

# Human-Autonomy Teaming Using Flexible Human Performance Models: An Initial Pilot Study

Christopher J. Shannon, David C. Horney, Kimberly F. Jackson  
and Jonathan P. How

**Abstract** Recent advances in autonomy have highlighted opportunities for tight coordination between humans and autonomous agents in many current and future applications. For operations involving cooperating humans and robots, autonomous teammates must have flexibility to respond to the inherently unpredictable behavior of their human counterparts. To investigate this issue in detail, this paper uses an unmanned aerial vehicle (UAV) simulation to evaluate flexible human performance models over traditional static modeling approaches for multi-agent task allocation and scheduling. Additional comparisons are drawn between adaptive human models, which are adjusted in real-time by the autonomous planner according to realized human performance, and adaptable human models, where the human operator is given sole authority over model adjustments. Results indicate that adaptive human performance models significantly increase total mission reward over both the baseline static modeling framework ( $p = 0.0012$ ) as well as the adaptable modeling technique ( $p = 0.0028$ ) for this system.

**Keywords** Human-Autonomy teaming · Multi-Agent planning · Human performance modeling · Adaptive and adaptable autonomy

---

C.J. Shannon (✉) · D.C. Horney · J.P. How  
Massachusetts Institute of Technology, Cambridge, MA, USA  
e-mail: cshannon@mit.edu

D.C. Horney  
e-mail: dchorney@mit.edu

J.P. How  
e-mail: jhow@mit.edu

K.F. Jackson  
Draper, Cambridge, MA, USA  
e-mail: kjackson@draper.com

## 1 Introduction

Traditional approaches to multi-agent planning typically view agents as independent, autonomous vehicles. However, a variety of modern and future applications require highly coordinated efforts between humans and autonomous systems [1]. Planning for humans in the loop presents several key challenges, including the difficulties associated with adequately modeling actual human capabilities. In particular, reliable prediction of human task performance can be hard, and even well-established models from the human factors community are subject to inter-personal differences—such as experience and skill—as well as dynamic intrapersonal factors like distraction and fatigue [2]. These challenges motivate human performance models that can be updated in real-time to ensure better consistency between predictive planning and actual execution.

A promising strategy for human modeling is to use an adaptive approach in which actual human performance is leveraged as feedback for autonomous adjustments to models throughout mission execution. Adaptive autonomy has been steadily gaining traction in human-robot interaction (HRI) applications over recent years [3]. State-of-the-art autonomous systems that are able to reason on, and respond to, the needs and intentions of a person present new opportunities for close synergy between people and autonomy [4]. Recent studies have produced planners that enable a single robot to dynamically execute a shared plan with a human [5], while other efforts have developed intelligent task management strategies for human operators that regulate and optimize over human workload parameters [6, 7]. Additionally, adaptive levels of automation have shown promise for human supervisory control of multiple robots [8].

Apart from feedback based on realized human performance, an adaptable autonomy construct can provide information to the planner directly from the user. Examples include actively changing mission objective functions in real-time [9] as well as adjusting levels of automation [8]. This can enhance the user’s control of the environment, which has been shown to help in terms of change detection, system transparency, and overall user satisfaction within a variety of domains [10]. However, since adaptable autonomy requires deliberate user feedback, it can increase workload and draw attention away from the main task at hand. In addition, relying on human interaction may be less optimal than a well-designed adaptive framework [3].

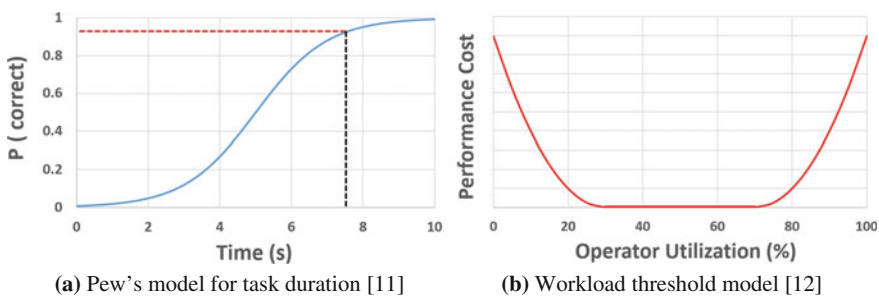
Drawing from these strategies, this pilot study focuses on an evaluation of adaptive and adaptable human performance models for the allocation and scheduling of joint human-robot tasks amongst a heterogeneous, dynamic human-robot team. Models for the human operators’ predicted task durations and workload thresholds are used to investigate the advantages and drawbacks of adaptive and adaptable human modeling in the context of these highly coordinated multi-agent systems. Results from the human subject experiment ( $n = 12$ ) provide insights into the use of flexible human performance models for human-autonomy teaming.

## 2 Background

For this pilot study, the human participant’s role within the multi-agent mission is to classify a series of discrete pieces of imagery. These imagery classifications are time-critical: the values of accurate imagery information acquisition are functions of time (e.g. time-windows and exponential time-decays). Therefore, the amount of time an operator needs for correct classification and the rate at which he can be presented with new demands are important for the allocation and scheduling of tasks.

Figure 1 shows the two human performance models used in this pilot study. Pew’s model (Fig. 1a) predicts that the correctness of a decision in a binary choice increases as a sigmoidal function of time spent on the task [11]. In order to predict or allocate task durations, probability of success may be optimized [7] or chosen (as in this case) depending on requirements for speed versus accuracy. The workload threshold model (Fig. 1b) moderates the rate at which tasks appear by incurring a cost on proposed schedules that exceed the prescribed threshold. This model, derived from the Yerkes-Dodson Law [13], uses the *utilization* metric defined as operator busy time over total mission time in order to keep the human at a desirable level of workload [12].

In our recent work, we have created a fast algorithm called ASSIST that integrates humans into the traditional task allocation and scheduling problem to produce plans for tightly coupled (i.e. synchronized joint-task) heterogeneous human-robot teams [14]. ASSIST can be embedded into a closed-loop replanning framework—in which humans are treated as dynamic agents—in order to adapt human and environment models in real-time. This pilot study evaluates the ASSIST planning framework and its various feedback mechanisms through actual human-autonomy teaming experiments.



**Fig. 1** Human performance models for operator’s imagery classification tasks. **a** Pew’s model for task duration [11]. **b** Workload threshold model [12]

### 3 Aim of the Pilot Study

The quantitative hypotheses tested in this experiment compare static (baseline open-loop approach), adaptive, and adaptable modeling strategies using an *overall mission reward* metric. This reward is calculated as the summation of all successfully completed tasks' values at their respective execution times, and incorrect imagery classifications result in zero reward for a task. Thus overall reward represents a measure of timely, accurate task completion as well as efficiency in the allocation and scheduling of the tasks among the multi-agent team to optimize execution in terms of specified mission goals. Subjective trends from post-experiment participant questionnaires—plus other recorded measures such as classification accuracy, mission makespan (total duration), and total vehicle distance traveled—supplement the evaluation.

We hypothesize that adaptive human performance modeling will result in the most overall mission reward, with adaptable modeling resulting in the second most reward, and static modeling resulting in the least. Autonomous adjustments of the models based on realized performance from actual mission execution in real-time should alleviate mismatches between the planning phase's models and the humans' actual ability. For the adaptable mode, the user's adjustment of these models should also mitigate model mismatches, albeit less efficiently than the autonomous adaptive mode. Additionally, the baseline static modeling case will likely illuminate the issues of rigid, one-size-fits-all open-loop approaches to modeling human agents.

We also hypothesize that the adaptable human performance modeling strategy will be the most preferred by users, with adaptive modeling being the second most preferred, and static modeling being the least preferred. We predict that giving the human operator authority over the models that predict his own performance (and thus drive the allocation and scheduling of his tasks) will result in higher user satisfaction than if the adjustments were made by the autonomous system instead. We further predict that the adaptive mode will be preferable to the static mode due to multi-agent plans that have flexibility and are more conformed to each participant's actual ability.

### 4 Experimental Methods

The human-autonomy teaming mission consists of discrete surveillance tasks in distinct locations throughout a map. Points of interest may have differing priorities or time-criticalities which correspond to their respective task values. The ASSIST planner allocates and schedules coupled human-robot pairs to the tasks, optimizing plans for maximum task value while adhering to human operator heuristics and constraints. The architecture uses the Robot Operating System [15] for a high-fidelity multi-vehicle simulation. Participants interact with a standard laptop

(2 GHz Intel Core i7 processor, 8 GB RAM) using a keyboard and mouse to accomplish missions.

### 4.1 User Interface

The full interface is illustrated in Fig. 2. Multi-agent plans from the ASSIST algorithm are executed by the human-robot team with the UAVs flying to task locations and providing imagery to the human operator. The top right portion of the interface provides an overhead mission view that allows the operator to anticipate vehicles arriving to their specified survey waypoints. The top-left command line prompts the user to alert him to incoming imagery. The bottom of the interface shows the live, top-down video feeds from simulated on-board UAV cameras that present the imagery to be classified.

The participant performs classification by inputting into the command line “c” for circle imagery or “s” for square imagery as well as the count of moving images. In all modes, this imagery flashes yellow to give the human operator an indication as to when he is exceeding the duration allotted for the task by the planner. Overall mission reward accumulates with timely, accurate completion of the joint human-robot tasks. The user continues the process of classifying images throughout the duration of the mission, cross-checking the overhead view for increased situational awareness and looking to the command line for incoming tasks and feedback on his performance.



Fig. 2 Full user interface for human-robot teaming mission

## 4.2 *Experiment Conditions*

There are three conditions in this experiment which relate to the modeling of human performance. Static modeling is a baseline control condition that uses open-loop modeling of humans with planning parameters set a priori. Adaptive modeling leverages actual human performance as feedback throughout each mission to autonomously adjust models. Finally, adaptable modeling uses inputs from the human operator to adapt his performance models during the mission. Replanning is conducted at 15 s intervals in all modes with their respective models.

**Static Modeling.** The static modeling mode provides a baseline open-loop approach to the representation of human performance. In this mode, humans are considered to be homogeneous agents whose predicted task performance is equivalent among different individuals. Moreover, they are assumed to be static, modeled by rigid functions throughout the length of the mission. As illustrated in Fig. 1, Pew's model for predicted task duration and the utilization model for projected workload threshold are specified a priori. For this pilot study, predicted task duration was set to 6.4 s, derived from the average length of time required for novice operators to complete a task in preliminary testing. Projected workload threshold was set to 70 %, consistent with previous studies involving human operators and multiple UAVs [12] and aligning with subjective responses from novice operators interacting with the current system.

**Adaptive Modeling.** The adaptive modeling mode leverages realized human performance throughout mission execution to adjust the planner's underlying models in real-time. Human operators are treated as heterogeneous, dynamic agents in that the specified models are adapted to fit the skills of each individual and remain flexible to track actual human performance throughout the mission. Pew's model is initially set to 6.4 s for predicted task duration and the utilization model originally specifies 70 % as the operator workload threshold. These values are then updated in real-time based on actual performance by the operator.

The two human performance models are autonomously adapted by the closed-loop system after each task. The actual time from the imagery's appearance to the operator's task completion becomes the new predicted task duration for future human-robot surveillance tasks. If the operator's inputs are correct, the utilization workload threshold is increased by 5 %, functioning as a heuristic to allow more demand on a successful operator. If the classification or count is incorrect, the utilization threshold is decreased by 15 %, decreasing overall workload on the operator in the near future. In this mode, replanning not only alleviates discrepancies between the current system state and past projections, but also leverages up-to-date realized agent performance to generate plans that more closely mirror the actual behavior of the multi-agent system. This closed-loop modeling strategy alleviates model mismatches arising from heterogeneous human operators (e.g. a model that is appropriate for one person may be ill-fitting for another) as well as dynamic, stochastic human performance.

**Adaptable Modeling.** As an alternative to the autonomous system providing feedback of realized performance to the planner, an adaptable mode was developed in which the human operator had sole authority over the adjustment of his performance models. This strategy shifts responsibility of model updates from the autonomy to the human operator in attempts at providing insights into both the objective effects on multi-agent teaming as well as the subjective perceptions of the user.

Adaptable adjustments are accomplished through the use of additional operator inputs into the user interface. First, each human-robot surveillance task requires the usual “c” or “s” classification input along with specifying the number of images. Following these two inputs, the command line prompts the subject to “Enter ‘z’ if too busy, ‘r’ if too bored (else, hit enter).” Inputting “z” decreases the operator utilization workload threshold by 10 %, while “r” increases it by 10 %. The command line then prompts the operator to “Enter ‘l’ for less allotted task time, ‘m’ for more allotted task time (else, hit enter).” Entering “l” shifts Pew’s model to decrease predicted task duration by three seconds, whereas “m” moves it in the opposite direction to increase predicted task duration by three seconds. These discrete adjustment values were chosen from a range of possibilities upon demonstrating satisfactory objective and subjective performance in preliminary testing with novice users.

### **4.3 Procedure**

Twelve individuals were selected to participate in the pilot study. Each experiment began with the subject reading an instruction document that explained the study’s purpose, the responsibilities of the participant, and the interaction procedures for the user interface. Each subject was then guided through two full mission trials (with three autonomous vehicle teammates and 15 surveillance tasks) for training purposes. The first practice trial was conducted in the adaptive human performance modeling mode to allow the subject to become familiar with his primary task. The second practice trial was performed in the adaptable human performance modeling mode in order for the subject to become familiar with providing feedback to the planner. Upon completion of the two practice missions, the participant completed a human-robot teaming mission in each of the static, adaptive, and adaptable human performance modeling modes. The ordering of modes was counterbalanced among the twelve participants (using the six possible combinations of mode ordering for two participants each) in order to mitigate potential confounding variables of learning and fatigue. The process required less than an hour of each participant’s time, and all testing took place over a two week period.

Objective mission metrics were recorded throughout each trial. Total makespan (or mission duration) and total vehicle distance traveled were noted. The start time, duration, and accuracy of each task was recorded and used to compute overall mission reward, average operator accuracy, and average task duration. All model

shifting inputs in the adaptable mode (“too busy,” “too bored,” “less time,” “more time”) and their associated input times were also logged. This level of detail into all relevant aspects of the mission allowed analysis of underlying reasons for human-robot team performance trends, such as temporally overloaded operators, model discrepancies between the planning phase and mission execution, and dynamic human performance.

Upon completing all trials, each participant filled out a questionnaire. The survey was used to evaluate situational awareness, team fluency, user interface satisfaction, and human performance modeling mode preference. Specific questions relating to the feedback modes (e.g. which was “most efficient in helping you complete the task” and which mode “you most enjoyed using”) were included in the survey. In order to garner as much information as possible from the pilot study, all questions and choices were open-ended and prompted the subject to specify why they felt a certain way. Participants’ subjective inputs provided insight into trends in the objective metrics and aided in future system development after pilot study completion.

## 5 Results

After all testing was completed, data were investigated both qualitatively and quantitatively to evaluate the effects of the various human modeling strategies on team performance and operator satisfaction. Primarily, statistical analysis of overall mission rewards and evaluations of user mode preferences were used to compare the three modes.

### 5.1 Objective Analysis

A one-way balanced ANOVA F-test was used to test whether there exists any statistically significant differences among the three modes’ associated mission rewards (Fig. 3). Based on the data and using a 95 % confidence interval, there is reason to believe that there exists significant variation between the three modes in terms of overall mission reward ( $F_{2,33} = 8.645$ ,  $p < 0.001$ ). Therefore, direct comparisons between each pair of modes using post hoc paired t-tests with Bonferroni corrections [16] are appropriate.

Comparing adaptive human performance modeling against the baseline static approach shows that average mission reward for the adaptive mode ( $M = 8116.25$ ,  $SD = 933.80$ ) is significantly different than the static mode’s ( $M = 6649.21$ ,  $SD = 1372.39$ ) average reward ( $t(11) = 4.353$ ,  $p = 0.0012$ ). With mean overall reward being much higher for the adaptive case, results indicate that adaptive human performance models significantly increase total mission reward over the static modeling framework in the context of this experiment. A second paired t-test



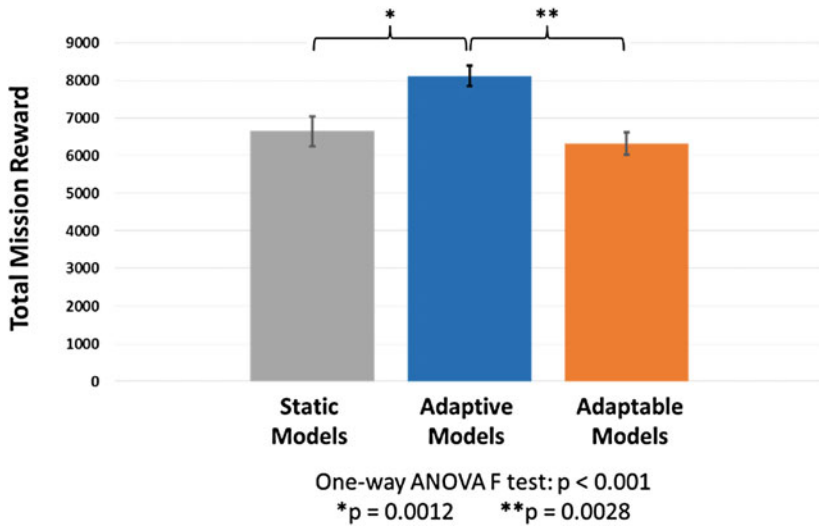


Fig. 3 Overall mission rewards between human modeling modes with standard error bars

comparing adaptable human operator modeling against the baseline static modeling strategy gives no indication that mission reward from adaptable modeling ( $M = 6311.36$ ,  $SD = 1036.45$ ) is significantly distinct from reward with static modeling ( $M = 6649.21$ ,  $SD = 1372.39$ ) in this pilot study ( $t(11) = 0.631$ ,  $p > 0.05$ ). A final paired t-test examines adaptive and adaptable human performance modeling, demonstrating that adaptive human performance models ( $M = 8116.25$ ,  $SD = 933.80$ ) result in statistically significant improvement of overall mission reward in comparison to adaptable models ( $M = 6311.36$ ,  $SD = 1036.45$ ) for this experiment ( $t(11) = 3.826$ ,  $p = 0.0028$ ).

The data suggest that adaptive human performance models can promote more effective coordination between humans and autonomous agents than traditional open-loop modeling approaches. Leveraging feedback from actual human performance can alleviate model mismatches at both the system level (e.g. workload thresholds for difficult vs. routine missions) and the individual level (e.g. different experience and skill between operators). Assuming humans to be homogeneous agents with one-size-fits-all models can be detrimental to overall team performance. In fact, average task duration varied by as much as 87.5 % between participants in this pilot study. A histogram of participants' average task durations in Fig. 4 shows a Gaussian-like distribution with substantial variance between subjects.

The advantages of adaptive modeling for heterogeneous human agents are further illustrated in Fig. 5, which shows mission timelines for a single participant in each of the three modes. This subject generally accomplishes tasks more quickly than the predicted task duration of 6.4 s set a priori, which causes a model mismatch for the static open-loop modeling approach. The adaptive human performance modeling mode, on the other hand, is able to adjust to the operator's realized

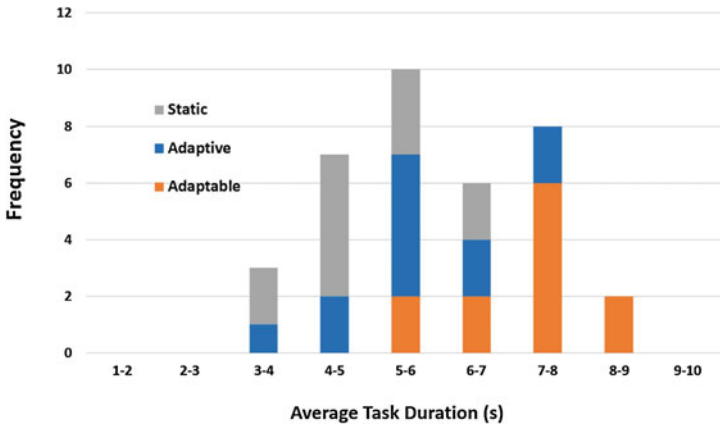


Fig. 4 Histogram of average task duration for each mission trial

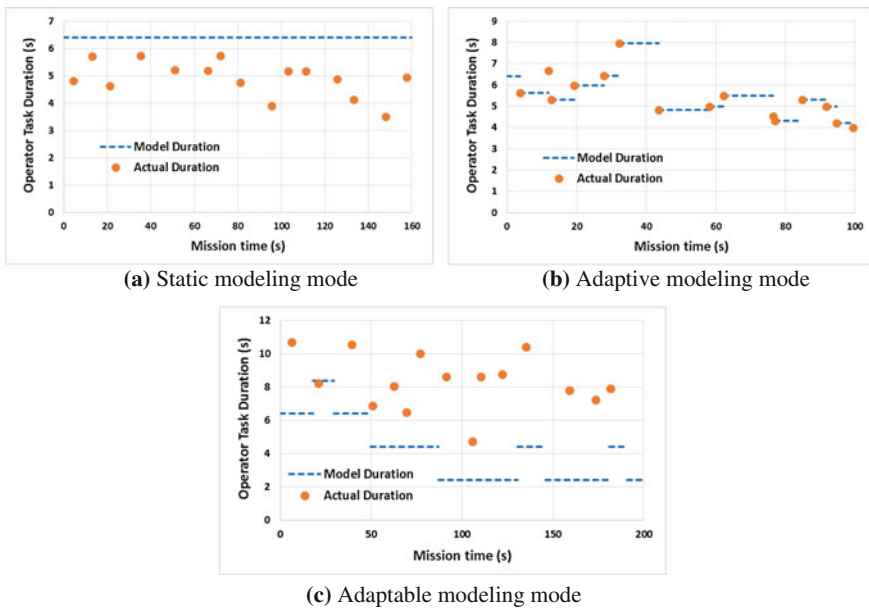


Fig. 5 Planning phase’s predicted human operator task duration from underlying models versus actual task duration over the course of one participant’s trials in each mode. a Static modeling mode. b Adaptive modeling mode. c Adaptable modeling mode

performance once the mission begins. As an additional advantage for adaptive modeling, the human’s dynamic behavior during this trial—which takes the form of decreasing durations as the mission progresses—is tracked by the closed-loop system.

Results also indicate (within the context of this system) that incorporating real-time autonomous adjustments of human performance models provides quantitative advantages over relying on human input for adjustments. This finding may be explained as the result of two main causes. First, the adaptable mode required slightly more workload than the static and adaptive modes. Figure 4 shows the adaptable trials favor the right side of the distribution, and a paired t-test confirms that the adaptable mode's additional inputs resulted in significantly longer average task durations ( $M = 7.246$ ,  $SD = 0.997$ ) over the adaptive approach ( $M = 5.617$ ,  $SD = 0.986$ ) in this pilot study ( $t(11) = 3.745$ ,  $p = 0.00324$ ). Second, the human operator may be worse than the autonomous system at adjusting his model appropriately to minimize model mismatches between the planning phase and actual mission execution. Figure 5c illustrates that the subject fails to track actual performance accurately. Even more so, at times throughout the trial his adjustments increase model discrepancies rather than mitigate them.

Other aspects of human-robot team performance among the feedback modes were analyzed to supplement comparisons of overall mission reward. Mission makespan, or total mission duration, indicates that adaptive human performance modeling significantly reduces total mission time ( $F_{2,33} = 10.729$ ,  $p < 0.001$ ) over both baseline static ( $t(11) = 6.710$ ,  $p < 0.001$ ) and adaptable modeling approaches ( $t(11) = 4.563$ ,  $p < 0.001$ ). This further signifies adaptive modeling's ability to mitigate model discrepancies between the planning and execution phases, resulting in more efficient task allocations and schedules amongst the human-robot team. Additionally, overall vehicle distance was longer in the adaptable modeling mode than both the static ( $t(11) = 1.976$ ,  $p = 0.0738$ ) and adaptive ( $t(11) = 2.012$ ,  $p = 0.0594$ ) strategies ( $F_{2,33} = 2.656$ ,  $p < 0.085$ ), pointing again to suboptimal multi-agent plans for the adaptable case.

## 5.2 Subjective Measures

Post-experiment questionnaires allowed subjects to provide qualitative feedback on their perceptions of the autonomous system, the user interface, and the various human performance modeling modes. All participants provided positive comments on the general system, specifically citing the efficiency of the interface under temporal pressure. In addition, real-time feedback was said to be especially useful in terms of helping operators stay motivated throughout the mission and allowing them to self-correct after errors. Finally, subjects enjoyed having the overhead view available for team-wide situational awareness and projections of incoming surveillance tasks.

Participant responses to questions on mode preference align with objective findings, as 92 % of subjects chose adaptive human performance modeling as the "most efficient" mode of the three tested. Additionally, 83 % of users stated that they "most enjoyed using" the adaptive scheme, describing it as "stimulating," "well-tuned," and "most engaging." This autonomous closed-loop modeling

approach was generally perceived as being more effective than the baseline static case, and its lower workload requirements relative to the adaptable strategy allowed users to focus on the primary classification task without distraction.

## 6 Discussion

The increase in overall mission reward for adaptive human performance modeling over both the baseline static case and the adaptable approach may be attributed to the adaptive strategy's ability to minimize model mismatches between the planning phase and actual human ability. Proper prediction of human operator performance within the human-robot teaming missions is important for effective multi-agent planning towards achieving specified mission goals. The ASSIST algorithm is allocating and scheduling tasks amongst the vehicles with the aim of maximizing mission reward. Again, this reward metric is calculated by totaling all successfully completed tasks' values at their execution times, and these time-varying task score functions may take various functional forms.

When the planning phase over-predicts operator task duration (Fig. 5a), ASSIST imposes a longer scheduling constraint on future tasks. This may lead to allocations in which other vehicles travel longer distances in order to reach higher-value tasks around the time the human is expected to become available. However, if the operator finishes the task more quickly than predicted, he may then be forced to wait idly for an undesirable amount of time before the next task arrival. If, on the other hand, the planning phase under-predicts operator task duration (Fig. 5c), ASSIST may allocate tasks in an overly-optimistic fashion. A vehicle may be required to stay committed to a task longer than expected due to the human operator, thus resulting in the vehicle arriving later-than-expected to subsequent tasks. This can lead to both inefficient scheduling throughout the multi-agent team (e.g. two vehicles each arriving to their tasks at the same time, requiring one to wait an extended period of time for the human) as well as "missed rendezvous" (e.g. missing a task time window).

In addition to task durations, predicting the human operator's workload threshold is important for the heuristics of the algorithm. Ideally, an operator would be able to achieve 100 % accuracy on imagery classification tasks while remaining as busy as possible in order to reduce scheduling constraints on task execution times and minimize overall mission duration. With this in mind, adaptive workload thresholds were increased following accurate imagery classifications to reflect the fact that the operator was able to succeed at the prescribed workload. Upon task failures, adaptive workload thresholds were decreased to allow the user to gather himself for subsequent demands. Instead of this heuristic approach, future experiments will incorporate secondary tasks to analyze and adjust the operator's workload threshold in a more direct fashion.

In terms of subjective data, the majority of participants preferring adaptive modeling over the adaptable approach is contrary to original predictions which

assumed that providing the human operator with more control over the system would result in greater user satisfaction. However, these results align with other findings in the literature. Specifically, Gombolay et al. [17] showed that humans generally prefer to work within an efficient team rather than have a heightened role in the planning process if that increased role is detrimental to overall team performance. In the current pilot study, allowing human operators to adjust their own performance models imposes additional workload requirements and generally produces larger model mismatches and thus less efficient multi-agent plans.

## 7 Conclusions

Closed-loop planning approaches in which the autonomous system adapts flexible human models throughout mission execution can improve overall mission performance for tightly coordinated human-robot teams. The challenges of unpredictable, heterogeneous human agents with dynamic, stochastic behavior can be addressed through responsive replanning that leverages realized execution information. Also, while adaptable human modeling failed to demonstrate improved mission performance over the baseline open-loop approach, future work concerning efficient input methodologies and comprehensive interface feedback—particularly as it relates to model mismatches—may reveal added benefits.

This human-robot teaming pilot study highlights many potential avenues for future work, such as testing longer missions to capture more dynamic human performance and utilizing more comprehensive subjective questionnaires and workload assessments. Additional work could be done to test more complicated feedback modes, such as improving adaptable adjustment mechanisms and incorporating stochastic averaging [18], filtering [19], and learning techniques [20] into the adaptive approach. A hybrid “adaptablive” mode could be created to investigate whether building on both adaptive and adaptable strengths can be synthesized. This could also distinguish effects of additional workload versus model discrepancies to help explain the adaptive mode’s advantageous performance over the adaptable strategy in this work.

**Acknowledgments** The authors would like to thank Draper (Cambridge, MA, USA) for funding this research as well as Professor Julie Shah for her insight and guidance throughout this effort.

## References

1. Unmanned Systems Integrated Roadmap: FY2013-2038. Washington, DC, USA (2013)
2. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multirobot control: a review of human factors issues. *IEEE Trans. Human-Mach. Syst.* **44**, 13–29 (2014)
3. Kaber, D.B., Prinzl, L.J.: Adaptive and Adaptable Automation Design: A Critical Review of the Literature and Recommendations for Future Research (2006)

4. Alami, R., Clodic, A., Chatila, R., Lemaignan, S.: Reasoning about humans and its use in a cognitive control architecture for a collaborative robot. In: *Human-Robot Interaction Workshop* (2014)
5. Shah, J.A., Wiken, J., Williams, B.C., Breazeal, C.: Improved human-robot team performance using Chaski, a human-inspired plan execution system. In: *6th International Conference on Human-Robot Interaction*, pp. 29–36 (2011)
6. Savla, K., Frazzoli, E.: A dynamical queue approach to intelligent task management for human operators. *Proc. IEEE* **100**, 672–686 (2012)
7. Srivastava, V., Surana, A., Bullo, F.: Adaptive attention allocation in human-robot systems. In: *American Control Conference* (2012)
8. Kidwell, B., Calhoun, G.L., Ruff, H.A., Parasuraman, R.: Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **56**, 428–432 (2012)
9. Clare, A.S., Cummings, M.L., How, J.P., Whitten, A.K., Toupet, O.: Operator object function guidance for a real-time unmanned vehicle scheduling algorithm. *J. Aerosp. Comput. Inf. Commun.* **9**, 161–173 (2012)
10. Sauer, J., Kao, C.-S., Wastell, D.: A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics* **55**, 840–853 (2012)
11. Pew, R.W.: The speed-accuracy operating characteristic. *Acta Psychol. (Amst)* **30**, 16–26 (1969)
12. Nehme, C.E.: *Modeling Human Supervisory Control in Heterogeneous Unmanned Vehicle Systems* (2009)
13. Yerkes, R.M., Dodson, J.D.: The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* **18**, 459–482 (1908)
14. Shannon, C.J., Johnson, L.B., Jackson, K.F., How, J.P.: Adaptive mission planning for coupled human-robot teams. In: *American Control Conference* (2016)
15. Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., Berger, E., Wheeler, R., Ng, A.: ROS: an open-source robot operating system. *ICRA Work.* **3** (2009)
16. Weisstein, E.W.: Bonferroni Correction (2004)
17. Gombolay, M.C., Gutierrez, R.A., Sturla, G.F., Shah, J.A.: Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams. *Robot. Sci. Syst. X.* (2014)
18. Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T.: Bayesian model averaging: a tutorial. *Stat. Sci.* **14**, 382–417 (1999)
19. Bertuccelli, L.F., Choi, H., Cho, P., How, J.P.: Real-time multi-UAV task assignment in dynamic and uncertain environments. In: *AIAA Conference on Guidance, Navigation, and Control* (2009)
20. Bishop, C.M.: *Pattern Recognition. Machine Learning* (2006)