Operator Choice Modeling for Collaborative UAV Visual Search Tasks

Luca F. Bertuccelli *Member, IEEE,* and Mary L. Cummings *Senior Member, IEEE*

***Abstract*—Unmanned Aerial Vehicles (UAVs) provide unprecedented access to imagery of possible ground targets of interest in real-time. The availability of this imagery is expected to increase with envisaged future missions of one operator controlling multiple UAVs. This research investigates decision models that can be used to develop assistive decision support for UAV operators involved in these complex search missions. Previous human-in-the-loop experiments have shown that operator detection probabilities may decay with increased search time. Providing the operators with the ability to requeue difficult images with the option of relooking at targets later was hypothesized to help operators improve their search accuracy. However, it was not well understood how mission performance could be impacted by operators performing requeues with multiple UAVs. This work extends a queueing model of the human operator by developing a retrial queue model (ReQM) that mathematically describes the use of relooks. We use ReQM to generate performance predictions through discrete event simulation. We validate these predictions through a human-in-the-loop experiment that evaluates the impact of requeueing on a simulated mutiple UAV mission. Our results suggest that while requeueing can improve detection accuracy and decrease mean search times, operators may need additional decision support to use relooks effectively.**

I. INTRODUCTION

An important aspect of ongoing and envisaged Unmanned Aerial Vehicle (UAV) missions is the visual search task, in which operators are responsible for finding a target in an image or a video feed. Due in part to advances in networked sensors, military analysts are becoming increasingly overwhelmed with the volume of incoming UAV imagery (both full motion video and static images) [1]. Given the future DoD vision of one operator supervising multiple UAVs, the amount of incoming imagery to analyze in real-time will grow [2]. Moreover, with recently announced wide area airborne sensors such as Gorgon Stare and Argus which can generate up to 64 images per single UAV camera concurrently, there is an urgent need to develop efficient approaches for human analysis of UAV-generated imagery [1, 3].

Given the complex interactions between the human and the automated sensors in these UAV missions, models of the human operator are necessary in order to develop more appropriate decision support systems that account for operator decision making inefficiencies, such as increased wait times for vehicle selection and loss of situation awareness [4]. Mathematical models for human operators interacting with multiple UAVs have been developed using a queueing framework [5, 6], where external tasks are generated from an underlying stochastic process, and the human supervisor, modeled as a server, services the stream of tasks. While analysis of realistic multi-UAV missions is analytically intractable, Discrete Event Simulation (DES) of operator queueing models has been used to generate accurate performance prediction of experimental results [6]. Operator models have also been developed for human information aggregation using 2-alternative choice (2-AC) models [7–11] and visual search formulations [12–19].

In difficult search environments, operators searching imagery in multi-UAV environments may desire more choice than determining if a target is present or absent in an image. More specifically, operators may seek additional information in order to find the target, possibly through another visit later on in the mission, or they may choose to ignore a task because there is not sufficient information to make a confident assessment. We hypothesized that instead of a two-choice model, operators would be better served by having a third option of reevaluating a search task at a later time by requeueing the image and taking another glance via a relook. Throughout this article, we make a distinction between the choice of requeueing, which is the abandonment of the current search task, and a relook, which is an additional glance at a previously searched image.

While it is straightforward to implement a requeue option in a multi-UAV simulator, the effect of providing requeue and relook capabilities must be investigated experimentally since there is a potential for undesirable effects such as increased operator workload. Nonetheless, these capabilities could be operationally important in minimizing collateral damage and reducing errors, and have been studied in the context of optimal stopping [20] and inspection problems [21]. While these works showed promising results in assessing the informational value of an additional look, these studies are limited in two main ways. First, previous work related to the speed-accuracy tradeoff shows that operator error rates may not be stationary, meaning that an operator's accuracy can change as a function of time spent searching [22]. This is supported by work in the vigilance literature as well, in which the operator's ability to discern a target of interest is dependent on the search time and difficulty [23]. In addition, multi-video visual search tasks with the possibility of a relook can increase operator workload.

This effort makes three novel contributions in presenting a choice model for a search task with requeues. First, we develop a Requeueing Model (ReQM) for visual search tasks that includes the possibility of requeuing difficult images, and

Manuscript received on XX XXXX, 2010. This work was supported by the Michigan/AFRL Collaborative Center in Control Science and under ONR Grant N00014-07-1-0230. L. F. Bertuccelli is with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA (email: lucab@mit.edu). M. L. Cummings is with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA (email: missyc@mit.edu).

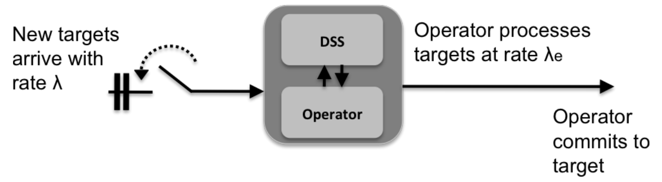


Fig. 1. A queueing model for the human operator in which new targets arrive at a known rate λ and they are processed at a rate λ_e.

pose ReQM as a variation of a retrial queue with feedback [24–27]. We next develop a DES with ReQM (DES-ReQM) that embeds operator models derived from previous experimental data of a simulated multi-UAV mission. We then present results in the predicted performance of multi-UAV visual search tasks using DES-ReQM. We build on previous work [28] by discussing the results of a human-in-the-loop experiment that confirm the predictions made by DES-ReQM, as well as presenting experimental observations of operator behavior in requeueing and relooking tasks. We conclude with a discussion on the implications for requeues and relooks with an actual operator in the loop.

II. Operator Modeling

Queueing models for human operators have been previously proposed in the context of air traffic control, where the human operator is treated as a serial controller, capable of handling one complex task at a time [5]. These queueing models of operators have been recently extended to account for operator workload and attention inefficiencies in the context of human supervisory control of multiple UAVs [6]. Figure 1 shows a general queueing model for a multi-UAV supervisory control problem for the visual search task. Search tasks are generated by a Poisson process at an average rate λ, and the human operator (with possible help from a decision support system, DSS) services the tasks at a rate λ_e. In complex tasks, operators may dedicate themselves only to a single task at time, allowing the incoming tasks to accumulate in the queue. The visual search task initiates when the operator begins examining the image feed once the UAV reaches the target, and concludes with a decision on the target location.

A. Decision models

An important feature of an operator queueing model is that a sub-model is needed to understand how humans accumulate information and ultimately make detection decisions in search tasks (the “Operator” block in Fig. 1). One common formulation uses a 2-alternative choice framework (2-AC) [7–11, 29]. 2-AC models originate from hypothesis testing models [7] and characterize information accumulation as a stochastic diffusion process. It can be shown [30] that sufficient statistics of the diffusion model can be summarized by two random variables that can be measured empirically: the probability of choosing one of the alternatives, P, and mean response time $\bar{T}$. For the visual search task in this article, P is the probability of detection.

Previous work in the visual search literature has also attempted to provide some insight in human response times and accuracy in a speed-accuracy tradeoff setting. For example, Signal Detection Theory (SDT) has been used to mathematically characterize human performance in visual search tasks. In most experiments, subjects are shown a sequence of images and both detection time and accuracy are measured. Earlier work has proposed generating a Receiver Operating Characteristic (ROC) for humans to understand the relationship between correct detection and false alarms [17], while other work extends the SDT framework to include attention [31]. King et al. [32] use realistic imagery and assumed an SDT model to describe the performance of the subjects. Waldman et al. [33] developed an empirical model for visual detection with search, based on identifying parameters of a probabilistic model, such as search time and accuracy. Sperling [34] investigated visual search with attention more specifically looking at reaction times, and investigating a so-called Attention Operating Characteristic (AOC). More recently, Huang [35] considered attention in visual search.

Extensive work in visual search has also emphasized the use of probabilistic models that relate mean decision time to the mean time to search [12–17]. Recent work has also moved beyond the mean decision times, and analyzed the role of parameter identification for parameterization of the search time distributions [36, 37]. While the characterization of the search is also important from a cognitive science perspective, the work in this article does not address the low-level details of how the search is accomplished but rather seeks to understand and quantify the effect of sequential searches within the context of supervisory control.

B. Motivation for requeues

We used experimental data obtained from a multi-UAV simulator including visual search tasks to determine the relationship between detection probability and search time in search tasks performed by single operators in multi-UAV simulated missions [6]. The search tasks contained imagery obtained from Google Maps and the participants were instructed to maximize the number of targets correctly found over the course of the mission. In multiple target searches, operator detection accuracy degraded with time [38] since difficult searches required additional cognitive effort from the operators. We used this data to determine that the empirical mission probabilities of detection decreased with increased search time and hypothesized that the increased likelihood of mistakes arises because people are forced to make a choice on a search task in order to move on to waiting search tasks.

Figure 2 shows the probability of detection modeled using a logistic regression for a visual search task obtained from previous experimental data in multi-UAV simulated missions [6] and is consistent with previous vigilance literature [23, 38]. The logistic model will be explained further in Section III, but given the decrease in operator accuracy with time, it appears that there could be a benefit to requeueing the current search task (and possibly abandoning it in the absence of new sources of information). While one of the key benefits of a requeue is

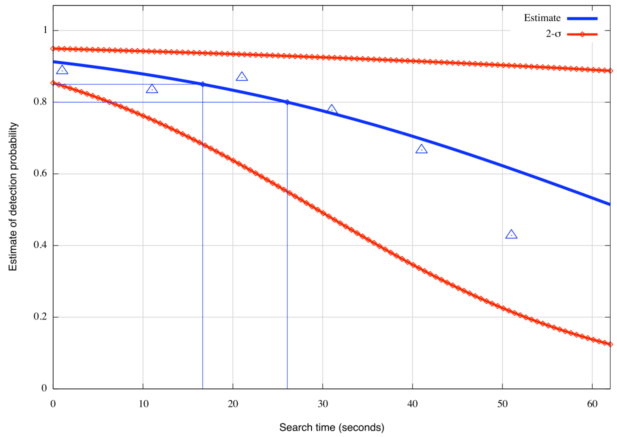


Fig. 2. Detection probability decreases as a function of time: best estimate is solid line, 2-σ regions shown in diamond-etched lines, Maximum Likelihood estimate in triangles. Vertical lines show illustrative examples of search times for detection probabilities of $P_d = 0.85$ and $P_d = 0.80$

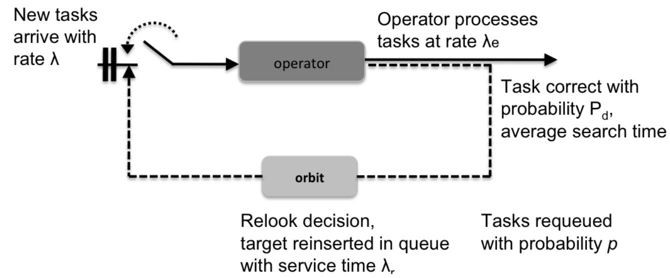


Fig. 3. In contrast to the queueing model of Fig. 1, retrial queueing allows for operator requeues in difficult tasks, which are then placed in an orbit queue. Targets are then reinserted in the available task list at some later time, and the operator can choose to search or abandon them.

that it frees the operator to pursue other searches, especially since the queueing model assumes that tasks are continually arriving. An additional benefit of a requeue is that a search task could be investigated at some later time in the mission via a "relook". Note when search tasks are requeued, they are simply reinserted in the queue, and there is no explicit provision for how a requeued task is searched again (e.g., First-come, First-served). For example, an operator may choose to take another look at a requeued task after having explored other tasks, or never search the task again for the remainder of the mission.

III. Retrial Queueing model ReQM

In order to account for requeueing in a queueing framework, this section first describes the formulation for requeueing a task using retrial queues [24, 25], and describes the ReQM developed for this effort. Figure 3 shows a visualization of ReQM. Just like in conventional queues for human operator models, ReQM treats the human as a server [6] and if the operator is free to initiate an available visual search task, the task is shown to the operator and can be serviced immediately. If the operator does not wish to complete the initiated task, and wishes to delay it to some other time (which could lead to never seeing it again), the task is inserted in a so-called *orbit* pool. ReQM is a slight variation of retrial queues in [26, 27] but differs fundamentally in attempting to model how requeues and target detections are made by the operator (which are not addressed in [26, 27]). Additional details of ReQM are provided in the following main components: 1) Task arrival and service rates, and 2) operator choice model.

A. Arrivals and service rates

ReQM assumes that new tasks follow a Poisson arrival with rate λ and that new tasks are serviced by the operator at a rate λ_e. Note that for queueing models of visual search tasks performed with UAVs, the UAVs are allocated in the environment to maximize the total number of targets found correctly. Therefore, λ_e has a strong dependence on numerous mission-specific factors, but the principal drivers are operator search time and, in the case of a multi-UAV setting, vehicle routing policy that likely requires operator intervention [39]. Under certain arrival and service rates, queue instability can occur, in which the number of outstanding targets will grow unbounded over time. Vehicle routing policies have been developed to provide guarantees under which queue instability can be prevented [39], but it is unclear whether these guarantees will hold with actual operators in the loop, especially if these operators have the capability of requeueing targets.

B. Choice model

The choice model is the underlying mechanism under which the operator can make a detection decision (e.g., whether there is a target in the image or not), or decide that a task needs to be requeued.

1) Detection decision: We abstract the operator choice model into a detection probability and search time distributions. For the detection probability, we derived a logistic regression model from previous experimental data [6]. The operator is assumed to make correct detections with probability

$$P_d(t_s) = \frac{1}{1+\exp(\hat{\beta}^T \mathbf{t})} \tag{1}$$

where $\mathbf{t} = [1, t_s]$, t_s denotes the search time, and β is a vector of parameters obtained from experimental data. In distinction to the work in [40], the detection probability is a non-stationary quantity, and is negatively dependent on the search time.

Search time distributions can likewise be estimated from previous experimental data regarding the visual search task in a simulated multi-UAV experiment [6]. We found that the log-normal distribution of Eq. (2) is a good approximation for the search time distribution, where $\bar{T}$ and σ^2 are the mean and variance of the search times

$$f(t_s; \bar{T}, \sigma^2) \propto \exp\left(\frac{-(\log(t_s) - \log(\bar{T}))^2}{2\sigma^2}\right), \quad t_s > 0 \tag{2}$$

2) Requeueing decision: If the operator is not willing to make a detection decision, then the operator can choose to requeue the task. However, the requeueing policy describing how the operator decides to requeue tasks may depend on a number of factors, including the total amount of time spent searching for a target, the total number of remaining tasks, and target arrival rates. As a first approximation, ReQM assumes that an operator will requeue the target with some probability p, which is the probability that the search time exceeds some critical search time T_{rl} (additional details on how T_{rl} is chosen are provided in Section IV).

In summary, the operator choice model in ReQM assumes that the operators are going to either make a correct detection, an incorrect detection, or ask to requeue the task. Search times are distributed according to $f(t_s;\bar{T},\sigma^2)$, and in the event of a detection decision given a realization t_s from this search time distribution, the operator makes a correct detection with probability $P_d(t_s)$.

C. ReQM analysis and the need for simulation

ReQM is an initial attempt to represent how a repeated visual search task with a task requeueing option can be properly formalized using retrial queues. However, analysis of this model is difficult without human-in-the-loop experimental data, as it is unclear how frequently subjects decide to requeue tasks, and previous work in retrial queueing theory does not provide insight into these choices for human operators.

In addition, even if we understood how operators requeue targets, it would be difficult to analyze ReQM in closed form since real models for retrial queues may deviate from some of the common assumptions necessary for analytical tractability. In ReQM, for example, requeueing invalidates the assumption of independent arrivals. In addition, [26] assumes that a task in the orbit queue can only be serviced if the nominal queue is empty. This is not a suitable representation for the multiple UAV relook problem, since a target can be requeued regardless of the remaining outstanding visual search tasks. Queueing theory also is concerned with queue stability, in the sense that the number of tasks does not grow unbounded over time, which may not be a valid assumption when human performance is considered. While it will be the topic of future work to investigate whether the analytical methods from retrial queueing theory may be applicable, a method for admitting less restrictive assumptions can be addressed by using Discrete Event Simulation (DES).

IV. Discrete Event Simulation of ReQM: DES-ReQM

DES provides a number of advantages. First, where analytical methods are not available for analyzing a queue in closed form, DES can help provide insight of the queue transient properties. Secondly, DES can be used for tuning the appropriate set of parameters to be used for human-in-the-loop experiments, such as determining appropriate task arrival rates. The ultimate goal of the DES environment in this effort is to provide a high-fidelity simulation of the experiment and this section discusses a DES model of ReQM, DES-ReQM , which is composed of three main parts: an environmental module, a routing policy module, and a requeue policy module.

A. Environmental module in DES-ReQM

The environment is assumed to be a bounded region, populated with stationary targets that are generated according to a Poisson process with arrival rate λ. Without loss of generality for the kinds of single operator, multi-UAV missions envisioned for this work, we assume that the low level vehicle control loops are closed by an onboard autopilot, and that low-level planning problems (such as satisfying turn rate constraints on UAVs) are not the responsibility of the operator, but of appropriate low-level control algorithms.

B. Operator planning module in DES-ReQM

In modeling the operator planning policy, we make the assumption in DES-ReQM that operators allocate UAVs to targets according to a policy that routes UAVs to the targets that are nearest geographically. While current research is investigating the role of different routing strategies [39], we will assume this greedy approach.

We are interested in a mission objective that maximizes the total number of targets correctly found (N_F) out of the total number of possible targets in the environment (N_T)

$$J_F = N_F/N_T \tag{3}$$

N_F is a function of numerous operator-specific parameters (such as target difficulty, search time, and requeue policy), but in the multi-UAV setting it also has a strong dependency on the routing policy for the vehicles. For example, operators could increase vehicle travel time by assigning a vehicle to visit a distant target, rather than allocating vehicles to service nearer tasks.

Upon reaching the targets, the UAVs are assumed to loiter around the target, and initiate a visual search task only when the operator has chosen an available UAV. For the ReQM, UAVs initiate a visual search task according to a First-Come, First-Served (FCFS) policy (in which the first UAV to reach the target initiates the search task first), which is a common assumption made in vehicle routing problems [39]. The search times were modeled by using the search time distributions from previous experiments performed in the RESCHU multi-UAV simulation environment [6]. Realizations of the search times for the i^{th} task, t_s^i, are generated by sampling a new random number from the log-normal distribution in Eq. (2) with mean $\bar{T}$ and variance σ^2,

$$t_s^i \sim f(t_s;\bar{T},\sigma^2) \tag{4}$$

In turn, the search outcome is generated by the realization of the random variable $P_d(t_s^i)$ from the logistic regression of Eq. (1). For the human-in-the-loop experiment discussed in the next section we determined, from previous experimental data [6], that the logistic regression parameters are given by $\hat{\beta} = [-2.300, -0.037]$. For the search time distributions, we found that $\log(\bar{T}) = 3.1$ and $\sigma^2 = 0.6$.

C. Requeue module in DES-ReQM

The requeueing model in DES-ReQM assumes that a task is requeued by the operator if the search times t_s^i exceed a

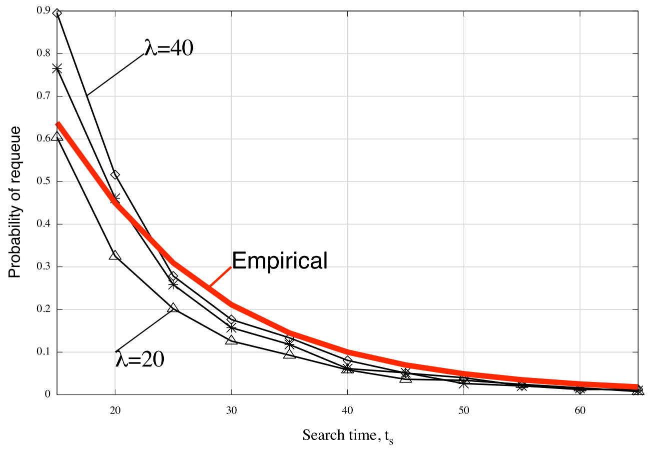


Fig. 4. Trend of empirical relook probabilities ($\triangle, *, \diamond$) agrees with theoretical relook probability (solid line) for 3 different arrival rates chosen for testing in DES-ReQM: $\lambda = \{20,\ 30,\ 40\}$ [sec/target].

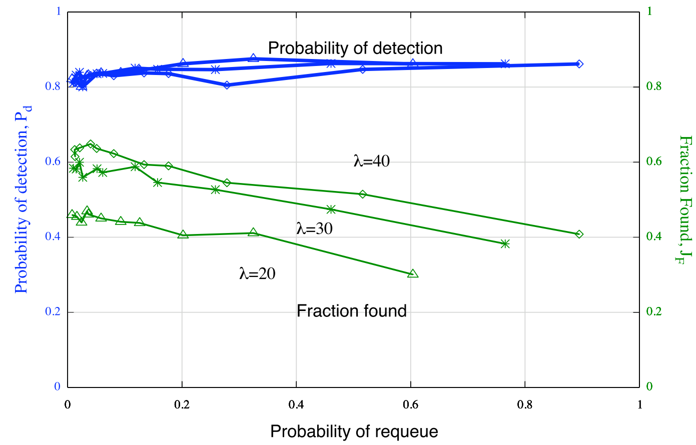


Fig. 5. Detection probability and fraction found J_F vs. requeue probability for 3 different arrival rates. Lower arrival rate decreases J_F from 0.62 to 0.42 as relook probability increases to 0.8, but detection probability increases from 0.82 to 0.86.

critical time T_{rl}. For example, if the critical time is chosen as $T_{rl} = 25$ seconds, the task is automatically requeued by the system when 25 seconds have elapsed. When a requeue takes place, the task is inserted in the orbit pool, and a new route is calculated for any available UAV in the simulation. The critical time was varied as a simulation parameter and discretized in the interval $T_{rl} \in \{20, 25, 30, \ldots, 60\}$. This interval was chosen since it described over 95% of the support of the search time data from previous experiments.

Note that the theoretical requeue probability $\bar{p}$ for each of the critical times T_{rl} can be found with the following integral (which is the red line labeled "Empirical" in Fig. 4)

$$\bar{p}(T_{rl}) = \Pr(t \geq T_{rl}) = 1 - \int_0^{T_{rl}} f(t_s \mid \mu, \sigma^2) dt_s \tag{5}$$

Unfortunately, the integral is not available in closed form, but numerical routines can evaluate the cumulative distribution function of the log-normal distribution.

D. Simulation results

This section presents simulation results of the performance using the previously developed operator choice models analyzing 100, 10-minute long simulated UAV missions. In this setting we analyzed the detection probability (P_d) and fraction found correctly (J_F, given by Eq. (3)).

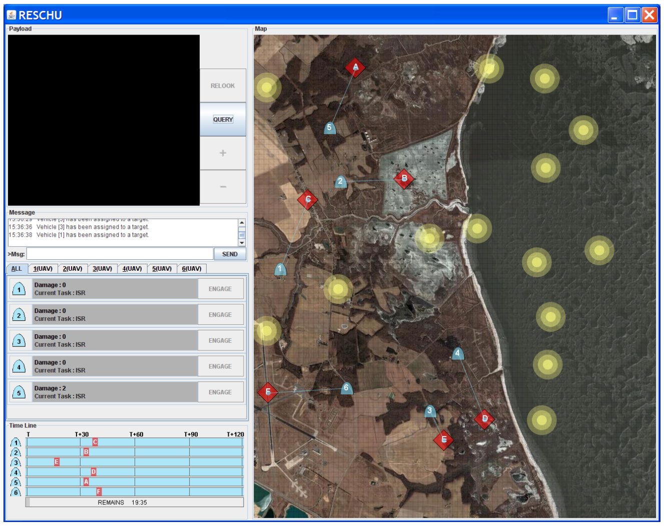
Targets were modeled given with 3 distinct average target arrival rates $\lambda \in \{20, 30, 40\}$ [sec/targ] that, given previous human-in-the-loop experimental data, were arrival rates representing different taskloads. Figure 4 shows the agreement between the theoretical prediction relating the requeue probability with the search time T_{rl}. For example, choosing a relook search time of $T_{rl} = 25$ seconds implies that the probability of requeue will be on the order of $p = 0.3$. Recall from Eq. (5) that as the search time T_{rl} threshold increases, we intuitively expect the probability of requeueing to decrease since operators will have more time to make decisions on the presence or absence of the target.

By varying the time T_{rl}, it is possible to, in turn, investigate the effect of requeueing on mission performance. Figure 5 shows the impact of varying the probability of requeueing on the detection probability and fraction correctly found J_F, averaged over the 50 Monte Carlo simulations. Figure 5 shows that the DES-ReQM predicts that tasks that were requeued with probability $p = 0.3$ for arrival rates of $\lambda = 30$ [sec/targets] resulted in a fraction found of $J_F = 0.5$, while requeueing with probability $p = 0.78$ resulted in a fraction found of $J_F = 0.4$ Predictably, the increase in requeue probability decreases the fraction found J_F since operators do not have enough time to find new targets. On the other hand, the probability of detection under all three arrival rates increases from 0.81 to 0.86, thereby demonstrating the potential for an improvement in overall probability of detection by virtue of moving on to less challenging tasks.

In summary, the DES-ReQM shows that there is an important tradeoff between maximizing the fraction of targets found and ensuring a high overall accuracy in target detection. Therefore, requeues may be beneficial in the context of improving probability of detection, but a human-in-the-loop experiment is needed to investigate operator requeueing strategies and the effect on mission performance.

V. Relook Experiment

Using the results from the DES-ReQM simulations, an experiment was conducted with the objective of investigating the performance of a retrial queue when users had an available requeue option. The experiment was performed in RESCHU (Research Environment for Supervisory Control of Heterogeneous Unmanned vehicles) [6], a simulation specifically tailored to investigate human-in-the-loop interaction with multiple UAVs. A typical RESCHU interface is shown in Fig. 6(a). A single operator is tasked with handling N UAVs in an environment where targets are non-moving, but appear at random intervals. When a UAV (shown in a blue bell shape) reaches a location of interest (shown as a red diamond), a visual search task is initiated by the operator in the top left panel of the interface, with a magnified version of the search



(a) RESCHU interface: UAVs (shown in a blue bell shape) are directed to fly to locations with targets (red diamond) while avoiding risky areas (yellow circles)

(b) RESCHU search panel during search task showing the timer counting down (top right), and the relook button

Fig. 6. RESCHU interface (a) and search panel (b)

panel shown in Fig. 6(b). Note that in the visual search task panel, the operator can zoom in and out of the display, while panning the image. In addition, the operator can query the system with the "Query" button, to find out how many residual search tasks still need to be processed in the mission.

A. *Experimental objective*

In this experiment, a single operator was responsible for the coordinated search of an area using six homogeneous Unmanned Aerial Vehicles (UAVs). The object of this experiment was to maximize the fraction found, which was explained to the participants as the total number of targets correctly found out of the total number of possible targets in the environment (Eq. 3). This experiment had two treatments: 1) a relook mode (a "within" subjects treatment); and 2) a timer condition to induce artificial time pressure (a "between" subjects treatment).

In the first treatment, the operators were tested under three different relook modes:

(a) Simple search task

(b) Complex search task

Fig. 7. Examples of imagery used in the HOSS interface. A simpler search task (a) requires that the participant find the fighter in the image. A more complex search task (b) requires a search for the helicopter pad.

- No relook (NR): Operators did not have the ability to relook. Operators were required to commit to the location of a target before returning to the UAV planning task;
- Relook with Consent (RWC): Operators had the option of requeueing at any time, but after T_{rl} seconds, a flashing message was displayed on the search screen to suggest to the operator to requeue;
- Relook without Consent (RWOC): Operators had the option of initiating a requeue at any time, but after T_{rl} seconds, the target was automatically requeued.

The second treatment involved the use of the timer, and was inserted to provide operator feedback on how much time had been spent searching. Previous work has shown that time pressure can cause different operator strategies, so the two experimental conditions were with and without a timer [4].

B. *Search task and visualization of requeued targets*

The search images were obtained from Google maps, and aerial views of different scenes were presented to the operator. Each scene contained a target that needed to be found by the operator, and hence errors were only in the class of missed detections. Different images of varied difficulty levels were presented to the operator. Figure 7 shows two examples of search tasks, where Fig. 7(a) shows a simpler search task with a clear objective of finding the fighter jet (located in the lower right corner of the image). Figure 7(b) shows an example of a more complex search task, where the task is more difficult since it requires finding a helicopter landing pad in a cluttered environment. In contrast to the simpler search task, the more complex search task places additional requirements on memory and attention, and were randomly placed in the mission. The more complex search tasks were generated intentionally to provide sources of error for the operators, and force them to make choices between continuing a very difficult search or moving on to simpler tasks. Pilot tests were used to populate the image database to ensure sufficient diversity.

In the RWC and RWOC conditions, a target became available for assignment T_r seconds after it was requeued. Tasks that were requeued changed color for $T_r = 15$ seconds to show that they were not available for search. When they became available again, the tasks were changed to their initial color,

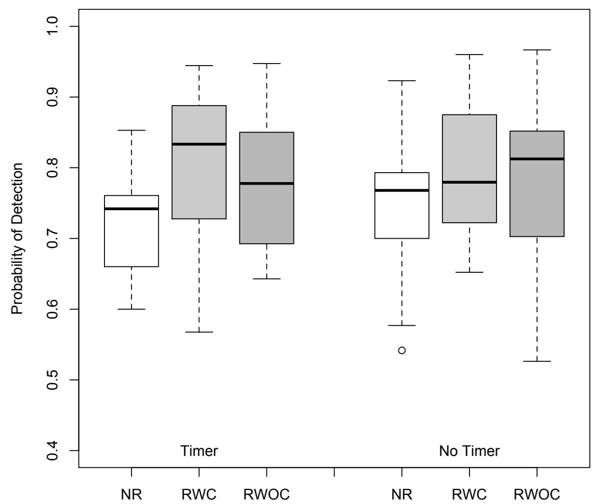


Fig. 8. Probability of detection increases most with timer using RWC and most with RWOC

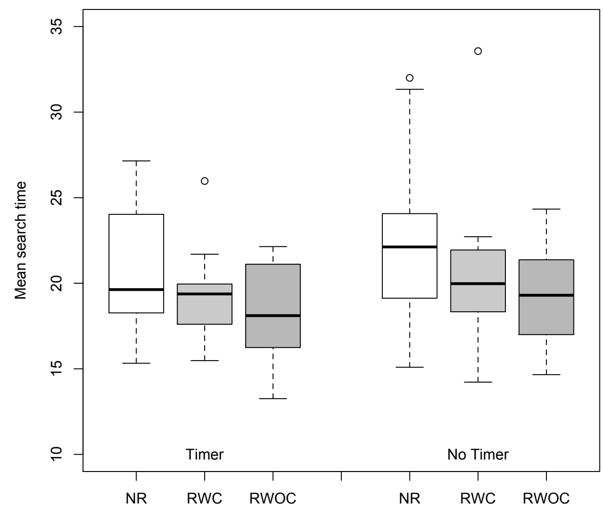


Fig. 9. Mean search time decreases for both RWC and RWOC

but were enclosed by an orange circle, to inform the operator that they had been previously visited. The relook counter was incremented by one if the targets with an encircling orange circle were searched again.

C. Experimental participants and procedures

A total of 36 participants took part in the experiment (8 female, 28 male): 30 participants were 18-25 years old, 4 participants were 25-35 years old, and 2 participants were older than 35 years. After a 10-minute training session, the participants were randomly assigned a relook mode (NR, RWC, RWOC) and timer setting (Timer, No Timer), and participated in 3 sessions with a fixed timer setting, and a counterbalanced order of relook modalities. The participants performed the first 20-minute test session, took a rest break, repeated a second 20-minute session (with another relook setting), took another rest break, and concluded with a third 20-minute experiment session. The order of the trials was randomized and counterbalanced on the timer conditions between subjects. Experimental parameters were chosen as follows: $N = 6$ UAVs; $\lambda = 30$ [sec/target]; $T_{rl} = 25$ seconds; $T_r = 15$ seconds.

VI. Experimental Results

A 2 (Timer, No Timer) $\times$ 3 (NR, RWC, RWOC) repeated measures analysis of variance was used for statistical analysis. Three participants had to be excluded since they were outside the 3-σ interval for the fraction found given by Eq. (3).

A. Performance with relook modalities

The results in this section present the probability of detection, mean search time, and fraction found for the timer condition and decision support modalities. Recall that since there is a target present in every image, then the probability of a missed detection is the only error for these search missions.

Figure 8 shows that probability of detection improved, on average, from 0.72 (with NR) to 0.80 (with RWC), for the case when a timer was used (left). Further, when no timer was used (right), probability of detection increased, on average, from 0.74 (with NR) to 0.78 (with RWOC). For overall probability of detection, mode was significant at the $\alpha = 0.05$ level, with RWOC and RWC showing better detection than none, $F(2,62) = 6.674, p = 0.002$. Posthoc comparisons using a Tukey HSD (Honest Significant Difference) test indicated that RWC and RWOC did not differ from each other, and there was neither a significant main effect of timer, nor a significant interaction between timer and modality. In particular, the change in the mean of the detection probability in the NR mode (Mean: 0.73, SD: 0.09) was significantly different ($p = 0.001$) from the RWC mode (Mean: 0.80, SD: 0.10). The change in the mean of the detection probability in the NR mode (Mean: 0.73, SD: 0.09) was also significantly different ($p = 0.03$) from the RWOC mode (Mean: 0.78, SD: 0.11). These results demonstrate that providing the operator with requeueing choices improved their accuracy in a statistically significant manner. Practically, an improvement of accuracy from 0.72 to 0.80 is significant for UAV operations, since it decreases the likelihood of collateral errors.

Next, we investigated the effect of operator search times for the tasks. Operator search times for repeated looks in the RWC and RWOC modes were aggregated together to ensure a fair comparison to the NR mode and are shown in Fig. 9, and hence are representative of the total time spent searching the images. It can be seen that mean search time decreased from 21.0 seconds (with NR) to 19.1 seconds (with RWC), for the case when a timer was used (left). Further, when no timer was used (right), mean search time decreased from 22.5 seconds (with NR) to 19.4 seconds (with RWOC). For overall mean search time ($\alpha = 0.05$), mode is significant: $F(2,62) = 7.032, p = 0.002$. The timer effect and interaction between mode and timer were not significant.

Posthoc comparisons using a Tukey HSD test indicated that the change in the mean search time in the NR mode (Mean: 21.82, SD: 4.38) was significantly different ($p = 0.02$) from the RWC mode (Mean: 19.79, SD: 3.49). The change in the

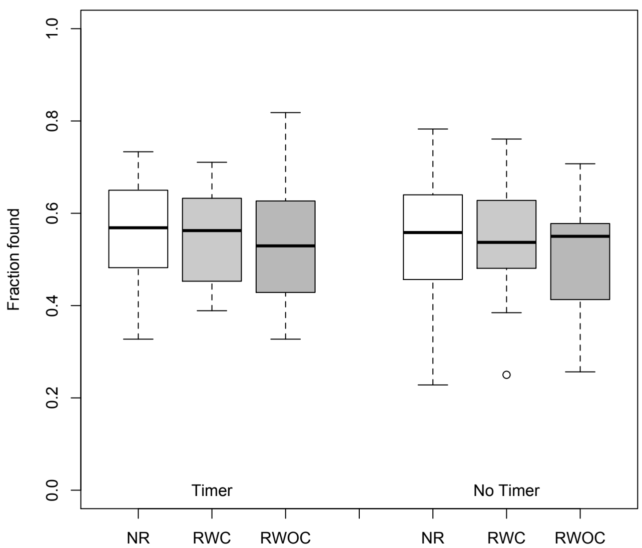


Fig. 10. Downward trend in fraction found is not statistically significant

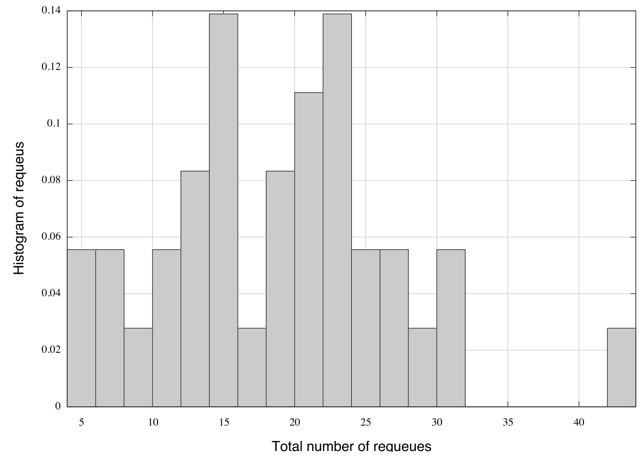


Fig. 11. Total number of requeues made by the 36 participants in the RWOC and RWC modes. Note the outlier with a total of 44 requeues.

mean search time in the NR mode (Mean: 21.82, SD: 4.38) was also significantly different ($p < 0.001$) from the RWOC mode (Mean: 18.91, SD: 2.87). These results also may have practical multi-UAV significance, where improvements in the speed of detection of the location of an adversary can have dramatic consequence for mission success.

Lastly, the objective of this experiment was to maximize the fraction correct, that is the quotient of the total number of targets correctly found and the total number of targets possible. The results for the fraction correct for each requeue mode and timer condition are shown in Fig. 10. As anticipated by the DES of Section III, there appeared to be a decrease in fraction found as relooks are employed: the mean fraction found decreased from 0.55 (with NR) to 0.51 (with RWOC). This change in fraction correct was not statistically significant with respect to the mode ($F(2,62) = 1.254, p = 0.29$) or timer ($F(2,62) = 0.03, p = 0.974$). However, a Pearson correlation revealed that as the probability of a relook increased, the fractions of targets found decreased (r=-.71; p<0.001). This highlights the cost of such an action in that the more frequently operators elected to reinsert tasks in the queue, the less time they had to prosecute other targets.

B. Behavioral analysis of use of relooks and requeues

The next step in the analysis was quantifying the number of times that requeueing was actually implemented by participants (Fig. 11). A total of 697 requeues were made by the subjects. The 12 subjects that had the highest number of requeues accounted for 47.9% of the total requeues made by all subjects. Interestingly, the subject that requeued the most targets had previous actual UAV experience, and may have been more inclined to relook at the targets for operational considerations.

In the RWC mode, a total of 269 requeues were made, with 149 requeues occurring after the time limit $T_{rl} = 25$ seconds had expired. This means that 55.3% of the participants ignored the recommendation made by the decision support algorithm, and chose to continue searching the image. In turn, 44.7% of the participants anticipated the automation's prompt.

In the RWOC mode, a total of 428 requeues were made, with 274 requeues being implemented by the decision support algorithm because the participants searched longer than $T_{rl} =$ 25 seconds. Out of all the participants in the RWOC mode, 10 out of 36 people never requeued voluntarily (the algorithm requeued the tasks for them all the time). Of the people that did requeue voluntarily at least once, they did so, on average, 41.6% of the time.

Target relooks occurred less frequently than requeues, and interestingly, there was a statistically significant difference between the probability of requesting a relook between the RWC and RWOC modes, $F(1,35) = 13.46,\ p < 0.001$. Out of all visual search tasks, a total of $N = 3499$ were analyzed in the first look, $N = 311$ were analyzed in the second look, $N = 72$ in the third look, $N = 17$ in the fourth look, and only $N = 2$ in the fifth look. Note that the first look takes into account tasks that were searched successfully, but also tasks that were skipped (either with or without consent). In the first searches, a total of $N = 2978$ target detections were made (either correct or incorrect), with a total of $N = 524$ requeues made after having investigated the targets. Out of the total looks that were made, only 17.8% of the tasks were looked at more than once. The discrepancy between the number of requeues and the number of looks is attributed to the fact that some targets were effectively skipped and never relooked again. This implies that operator were using the requeueing function predominantly to move on from difficult images.

To investigate further the existence of a benefit to relooking at targets later in the mission, we show the mean search times for correct detections and probability of errors associated with the different number of looks in Fig. 12. Note that the mean search time for correct detection slightly decreased from 17 seconds to 15 seconds with an increased number of relooks. This trend was also visible for targets that were incorrectly detected. However, the overall probability of error increased as a function of number of looks: from 22.7% in the first look,

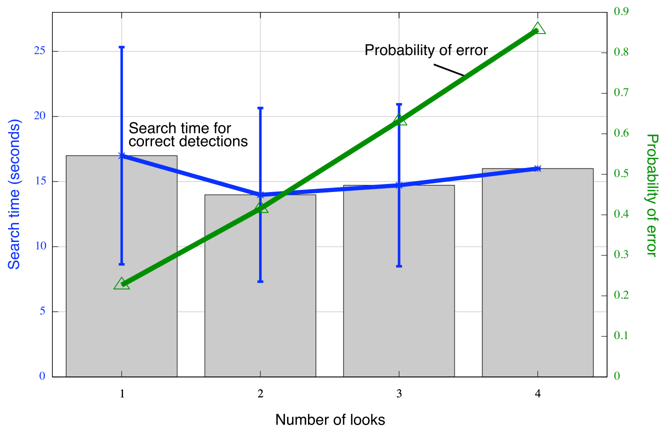


Fig. 12. Mean search time for targets that were correctly detected decreases with additional relooks. However, the probability of error increases, suggesting the subjects were not receiving additional information with each relook.

to 41.5% in the second look, 63.2% in the third look, and only 85.7% in the last look.

A logistic regression model was generated by using the number of relooks as a categorical variable, and the overall effect of the number of relooks was statistically significant with a χ^2 test ($\chi^2 = 64.5, df = 4, p < 0.001$). The difference in the coefficients of the logistic regression model from one look to two looks was statistically significant ($\chi^2 = 5.7, df = 1, p = 0.02$), but the difference from 2 to 3 looks and 3 to 4 looks was not significant. Nonetheless, this apparent increase in error, coupled with the fact that only 17.8% of targets were looked again two or more times, hints further at the possibility that the operators that were being presented with the same imagery later in the mission were pressured to make an assessment, and frequently made this assessment erroneously. Additional discussion on this observed effect is presented in Section VII.

Increased error with relooks of the same image has profound ramifications for supervisory control of UAV missions, because it suggests that operators may be willing to make a mistake to avoid repeating the same searches, and it is therefore necessary to create decision support systems that can avoid this disastrous consequence. Further, it also suggests that the perceived benefit of the requeuing methodology is the freedom to keep exploring other tasks, rather than being forced to make a choice on a difficult image. (Recall that targets that were requeued by the operators were enclosed by an orange circle, and were clearly identified to the operator.)

C. Subjective assessment of confidence and workload

Subjective assessment of the different requeue models is important, since the operator must ultimately accept or reject the recommendations set forth by such automated decision support systems. In developing decision support systems for complex mission planning involving the visual search task, two key subjective assessments are confidence and workload.

The subjective confidence assessment was a reflection of how accurately people felt that their performance was in detecting the targets in the image. Upon completion of a search task, participants were asked about their subjective

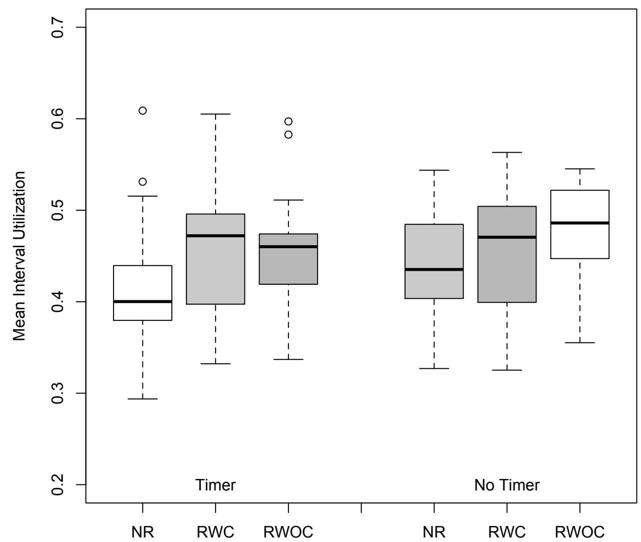


Fig. 13. Interval utilization increases with RWC and RWOC

confidence on the accuracy of their detection. A 5-pt Likert scale was used, where 1 indicated "Not very confident" and a 5 indicated "Very confident". The probability of detection was averaged for each condition the participant performed, and a correlation between confidence and average probability detection yielded a Pearson correlation coefficient of $r = 0.42$ (p <0.001), demonstrating that participant confidence was significantly correlated with their performance.

The three different modalities showed a statistically significant difference in confidence, at the $\alpha = 0.05$ level ($F(2,62) = 4.39,\ p = 0.02$). However, neither the interaction between mode and timer were significant. Posthoc comparisons using a Tukey HSD test indicated that the change in the mean of the self-assessed confidence in the NR mode (Mean: 3.79, SD: 0.59) was significantly different ($p = 0.01$) from the RWC mode (Mean: 3.95, SD: 0.48), indicating that operators felt that they performed better in the RWC mode.

Workload was measured in two distinct ways: in the first subjective method, participants were asked to rate their own level of workload. For the self-assessed workload and at the end of each search task, participants could enter a 1 (Not very loaded) to 5 (very loaded) on a 5-pt Likert scale. Changes in self-assessed workload across mode were not statistically significant at $\alpha = 0.05$ level $F(2,62) = 1.92, p = 0.15$. In addition, interaction between mode and timer was not significant ($F(2,62) = 1.10, p = 0.34$), and neither was timer treatment significant, $F(1,31) = 2.10, p = 0.16$. In the second method, interval utilization (or percent busy time) was used to gauge the objective workload in the experiment [41, 42]. Interval utilization has been shown in previous work to be a reliable assessment for workload [6]. For the mean interval utilization (Fig. 13), participant interaction time with the user interface was normalized by the time between detection tasks, and averaged for each participant. Increase in mean utilization across mode was statistically significant at $\alpha = 0.05$ level ($F(2,62) = 6.49, p = 0.003$). Interaction between mode and

TABLE I
FRACTION FOUND DES-REQM COMPARISON TO EXPERIMENT

	DES-ReQM		Exp		p-value
	μ	σ	μ	σ	
NR	0.58	0.14	0.54	0.14	p=0.16
RWC	0.53	0.16	0.53	0.12	p=0.61
RWOC	0.55	0.15	0.51	0.13	p=0.23

timer was not significant: $F(2,62) = 0.57, p = 0.57$) and timer treatment was also not significant, $F(1,31) = 0.55, p = 0.47$. While the differences in participant interaction time were statistically significant, the small practical difference in percent utilization (from 43% in the NR mode to 46% in the RWC mode) may have made this difference imperceptible to the subjects, and hence may be a possible reason for the lack of statistical significance in the subjective assessment.

Posthoc comparisons using a Tukey HSD test indicated that the change in the mean of the interval utilization in the NR mode (Mean: 0.43, SD: 0.07) was significantly different ($p = 0.02$) from the RWC mode (Mean: 0.46, SD: 0.07). Also, change in the mean interval utilization in the NR mode (Mean: 0.43, SD: 0.07) and RWOC mode (Mean: 0.46, SD: 0.07) were also significant ($p = 0.002$). While interval utilization increase is expected in the RWC and RWOC modes since the operators had an additional tool to use, utilizations on the order of 40-70% are still within the acceptable range for human supervisory control [43].

D. Comparison to DES model

Finally, we validate the predictions made by DES-ReQM by comparing the results obtained in the DES simulation and the experiment. Table I presents results comparing the use of requeues in terms of fraction found metric as predicted by DES-ReQM and obtained experimentally. The first principal column shows the mode condition (NR, RWC, and RWOC), the second principal column shows the mean and standard deviation of the fraction found predicted by DES-ReQM, while the third principal column shows the experimental results. Note that in order to compare the RWC experimental and simulated condition, we had to first determine empirical average probability of requeueing which we found to be $p = 0.38$. Non-parametric Mann-Whitney U-tests showed good agreement between the predictions made by the DES-ReQM model and the experiment for all the three conditions: $z =, 1.41, p = 0.16$ for NR, $z = -0.51, p = 0.11$ for RWC, and $z = 1.21, p = 0.23$ for RWOC.

VII. DISCUSSION ON THE USE OF REQUEUES

In this experiment, the probability of detection and mean search time improved with the presence of the requeue option, whether mandated or not. The cost of such requeues meant somewhat increased objective workload (but no increase in subjective workload), and the more participants accessed the relook feature, they were likely to find fewer targets. Such results highlight the cost-benefit issues surrounding any new decision support tool in that it can often provide benefit, but there are also possible negative consequences if such a tool is invoked too often.

One of the interesting results from this experiment was that detection probability increases as subjects are provided with the capability to requeue. However, relooks actually increases the probability of making an error. This seemingly counterintuitive result arises from the total number of targets that were searched in the NR mode compared to the RWC and RWOC conditions. In fact, subjects in the NR mode made an average of 32.6 total searches, while 31.2 and 30.8 were made in the RWC and RWOC modes. However, in the RWC mode, only 25.8 and 22.0 total decisions were made, and of these decisions, the respective error probabilities were 19.2% and 19.1% for for RWC and RWOC. In contrast, the NR mode had an error rate of 27.8%, suggesting that the benefit of the relook was to effectively allow people to skip difficult targets. Nonetheless, it appears that people felt implicit pressure to make a decision for repeated targets. Thus this research suggests that allowing operators to requeue (i.e., skip) a target, but not relook at it, may be a more effective strategy. This results also has broader implications for teaming of operators as it may be advantageous to requeue skipped targets for other teammates to avoid the increased likelihood of a mistake.

VIII. CONCLUSION AND FUTURE WORK

This article has developed a choice model for an operator performing visual search tasks generated from multiple unmanned vehicles. Using previous experimental data that demonstrated that human search accuracy could decay with time, we have developed a novel retrial queueing model of an operator that provides the operator with an additional choice by allowing the operator to requeue challenging targets.

A human-in-the-loop experiment was performed under different requeue conditions, and showed that the use of requeues increases overall probability of detection and operator confidence, ultimately improving operator accuracy. However, the additional use of requeues increased operator interval utilization (percent busy time) with a slightly (but not significantly) decreasing effect on the total number of targets found.

These observations open up an interesting area of work that should seek to understand the value of information of an image, and under what circumstances an operator may require additional imagery to reach a conclusion. Another important conclusion of this paper is that it has shown that supplying the operators with one additional choice of requeueing, rather than constraining them to a forced choice context, improves accuracy and confidence. This has important ramifications, not just for the external validity of the 2-AC models, but also for practical considerations in actual missions, in designing decision support systems that can provide additional flexibility to stressed operators.

Future work will include developing “optimal” relook policies, understanding that, in practicality, generating satisficing parameters is more realistic since optimality may be difficult to quantify in dynamic, uncertain command and control settings [44]. Moreover, additional work is needed to more fully understand the information processing ramifications of relooks as it is not clear whether the success of the relook mechanism is due to scene complexity, a possible attention filtering bias, or that operators had more confidence knowing they had such a tool available. Such understanding could possibly lead to identification of images in advance that could cause operators difficulty, possibly allowing them to be inserted in the queue at a more opportune time.

An additional consideration would be to quantify what kind of additional imagery information would be desired by an operator to increase the likelihood of detection in the event of a relook. Finally, a tighter coupling between the role of requeueing and the mission parameters needs to be made. For example, it will be beneficial to understand precisely what the role of vehicle routing is for the purposes of aiding the relook tasks (e.g., with different path planners), as well as the number and heterogeneity of UAVs.

References

[1] C. Drew, "Military Is Awash in Data From Drones." http://www.nytimes.com/2010/01/11/business/11drone.html.

[2] J. Clapper, J. Young, J. Cartwright, and J. Grimes, "Unmanned Systems Roadmap 2007-2032," 2007.

[3] M. Hoffman, "New Reaper sensors offer a bigger picture." http://www.airforcetimes.com/news/2009/02/airforce_WAAS_021609/.

[4] M. Cummings and P. Mitchell, "Predicting controller capacity in remote supervision of multiple unmanned vehicles," *IEEE Systems, Man, and Cybernetics, Part A Systems and Humans*, vol. 38, no. 2, pp. 451–460, 2008.

[5] D. K. Schmidt, "A queuing analysis of the air traffic controller's workload," in *IEEE Systems, Man, and Cybernetics*, vol. 8, pp. 492–498, 1978.

[6] C. Nehme, *Modeling Human Supervisory Control In Heterogeneous Unmanned Vehicle Systems*. PhD thesis, MIT, 2008.

[7] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The Physics of Optimal Decision making: A Formal Analysis of Models of Performance in Two-Alternative Forced Choice Tasks," in *Psychological Review*, vol. 113, pp. 700–765, 2006.

[8] J. R. Busemeyer and A. Diederich, "Survey of Decision Field Theory," in *Mathematical Social Sciences*, vol. 43, pp. 345–370, 2002.

[9] J. R. Busemeyer and J. T. Townsend, "Fundamental Derivations From Decision Field Theory," in *Mathematical Social Sciences*, vol. 23, pp. 255–282, 2000.

[10] A. Diederich and J. R. Busemeyer, "Simple Matrix Methods for Analyzing Diffusion models of Choice Probability, Choice Response Time, and Simple Response Time," in *Journal of Mathematical Pschology*, vol. 47, pp. 304–322, 2003.

[11] R. J. Herrnstein, "Rational Choice Theory: Necessary but Not Sufficient," in *American Psychologist*, vol. 45, pp. 356–357, 1990.

[12] J. Wolfe, "Guided search 2.0 a revised model of visual search," *Psychonomic bulletin & review*, vol. 1, no. 2, pp. 202–238, 1994.

[13] T. Horowitz and J. Wolfe, "Visual search has no memory," *Nature*, vol. 394, no. 6693, pp. 575–577, 1998.

[14] J. Wolfe, "What can 1 million trials tell us about visual search?," *Psychological Science*, vol. 9, no. 1, pp. 33–39, 1998.

[15] L. Itti, C. Gold, and C. Koch, "Visual attention and target detection in cluttered natural scenes," *Optical Engineering*, vol. 40, pp. 1784–1793, 2001.

[16] M. Chun and M. Potter, "A two-stage model for multiple target detection in rapid serial visual presentation.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, no. 1, pp. 109–127, 1995.

[17] J. Irvine, "Assessing target search performance: The free-response operator characteristic model," *Optical Engineering*, vol. 43, pp. 2926–2934, 2004.

[18] M. Peterson, A. Kramer, R. Wang, D. Irwin, and J. McCarley, "Visual search has memory," *Psychological Science*, vol. 12, no. 4, pp. 287–292, 2001.

[19] J. McCarley, A. Kramer, C. Wickens, E. Vidoni, and W. Boot, "Visual skills in airport-security screening," *Psychological Science*, vol. 15, no. 5, pp. 302–306, 2004.

[20] D. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont: Athena Scientific, 2005.

[21] M. Pachter, P. R. Chandler, and S. Darbha, "Optimal sequential inspection," in *IEEE Conference on Decision and Control*, 2006.

[22] R. Swensson, "The elusive tradeoff: Speed vs accuracy in visual discrimination tasks," *Perception & Psychophysics*, vol. 12, no. 1-A, pp. 16–32, 1972.

[23] D. Davies and R. Parasuraman, *The Psychology of Vigilance*. London: Academic Press, 1982.

[24] G. I. Falin and J. G. C. Templeton, *Retrial Queues*. Chapman and Hall, 1990.

[25] J. Artalejo and A. Gomez-Corral, *Retrial Queueing Systems*. London: Springer, 2008.

[26] B. D. Choi and K. K. Park, "The M/G/1 Retrial Queue with Bernoulli Schedule," in *Queueing Systems*, vol. 7, pp. 219–228, 1990.

[27] N. Sherman and J. Kharoufeh, "An M/M/1 Retrial Queue with Unreliable Server," in *Operations Research Letters*, vol. 34, pp. 697–705, 2006.

[28] L. Bertuccelli, N. Pellegrino, and M. Cummings, "Choice Modeling of Relooks Tasks in UAV Search Missions," *IEEE American Control Conference*, 2010.

[29] A. Diederich, "Dynamic Stochastic Models for Decision Making under Time Constraints," in *Journal of Mathematical Pschology*, vol. 41, pp. 260–274, 1997.

[30] C. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. New York: Springer-Verlag, 1985.

[31] P. Verghese, "Visual Search and Attention:: A Signal Detection Theory Approach," *Neuron*, vol. 31, no. 4, pp. 523–535, 2001.

[32] M. King, G. Stanley, and G. Burrows, "Visual search processes in camouflage detection," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 26, no. 2, pp. 223–234, 1984.

[33] G. Waldman, J. Wootton, and G. Hobson, "Visual detection with search: an empirical model," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 21, no. 3, pp. 596–606, 1991.

[34] G. Sperling and M. Melchner, "The attention operating characteristic: Examples from visual search," *Science*, vol. 202, no. 4365, p. 315, 1978.

[35] L. Huang and H. Pashler, "Attention capacity and task difficulty in visual search," *Cognition*, vol. 94, no. 3, pp. B101–B111, 2005.

[36] J. Wolfe, E. Palmer, and T. Horowitz, "Reaction time distributions constrain models of visual search," *Vision research*, vol. 50, no. 14, pp. 1304–1311, 2010.

[37] E. Palmer, T. Horowitz, A. Torralba, and J. Wolfe, "What are the shapes of response time distributions in visual search?," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 1, pp. 58–71, 2011.

[38] M. Mouloua and R. Parasuraman, *Automation and Human Performance: Theory and Applications (Ch. 9)*. New Jersey: Lawrence Erlbaum Associates, 1996.

[39] K. Savla, T. Temple, and E. Frazzoli, "Human-in-the-loop vehicle routing policies for dynamic environments," in *IEEE Conference on Decision and Control*, 2008.

[40] A. Girard, S. Dharba, M. Pachter, and P. R. Chandler, "Stochastic Dynamic Programming for Uncertainty Handling in UAV Operations," in *IEEE American Control Conference*, 2007.

[41] W. B. Rouse, *Systems Engineering Models of Human-Machine Interaction*. New York: North Holland, 1983.

[42] M. Cummings and S. Guerlain, "Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles," in *Human Factors*, vol. 49, pp. 1–15, 2007.

[43] M. Cummings, A. Clare, and C. Hart, "The role of human-automation consensus in multiple unmanned vehicle scheduling," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 1, pp. 17–27, 2010.

[44] W. C. Stirling, *Satisficing Games and Decision Making: With Applications to Engineering and Computer Science*. New York, NY, USA: Cambridge University Press, 2003.