Optimal Fidelity Selection for Human-in-the-loop Queues using Semi-Markov Decision Processes

Piyush Gupta Vaibhav Srivastava

***Abstract*— We study optimal fidelity selection for a human operator servicing a queue of homogeneous tasks. The service time distribution of the human operator depends on her cognitive dynamics and the level of fidelity selected for servicing the task. Cognitive dynamics of the operator evolve as a Markov chain in which the cognitive state increases (decreases) with high probability whenever she is busy (resting). The tasks arrive according to a Poisson process and each task waiting in the queue loses its value at a fixed rate. We address the trade-off between high quality service of a task and consequent loss in value of future tasks using a Semi-Markov Decision Process (SMDP) framework. We numerically determine an optimal policy and establish its structural properties.**

I. Introduction

Advances in technology and automation has led towards an era of intelligent machines capable of performing complex tasks under extreme environments. However, to ensure safe and efficient operation for many safety-critical systems, human operators are often kept in the loop with autonomous robots due to their perception and intelligent decision making skills. Human-in-the-loop systems are pervasive in areas such as search and rescue [1], [2], semi-autonomous vehicle systems [3], and robot-assisted surgery [4]. Human-robot collaboration allows for the integration of human knowledge and perception skills with autonomy. In such systems, it is often of interest to maximize the ratio of the number of robots to the number of human operators, which leads to increased workload for human operators. Thus, to facilitate the effective use of cognitive resources of human operators, the optimal fidelity selection is critically important.

We study optimal fidelity selection for a human operator servicing a stream of homogeneous tasks. We incorporate human cognitive dynamics into fidelity selection problem and study its influence on an optimal policy. Our results can provide insight into efficient design of human decision support systems.

Recent years have seen significant efforts in integrating human knowledge and perception skills with autonomy for sophisticated tasks such as search and rescue; see [5] for an overview. A key research theme within this area concerns systematic allocation of human cognitive resources for efficient overall performance. Therein, some of the fundamental questions studied include optimal scheduling of the tasks to be processed by the operator [6]–[9], enabling shorter operator reaction times by controlling when to release a task for the operator to process [10], efficient work-shift design to counter fatigue or interruption effects [11], determining optimal operator attention allocation [12]–[14], and managing operator workload to enable better performance [15].

The optimal control of queueing systems [16]–[18] is a classical problem in queueing theory. Of particular interest are the works [19], [20], where authors study the optimal servicing policies for a M/G/1 queues by formulating an SMDP and describing its qualitative features. In this paper, we study optimal fidelity selection for a human operator servicing a queue of homogeneous tasks using an SMDP framework. In contrast to a standard control of queues problem, the server in our problem is a human operator, with her own cognitive dynamics that needs to be incorporated into the problem formulation.

We use a model for the time required to service a task that is inspired by experimental psychology literature and incorporates the influence of cognitive state as well as fidelity on the service time. For servicing each task, the human operator receives a reward based on the level of fidelity selected for the task. However, with higher fidelity, the cognitive state quickly rises to the high sub-optimal levels, thereby requiring a longer time to process the next task. Hence, there is a trade-off between the reward obtained by servicing a task with high fidelity, and the penalty incurred due to the resulting delay in processing other tasks. We elucidate on this trade-off and find an optimal policy for fidelity selection. The major contributions of this work are threefold: (i) we pose the fidelity selection problem in an SMDP framework and compute an optimal policy, (ii) we show the influence of cognitive dynamics on the optimal policy, and (iii) we establish structural properties of the optimal policy and show the existence of thresholds on queue lengths at which optimal policy switches different fidelity levels.

The rest of the paper is structured in the following way. Section II presents the problem setup and formulates the fidelity selection problem using an SMDP framework. In Section III, we numerically illustrate an optimal fidelity selection policy. In Section IV, we establish some of the structural properties of the optimal policy. Our conclusions and future directions are discussed in Section V.

II. Background and Problem Formulation

A. Problem Setup

We consider a human supervisory control system in which a human operator is tasked with servicing a stream of homogeneous tasks. The human operator may service these tasks with different level of fidelity. The time human spends on servicing a task depends on the level of fidelity with

This work has been supported by NSF Award IIS-1734272.

Piyush Gupta and Vaibhav Srivastava are with Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan, 48824, USA.

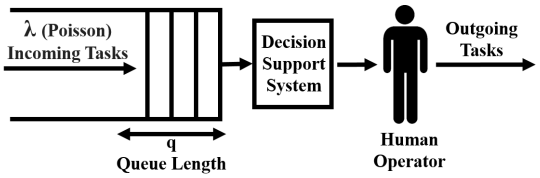


Fig. 1: Overall schematic of the problem setup. The incoming tasks arrive as a Poisson process with rate λ. The tasks are processed by the human operator based on the recommended fidelity level by the decision support system. Each task loses its value while waiting in the queue.

which she services the task as well as her cognitive state. We assume that the mean service time of the human operator increases with the level of fidelity to service the task, e.g., when human selects high fidelity, she may look into deeper details and consequently take longer time.

For a fixed level of fidelity, we model the service time as a unimodal function of human cognitive state. We treat cognitive state as a lumped parameter that captures various psychological factors such as workload, stress etc. The unimodal behavior of the mean service time with the human cognitive state is motivated by the Yerkes-Dodson law [21], [22]: excessive stress (high cognitive state) overwhelms the operator and too little stress (low cognitive state) leads to boredom and reduction in vigilance. Specifically, the mean service time is minimal corresponding to an intermediate optimal cognitive state.

We are interested in optimal fidelity selection policy for the human operator. To this end, we formulate a control of queue problem, where in contrast to a standard queue, the server is a human operator with her own cognitive dynamics. The incoming tasks arrive according to a Poisson process at a given rate $\lambda \in \mathbb{R}_{>0}$ and are processed by the human operator based on the fidelity level recommended by a decision support system (see Fig. 1). We consider a dynamic queue of homogeneous tasks with a maximum capacity $L \in \mathbb{N}$. Let each task waiting in the queue lose value at a constant rate $c \in \mathbb{R}_{>0}$ per unit delay in its processing. The set of possible actions available for the human operator corresponds to: (i) ***Waiting (W)***, when the queue is empty, (ii) ***Resting (R)***, which provides the resting time for the human operator to reach the optimal level of cognitive state, (iii) ***Skipping (S)***, which allows the operator to skip a task to reduce the queue length and thereby focus on newer tasks, (iv) ***Normal Fidelity (N)*** for processing the task with normal fidelity, and (v) ***High Fidelity (H)*** for processing the task more carefully with high precision.

Let $s \in \mathcal{S}$ be the state of the system and $\mathcal{A}_s$ be the set of admissible actions in state s, which we define formally in II-B. The human receives a reward $r : \mathcal{S} \times \mathcal{A}_s \to \mathbb{R}_{\geq 0}$ defined by

$$r(s,a) = \begin{cases} r_H, & \text{if } a = H, \\ r_N, & \text{if } a = N, \\ 0, & \text{if } a \in \{W,\ R,\ S\}, \end{cases} \tag{1}$$

where, $r_H, r_N \in \mathbb{R}_{\geq 0}$ and $r_H > r_N$. For such a dynamic queue setting, we intend to design a decision support system that assists the human operator by recommending optimal fidelity level to process each task. The recommendation is

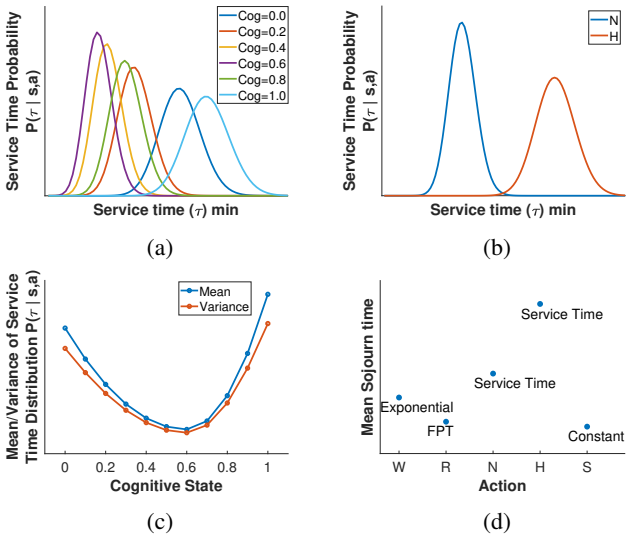


Fig. 2: Service time distribution of the human operator with (a) varying cognitive state and normal fidelity, (b) varying fidelity level and fixed cognitive state. (c) mean and variance of the service time are unimodal functions the cognitive state (d) The mean sojourn time distribution takes on different forms based on the selected action.

made based on the queue length and the cognitive state of the human operator which we assume to have real time access using, e.g., Electroencephalogram (EEG) measurements (see [23] for measures of cognitive load from EEG data).

B. Mathematical Modeling

We formulate the control of queue problem as an SMDP Γ defined by six components described below:

(i) A finite state space $\mathcal{S} := \{(q, \text{cog}) |\ q \in \{0, 1, ..., L\}, \text{cog} \in \mathcal{C} := \{(i-1)/N\}_{i \in \{1,...,N+1\}}\}$, for some $N \in \mathbb{N}$, where q is the queue length and cog represents the lumped cognitive state, which increases when the operator is busy, and decreases when the operator is idle.

(ii) A set of admissible actions $\mathcal{A}_s$ for each state $s \in \mathcal{S}$ which is given by: (i) $\mathcal{A}_s := \{W \mid s \in \mathcal{S},\ q = 0\}$ when queue is empty, (ii) $\mathcal{A}_s := \{\{R,\ S,\ N,\ H\ \} |\ s \in \mathcal{S},\ q \neq 0\}$ when queue is non-empty and $\text{cog} > \text{cog}^*$, where $\text{cog}^* \in \mathcal{C}$ is the optimal cognitive state, and (iii) $\mathcal{A}_s := \{\{S,\ N,\ H\ \} |\ s \in \mathcal{S},\ q \neq 0\}$ when queue is non-empty and $\text{cog} \leq \text{cog}^*$.

(iii) A state transition probability distribution $\mathbb{P}(s' |\ \tau, s, a)$ from state s to s' for each action $a \in A_s$ conditioned on sojourn time τ (time spent in each state before transitioning into next state). $\mathbb{P}(s' |\ \tau, s, a)$, which represents a transition from $s = (q,\ \text{cog}) \to s' = (q',\ \text{cog}')$ consists of two independent transition processes which are given by (i) a Poisson process for transition from $q \to q'$ (ii) human cognitive dynamics for a transition from $\text{cog} \to \text{cog}'$. We model the cognitive dynamics of the human operator as a Markov chain in which the probability of increase in cognitive state in small time $\delta t \in \mathbb{R}_{>0}$ is greater than the probability of decrease in cognitive state and increases with the level of fidelity selected for servicing the task. Similarly, while waiting or resting, the probability of decrease in cognitive

state in small time δt is higher than the probability of increase in cognitive state. Sample parameters of the model used in our numerical simulations are shown in Table I. This model of cognitive state dynamics is a stochastic equivalent of deterministic models of the utilization ratio considered in [10], [15]. It is considered that the cognitive state remains unchanged when the human operator chooses to skip the task.

TABLE I: Cognitive Dynamics modeled as Markov chain

Action	*Forward Probability*[a] ($\lambda_f \delta t$)	*Backward Probability*[b] ($\lambda_b \delta t$)	*Stay Probability*[c] (*1-*$\lambda_f \delta t - \lambda_b \delta t$)
W	$\lambda_f = 0.02$ (Noise)	$\lambda_b = 0.5$	$1 - 0.52\delta t$
R	$\lambda_f = 0.02$ (Noise)	$\lambda_b = 0.5$	$1 - 0.52\delta t$
N	$\lambda_f = 0.6$	$\lambda_b = 0.02$ (Noise)	$1 - 0.62\delta t$
H	$\lambda_f = 1.1$	$\lambda_b = 0.02$ (Noise)	$1 - 1.12\delta t$
S	$\lambda_f = 0$	$\lambda_b = 0$	1

[a]Forward Probability does not exist for $\mathrm{cog} = 1$ (reflective boundary)
[b]Backward Probability does not exist for $\mathrm{cog} = 0$ (reflective boundary)
[c]Stay Probability is $1 - \lambda_f \delta t$ for $\mathrm{cog} = 0$ and $1 - \lambda_b \delta t$ for $\mathrm{cog} = 1$

(iv) Sojourn time distribution $\mathbb{P}(\tau |\ s, a)$ of (discrete) time $\tau \in \mathbb{R}_{>0}$ spent in state s until the next action is chosen takes on different forms depending on the selected action (see Fig 2d). The sojourn time is the service time while processing the task (normal/ high fidelity), resting time while resting, constant time of skip while skipping, and time until the next task arrival while waiting in case of an empty queue. We model the rest time as the time required to reach from current cognitive state to optimal cognitive state cog^*. We model the service time distribution (see Fig 2a and 2b) while processing the task using a hypergeometric distribution, where the parameters of the distribution are chosen such that the mean service time of the human operator has the desired characteristics i.e. it increases with the fidelity level (see Fig 2d) and is a unimodal function of the cognitive state (see Fig 2c). While resting, sojourn time distribution is the first passage time (FPT) distribution for transitioning from current cognitive state cog to cog^*. We determine this distribution using matrix methods [24] applied on the Markov chain used to model the cognitive dynamics.

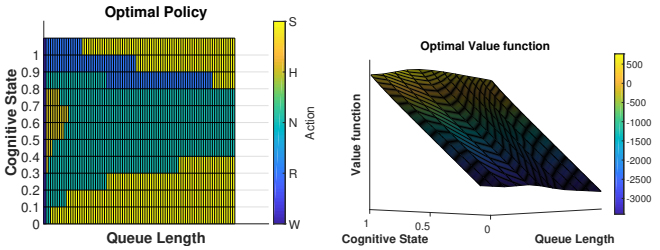
(v) For selecting action a at state s, the human receives a bounded immediate reward $r(s, a)$ defined in (1). We assume that each task waiting in the queue loses its value continuously. Hence, the human incurs a penalty at a constant cost rate $c \in \mathbb{R}_{>0}$ due to each task waiting in the queue, and consequently, the cumulative expected cost for choosing action a at state $s = (q, \mathrm{cog})$ is given by:

$$\sum_{\tau} \mathbb{P}(\tau|s,a) c\tau \Big(\mathbb{E}\Big[\frac{q+q'}{2} \Big|\ \tau, s, a \Big] \Big) = \sum_{\tau} \mathbb{P}(\tau|s,a) c\tau \Big(\frac{2q + \lambda\tau}{2} \Big).$$

The expected net immediate reward received by the operator for selecting action a in state s is given by:

$$R(s,a) = r(s,a) - \sum_{\tau} \mathbb{P}(\tau|s,a) c \Big(\frac{2q + \lambda\tau}{2} \Big) \tau. \quad (2)$$

(vi) A discount factor $\alpha \in [0, 1)$, which we choose as 0.98 for the purposes of numerical illustration.



(a) (b)

Fig. 3: (a) Optimal Policy π^* and the (b) Optimal Value Function V^* for SMDP Γ where time required to skip the tasks is not too small compared to the mean service time required to process the task

C. Solving SMDP for Optimal Policy

For SMDP Γ, the optimal value function $V^* : \mathcal{S} \to \mathbb{R}$ satisfies the following Bellman equation [25]

$$V^*(s) = \max_{a \in A_s} \Big[R(s,a) \ + \sum_{s',\tau} \gamma^{\tau} \mathbb{P}(s', \tau|s, a) V^*(s') \Big], \quad (3)$$

where $\mathbb{P}(s', \tau|s, a)$, which is the joint probability that a transition from state s to state s' occurs after time τ when action a is executed, can be rewritten as:

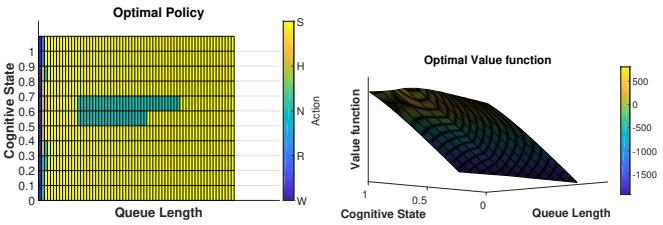
$$\mathbb{P}(s', \tau|s, a) = \mathbb{P}(s'|\tau, s, a) \mathbb{P}(\tau|s, a), \quad (4)$$

where $\mathbb{P}(s'|\tau, s, a)$ and $\mathbb{P}(\tau|s, a)$ are given by the state transition probability distribution and the sojourn time probability distribution, respectively. An optimal policy $\pi^* : \mathcal{S} \to \mathcal{A}_s$ at each state s selects an action that achieves the maximum in (3). We utilize the value iteration algorithm [26] to compute an optimal policy in our numerical illustrations.

III. Numerical Illustration of Optimal Fidelity Selection

We now numerically illustrate optimal policies for SMDP Γ. Figs. 3a and 3b show an optimal policy π^*, and the optimal value function V^*, respectively, for the case in which the time required for skipping the task is not too small compared to the mean service time required to process the task. For sufficiently high arrival rate λ such that there is always a task in the queue after the human finishes processing the current task (queue is never empty), we observe that for any given cognitive state cog, the optimal value function is monotonically decreasing with the queue length.

Additionally, we also observe that for a given queue length q, the optimal value function is a unimodal function of the cognitive state, with its maximum value corresponding to the optimal cognitive state ($\mathrm{cog}^* = 0.6$ for numerical illustrations). For the optimal policy π^*, we observe that the optimal policy selects the highest fidelity around the optimal cognitive state for low queue length and thereafter transitions to normal fidelity for higher queue lengths. We also observe that in low cognitive states, the optimal policy is to keep skipping the tasks until the queue length reduces to small number, and then processing the tasks. In higher cognitive states, we observe resting at small queue lengths followed by skipping of tasks at large queue lengths. Additionally, we observe the effect of cognitive state on the optimal policy. In particular, we observe that the optimal policy switches from



(a) (b)

Fig. 4: (a) Optimal Policy π^* and the (b) Optimal Value Function V^* when the time required to skip the task is too small compared to the mean service time required to process the task

H to *N*, *N* to *R*, and *R* to *S* at certain thresholds on the queue length and these thresholds appear to be a unimodal function of the cognitive state. This behavior can be attributed to the mean service time being a unimodal function of the cognitive state. We prove some of these properties of the optimal policy in the next section based on assumptions on the structure of the optimal value function.

Figs 4a and 4b show an optimal policy π^* and the optimal value function V^*, respectively, for the case in which the time required to skip the task is very small compared to the mean service time required to process the task. Here, we observe that in general, the optimal policy is to reduce the queue length by skipping all the tasks, and then processing the most recent tasks. Around the optimal cognitive state, we observe that we get a small region of skipping in between the region of high fidelity and normal fidelity. This can be attributed to the higher expected future rewards by skipping some tasks followed by receiving a future high reward for processing the recent tasks with high fidelity.

IV. Structural Properties of the Optimal Policy

In this section, we characterize structural properties of the optimal policy for SMDP Γ. We first introduce the following notation. Let $q_j^* : \mathcal{C} \to \mathbb{Z}_{\geq 0} \cup \{+\infty\}$, for $j \in \{1, 2, 3\}$ be some functions of the cognitive state. Let $\mu^1 : \mathcal{S} \times \mathcal{A}_s \to \mathbb{R}_{>0}$ and $\mu^2 : \mathcal{S} \times \mathcal{A}_s \to \mathbb{R}_{>0}$ be function defined by $\mu^1(s,a) = \mathbb{E}(\tau|s,a)$ and $\mu^2(s,a) = \mathbb{E}(\tau^2|s,a)$, where τ is the sojourn time. We study the properties of SMDP Γ in the limit $L \to +\infty$ and under following assumptions:

(A1) We assume that the optimal value function V^* satisfies: (i) $q \mapsto V^*(q, \cdot)$ is a monotonically decreasing function, and (ii) $\mathrm{cog} \mapsto V^*(\cdot, \mathrm{cog})$ is a unimodal function and admits the maximum at cog^* for any q.

(A2) The task arrival rate λ is sufficiently high so that the queue is never empty.

(A3) For any state $s = (q, \mathrm{cog})$ such that $\mathrm{cog} > \mathrm{cog}^*$:

$$\begin{aligned} &\mu^1(s,S) < \mu^1(s,R) < \mu^1(s,N) < \mu^1(s,H), \text{ and} \\ &\mu^2(s,S) < \mu^2(s,R) < \mu^2(s,N) < \mu^2(s,H). \end{aligned} \tag{5}$$

(A4) The jump in the optimal value function V^* over one extra task in the queue is upper bounded by α i.e., $V^*(\mathrm{cog}, q) - V^*(\mathrm{cog}, q+1) < \alpha$, where $\alpha \in [0, k)$, with $k = \min\{\mathbb{E}(\tau|\mathrm{cog}^*, H) - \mathbb{E}(\tau|\mathrm{cog}^*, N),\ \mathbb{E}(\tau|\mathrm{cog}^*, N) - \mathbb{E}(\tau|\mathrm{cog}^*, S)\}$.

Our numerical investigation suggests that assumptions (A1) and (A4) hold for SMDP Γ. We will seek to prove that these assumptions indeed hold in our future work. We make assumption (A2) for convenience. Indeed, if queue is allowed to be empty, then we will need to deal with an extra "waiting" action. Also, high arrival rate is the most interesting setting to study optimal fidelity selection. Assumption (A3) is true for a broad range of interesting parameters that define sojourn time distribution(s).

*Theorem 1 (**Structure of optimal policy**):* For SMDP Γ under assumptions (A1-A4) and an associated optimal policy π^*, the following statements hold

(i) there exists unique threshold functions $q_1^*(\mathrm{cog}), q_2^*(\mathrm{cog}), q_3^*(\mathrm{cog})$ such that for each $\mathrm{cog} > \mathrm{cog}^*$:

$$\pi^*(s = (q, \mathrm{cog})) = \begin{cases} H, & q \leq q_1^*(\mathrm{cog}), \\ N, & q_1^*(\mathrm{cog}) < q \leq q_2^*(\mathrm{cog}), \\ R, & q_2^*(\mathrm{cog}) < q \leq q_3^*(\mathrm{cog}), \\ S, & q > q_3^*(\mathrm{cog}); \end{cases} \tag{6}$$

(ii) there exists unique threshold functions $q_1^*(\mathrm{cog}), q_2^*(\mathrm{cog})$ such that for any $\mathrm{cog} \leq \mathrm{cog}^*$:

$$\pi^*(s = (q, cog)) = \begin{cases} H, & q \leq q_1^*(\mathrm{cog}), \\ N, & q_1^*(\mathrm{cog}) < q \leq q_2^*(\mathrm{cog}), \\ S, & q > q_2^*(\mathrm{cog}). \end{cases} \tag{7}$$

We prove Theorem 1 with the help of following lemmas.

Lemma 2: For SMDP Γ under assumption (A3), the immediate expected reward $R(s,a)$, for each action $a \in A_s$

(i) is linearly decreasing with queue length q for any fixed cognitive state cog;

(ii) is a unimodal function[1] of the cognitive state cog for any fixed queue length q with its maximum value achieved at the optimal cognitive state.

Proof: We start by establishing the first statement. The expected net immediate reward received by the human operator for selecting action a in state s (see Eq. (2)) can be rewritten as:

$$R(s,a) = -c\,\mathbb{E}(\tau|s,a)q + r(s,a) - \frac{c\lambda}{2}\,\mathbb{E}(\tau^2|s,a), \tag{8}$$

where $\mathbb{E}(\tau|\ s,a)$ and $\mathbb{E}(\tau^2|s,a)$ represents the first and the second moment of the sojourn time distribution, respectively. We note that the moments of the sojourn time distribution is independent of the queue length q.

$$R(s,a) = -a_1(\mathrm{cog}, a)q + a_2(\mathrm{cog}, a), \tag{9}$$

where $a_1(\mathrm{cog}, a) = c\,\mathbb{E}(\tau|s,a)$ and $a_2(\mathrm{cog}, a) = r(s,a) - \frac{c\lambda}{2}\,\mathbb{E}(\tau^2|s,a)$.

For a fixed cognitive state cog and action a, both a_1 and a_2 are constants and therefore, the expected immediate reward linearly decreases with the queue length q and the first statement follows.

The second statement follows by observing that, for a given queue length q, the mean and variance of the sojourn time for each action $a \in A_s$ are unimodal functions of the cognitive state with their peaks at cog^*. ■

[1]The expected immediate reward under action *S* is a constant, which we treat as a unimodal function.

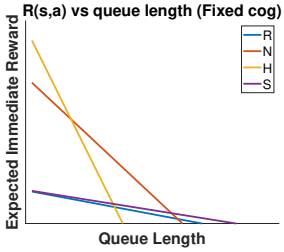


Fig. 5: Plot of Expected Immediate Reward R(s,a) for each action vs the queue length (fixed $\text{cog} > \text{cog}^*$)

Lemma 3: For the SMDP Γ under assumptions (A1-A3), and an associated optimal policy π^*, the following statements hold for each $\text{cog} > \text{cog}^*$:

(i) there exists a threshold function $q_1^*(\text{cog})$, such that the action *N* strictly dominates action *H*, for each $q > q_1^*(\text{cog})$;
(ii) there exists a threshold function $q_2^*(\text{cog})$, such that for each $q > q_2^*(\text{cog})$, the action *R* strictly dominates action *N*,
(iii) there exists a threshold $q_3^*(\text{cog})$, such that for each $q > q_3^*(\text{cog})$, action *S* is optimal.

Proof: We start with the first statement. We first analyze the expected immediate rewards. It follows from Lemma 2 that for a fixed cognitive state and for a given action the expected immediate reward (9) is a linear function. Since these linear functions have different slopes due to (5), it follows that the graphs of the expected immediate rewards for different actions intersect (see Fig. 5 for an illustration). If the lines associated with the expected immediate reward for action *H* and *N* intersect for $q \geq 0$, then let $\tilde{q_1} \in \mathbb{Z}_{\geq 0}$ be the unique point of intersection. It follows that for $q > \tilde{q_1}$, the expected immediate reward from *N* is strictly higher than that of *H*. If the lines associated with the expected immediate reward for action *H* and *N* intersect for $q < 0$, the expected immediate reward from *N* is strictly higher than that of *H* for any $q \geq 0$.

Let $F(s,a)$ denote the expected future rewards received in state s for taking action a (the second term in the Bellman equation (3)).

$$F(s,a) = \sum_{\text{cog}',q',\tau} \gamma^\tau \mathbb{P}(q'|\tau,q) V^*(\text{cog}',q') \times \mathbb{P}(\text{cog}'|\tau,\text{cog},a)\mathbb{P}(\tau|\text{cog},a) =: \sum Ax, \quad (10)$$

where $A = \gamma^\tau \mathbb{P}(q'|\tau,q)V^*(\text{cog}',q')$ & $x = \mathbb{P}(\text{cog}',\tau|\text{cog},a)$. Note that $F(s,a)$ in (10) represents a linear affine function over a probability simplex. Since V^* is strictly decreasing function of the queue length and is a unimodal function of the cognitive state, for each $\text{cog} > \text{cog}^*$, the expected future rewards received from *N* is strictly higher than H. This is due to the fact that action *H*, which has higher mean service time, leads to the higher queue length (q') with higher probabilities. Furthermore, it also leads to cognitive states farther from cog^* with higher probabilities. Therefore, it follows from assumption (A1) that $F(s,N) > F(s,H)$. Hence, for each $\text{cog} > \text{cog}^*$, there exists a unique threshold $q_1^* \leq \tilde{q_1}$, such that for $q > q_1^*$, the action *N* dominates the action *H*, for a given cognitive state cog. The proof of the second statement follows analogously to the first statement.

To establish the last statement, we note that the action *S* has the largest expected immediate reward, when queue length is sufficiently high, it is the action with minimum sojourn time and it is the only action that reduces the expected queue length. With these observations, the last statement follows analogously to the first statement. ■

Proof of Theorem 1: The proof of the first statement follows immediately from Lemma 3. We now focus on the second statement. Recall that action *R* is not an admissible action for $\text{cog} \leq \text{cog}^*$. Unlike the case with $\text{cog} > \text{cog}^*$, where high sojourn time leads to high sub-optimal cognitive state and high queue length; for $\text{cog} < \text{cog}^*$, high sojourn time might take the state towards optimal cognitive state and high queue length. Although the expected immediate rewards follows the trend similar to $\text{cog} > \text{cog}^*$, there is a trade-off between moving towards optimal cognitive state and reaching high queue length. In the following, using assumption (A4), we show that if action *N* is the optimal choice at queue length q_1^* for a given cognitive state, then for all $q > q_1^*$, *N* dominates *H*. We further show that if action *S* is the optimal choice at queue length q_2^* for a given cognitive state, then for all $q > q_2^*$, *S* dominates both *H* and *N*.

Let *N* be the optimal action in state $s = \{q_1^*, \text{cog}\}$. Then,

$$\begin{aligned} &R(s,N) - R(s,H) + F(s,N) - F(s,H) > 0, \\ &\quad \implies (\mathbb{E}(\tau|s,H) - \mathbb{E}(\tau|s,N))q_1^* + \\ &\quad (\mathbb{E}(\tau^2|s,H) - \mathbb{E}(\tau^2|s,N)) + \\ &\sum_\tau \sum_{\text{cog}'} \sum_{\overline{q}} \gamma^\tau \mathbb{P}(q_1^* + \overline{q}|q_1^*,\tau) V^*(\text{cog}', q_1^* + \overline{q} - 1) \times \\ &\quad (\mathbb{P}(\text{cog}',\tau|\text{cog},N) - \mathbb{P}(\text{cog}',\tau|\text{cog},H)) > 0, \quad (11) \end{aligned}$$

where $\overline{q}$ is the number of arrivals during service time τ.

Now for the state $s' = \{q_1^* + 1, \text{cog}\}$, under the assumption (A1) we show that:

$$R(s',N) - R(s',H) + F(s',N) - F(s',H) > 0. \quad (12)$$

To show (12), we prove that the difference between LHS of (12) and (11) is positive. The difference between the LHS of (12) and (11) is given by:

$$\begin{aligned} &(\mathbb{E}(\tau|s,H) - \mathbb{E}(\tau|s,N)) + \sum_\tau \sum_{\text{cog}'} \sum_{\overline{q}} \gamma^\tau \mathbb{P}(q_1^* + \overline{q}|q_1^* + 1,\tau) \times \\ &\quad (V^*(\text{cog}', q_1^* + \overline{q} - 1) - V^*(\text{cog}', q_1^* + \overline{q})) \times \\ &\quad (\mathbb{P}(\text{cog}',\tau|\text{cog},N) - \mathbb{P}(\text{cog}',\tau|\text{cog},H)). \quad (13) \end{aligned}$$

To find a lower bound on (13), we use assumption (A4) and replace $V^*(\text{cog}', q_1^* + \overline{q} - 1) - V^*(\text{cog}', q_1^* + \overline{q})$ by $\alpha \mathbf{1}(\mathbb{P}(\text{cog}',\tau|\text{cog},N) \leq \mathbb{P}(\text{cog}',\tau|\text{cog},H))$, where $\mathbf{1}(\cdot)$ is

the indicator function. Hence, (13) is lower bounded by

$$
\begin{aligned}
&(\mathbb{E}(\tau|s,H)-\mathbb{E}(\tau|s,N))+\\
&\qquad\sum_{\tau}\sum_{\mathrm{cog}'}\sum_{\overline{q}}\gamma^{\tau}\mathbb{P}(q_1^*+\overline{q}|q_1^*+1,\tau)\alpha\times\\
&\qquad(\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},N)-\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},H))\\
&\qquad\times\mathbf{1}(\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},N)\leq\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},H)),\\
&\qquad\geq(\mathbb{E}(\tau|s,H)-\mathbb{E}(\tau|s,N))+\\
&\alpha\sum_{\tau}\sum_{\mathrm{cog}'}(\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},N)-\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},H))\\
&\qquad\times\mathbf{1}(\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},N)\leq\mathbb{P}(\mathrm{cog}',\tau|\mathrm{cog},H))\\
&\qquad\geq(\mathbb{E}(\tau|s,H)-\mathbb{E}(\tau|s,N))-\alpha\\
&\qquad\geq(\mathbb{E}(\tau|\mathrm{cog}^*,H)-\mathbb{E}(\tau|\mathrm{cog}^*,N))-\alpha.
\end{aligned}
$$

The last term is positive under the assumption (A4). Hence, for a given cognitive state cog, *N* strictly dominates *H*, for each $q > q_1^*$, given that *N* is the optimal action at $q = q_1^*$.

Similarly, under the assumption (A4), i.e. with $k \leq \mathbb{E}(\tau|\mathrm{cog}^*,N)-\mathbb{E}(\tau|\mathrm{cog}^*,S)$, for a given cognitive state cog, *S* strictly dominates *H* and *N*, for each $q > q_2^*$, given that *S* is the optimal action at $q = q_2^*$. □

V. Conclusions and Future Directions

We studied optimal fidelity selection for a human operator servicing a stream of supervision tasks using a SMDP framework. In particular, we studied the influence of human cognitive dynamics on an optimal policy. We modeled cognitive dynamics using Markov chains and incorporated their influence on servicing time using well-established models in psychology. We presented numerical illustrations of the optimal policy and showed that the value function is a monotonically decreasing function of the queue length and a unimodal function of the cognitive state. Assuming that the value function satisfies these properties for generic parameters, we established some of the key structural properties of the optimal policy for fidelity selection. We seek to prove these properties in our future work. We showed the existence of thresholds on queue lengths for each cognitive state at which the optimal policy switches from selecting high fidelity to selecting normal fidelity to resting to skipping.

There are several possible avenues of future research. An interesting open problem is to control the task arrival rate in order to control the thresholds for these transitions. Another interesting direction is to conduct experiments with the human operators servicing a steam of tasks, measure EEG signal to assess their cognitive state and test the benefit of recommending optimal fidelity. It is of interest to extend this work to a team of human operators processing stream of heterogeneous tasks. In such a setting, finding the optimal strategies for routing and scheduling of these heterogeneous tasks is also of interest.

References

[1] I. R. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion, "Human-robot teaming for search and rescue," *IEEE Pervasive Computing*, no. 1, pp. 72–78, 2005.

[2] M. A. Goodrich, J. L. Cooper, J. A. Adams, C. Humphrey, R. Zeeman, and B. G. Buss, "Using a mini-UAV to support wilderness search and rescue: Practices for human-robot teaming," in *IEEE International Workshop on Safety, Security and Rescue Robotics*. IEEE, 2007, pp. 1–6.

[3] S. A. Seshia, D. Sadigh, and S. S. Sastry, "Formal methods for semi-autonomous driving," in *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, June 2015, pp. 1–5.

[4] A. M. Okamura, "Methods for haptic feedback in teleoperated robot-assisted surgery," *Industrial Robot: An International Journal*, vol. 31, no. 6, pp. 499–508, 2004.

[5] J. Peters, V. Srivastava, G. Taylor, A. Surana, M. P. Eckstein, and F. Bullo, "Human supervisory control of robotic teams: Integrating cognitive modeling with engineering design," *IEEE Control System Magazine*, vol. 35, no. 6, pp. 57–80, 2015.

[6] C. Nehme, B. Mekdeci, J. W. Crandall, and M. L. Cummings, "The impact of heterogeneity on operator performance in futuristic unmanned vehicle systems," *The International C2 Journal*, vol. 2, no. 2, pp. 1–30, 2008.

[7] L. F. Bertuccelli, N. Pellegrino, and M. L. Cummings, "Choice modeling of relook tasks for UAV search missions," in *American Control Conference*, Baltimore, MD, USA, Jun. 2010, pp. 2410–2415.

[8] J. W. Crandall, M. L. Cummings, M. Della Penna, and P. M. A. de Jong, "Computing the effects of operator attention allocation in human control of multiple robots," *IEEE Transactions on Systems, Man & Cybernetics. Part A: Systems & Humans*, vol. 41, no. 3, pp. 385–397, 2011.

[9] J. R. Peters, A. Surana, and F. Bullo, "Robust scheduling and routing for collaborative human/unmanned aerial vehicle surveillance missions," *Journal of Aerospace Information Systems*, pp. 1–19, 2018.

[10] K. Savla and E. Frazzoli, "A dynamical queue approach to intelligent task management for human operators," *Proceedings of the IEEE*, vol. 100, no. 3, pp. 672–686, 2012.

[11] N. D. Powel and K. A. Morgansen, "Multiserver queueing for supervisory control of autonomous vehicles," in *American Control Conference*, Montréal, Canada, Jun. 2012, pp. 3179–3185.

[12] V. Srivastava, R. Carli, C. Langbort, and F. Bullo, "Attention allocation for decision making queues," *Automatica*, vol. 50, no. 2, pp. 378–388, 2014.

[13] V. Srivastava and F. Bullo, "Knapsack problems with sigmoid utility: Approximation algorithms via hybrid optimization," *European Journal of Operational Research*, vol. 236, no. 2, pp. 488–498, 2014.

[14] M. Majji and R. Rai, "Autonomous task assignment of multiple operators for human robot interaction," in *American Control Conference*, Washington, DC, Jun. 2013, pp. 6454–6459.

[15] V. Srivastava, A. Surana, and F. Bullo, "Adaptive attention allocation in human-robot systems," in *American Control Conference*, Montréal, Canada, Jun. 2012, pp. 2767–2774.

[16] L. I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems*. John Wiley & Sons, 2009, vol. 504.

[17] S. S. Jr., *Optimal Design of Queueing Systems*. CRC Press, 2009.

[18] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Springer Science & Business Media, 2009.

[19] S. Stidham Jr and R. R. Weber, "Monotonic and insensitive optimal policies for control of queues with undiscounted costs," *Operations Research*, vol. 37, no. 4, pp. 611–625, 1989.

[20] L. I. Sennott, "Average cost semi-markov decision processes and the control of queueing systems," *Probability in the Engineering and Informational Sciences*, vol. 3, no. 2, pp. 247–272, 1989.

[21] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology*, vol. 18, no. 5, pp. 459–482, 1908.

[22] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman, *Engineering Psychology & Human Performance*. Psychology Press, 2015.

[23] R. P. Rao, *Brain-Computer Interfacing: An Introduction*. Cambridge University Press, 2013.

[24] A. Diederich and J. R. Busemeyer, "Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time," *Journal of Mathematical Psychology*, vol. 47, no. 3, pp. 304–322, 2003.

[25] A. G. Barto and S. Mahadevan, "Recent advances in hierarchical reinforcement learning," *Discrete Event Dynamic Systems*, vol. 13, no. 1-2, pp. 41–77, 2003.

[26] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT press, 2018.