction, we solve the combined POMDP model for trust and workload. Although the exact optimal solution for a POMDP can be obtained using value iteration, the time complexity of solving POMDP value iteration is exponential in actions and observations. Considering that one may need to define even larger sets of actions and observations for a real scenario, using exact value iteration is not feasible. Therefore, we use a greedy approach called the Q-MDP method to obtain a near-optimal solution (Cassandra et al., 1994). This involves solving the underlying MDP to obtain the Q-function $Q_{\mathrm{MDP}} : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and choosing the action based on the current belief state as

$$a^* = \underset{a}{\operatorname{argmax}} \sum_{s \in \mathcal{S}} b(s) Q_{\mathrm{MDP}}(s, a) \ , \tag{12}$$

where the belief state $b(s)$ can be iteratively calculated as

$$b'(s') = \Pr(s'|o, a, b) = \frac{\Pr(o|s', a) \sum_{s \in \mathcal{S}} \Pr(s'|s, a) b(s)}{\sum_{s' \in \mathcal{S}} \Pr(o|s', a) \sum_{s \in \mathcal{S}} \Pr(s'|s, a) b(s)}. \tag{13}$$

Fundamentally, the Q-MDP method finds the optimal solution assuming that the POMDP were to become observable after the next action. The underlying MDP can be solved directly using value iterations to obtain the Q-function (Puterman, 2014). However, the solution obtained assumes that the decision-aid system can take any action $a \in \mathcal{A}$, while in this case, the system can only control the transparency because the recommendation and experience depend on the true situation and machine reliability. In other words, the transparency   is the controllable action and the recommendation S and experience E are uncontrollable actions; the latter are analogous to disturbance inputs in a typical control system. In order to accommodate uncontrollable actions during value iterations, we calculate the expected Q-function that is only dependent on the controllable action considering the probabilities of the uncontrollable actions. We calculate an intermediate Q-function of the form $Q^\tau : \mathcal{S} \times \quad \to \mathbb{R}$ and iterate (14) until convergence to obtain $Q_{\mathrm{MDP}}(s, a)$.

$$\begin{aligned}Q_{\mathrm{MDP}}(s, a) &= \sum_{s' \in \mathcal{S}} \mathcal{T}(s'|s, a)\left(\mathcal{R}(s'|s, a) + \gamma V(s')\right) \\ Q^\tau(s, \quad) &= \sum_{S_A, E} \Pr(S_A, E) Q_{\mathrm{MDP}}(s, a = [S_A, E, \tau]) \\ V(s) &= \max_\tau Q^\tau(s, \quad)\end{aligned} \tag{14}$$

Here, $S_A \in \{S_A^-, S_A^+\}$ and $E \in \{E^-, E^+\}$. Furthermore, the present recommendation S_A and experience E due to the reliability of the last recommendation are independent, that is, $\Pr(S_A, E) = \Pr(S_A)\Pr(E)$. $\Pr(S_A)$ and $\Pr(E)$ can be calculated as

$$\begin{aligned}\Pr(S_A^-) &= \ d \ + (1-\alpha)(1-d) \ , \\ \Pr(S_A^+) &= 1 - \Pr(S^-) \ , \\ \Pr(E^-) &= \alpha(1-d) + \ d \quad , \\ \Pr(E^+) &= 1 - \Pr(E^-) \ .\end{aligned} \tag{15}$$

For the human subject study described in this paper, $d = 0.5$, $\alpha = 0.2$, and $\quad = 0.2$. For implementation, once S and E are known in a trial, near-optimal transparency * can be determined as

$$^* = \underset{\tau}{\operatorname{argmax}} \sum_{s \in \mathcal{S}} b(s) Q_{\mathrm{MDP}}(s, a = [S_A, E, \tau]) \ . \tag{16}$$

We now obtain the solutions for two sets of reward functions.

4.1 Case 1: Considering State Rewards Only

We first consider only state rewards with equal importance given to trust and workload. Therefore, the parameters of the reward function are: $\zeta = 1$, $\eta = 1$, $\kappa = 0$, and $\delta = 0$. The solutions are represented in Fig. 1. We first consider the case when the recommendation indicates no danger (Stimulus Absent S_A^-) as shown in Fig. 1(a) and 1(b). This case represents a high risk situation in that it can cause a human to make beta-errors and can lead to injury, resulting in a penalty of 2 minutes. When Low Trust $T_\downarrow$ is more probable, high transparency should be used to increase trust. However, high transparency should be avoided at high workload as shown in Fig. 1(b). When the recommendation indicates danger S_A^+ (see Fig. 1(c) and 1(d)), low transparency is better at maintaining high trust and low workload as the associated risk is low.

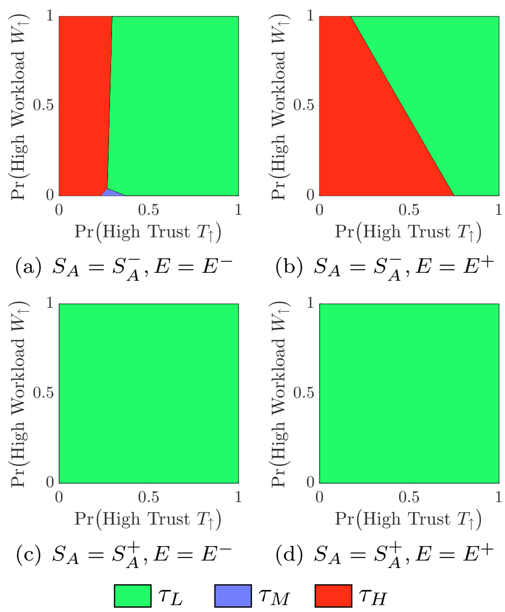

Fig. 1. Solution considering state rewards only with $\zeta = 1$, $\eta = 1$, $\kappa = 0$, and $\delta = 0$.

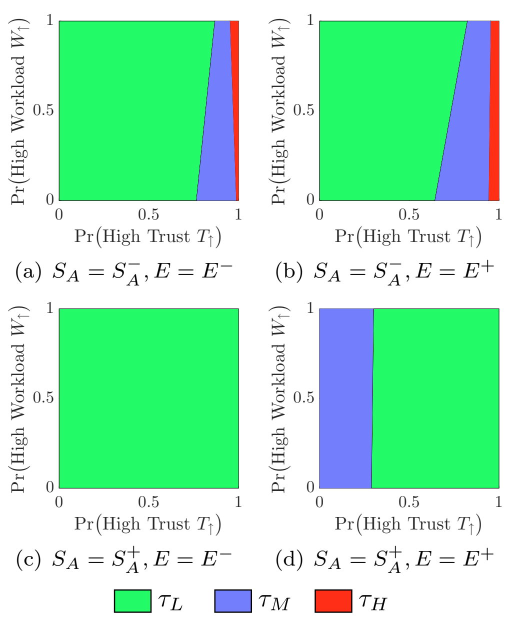

Fig. 2. Solution considering state and performance rewards with $\zeta = 1$, $\eta = 1$, $\kappa = 25$, and $\delta = 250$.

4.2 Case 2: Considering State and Performance Rewards

We now consider performance rewards along with state rewards with higher weights given to performance. We also penalize the human more for making beta-errors than for making alpha-errors as those have greater consequences in the specific context considered in our human subject study. Therefore, the parameters for the reward function are: $\zeta = 1$, $\eta = 1$, $\kappa = 25$, and $\delta = 250$. The solutions are presented in Fig. 2.

We first consider the high risk case when the recommendation indicates no danger S_A^- (see Fig. 2(a) and 2(b)). The solution aims to avoid over-trust by providing higher levels of transparency when high trust is more probable. This allows the human to make a more informed choice and avoid the chances of a beta-error. When the recommendation indicates danger S_A^+ (see Fig. 2(c) and 2(d)), lower transparencies are more effective for maintaining high trust and low workload as the associated risk is low.

In the next section, these solutions are used to implement transparency-based feedback based on the participant's current trust and workload in a reconnaissance mission study.

5. HUMAN SUBJECT STUDY

The goal of the following human subject study is to experimentally validate the performance of the proposed control policy for transparency-based feedback in interactions between humans and an intelligent decision-aid system. The experiment described below is identical to that used in our companion paper but with transparency controlled using feedback between the machine and human based on the solution of the POMDP.

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 or unsafe based on the presence of danger. 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 the first mission, the robot reported to the human with a transparency level randomly chosen from the three levels. Each transparency level was chosen approximately an equal number of times. This acted as a baseline case against which the closed-loop interactions could be compared. In the subsequent two missions, the machine determined the transparency with which to communicate to the human based on each of the POMDP solutions described in Section 4.1 and 4.2, respectively. The choice of which POMDP solution (Case 1 or Case 2) to apply in mission 2 versus 3 was randomized to avoid ordering effects.

Participants: Eighty-one participants (36 males and 45 females) recruited using Amazon Mechanical Turk (Amazon, 2005), ranging in age from 24-71 (mean 39.68 and standard deviation 10.76) participated in the study. 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.

6. RESULTS AND DISCUSSIONS

Using the collected human subject data, we define five metrics to quantify and evaluate each participant's performance in both the baseline case as well as in the cases that included transparency-based feedback. We removed outlying values for each of the metrics determined by the interquartile range (IQR) rule (the 1.5×IQR rule) (Rousseeuw and Hubert, 2011). We use repeated measures one-way analysis of variance (ANOVA) tests to determine whether the use of feedback had any significant effect on these metrics. Post hoc analyses are conducted if

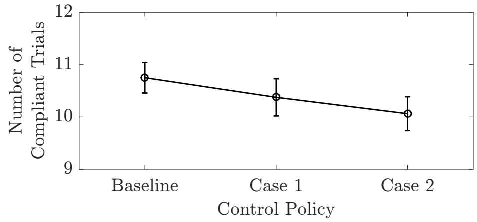

Fig. 3. Effect of feedback control policies on the number of compliant trials. Error bars represent standard error of the mean.

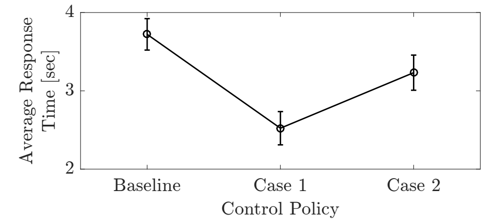

Fig. 4. Effect of various feedback control policies on the average response time of the participants after receiving robot's recommendation. Error bars represent standard error of the mean.

the significant ANOVA F test is obtained. Specifically, we conduct paired t-tests on all possible pairwise contrasts.

1. ***Number of compliant trials.*** The number of compliant trials is defined as the number of trials in which a participant agreed with the robot's recommendation. Agreeing with the robot is an indicator of the participant's trust level with more compliant trials implying higher trust. Figure 3 shows the effect of using a feedback control policy on the number of compliant trials. An ANOVA test showed that the use of feedback did not have a significant effect on the number of compliant trials, $F(2, 158) = 1.8080$, $p = 0.1673$. Nonetheless, we observe fewer compliant trials when the control policy based on the Case 2 rewards was used as compared to that of the Case 1 rewards and the baseline case. In other words, the control policy based on the Case 2 rewards decreased the trust of participants. This is expected as participants needed to distrust the robot in order to avoid injuries when the robot made an erroneous recommendation.

2. ***Average response time.*** The average response time is defined as the average time a participant took to respond to the robot's recommendation. Response time is an indicator of the participant's workload, with higher response time implying higher workload. Figure 4 shows the effect of the use of a feedback control policy on the average response time. An ANOVA test showed that the use of feedback had a significant effect on average response time, $F(2, 126) = 20.3223$, $p \approx 0.0000$. Specifically, the average response time with the use of control policy based on the Case 2 rewards was significantly lower as compared to that of the baseline case ($p = 0.0020$), but was significantly higher as compared to that of Case 1 rewards ($p = 0.0008$). This shows that both control policies were able to reduce the workload as compared to the baseline case. However, the control policy based on Case 2 rewards was worse at reducing the workload as compared to the Case 1 rewards. This is expected as participants saw a higher transparency user interface more often in Case 2 as recommended by the performance and state reward-dependent control policy.

3. ***Number of injured trials.*** The number of injured trials is defined as the number of trials in which a participant made beta-errors and received a penalty of 2 minutes. Figure 5 shows the effect of the use of a feedback control policy on the number of injured trials. An ANOVA test showed that the use of feedback had a significant effect on the number of injured trials, $F(2, 134) = 15.7611$, $p \approx 0.0000$. Specifically, the number of injured trials decreased significantly when the control policy based on the Case 1 rewards was used as compared to that of the baseline

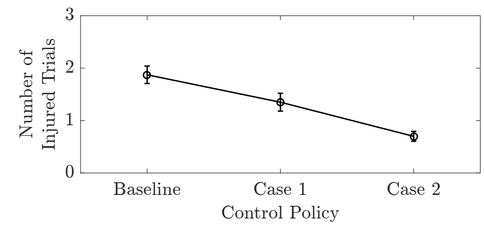

Fig. 5. Effect of various feedback control policies on the number of trials participants got injured due to beta-errors. Error bars represent standard error of the mean.

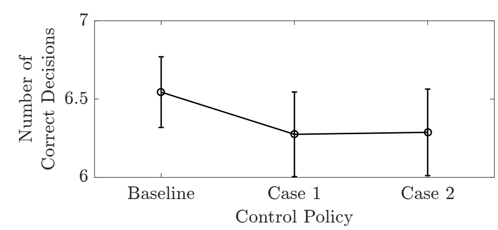

Fig. 6. Effect of various feedback control policies on the number of correct decisions made by participants. Error bars represent standard error of the mean.

($p = 0.0084$). Furthermore, the number of injured trials decreased significantly for Case 2 rewards as compared to Case 1 rewards ($p = 0.0017$). This shows that both the control policies were able to significantly reduce beta-errors made by the participants. Moreover, the control policy based on the Case 2 rewards was significantly better in reducing the beta-errors made by the participants. This is expected as the control policy based on the Case 2 rewards prioritized performance rewards with a high penalty for beta-errors.

4. ***Number of correct decisions.*** The number of correct decisions is defined as the number of correct decisions made by the participants, i.e., the trials in which they avoid both alpha- and beta-errors. Figure 6 shows the effect of the use of a feedback control policy on the number of correct decisions. An ANOVA test showed that the use of feedback did not have a significant effect on the number of correct decisions, $F(2, 156) = 0.32349$, $p = 0.7241$.

5. ***Total mission time.*** The total mission time is defined as the total time a participant took to complete the mission, which includes lost time due to alpha-errors (15 seconds) and beta-errors (2 minutes) made by the participants along with their response times. This is an overall indicator of the participants' performance in the mission as their objective was to complete the mission in the least possible time. Figure 7 shows the effect of the

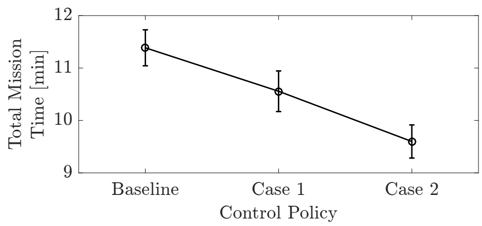

Fig. 7. Effect of various feedback control policies on the total mission time. Error bars represent standard error of the mean.

use of a feedback control policy on the total mission time. An ANOVA test showed that the use of feedback had a significant effect on the total mission time, $F(2, 128) = 9.2173$, $p = 0.0002$. Specifically, we observe a significantly lower total mission time as compared to the baseline case with the control policy based on Case 1 rewards ($p = 0.0438$) and Case 2 rewards ($p = 0.0002$). The decrease in the total mission time was more apparent using the control policy based on the Case 2 rewards.

Though published studies have shown that transparency *affects* both human trust (Helldin, 2014; Mercado et al., 2016) and workload (Lyu et al., 2017; Bohua et al., 2011), we have used it to systematically *control* the trust-workload behavior of humans. We observe that the control policy based on the Case 1 rewards, which focused on improving human trust and reducing workload, was not able to increase participants compliance but was better at reducing their response time. However, the control policy based on the Case 2 rewards was significantly better at reducing the beta-errors made by human and at improving the overall performance. Therefore, we conclude that when designing intelligent systems to affect human trust-workload behavior, overall improvement in the collaborative performance should be considered in addition to objectives related to increased trust and decreased workload.

7. CONCLUSION

To attain improved human-machine collaboration, it is necessary for autonomous systems to infer human trust and workload and respond accordingly. In turn, autonomous systems require models that capture both human trust and workload dynamics. In a companion paper, we developed a trust-workload POMDP model framework that captured changes in human trust and workload for contexts that involve interaction between a human and an intelligent decision-aid system. In this paper, we defined intuitive reward functions and showed that these could be readily transformed for integration with the proposed POMDP model. We synthesized a near-optimal control policy using transparency as the feedback variable based on solutions for two cases: 1) increasing human trust and reducing workload, 2) improving overall performance along with the aforementioned objectives for trust and workload. We implemented these solutions in a reconnaissance mission study in which human subjects were aided by a virtual robotic assistant. We found that it is not always beneficial to increase trust; instead, the control objective should be to optimize a context-specific performance objective when designing intelligent decision-aid systems that influence trust-workload behavior. Future work will include considering coupled trust-workload dynamics and validating it in other contexts.

ACKNOWLEDGMENT

The authors sincerely acknowledge the contribution of Katelyn Polson in the experiment design and collection of human subject data.

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., Polson, K., Reid, T., and Jain, N. (2018). Improving Human-Machine Collaboration Through Transparency-based Feedback – Part I: Human Trust and Workload Model. 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.

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).

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.

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. 58(3), 401–415.

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)*.

Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, 43(7), 931–951.

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

Rousseeuw, P.J. and Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 73–79.

Sheridan, T.B. and Parasuraman, R. (2005). Human–automation interaction. *Reviews of human factors and ergonomics*, 1(1), 89–129.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human factors*, 50(3), 449–455.