On-the-Fly Scheduling as a Manifestation of Partial-Order Planning and Dynamic Task Values

Samuel D. Hannah and Andrew Neal

Human Factors: The Journal of the Human Factors and Ergonomics Society 2014 56: 1093 originally published online 5 March 2014
DOI: 10.1177/0018720814525629

On-the-Fly Scheduling as a Manifestation of Partial-Order Planning and Dynamic Task Values

Samuel D. Hannah, University of Saskatchewan, Saskatoon, Canada, and **Andrew Neal**, University of Queensland, Brisbane, Australia

Objective: The aim of this study was to develop a computational account of the spontaneous task ordering that occurs within jobs as work unfolds ("on-the-fly task scheduling").

Background: Air traffic control is an example of work in which operators have to schedule their tasks as a partially predictable work flow emerges. To date, little attention has been paid to such on-the-fly scheduling situations.

Method: We present a series of discrete-event models fit to conflict resolution decision data collected from experienced controllers operating in a high-fidelity simulation.

Results: Our simulations reveal air traffic controllers' scheduling decisions as examples of the partial-order planning approach of Hayes-Roth and Hayes-Roth. The most successful model uses opportunistic first-come-first-served scheduling to select tasks from a queue. Tasks with short deadlines are executed immediately. Tasks with long deadlines are evaluated to assess whether they need to be executed immediately or deferred.

Conclusion: On-the-fly task scheduling is computationally tractable despite its surface complexity and understandable as an example of both the partial-order planning strategy and the dynamic-value approach to prioritization.

Keywords: decision making, scheduling, planning, air traffic control, temporal discounting

Address correspondence to Samuel D. Hannah, Department of Psychology, University of Saskatchewan, Arts 151, 9 Campus Drive, Saskatoon, Saskatchewan, S7N 5A5, Canada; e-mail: hannah.sam@gmail.com.

HUMAN FACTORS
Vol. 56, No. 6, September 2014, pp. 1093–1112
DOI: 10.1177/0018720814525629

INTRODUCTION

A common decision-making problem in most jobs concerns the need to schedule tasks. This problem is particularly acute for jobs in which the flow of tasks is only partially predictable. Paramedic response crews, for example, do not know when the next call will come in, what the problem will be, or how urgent the job will be, and thus what deadline will accompany it. Such unpredictable demands and deadlines are typically mixed in with a number of regular tasks with known deadlines—equipment inventory must be taken at regular intervals, supplies ordered within administrative deadlines, monthly work schedules submitted, and so on. The unpredictable nature of work flow in these environments, however, means that the scheduling of even these stable tasks must be done provisionally and often on an ad hoc basis. Such on-the-fly task scheduling—the scheduling of tasks in the course of conducting other tasks—thus poses an interesting set of cognitive demands, the mastery of which is a necessity for developing expertise within dynamic and uncertain work environments, particularly when human life is at risk. The human factors literature, however, has relatively little to say regarding the psychological mechanisms responsible for task scheduling, and to the best of our knowledge, formal psychological models of this process have not yet been developed.

In this paper, we develop a series of formal models that describe some of the mechanisms by which people may carry out on-the-fly task scheduling. We examine whether these models are capable of explaining the scheduling behavior of experts performing a complex task, with and without the assistance of automation. The specific context in which we test the models is air traffic control. We fit the model to data generated by two expert air traffic controllers performing a series of demanding exercises in a training simulator. The air traffic controllers performed the exercises with and without the assistance of an automated conflict detection tool. The simulations demonstrate that the order and timing of the controllers' actions can be accounted for by a model that incorporates dynamic task values (e.g., Vancouver, Weinhardt, & Schmidt, 2010) and what Ratterman, Spector, Grafman, Levin, and Harward (2001) called "partial-order" planning (e.g., Hayes-Roth & Hayes-Roth, 1979).

Despite the small number of participants generating the original data, the value of this work lies as a proof of principle. It is proof that despite the complex, dynamic, and uncertain environment in which people perform on-the-fly task scheduling, the issue is computationally tractable. There has been very little progress made on the study of the way that humans schedule tasks in the past 25 years, at least in part because it is seen as a complex and intractable problem. Scheduling is seen as an engineering problem rather than a psychological problem. The work presented in the current paper demonstrates not only that this process can be formally defined psychologically, but that it can be done so using relatively simple models. As we discuss at greater length in the General Discussion, the strategy of testing a formal model on a small number of participants is well established in cognitive science, especially when the work represents an initial exploration of a phenomenon.

In the next sections, we first explain why we chose air traffic control as an environment to study on-the-fly task scheduling, then describe the existing state of psychological research on task scheduling, and develop a series of formal models. We then report the results of a series of simulations designed to test those models and discuss the implications.

Air Traffic Control

Air traffic control is a useful environment to study on-the-fly scheduling, because air traffic controllers need to manage conflicting deadlines in a dynamic and uncertain environment (Durso & Manning, 2008). En route air traffic controllers are responsible for establishing the safe,

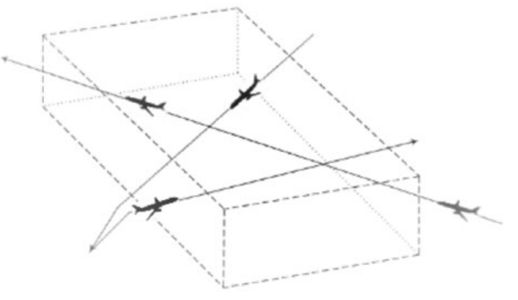

Figure 1. Aircraft approaching, flying through, and exiting an en route sector.

orderly, and expeditious flow of traffic through their sector. A sector is a three-dimensional volume of airspace with a boundary around it. For any given aircraft, the typical sequence of activity starts as the aircraft approaches the sector (see Figure 1). The controller will check the trajectory to establish that the aircraft will be legally separated from other aircraft before it enters the sector. Depending on the technology that is in place, this check may be done manually by the controller or may be done with the assistance of automation (Rovira & Parasuraman, 2010). When the aircraft enters the sector, the controller and pilot complete a handover sequence. Once the aircraft is under jurisdiction, the controller will periodically check the trajectory of the aircraft, perform routine data management tasks, and respond to problems or issues as they arise. As the aircraft leaves the sector, the controller may coordinate the handover of the aircraft to the downstream sector. A controller may have up to 10 or 20 aircraft at various stages in this sequence at any point in time, which means that there can be a large number of tasks that have to be completed within a given time interval. One of the most challenging aspects of the controller's job is ensuring that all tasks are completed by their respective deadlines.

The scheduling problem is complex because tasks can vary in importance and deadline (Averty, Collet, Dittmar, Vernet-Maury, & Athenes, 2004). Separation assurance is the most important task that a controller has to perform. If the controller cannot be assured that an aircraft will meet the applicable separation standards as it flies through the sector, then the controller needs to intervene (Loft, Bolland, Humphreys, & Neal, 2009). A controller typically intervenes by changing the route or level of one of the aircraft. However, the controller may not have to act immediately.

The study reported in the current paper was conducted in an oceanic sector, in which controllers are often able to detect potential breakdowns of separation between 1 and 2 hr in advance. In this situation, there is a cost associated with intervening too early. Changes to the trajectories of aircraft are disruptive because they can create workload for pilots and controllers and increase fuel burn. Furthermore, an intervention on one aircraft may make the resolution of subsequent separation problems more difficult because it limits the degrees of freedom available to the controller. Finally, given the inherent uncertainties associated with flight, many separation problems resolve themselves without the need for intervention. There is, therefore, a trade-off between acting early to assure separation and delaying intervention to reduce uncertainty and thereby minimize workload and disruption (Averty et al., 2004). These decisions need to be made in the context of more immediate demands, such as the need to respond to pilot requests and to complete outstanding handovers and data management tasks.

An additional factor that needs to be taken into account when scheduling tasks is the need for flexibility. Unexpected events can occur that disrupt existing plans. For example, a wave of incoming traffic may cause a surge in workload, which means that it is no longer safe to delay intervention. Alternatively, additional aircraft may appear that block the expected solution to the problem or create new problems that need to be solved urgently. For these reasons, the scheduling strategies that controllers use have to be sufficiently flexible to accommodate unexpected changes in the environment.

Task Scheduling

The cognitive literature has little to say about on-the-fly task scheduling. Although there is a large literature on task switching (e.g., Rogers & Monsell, 1995; Wylie & Allport, 2000; see Kiesel et al., 2010, for a recent review), voluntary switching has only recently become a recent topic of investigation (e.g., Arrington, 2008; Arrington & Logan, 2004; Arrington, Weaver, & Parker, 2010; Mayr & Bell, 2006). In voluntary task-switching experiments, however, participants are typically allowed to switch only between two tasks, usually with the constraint that they maintain an approximately even split across the two tasks. No rewards or penalties are provided for satisfactory or unsatisfactory performance, and tasks do not vary in either difficulty or deadlines. The organization of tasks is not the subject of attention so much as the costs of switching and the mechanisms underlying a switch decision.

There is, however, a body of scheduling research within the human factors literature (e.g., Dessouky, Moray, & Kijowski, 1995; Moray, Dessouky, Kijowski, & Adapathya, 1991). Most of the research on human scheduling has been carried out within manufacturing, where workers make decisions regarding the allocation of work to machines (see Sanderson, 1989, for a review). This body of research suggests that a wide range of scheduling rules can be used in different settings. For example, workers can use simple priority rules, such as "first-come-first-served," "earliest due date," or "highest dollar cost" (Sanderson, 1989). First-come-first-served involves performing tasks or operations in the order they arrive, whereas earliest due date involves performing tasks or operations in the order that they must be completed. Highest dollar cost involves performing tasks in order of importance. The advantage of simple priority rules is that they require relatively little in the way of cognitive effort to apply. However, workers may also use more complex rules involving a number of simple priority rules or a weighted combination of different factors (Sanderson, 1989). Although these strategies are more complex and require more cognitive effort to apply, they are also more flexible, because they take into account a broader range of considerations.

Although the scheduling literature may provide insights into the different strategies that workers might use when making scheduling decisions, the research is mostly descriptive and says little regarding the psychological mechanisms responsible for scheduling (Fransoo & Wiers, 2006; Sanderson, 1989). Furthermore, the scheduling literature largely deals with manufacturing and other systems in which the timing of events is predictable. Air traffic controllers, by contrast, need to be able to change the order in which they do things as events unfold, and they need to ensure that prior decisions do not unnecessarily constrain future choices.

The planning research inspired by Hayes-Roth and Hayes-Roth (1979) may provide insight into the strategies that people use in a more dynamic and uncertain environment. Hayes-Roth and Hayes-Roth used errand running as their paradigm. They found that their participants were highly opportunistic. Their participants typically created a general plan consisting of a series of clusters of errands needing to be connected, which they did as they went along, reorganizing the plan to accommodate unexpected opportunities or problems offered by the environment. This so-called partial-order planning approach has been popular in route-planning research (Ettema, Borgers, & Timmermans, 1993; Gärling, Säisä, Böök, & Lindberg, 1986), and Fransoo and Wiers (2006) have applied it to an order-filling problem. Partial-order planning is appropriate in a dynamic and uncertain environment, such as air traffic control, because it minimizes cognitive effort while also providing the flexibility to respond to opportunities as they occur.

Partial-Order Planning in Air Traffic Control

In the current section, we present a series of models of on-the-fly task scheduling based on the partial-order planning approach, featuring a mix of simple priority rules and more flexible, but complex, strategies. These models all share the assumption that outstanding tasks are stored in a queue—in other words, that controllers have a memory. They assume that the order in which tasks are done is an emergent property of the way that the queue is managed, that is, of the moment-to-moment decisions that are taken as events unfold dynamically over time. The models vary on two dimensions, namely, how tasks are selected from the queue and whether tasks can be deferred once they are selected.

Selection from the queue. Turning to the first dimension, we assume that controllers select the task with the greatest value to act on. However, there are a number of ways in which the value of a task may be assessed. The simplest approach is to use first-come-first-served. Under this approach, the tasks in the queue are ordered according to the time they arrive, and the value of task *i* at a given point in time (V_t) is equal to the order of the task in the queue (Q_i):

$$V_t = Q_i. \tag{1}$$

An alternative approach is to select the most important task from the queue. The importance of a task is the perceived impact of success or failure, regardless of context. For example, separation assurance is more important than transferring jurisdiction in air traffic control, all else being equal, because a violation of the separation standards has more serious consequences than a failure to complete a handover. Under this approach, the value of task *i* is equal to its importance (I_i):

$$V_t = I_i. \tag{2}$$

Both of these approaches assume that value is static. However, research in the motivation literature suggests that value may be dynamic (Kernan & Lord, 1990; Schmidt & DeShon, 2007; Vancouver et al. 2010). Prior research suggests at least two ways in which the value of completing a task may change over time. The first is in response to the level of workload that the person has to deal with, and the second is in response to the deadline for the task.

Raby and Wickens (1994) examined the effect of workload on task prioritization among pilots. Raby and Wickens had pilots perform a series of tasks in flight under different levels of workload. A group of subject matter experts rated the importance of the different tasks prior to the experiment. The key result was that pilots were more likely to carry out tasks in order of importance as workload increased. These results suggest that importance may interact with workload. According to this account, V_t is a multiplicative function of I_i and the current workload (W_t):

$$V_t = I_i \times W_t. \tag{3}$$

We define workload as a hypothetical construct that represents the level of demand that a person experiences while performing a task (Rouse, Edwards, & Hammer, 1993). Following Rouse et al. (1993), we assume that workload is a dynamic variable that is subject to inertia, meaning an individual's perception of workload is in part a function of its previous level; that is, workload is characterized by "hysteresis" (Morgan & Hancock, 2011). We also assume such inertia is asymmetrical, such that workload drops more slowly than it rises; this assumption is broadly consistent with Morgan and Hancock's (2011) findings. The perceived workload at a given point in time is, therefore, the sum of two components, the momentary workload (W_M) and a dynamic baseline (W_B). The momentary workload (W_M) represents the current level of demand. In contrast, the dynamic baseline (W_B) represents the residual effect of past demands. Because it is dynamic, the baseline responds at a certain rate to changes in demand from the previous point in time to the current point in time, capturing the inertial properties of workload. Together, the momentary and baseline components combine to determine the level of workload experienced at any given time t. More formally, $W_t = W_M + W_B$.

We assume that the momentary workload is determined by the number and importance of the tasks in the queue. This value is obtained by summing the importance of each task i of the j tasks in the queue ($W_M = \sum_{i=1}^{j} I_i$). This is, of course, a simplification, as there are many factors that contribute to the current level of demand that an air traffic controller experiences. However, the number and importance of tasks is a reasonable proxy for current purposes (Loft, Sanderson, Neal, & Mooij, 2007). The dynamic baseline depends upon the previous time point's baseline level ($W_{B[t-1]}$) and the change between the previous actual workload (W_{t-1}) and the momentary workload. The change between the actual workload of the previous time point and the momentary workload (W_Δ) is scaled by multiplying it by a weight (s) between zero and one, where s represents the sensitivity to changes in workload:

$$W_\Delta = s(W_M - W_{t-1}). \quad (4)$$

Equation 4's difference operation yields a positively signed difference if workload is increasing and a negatively signed difference if it is decreasing. The value of s in Equation 4 is smaller for a negatively signed difference (s^-) than for a positively signed difference (s^+), so that workload declines are smaller than workload gains—implementing asymmetric inertia. That is,

$$\begin{gathered} \text{if } W_M - W_{t-1} < 0, \text{ then } s = s^-, \\ \text{if } W_M - W_{t-1} > 0, \text{ then } s = s^+, \\ 0 < s^- < s^+ < 1. \end{gathered} \quad (5)$$

The scaled difference is added to the previous value of the baseline to give the current value of the baseline at time t,

$$W_{B(t)} = W_{B(t-1)} + W_\Delta. \quad (6)$$

The updated baseline is then added to the momentary workload to produce the full equation defining workload at time t:

$$W_t = W_M + [W_{B(t-1)} + s(W_M - W_{t-1})]. \quad (7)$$

An alternative way to conceptualize the selection of tasks from a queue is in terms of temporal discounting. Temporal discounting refers to the robust finding that the subjective values of rewards and penalties reduce as the delay to obtain them increases (e.g., Caruso, Gilbert, & Wilson, 2008; Chapman, 1996; Grace, 1999; Mazur & Biondi, 2009; Myerson & Green, 1995; Rachlin & Green, 1972; Rachlin, Raineri, & Cross, 1991). Steel and König (2006) have recently produced an account of choice decision making that incorporates temporal discounting. Temporal discounting allows tasks to be revalued based on the time remaining to deadline—the further in time a task can be deferred, the lower the task's perceived failure cost.

Most models of temporal discounting use a hyperbolic discounting function (Green, Fry, & Myerson, 1994; Myerson & Green, 1995; Simpson & Vuchinich, 2000). Under hyperbolic discounting (we are here collapsing several variants of the basic hyperbolic discounting function, some of which add more parameters to the basic function used in this report), V_t for task i is a function of the importance and the reciprocal of the time to deadline for the task (D_i), weighted by a subject-dependent weight (k):

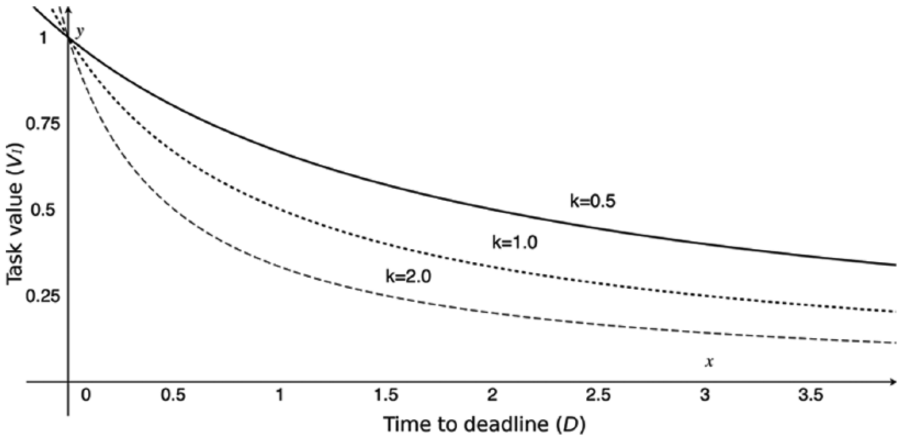

Figure 2. Changes in task value (V_i) with increases in time to deadline (D) as a function of hyperbolic discounting for different values of k.

$$V_t = \frac{I_i}{1 + kD_i} \quad (8)$$

The k parameter scales the sensitivity to discounting: The larger k is, the steeper and less linear is the discounting (see Figure 2), so that higher values of k lead to a greater willingness to ignore tasks with long deadlines. Incorporating hyperbolic discounting into the assessment of value ensures that tasks with shorter deadlines tend to get selected before tasks with longer deadlines.

Deferral. Turning to the second dimension, the simplest way to schedule a set of tasks is to act on each task immediately, once it is selected from the queue. A worker using this strategy will complete all tasks in the queue in order of priority, until there is nothing left in the queue. If there is a task with a long deadline, then it will be completed once it reaches the top of the queue, regardless of how much time remains before it has to be completed. A more flexible approach is to defer tasks with long deadlines. As noted earlier, air traffic controllers in oceanic sectors may detect conflicts up to 2 hr in advance. Deferring intervention can be beneficial, because it gives the controller more flexibility. The controller can reassess the conflict in the future, when there is less uncertainty regarding the trajectories of the aircraft, and the controller can assess how those aircraft interact with other traffic in the area. The models that incorporate deferral add an additional step after a conflict is selected from the queue. These models assume that when controllers select a conflict from the queue, they assess the value of resolving that conflict (V_R) at the current point in time.

There are a number of ways that the value of resolving a conflict immediately may be assessed. The simplest model assumes that the value of resolving a conflict immediately is a function of the amount of time that has elapsed since the conflict was detected (termed "conflict duration" [CD]), and airspace complexity. Both of these factors are relevant, because resolving outstanding conflicts immediately eliminates the need for ongoing monitoring and helps to reduce complexity (Loft et al., 2007). A commonly used measure of complexity is the number of conflicts in a given region of airspace (Durso & Manning, 2008; Loft et al., 2007). We approximate complexity, therefore, by using the mean number of conflicts that each pair member has been in ($\bar{C}$), including the current conflict. The basic algorithm for calculating the value of resolving conflict i immediately is given by

$$V_{\mathrm{R}(i)} = I_R \times CD_i \times \bar{C}_i, \quad (9)$$

where I_R is the importance of resolving a conflict and, thus, like all forms of task importance in our approach, is a constant (see the discussion introducing Equation 2).

The first alternative to the model described in Equation 9 incorporates workload. Controllers report that they prefer to intervene early when workload is heavy, because it minimizes the risk of a breakdown in separation and helps to manage their workload (Kallus, Van Damme, & Dittman, 1999; Loft et al., 2009). According to this model, the value of resolving a conflict immediately is

$$V_{\mathrm{R}(i)} = I_R \times CD_i \times \bar{C}_i \times W_t. \qquad (10)$$

The second alternative to the model described in Equation 9 incorporates temporal discounting. According to this account, the value of resolving a conflict immediately is dependent upon time to deadline and sensitivity to discounting:

$$V_{\mathrm{R}(i)} = \frac{I_R \times CD_i \times \overline{C_i}}{1 + kD_i}. \qquad (11)$$

The Data Set: Air Traffic Control Exercises

We tested the competing models described in the preceding section on four archival data sets derived from a series of exercises conducted in the training simulator at Brisbane Air Traffic Services Centre. The simulator is a fully functional air traffic management system. The hardware and software are identical to those used in the operations center, and the simulator consoles can provide operational air traffic control on an emergency basis.

The exercises were conducted in an oceanic sector, without radar coverage. Approximately 40% of aircraft were equipped with technology enabling satellite-based surveillance (Automatic Dependent Surveillance, Broadcast [ADS-B]). ADS-B enables the position and level of the aircraft to be displayed on the controller's air situation display with a high level of accuracy. Aircraft without ADS-B are required to provide position reports to controllers as they cross waypoints. Controllers enter these position reports into an electronic flight data record. The system extrapolates the likely position and level of the aircraft from the information contained in the flight data record.

Two licensed air traffic controllers (1 male and 1 female, ages 40 and 43, respectively, with 6 and 12 years experience, respectively) completed the exercises, both of whom held an endorsement for the sector being simulated. Each controller completed one exercise with the assistance of an automated conflict detection tool and one exercise without the assistance of the automated conflict detection tool. The automated conflict detection tool was new for controllers, with each controller having received 5 days training in the use of the tool prior to the exercises. Each exercise lasted 2 hr; workload peaked twice during each exercise. The first peak occurred approximately 30 min into the exercise, as a wave of inbound traffic entered the sector from the east. The second peak occurred approximately 1 hr later when a wave of inbound traffic entered the sector from the west, crossing the earlier wave of traffic from the east, which was now approaching the western boundary of the sector. Subject matter experts were consulted with regard to timing and all other exercise details.

We used electronic data records from the simulator to reconstruct the timing of the appearance of aircraft in each exercise, along with deadlines for aircraft entering or leaving the sector, conflict start times, and resolution deadlines. A conflict resolution deadline is the time at which a pair of aircraft would breach the applicable separation standard if no resolution were implemented. We coded the timing of controllers' resolutions by viewing video of the simulation sessions and identifying the point when they intervened to assure separation by changing the route or level of an aircraft or by placing a requirement on an aircraft.

Model fitting: Data and methods. Models were fit separately to each exercise for each controller. The number of conflicts varied from 12 to 22 across the four sets. Our main goal was to model the timing of conflict resolutions, but we also had information about the order of resolutions. To constrain the model fitting to maximize the generalizability of parameter estimates and model behavior, we found parameter estimates that optimized the fit to both the timing and order information. Models containing no more than one free parameter are thus fit to data consisting of 24 to 44 data points generated by individual subjects, as each conflict has one timing and one order data point.

Given that there are two data points for each conflict, overall goodness-of-fit estimates are based on a mixture of two different measures, timing and order, and thus not readily interpretable. After finding parameters maximizing overall fit, we separately calculated goodness-of-fit measures for timing and order data; both timing and order fits are reported, along with parameter estimates when models contain free parameters. We evaluated goodness of fit using both the root mean square deviation (RMSD) and r^2, where r^2 is calculated as $1.0 - (SS_{\text{error}} / SS_{\text{total}}$; SS = sum of squares), rather than the squared Pearson correlation coefficient. (Although both forms of r^2 are usually very close, r^2 based on the *SS* ratio is slightly more conservative, being sensitive not only to differences in the relations between two sets of values but also to the absolute offset between them.)

We used the Nelder-Mead simplex algorithm (Nelder & Mead, 1965) to find optimal parameter estimates, using the initialization and fitting algorithms found in Press, Teukolsky, Vetterling, and Flannery (1992). The Nelder-Mead simplex is a widely used algorithm that finds parameters minimizing some function. When the function is an error function, such as RMSD, the Nelder-Mead simplex usually finds parameters that yield the best fit.

SECTION 1: SELECTION FROM THE QUEUE

In the first section, we tested four models that describe different ways of selecting tasks from the queue. All of these models assume that the controller acts on each task immediately, once it is selected from the queue. In subsequent sections, we allow conflict resolution to be deferred. We tested competing models that describe different ways in which the controller may make this deferral decision and assess whether these models produce a better fit to the data than the models that do not allow deferral. In the final section, we combine the features of all models. (Fortran code for the models described in this paper can be obtained from the first author upon request.)

Discrete-Event Simulation Details

We implemented all models as discrete-event simulations (e.g., Wainer, 2009). Discrete-event simulations define a system as a sequence of chronological actions or events performed on some entities, each of which takes a certain amount of time to complete. Our simulations describe the way that tasks are added and removed from a queue, and executed. At the beginning of each cycle, the model consults the task queue. If the queue is empty, the model will scan the airspace to identify any new tasks that need to be performed. If the queue is not empty, the model selects the task with the greatest value to act on, using one of the algorithms described earlier (Equations 1 through 3 and 8). We assumed that task values are fuzzy and implement fuzziness by truncating task values prior to selection by rounding them down to the nearest integer or taking the floor of the functions in Equations 1 through 3 and 8. All models revert to first-come-first-served to resolve ties.

Completion times for each task are determined by randomly drawing from a Gaussian distribution defined by a mean and standard deviation (see Appendix A). The task is removed from the queue once it has been completed. Depending on the situation, a new task may be added to the queue at this point. The workload level is also updated, as are the deadlines for any outstanding tasks. The model then consults the task queue again and either scans the airspace or selects the next task to complete. The model continues in this manner until the end of the simulation (120 min).

Figure 3 illustrates how tasks are added and removed from the queue. The tasks are grouped into three broad activities: *scan airspace, check trajectories*, and *do task*. When the queue is empty, the model scans the airspace to identify new aircraft approaching the sector, aircraft that are entering the sector, aircraft that are passing waypoints, and aircraft that are leaving the sector. If a new aircraft is identified, a new task (*check-trajectory*) is added to the queue (see Task 1 in Figure 3). At some point, the controller may subsequently select this task from the queue and carry it out. This activity removes *check-trajectory* from the queue. If the check reveals that separation is not assured, a new task

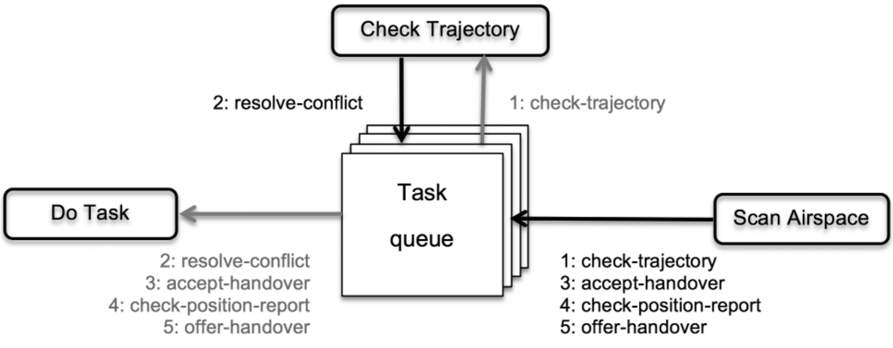

Figure 3. Addition and deletion of tasks from the task queue. Each activity can amend the task queue by writing a new task to it (black arrows) or by deleting an existing task (gray arrows).

(*resolve-conflict*) is added to the queue (see Task 2 in Figure 3).

At some point, the aircraft will enter the sector, and a new task (*accept-handover*) will be added to the queue (see Task 3 in Figure 3). The controller may select this task from the queue and complete the handover sequence, removing *accept-handover* from the queue. If the aircraft is in conflict, the controller will act to resolve the conflict while the aircraft is flying through the sector. Doing so removes *resolve-conflict* from the queue.

As the aircraft flies through the sector, it will pass a series of waypoints before leaving the sector. When the aircraft passes a waypoint, a new task (*check-position-report*) is added to the queue (see Task 4 in Figure 3). At some point, the controller may select this task from the queue and check the position report. This activity removes *check-position-report* from the queue. As aircraft leaves the sector, another task (*offer-handover*) is added to the queue (see Task 5 in Figure 3). The controller may select this task from the queue and complete the transfer of jurisdiction to the next sector. This activity removes *offer-handover* from the queue.

The values for I for the different tasks are listed in Appendix A. We estimated I_i for each of i tasks by doing an initial fit of the data to each of our four data sets, treating I_i as a free parameter. We then averaged the resulting four sets of I_i, using these averaged estimates to fix I_i for all models.

The temporal discounting model assumes that task values change as a function of the time remaining until deadline for each task. The deadlines for *accept-handover, offer-handover* were set at 4 min. The deadline for *resolve-conflict* was the point at which the aircraft would violate the applicable separation standard. The deadline for *check-trajectory* is based on the deadline of the most immediate conflict. Aircraft that are not involved in conflicts are given a deadline of 180 min, as the simulation is only 120 min. For *check-position-report*, an arbitrary long deadline of 60 min is given when the task is initially entered into the task queue.

Results and Discussion

The first-come-first-served model accounted for roughly half of the variance in the timing of conflict resolutions for all of the data sets, as Table 1 reveals (see Model 1.1). None of the other models seemed to provide any substantial improvement in fit for the timing data. The only exception is the temporal discounting model's reproduction of Controller 1's data when the exercise was conducted without the assistance of automated conflict detection (see Model 1.4). The temporal discounting model produced a better fit for timing data, and a nearly identical fit to the ordinal data, in this exercise. All models fit Controller 2's timing data better than they fit Controller 1's timing data.

All models produced better fits for the exercise without the automated conflict detection tool. All of the models examined in this section assume that controllers act on tasks immediately upon selection from the queue. One of the benefits of automated conflict detection is that it provides the controller with a list of ongoing conflicts. This characteristic is likely to reduce the memory load for the controller and may

TABLE 1: Fit Statistics and Parameter Estimates Across the Four Exercises for the Models Describing Different Ways of Selecting Tasks From the Queue, Assuming Each Task Is Acted on Immediately

	Controller 1						Controller 2					
	Without Detection Tool			With Detection Tool			Without Detection Tool			With Detection Tool		
Model	Timing	Ordinal	k	Timing	Ordinal	k	Timing	Ordinal	k	Timing	Ordinal	k
1.1 First-come-first-served	0.554 (23.83)	0.873 (1.77)	NA	0.480 (27.70)	0.194 (5.66)	NA	0.743 (23.23)	0.710 (2.23)	NA	0.586 (29.98)	0.553 (2.37)	NA
1.2 Importance	0.556 (23.77)	0.815 (2.14)	NA	0.476 (27.80)	0.195 (5.65)	NA	0.729 (23.84)	0.711 (2.23)	NA	0.583 (30.11)	0.553 (2.37)	NA
1.3 Workload	0.585 (22.98)	0.866 (1.82)	NA	0.478 (27.75)	0.194 (5.66)	NA	0.728 (23.87)	0.710 (2.23)	NA	0.583 (30.10)	0.555 (2.37)	NA
1.4 Temporal discounting	0.682 (20.10)	0.870 (1.79)	0.001	0.483 (27.67)	0.301 (5.27)	2.14	0.744 (23.17)	0.712 (2.22)	0.07	0.593 (29.74)	0.575 (2.31)	10.1

Note. Root mean square deviation is shown in parentheses underneath r^2; r^2 is calculated as $(1 - SS_{error} / SS_{total})$. k = estimated value of discounting parameter.

make it easier to defer intervention. It is possible that this may explain the poor fit of the models to the data when controllers had the benefit of automated conflict detection. In the next section, we report the results of models incorporating a deferral mechanism, both when automated conflict detection is present and when it is absent.

SECTION 2: DEFERRAL MODELS

In the current section, we tested a series of models that allow conflict resolution to be deferred. The models all assume that when a conflict is selected from the queue, the controller assesses the value of resolving that conflict immediately (V_R). The resolution task value for a given conflict is compared to a threshold arbitrarily set to 10.0. Any *resolve-conflict* task with a V_R below threshold is dropped from the task queue—that is, deferred—until the model identifies the conflict during the next scan of the airspace. The models differ in the way that V_R is assessed. The first model assumes that the value of resolving a conflict immediately is a function of conflict duration and airspace complexity (Equation 9). The second model incorporates workload (Equation 10), and the third incorporates temporal discounting (Equation 11). For all models, selection of tasks from the queue is based on task value, V_t. The simplest form of task value outside of task order is based on a task's importance. The importance model from Section 1, therefore, provides the baseline against which improvements in model fit can be judged.

Results and Discussion

The introduction of a mechanism to allow deferral produced an overall improvement in fit across most scenarios, as Table 2 makes clear. The most substantial gains were obtained for the deferral model that incorporates discounting (see Model 2.3). The deferral discounting model yields the best fit for the timing data for all four data sets, and the fits are substantially better than the relevant benchmark from Section 1 (see Model 1.2). The worst fit of this model accounts for nearly 70% of the variance in timing data, compared to just under 50% of timing variance for the importance model in Section 1. In all other simulations, the deferral discounting model accounted for over 80% of the variance in the timing data. The same model, moreover, provides fits to the order data that are either the best fit or close to the best fit. As in the previous section, the fits for Controller 2's timing data are generally better than for Controller 1.

SECTION 3: COMBINING SELECTION AND DEFERRAL

The results from Section 2 suggest that deferral plays an important role in scheduling. The

TABLE 2: Fit Statistics and Parameter Estimates for Timing and Order Data Across the Four Exercises for the Models Allowing Deferral

	Controller 1						Controller 2					
	Without Detection Tool			With Detection Tool			Without Detection Tool			With Detection Tool		
Model	Timing	Ordinal	*k*	Timing	Ordinal	*k*	Timing	Ordinal	*k*	Timing	Ordinal	*k*
2.1 Duration and complexity	0.592 (22.79)	0.865 (1.82)	NA	0.602 (24.23)	0.356 (5.06)	NA	0.734 (23.64)	0.709 (2.24)	NA	0.671 (26.74)	0.591 (2.27)	NA
2.2 Deferral workload	0.565 (23.53)	0.884 (1.70)	NA	0.495 (27.30)	0.354 (5.06)	NA	0.735 (23.56)	0.733 (2.14)	NA	0.604 (29.33)	0.576 (2.31)	NA
2.3 Deferral discounting	0.847 (13.93)	0.884 (1.69)	5.46	0.679 (21.75)	0.504 (4.44)	0.02	0.827 (19.07)	0.695 (2.29)	0.04	0.816 (20.01)	0.741 (1.81)	0.08

Note. Root mean square deviation is shown in parentheses underneath r^2; r^2 is calculated as $(1 - SS_{error} / SS_{total})$. k = estimated value of discounting parameter. All models assume that selection of tasks from the queue is based on importance.

best-fitting model suggests that the value of resolving a conflict immediately is a product of importance, conflict duration, and airspace complexity and that this value is subject to hyperbolic discounting. However, it is possible that a model combining both sources of dynamics—discounting and workload—may perform even better. In Section 3, therefore, we extended the flexible discounting model to include workload in the V_R function, yielding the full dynamic model for selection and deferral:

$$V_R = \frac{I_R \times CD_i \times \overline{C_i} \times W_t}{1 + kD_i}. \tag{12}$$

However, one final assumption built into all models remains to be tested. All models but the pure first-come-first-served model have a mechanism for selecting tasks from the queue that combines both importance and first-come-first-served selection. The need for importance has not been demonstrated. Eliminating importance would simplify the mechanism for selecting tasks from the queue, reducing it to a pure first-come-first-served algorithm. In Section 3, therefore, we tested another version of the full dynamic model for selection and deferral, in which the importance of all tasks was set to 1.0, yielding the reduced dynamic model. The reduced dynamic model uses first-come-first-served to select tasks from the queue and uses hyperbolic discounting, with workload, to assess the value of immediately resolving conflicts.

Results and Discussion

As can be seen in Table 3, the addition of workload to the V_R function of the deferral discounting model improved the fit to most data sets without any increase in the number of free parameters, although gains were quite modest (see Model 3.1 vs. Model 2.3). The elimination of importance in the reduced dynamic model had some slight effect on fits, but its effects were inconsistent (see Model 3.1 vs. Model 3.2). For two simulations, eliminating importance slightly lowered the fit to both timing and order data compared to the full dynamic model. For one simulation (Controller 2, without automated conflict detection), eliminating importance slightly improved the fit to both the timing and order data; for a second simulation (Controller 2, with automated conflict detection), the changes were mixed, with a slight improvement in fit to the timing data and a slight decrement in fit to the order data. Overall, it appears that importance does not play an important role in the selection of tasks from the queue. However, to obtain a more precise description of the relative performance of these models, we can compute the corrected Akaike information criterion (AIC_C; Hurvich & Tsai, 1989), a variant of the AIC. AIC_C corrects for small data size, a reasonable precaution given that the number of observations for each exercise ranges from 12 to 22. If data size is not a problem, then the AIC_C and AIC values will be identical.

TABLE 3: Fit Statistics and Parameter Estimates for Timing and Order Data Across the Four Exercises for the Full Dynamic Model and the Reduced Dynamic Model

	Controller 1						Controller 2					
	Without Detection Tool			With Detection Tool			Without Detection Tool			With Detection Tool		
Model	Timing	Ordinal	k	Timing	Ordinal	k	Timing	Ordinal	k	Timing	Ordinal	k
3.1 Full dynamic	0.852 (13.74)	0.883 (1.70)	39.63	0.723 (20.22)	0.623 (3.87)	0.65	0.851 (17.67)	0.731 (2.15)	0.77	0.850 (18.06)	0.730 (1.84)	0.65
3.2 Reduced dynamic	0.833 (14.58)	0.879 (1.73)	9.64	0.714 (20.54)	0.599 (3.99)	0.18	0.877 (16.09)	0.758 (2.04)	0.12	0.876 (16.45)	0.717 (1.8)	0.19

Note. Root mean square deviation is shown in parentheses underneath r^2; r^2 is calculated as $(1 - SS_{error} / SS_{total})$. k = estimated value of discounting parameter.

TABLE 4: AIC_C and w_i Values for Timing Data Across the Four Exercises for the First-Come, First-Served, Deferral Discounting, Reduced Dynamic, and Full Dynamic Models

		Controller 1				Controller 2			
		Without Detection Tool		With Detection Tool		Without Detection Tool		With Detection Tool	
Model	Parameters	AIC_C	w_i	AIC_C	w_i	AIC_C	w_i	AIC_C	w_i
1.1 First-come, first-served	0	46.82	0.021	63.47	0.059	38.25	0.152	35.45	0.069
2.3 Deferral discounting	1	41.16	0.355	61.05	0.199	38.18	0.157	33.63	0.170
3.1 Full dynamic	1	40.96	0.392	59.65	0.399	37.26	0.250	32.56	0.290
3.2 Reduced dynamic	1	41.83	0.253	59.95	0.343	36.12	0.441	31.59	0.472

Note. AIC = Akaike information criterion; AIC_C = corrected AIC; w = weight.

Table 4 gives the AIC_C values for the fits to the timing data across each of the four scenarios for the following models: full dynamic (Model 3.1), reduced dynamic (Model 3.2), deferral discounting (Model 2.3), and first-come, first-served (Model 1.1)—the best performing zero-parameter model. The AIC_C is similar to the RMSD in that the lower an AIC_C value, the better the fit. Table 4 also provides a weight value, w_i; this weight represents the likelihood that a given model, i, is the best-fitting model. The results show that the full dynamic and reduced dynamic models are the most likely models for two scenarios each. Augmenting the discounting model with a workload interaction term does seem to provide a better description of timing decisions, but task importance seems to add little or nothing.

GENERAL DISCUSSION

We have shown that on-the-fly scheduling is in principle describable by relatively simple models that treat scheduling as an emergent property of the way that a task queue is managed. Despite having only a single free parameter, the reduced dynamic model accounted for 71% to 88% of the variance in the timing of air traffic controllers' conflict resolutions and for 60% to 88% of the variance in their ordering. This result marked a substantial improvement on a pure first-come-first-served model, which accounted for 48% to 74% of the variance in resolution timings and 19% to 87% in the ordering of resolutions. The results suggest that there is promise in treating on-the-fly scheduling as a psychological process involving the selection and deferral of tasks in a queue. Perhaps just as importantly, we found that some factors had little or no weight. Workload seemed to affect only the deferral of aircraft conflict resolutions, and intrinsic task importance had no effect at all.

The reduced dynamic model represents a form of partial-ordering planning using dynamic task values. According to this model, the selection of tasks from the queue is done opportunistically, on a first-come-first-served basis. Tasks with a short deadline are executed as soon as they are selected. This strategy is a simple and efficient way of ensuring that routine tasks are completed quickly, without the need for complex assessments of the value of competing tasks. Tasks with longer deadlines are evaluated to assess whether they need to be executed immediately or whether they can be deferred. The ability to defer intervention is important in a dynamic and uncertain environment because it provides flexibility. When a conflict needs to be resolved, there can be several options available, some of which may not be obvious and all of which may have uncertain consequences. Deferral allows the decision maker to reassess the need for intervention when there is less uncertainty regarding the need for that intervention and the consequences of the various options.

Temporal Discounting

In our most successful model, the reduced dynamic model, the value of acting immediately to resolve a particular conflict is dependent on airspace complexity, conflict duration, workload, and time to deadline. At first glance, it may appear surprising that so many factors are considered when deciding whether to defer intervention. However, conflict resolution is a complex problem (Durand & Alliot, 1997), and any decision to act now or defer involves a mix of risks and benefits. One of the primary benefits of acting early is to reduce workload and complexity, yet this may be offset by uncertainty regarding the trajectories of the aircraft (Averty et al., 2004). Our study provides evidence that the assessment of the value of acting immediately may change as the conditions of work and the worker change, validating recent approaches in the work motivation literature emphasizing the dynamic, temporally sensitive nature of task values (e.g., Steel & König, 2006; Vancouver et al., 2010).

Steel and König's (2006) temporal motivation theory gives hyperbolic discounting an important place in the dynamics of motivation and decision making. To date, the predictions of temporal motivation theory have been tested only in simple laboratory tasks, in which participants have limited discretion regarding the timing and order of tasks and are given complete information regarding the consequences of their actions (Schmidt & DeShon, 2007). Our study provides evidence that hyperbolic discounting may shape decision making in a much more complex environment outside of the laboratory, when decision makers have considerable discretion over the timing and order of the tasks that they perform.

The incorporation of hyperbolic discounting into models of task scheduling may also provide insight into individual differences and between experts and novices in particular. The larger the value of k, the larger the denominator of the discounting function for a given deadline and the greater the reduction in a conflict's perceived urgency. Thus, differences in k index the degree to which a controller is sensitive to deadlines when evaluating a task. Expert controllers seem to intervene more readily than novices with a year or less of training (Loft et al., 2009), suggesting that experts are more sensitive to deadlines than are novices, discounting conflicts less than do novices for a given deadline. That is, we would predict systematically smaller values of k for experts than for novices. Understanding temporal discounting may help in understanding the acquisition of expertise.

The value of k was generally quite low in the reduced dynamic models, except for Controller 1 when there was no automated conflict detection to help reduce workload. Thus the degree of discounting was always rather moderate for our experts. The sharp reduction in k when Controller 1 used a detection tool suggests that discounting might be a strategy to help cope with workload and becomes less important when the work is made easier.

Workload

We were initially motivated to explore the effects of workload on scheduling because of the findings of Raby and Wickens (1994) that pilots adhered more closely to an a priori rating of task importance as workload increased. However, workload appears to play a different role in the current study. The results suggest that workload influences the deferral of tasks with long deadlines rather than the selection of tasks from a queue. This finding may reflect differences in the work environment across the two studies. The pilots in Raby and Wickens' (1994) study needed to make choices among tasks with relatively short deadlines, whereas the air traffic controllers in our study had a mix of tasks, some with short deadlines and some with deadlines of an hour or more. We suspect that a model in which workload and importance jointly influence the selection of tasks from the queue may provide a better fit in an environment in which the worker has to make choices among tasks with short deadlines and deferral is either not possible or has limited benefit.

The effect of workload in the current study is similar in some respects to that previously reported by Loft et al. (2009). Loft et al. presented controllers with a series of pairwise conflicts while instructing them to imagine that they were under high or low levels of workload. Controllers were asked to indicate whether they would intervene to assure separation. Loft et al. found that controllers were more likely to intervene in the high-workload condition than in the low-workload condition and interpreted this as evidence that controllers were less tolerant of uncertainty and applied greater safety margins under high workload. The result was an elevated false alarm rate. However, the task used by Loft et al. was static rather than dynamic. It is possible that in a dynamic environment, the tendency to be less tolerant of uncertainty when workload increases may cause controllers to act earlier, rather than changing their criterion for intervention.

An additional contribution of the current study lies in the treatment of workload as inertial. Although not an entirely new idea, workload inertia has received limited attention. Morris and Rouse (1988) discussed it in a technical report commission for NASA, and Rouse et al. (1993) developed a dynamic model of workload to account for workload inertia but used a model that assumes symmetric changes in workload, unlike our asymmetric approach. There is research showing that workload rises faster than it falls (Morgan & Hancock, 2011), yet to our knowledge, computational models of workload do not yet incorporate this asymmetry.

Finally, we should note that our approach to workload ignores memory load, especially that prospective memory load that may arise from interrupted tasks (see Dismukes & Nowinski, 2007, for a review). Such task interruptions are likely to be especially characteristic of the kind of open-ended work situations in which on-the-fly scheduling is important. Indeed, a controller's decision to defer intervention can be seen as a task interruption. (We thank an anonymous reviewer for pointing out the relevance of task interruptions to on-the-fly scheduling.) For this reason, on-the-fly scheduling decisions may be an important contributor to prospective memory load.

Limitations

There are a number of limitations of the current study. First, our simulations provide a highly simplified representation of the tasks that the controllers carry out. Most of the tasks involve a series of steps that are not represented in the model and involve interactions with other actors in the system. Furthermore, we have tried to fit only the timing of resolution decisions rather than the timing of all tasks. However, the level of abstraction that we have chosen is appropriate for the study of scheduling behavior, because our focus on the scheduling of conflict resolutions decisions requires only that the time taken for rival tasks be reasonably approximated so that the model blocks off an appropriate amount of time when "doing" any task. Although a more fine-grained task analysis may yield more accurate timing estimates, the quality of the obtained fits suggests that our estimates are adequate.

A second limitation is that this model confined scheduling to an immediate decision whether to act now or defer. Our approach did not allow for organizing sequences of actions. Hayes-Roth and Hayes-Roth (1979) originally conceived of partial-order planning as planning of tentative sequences of tasks that would be readily changed as events unfolded. Controllers, moreover, often speak of using lulls in activity to plan a sequence of actions. Such sequencing would require a more detailed set of decision data as well as a more complex model and a deeper understanding of workload management by air traffic controllers. Vortac and colleagues have quantified activity sequences using multivariate and regression-based techniques (e.g., Edwards, Fuller, Vortac, & Manning, 1995; Vortac, Edwards, & Manning, 1994). Whether such quantification of organization can be integrated with the current modeling approach, however, is currently unclear.

A third limitation is that the models have been tested using data from just two controllers. As a result, we do not know how well these models will generalize. However, the current paper represents an important first step in the development of formal models of human scheduling. We have demonstrated that the problem is tractable and that it is possible to develop models that are capable of explaining scheduling behavior in a complex environment. This approach is commonly used in cognitive science when fitting individual data, especially when the model is intended as an initial demonstration of some approach or existence of proof. For example, Nosofsky's (1986) introduction of the generalized context model consisted of fits to individual data sets consisting of 16 data points produced by two subjects, fit by a model with three free parameters. Ratcliff, Van Zandt, and McKoon (1999) tested the diffusion model against connectionist models of response time using data from just 4 participants performing a signal detection task. In the current study, we fitted the models to individual data sets, with anywhere from 24 to 44 data points, and only a single free parameter to account for the considerable variance, all of which imposes tight constraints on overfitting.

Having developed the approach, the next step is to assess the generalizability of the models across people and across work environments. It is almost certainly the case that there are individual differences in the way that people perform scheduling as well as differences across work environments. For example, air traffic controllers managing tightly spaced arrival flows in a terminal area may be reluctant to defer intervention because the deadlines for intervention are much shorter than in an oceanic environment. An objective for future research should be to identify the extent of variability in scheduling behavior across these dimensions and assess whether the same models are capable of accounting for this variability or whether different models are required for different people or different environments. This is a challenging problem and will most likely require the use of cutting-edge modeling techniques, such as hierarchical Bayesian modeling (e.g., Lee, 2011). Hierarchical Bayesian models enable researchers to examine how the parameters of a model (e.g., a model of decision making) vary across people and environments and to assess whether a mixture of models is needed to explain the data from different individuals. It is difficult to make these kinds of inferences using conventional modeling techniques.

Last, this work does not address the question of optimality. Whether the mix of first-come-first-served for routine tasks with short deadlines and possible deferral for longer-running tasks represents an optimal scheduling approach in an on-the-fly scheduling environment is unaddressed here. Air traffic is growing rapidly worldwide, and new capabilities are needed to meet projected growth (International Civil Aviation Organization [ICAO], 2011). Capabilities such as integrated arrivals and departure management, which is intended to optimize the flow of traffic through the system as a whole, are currently under development. The ultimate goal is to enable the introduction of so-called trajectory-based operations, in which airspace users negotiate and fly mutually agreed trajectories from departure to destination (ICAO, 2005). The effect of these changes will be to increase the complexity of traffic flows and make the system more tightly coupled and interdependent (Durso & Manning, 2008; Neal, Flach, Mooij, Lehmann, Stankovic, & Hasenbosch, 2011). As the system becomes more tightly coupled, it will become increasingly important for tasks to be scheduled optimally.

Applications and Conclusions

There is growing interest in the human factors community in the use of formal models of human performance for the development and acquisition of systems, the redesign of equipment and procedures, and the development of training simulators and decision support systems. However, there is recognition that the current generation of models lacks the flexiblity and adpatability of human behavior and generates predictions that are brittle and unrealistic (Pew & Mavor, 1998, 2007). For this reason, there have been calls to incorporate higher-fidelity models of human cognition into simulations (Lotens et al., 2009).

Substantial progress has been made in the development of formal theories of decision making that are capable of accounting for a broad range of empirical phenomena in the laboratory (e.g., Roe, Busemeyer, & Townsend, 2001), and these approaches are starting to be applied in more complex environments, such as air traffic control (Neal & Kwantes, 2009; Vuckovic, Kwantes, Humphreys, & Neal, in press). The focus of this work has been on the way that people make choices among decision alternatives. However, much less is known about the way that people make decisions regarding the timing and order of tasks that they have to perform. Any attempt to simulate performance in a complex environment, like air traffic control, requires a theory that explains how people do this. The model of task selection and deferral developed in the current paper provides a way of approaching this problem.

A further potential application lies in the development of predictive models of workload. There is a long history of research examining the causes and consequences of workload in military and industrial settings (e.g., see Durso & Manning, 2008). However, it has proven to be extremely challenging to develop models that are capable of predicting the level of workload that an operator will experience in the future, given projected workflows (Neal et al., 2014). In principle, it should be possible to predict future workload, given knowledge of the tasks that have to be performed and the time available to complete them (e.g., Hendy, Liao, & Milgram, 1997). However, operators frequently have discretion over the timing and order of the tasks they perform, making prediction difficult. Scheduling theory has been used as a normative model to determine the optimal sequence in which to perform a set of tasks; however, humans do not schedule optimally, and knowing the optimal rule does not help to reduce operator workload (Moray et al., 1991). The development of a psychological theory that explains how operators make scheduling decisions may enable more accurate prediction and management of workload. Ultimately, by understanding how operators make scheduling decisions, we should be able to design tools and processes that enhance the safety and effectvieness of the systems that they control.

APPENDIX A

Importance Values and Completion Times

TABLE A1: Importance Values for Five Tasks Across All Models

Task	Importance
Check-position-report	0.708
Accept-handover	2.154
Offer-handover	1.729
Check-trajectory	3.223
Resolve-conflict	1.600

Note. "Accept-handover" = accept jurisdiction over new aircraft; "offer handover" = pass on jurisdiction to controller in adjacent sector.

TABLE A2: Mean Completion Times in Minutes for Each State

State	Mean Completion Time
Scan airspace	0.10 (0.05*N)
Check trajectory	0.07 (0.07*N)
Decide resolve	0.20 (0.5*N)
Act	
Check-position-report	0.5 (0.15)
Accept-handover	0.25 (0.05)
Offer-handover	0.25 (0.05)
Resolve-conflict	1.5 (0.5)

Note. Standard deviations in parentheses. N = number of aircraft under jurisdiction. The Act state comprises four actions, each with its own mean completion time and standard deviation. The Check Trajectory activity is identical with a single task, check-trajectory, repeated across N aircraft trajectories. Values were derived by observation of and discussions with air traffic controllers.

APPENDIX B

Task Queue and Aircraft Matrix

Task queue. The task queue is an $n \times 8$ matrix, with eight columns describing the properties of each of n tasks. Columns 1 through 5 indicate the task; entries in these columns index craft by reference to their row number in the aircraft matrix. Column 6 indicates the conflict partner of a craft initiating a conflict, again via reference to its row number in the aircraft matrix. Column 7 indicates time remaining to deadline for the task, and column 8 gives the task value for the task.

An example task queue is illustrated in Table B1. This task queue holds three tasks: handing off Aircraft 1, which must be done in 3 min; resolving a conflict between Aircraft 23 and Aircraft 17, which has 54 min until loss of separation; and updating flight information for Aircraft 18, which has a deadline of 60 min. The time to deadline for an uncompleted task is reduced whenever a model's clock is updated.

Aircraft matrix. The aircraft matrix tracks aircraft states or what the controller has discovered of those states. The matrix is an $n \times 9 \times m$ matrix, where each of the n rows is a single craft defined along the nine properties composing the columns, such as time of appearance in the simulation or time until a jurisdiction change (enter or leave sector). The m pages replicate the craft information for each of the conflicts a craft is in. An aircraft that is not involved in any conflict has nonzero entries only in the first page. The matrices for simulations with and without use of an automated conflict detection tool differ slightly. Both are illustrated in Tables B2 and B3.

TABLE B1: Example of a Task Queue Used Across All Models and Simulations

Check Flight Report	Accept Handover	Offer Handover	Check Trajectory	Resolve Conflict	Conflict Partner	Time Remaining	Task Value
	1					3	18
				23	17	54	6
18						60	1
…							

Note. "Accept handover" is to accept an incoming aircraft as under jurisdiction; "offer handover" is to transfer jurisdiction to another controller for an exiting aircraft. Time is given in minutes. "Check flight report" is initially given an arbitrary deadline of 60 min and is added to the queue probabilistically during Scan Airspace.

TABLE B2: Example Aircraft Matrix for Simulating a No-Tool Scenario

Time of Appearance	Status at Start of Scenario	Jurisdiction 1 Deadline	Jurisdiction 2 Deadline	Number of Conflicts	Trajectory Checked	Conflict Partner	Conflict Detected	Conflict Deadline
0.0	2.0	98.0	180.0	1	1.0	7	1.0	32.0
0.0	1.0	30.0	90.0	0.0	1.0	0.0	0.0	180.0
0.0	2.0	81.5	180.0	0.0	1.0	36	0.0	70.0
…								
87.0	0.0	113.0	180.0	0.0	0.0	0.0	0.0	180.0
90.0	0.0	115.0	180.0	0.0	0.0	12	0.0	105.0

Note. For Status, 2 = within sector, 1 = outside of sector, 0 = unannounced. Jurisdiction 1 Deadline and Jurisdiction 2 Deadline indicate time to sector entry or exit from the start of the scenario. Number of Conflicts is a running total of detected conflicts. Trajectory Checked is a binary indicating whether a craft has had its trajectory checked for conflicts. Conflict Partner indicates the row corresponding to a conflict partner for a given craft. Detected Conflict is a binary indicating whether the conflict with the craft indexed in Conflict Partner has been detected. Conflict Deadline indicates the time from the start of the scenario for collision between the current craft and the craft indexed in Conflict Partner. All times are in minutes; times of 180 indicate events that never occur.

TABLE B3: Example Aircraft Matrix for Simulating a Conflict Detection Tool Scenario

Time of Appearance	Status at Start of Scenario	Jurisdiction Deadline	Number of Conflicts	Trajectory Checked	Conflict Partner	Conflict Detected	Conflict Deadline	Registered in SCW
0.0	2.0	89.0	1	1.0	3	1.0	32.0	18.0
0.0	1.0	25.0	0.0	1.0	0.0	0.0	180.0	180.0
0.0	2.0	81.5	0.0	1.0	38	1.0	78.0	43.0
…								
87.0	0.0	100.5	0.0	0.0	0.0	0.0	180.0	180.0
90.0	0.0	105.0	0.0	0.0	12	0.0	105.0	72.0
97.0	0.0	180.0	0.0	0.0	0.0	0.0	180.0	180.0

Note. SCW = sector conflict window. "Registered in SCW" indicates the time from the start of scenario when a conflict is "detected" by a model's conflict detection tool. All times are in minutes; times of 180 indicate events that never occur.

For the no-tool simulations, the nine columns of the matrix specify the time in minutes that an aircraft appears within the simulation airspace (or "announced," zero if present already), current status (under controller's jurisdiction, announced but not under jurisdiction, not yet announced, departed controller's jurisdiction), time from start of scenario to initial jurisdiction change (entering or leaving sector) if any, time from start of scenario to second jurisdiction change if any, the number of conflicts the craft has been involved with so far, the row number (aircraft index) of the partner for the current (mth) conflict, whether this current (mth) conflict has been detected, time remaining to conflict, and whether the craft's current trajectory is checked.

With the use of the conflict detection tool, conflict detection is identified as beginning with the registering of the conflict in the tool's sector conflict window (SCW) rather than when the second aircraft of a conflict pair is announced. The use of the SCW requires an extra column to simulate the tool-use condition; however, the aircraft in this condition seem to engage in only one jurisdiction change, thus the total number of columns is unchanged at nine.

ACKNOWLEDGMENTS

The work reported in the current paper was supported by Airservices Australia and NICTA. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications, and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. We thank Roger Remington, Mike Humphreys, Stephan Lehmann, and Jeffrey Vancouver for valuable comments and critique. We thank Sean Murphy, Kim Loong Lim, Bo-Han Hsieh, and John Guimelli for help in video coding.

KEY POINTS

- The spontaneous scheduling of tasks during the execution of work ("on-the-fly task scheduling") is examined in the context of air traffic control.
- Hypotheses regarding the control of on-the-fly scheduling are tested using a series of discrete-event models.
- The models are fit to conflict resolution decision data collected in a high-fidelity simulation involving two expert air traffic controllers.
- The best-fitting models incorporate a mixture of opportunistic scheduling driven by task order and a more complex but flexible mechanism enabling deferral of tasks with long deadlines, consistent with the mixed planning approach of Hayes-Roth and Hayes-Roth (1979).
- The deferral mechanism involves computing task values that vary with workload and time to deadline, where the time to deadline weights tasks following a hyperbolic discounting function.

REFERENCES

Arrington, C. M. (2008). The effect of stimulus availability on task choice in voluntary task switching. *Memory & Cognition*, *36*, 991–997.

Arrington, C. M., & Logan, G. D. (2004). The cost of a voluntary task switch. *Psychological Science*, *15*, 610–615.
Arrington, C. M., Weaver, S. M., & Parker, R. L. (2010). Stimulus-based priming of task choice during voluntary task switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1060–1067.
Averty, P., Collet, C., Dittmar, A., Vernet-Maury, E., & Athenes, S. (2004). Mental workload in air traffic control: An index constructed from field tests. *Aviation, Space and Environmental Medicine*, *75*, 333–341.
Caruso, E. M., Gilbert, G. T., & Wilson, T. D. (2008). A wrinkle in time: Asymmetric valuation of past and future events. *Psychological Science*, *19*, 796–801.
Chapman, G. B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 771–791.
Dessouky, M. I., Moray, N., & Kijowski, B. (1995). Taxonomy of scheduling systems as a basis for the study of strategic behavior. *Human Factors*, *37*, 443–472.
Dismukes, K., & Nowinski, J. (2007). Prospective memory, concurrent task management, and pilot error. In A. F. Kramer, D. Wiegmann, & A. Kirlik (Eds.), *Attention from theory to practice* (pp. 225–236). Oxford, UK: Oxford Press.
Durand, N., & Alliot, J. (1997). *Optimal resolution of en route conflicts*. Toulouse, France: Laboratoire d'Optimisation Globale, CENA at ENAC.
Durso, F. T., & Manning, C. A. (2008). Air traffic control. *Reviews of Human Factors and Ergonomics*, *4*, 195–244. doi:10.1518/155723408X342853
Edwards, M., Fuller, D., Vortac, O. U., & Manning, C. A. (1995). The role of flight progress strips in en route air traffic control: A time-series analysis. *International Journal of Human-Computer Studies*, *43*, 1–13.
Ettema, D., Borgers, A., & Timmermans, H. (1993) Simulation model of activity scheduling behavior. *Transportation Research Record*, *1413*, 1–11.
Fransoo, J. C., & Wiers, V. C. S. (2006). Action variety of planners: Cognitive load and requisite variety. *Journal of Operations Management*, *24*, 813–821.
Gärling, T., Säisä, J., Böök, A., & Lindberg, E. (1986). The spatiotemporal sequencing of everyday activities in the large-scale environment. *Journal of Environmental Psychology*, *6*, 261–280.
Grace, R. C. (1999). The matching law and amount-dependent exponential discounting as accounts of self-control choice. *Journal of the Experimental Analysis of Behavior*, *71*, 27–44.
Green, L., Fry, A., & Myerson, J. (1994). Discounting of delayed rewards. *Psychological Science*, *5*, 33–36.
Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science*, *3*, 275–310.
Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors*, *39*, 30–47.
Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297–307.
International Civil Aviation Organization. (2005). *Global air traffic management operational concept* (No. 9854). Montreal, Canada: Author.
International Civil Aviation Organization. (2011). *Working document for the aviation systems block upgrades: The framework for global harmonization*. Montreal, Canada: Author.
Kallus, K. W., Van Damme, D., & Dittman, A. (1999). I*ntegrated job and task analysis of air traffic controllers: Phase 2. Task analysis of en-route controllers* (European Air Traffic Management Programme Rep. No. HUM.ET1.ST01.1000-REP-04). Brussels, Belgium: Eurocontrol..
Kernan, M. C., & Lord, R. G. (1990). Effects of valence, expectancies, and goal-performance discrepancies in single and multiple goal environments. *Journal of Applied Psychology*, *75*, 194–203.
Kiesel, A., Wendt, M., Jost, K., Steinhauser, M., Falkenstein, M., Phillip, A. M., & Koch, I. (2010). Control and interference in task switching: A review. *Psychological Bulletin*, *136*, 849–874.
Lee, M. D. (2011). How cognitive modeling can benefit from Hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*, 1–7.
Loft, S., Bolland, S., Humphreys, M. S., & Neal, A. (2009). A theory and model of conflict detection in air-traffic control: Incorporating environmental constraints. *Journal of Experimental Psychology: Applied*, *15*, 106–124.
Loft, S., Sanderson, P. M., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors*, *49*, 376–399. doi:10.1518/001872007X197017
Lotens, W., et al. (2009). *Human behavior representation in constructive simulation: Final report of Task Group 128, consolidating the findings of Task Group 143* (RTO Tech. Rep. TR-HFM-128). Brussels, Belgium: North Atlantic Treaty Organization.
Mayr, U., & Bell, T. (2006). On how to be unpredictable: Evidence from the voluntary task-switching paradigm. *Psychological Science*, *17*, 774–780.
Mazur, J. E., & Biondi, D. R. (2009). Delay-amount tradeoffs in choices by pigeons and rats: Hyperbolic versus exponential discounting. *Journal of the Experimental Analysis of Behavior*, *91*, 197–211.
Moray, N., Dessouky, M. I., Kijowski, B. A., & Adapathya, R. (1991). Strategic behavior, workload and performance in task scheduling. *Human Factors*, *33*, 607–629.
Morgan, J. F., & Hancock, P. A. (2011). The effect of prior task loading on mental workload: An example of hysteresis in driving. *Human Factors*, *35*, 75–86.
Morris, N., & Rouse, W. B. (1988). *Human operator response to error-likely situations in complex engineering systems* (Report 177484). Moffett Field, CA: Ames Research Center, NASA.
Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior*, *64*, 263–276.
Neal, A., Flach, J., Mooij, M., Lehmann, S., Stankovic, S., & Hasenbosch, S. (2011). Envisaging the future air traffic management system. *International Journal of Aviation Psychology*, *21*, 16–34. doi:10.1080/10508414.2011.537557
Neal, A., Hannah, S. D., Sanderson, P., Bolland, S., Mooij, M., & Murphy, S. (2014). Development and validation of multilevel model for predicting workload under routine and non-routine conditions in an air traffic management center. *Human Factors*, *56*, 287–305.
Neal, A., & Kwantes, P. J. (2009). An evidence accumulation model for conflict detection performance in a simulated air traffic control task. *Human Factors*, *51*, 164–180. doi:10.1177/0018720809335071
Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *8*, 308–313.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
Pew, R., & Mavor, A. (1998). *Modeling human and organizational behavior: Application to military simulation*. Washington, DC: National Academy Press.

Pew, R., & Mavor, A. (2007). *Human-system integration in the system development process: A new look*. Washington, DC: National Academies Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2nd ed.). New York, NY: Cambridge University Press.

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *International Journal of Aviation Psychology*, *4*, 211–240.

Rachlin, H., & Green, L. (1972). Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior*, *17*, 15–22.

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior*, *55*, 233–244.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*, 261–300.

Ratterman, M. J., Spector, L., Grafman, J., Levin, H., & Harward, H. (2001). Partial and total-order planning: Evidence from normal and prefrontally damaged populations. *Cognitive Science*, *25*, 941–975.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*, 370–392.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.

Rouse, W. B., Edwards, S. L., & Hammer, J. M. (1993). Modeling the dynamics of mental workload and human performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*, 1662–1671.

Rovira, E., & Parasuraman, R. (2010). Transitioning to future air traffic management: Effects of imperfect automation on controller attention and performance. *Human Factors*, *52*, 411–425.

Sanderson, P. (1989). The human planning and scheduling role in advanced manufacturing systems: An emerging human factors domain. *Human Factors*, *31*, 635–666.

Schmidt, A. M., & DeShon, R. P. (2007). What to do? The effects of discrepancies, incentives, and time on dynamic goal prioritization. *Journal of Applied Psychology*, *92*, 928–941.

Simpson, C. A., & Vuchinich, R. E. (2000). Reliability of a measure of temporal discounting. *Psychological Record*, *50*, 3–16.

Steel, P., & König, C. J. (2006). Integrating theories of motivation. *Academy of Management Review*, *31*, 889–913.

Vancouver, J. B., Weinhardt, J. M., & Schmidt, A. M. (2010). A formal, computational theory of multiple-goal pursuit: Integrating goal-choice and goal-striving processes. *Journal of Applied Psychology*, *95*, 985–1008.

Vortac, O. U., Edwards, M., & Manning, C. A. (1994). Sequences of actions for individual and teams of air traffic controllers. *Human-Computer Interaction*, *9*, 319–343.

Vuckovic, A., Kwantes, P., Humphreys, M., & Neal, A. (in press). A sequential sampling account of response bias and speed-accuracy tradeoffs in a conflict detection task. *Journal of Experimental Psychology: Applied*.

Wainer, G. A. (2009). *Discrete-event modeling and simulation: A practitioner's approach*. Boca Raton, FL: CRC Press.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs." *Psychological Research*, *63*, 212–233.

Samuel D. Hannah is an assistant professor of cognitive psychology in the Department of Psychology at the University of Saskatchewan. He received his PhD in 2004 from McMaster University.

Andrew Neal is a professor of organizational psychology and human factors in the School of Psychology at the University of Queensland. He received his PhD in 1996 from the University of New South Wales.

Date received: April 2, 2013
Date accepted: January 30, 2014