Modeling Capabilities and Workload in Intelligent Agents for Simulating Teamwork

Thomas R. Ioerger, Linli He, Deborah Lord
Department of Computer Science, Texas A&M University
College Station, TX 77843-3112
{ioerger,linli,drl4468}@cs.tamu.edu
Pamela Tsang
Department of Psychology, Wright State University
Dayton, OH 45435-0001
(pamela.tsang@wright.edu)

Abstract

The ability of members on a team to reason about each others' capabilities and workload is important for effective teamwork. This is required for proper task allocation and load balancing, as well as many other team processes such as adaptiveness, proactive assistance, and backing-up behavior. The present work proposes to incorporate capability reasoning into intelligent agents to produce better teamwork simulations, to work better with humans as virtual team members, and to facilitate team training. However, classical models of capabilities in computational systems and intelligent agents are inadequate for representing the more complex aspects of human performance, such as the ability to perform multiple tasks in parallel, interference among these tasks, effects of limits on attention and other cognitive resources, and the ability of humans to dynamically adjust their level of effort on tasks. In this paper, we present a formal mathematical model of capabilities that accounts for these effects. The model posits finite pools of internal resources, for which tasks compete; quality of performance depends on the amount of resources allocated. Capabilities are defined according to whether a feasible schedule can be found that allows a set of tasks to be completed within given constraints (e.g. deadlines) while not exceeding the capacity of any internal resource. An extension of the model is then proposed to incorporate multiple resources.

Introduction

Many studies have suggested that the ability to distribute tasks appropriately and to adaptively balance the workload within teams is essential for producing effective teamwork (Kleinman et al., 1992; Kozlowski, 1997). In order to do this, team members must be able to reason about their own and each others' capabilities. For example, they must be able to know when to accept or reject new tasks, based on how they might interfere with current on-going tasks, to delegate sub-tasks to the (best) team members who are not overloaded, and to offer assistance to those who are. Even in intra-team communication and coordination, assessment of capabilities and workload have an impact; in one study, it was found that communications among team members in the best-performing teams actually decreased in high-tempo situations (Serfaty et al., 1997), presumably due to a recognition that excessive communication activities place a demand for attention on both the sender and receiver that competes with processing of intense taskwork. Therefore reasoning about capabilities, including knowledge of task demands, skill levels of individual team members, and momentary workload across the team in a given situation, must be considered an essential component of team cognition.

Recently there has been a rise of interest in incorporating intelligent agents into automated team-training systems (Rickel and Johnson, 1997). These agents could be used in a variety of ways, from automated assistants (decision aids), to virtual role players, to coaches. In order for agents to monitor, understand, critique, or participate in teamwork with human trainees, the agents must also be endowed with the ability to reason about capabilities and workload of individuals on the team. Agents in the simulation must be able to assess the workload of humans with whom they interact in order to make decisions about when and how to interact in a way that is not disruptive or unnatural. (This is an additional constraint that purely agent-based systems do not have to be concerned with.) However, most existing formal models of capability reasoning in agents do not adequately address the kind of reasoning that is required in these agent-based team-training systems. Typically, these prior models treat capabilities as a simple association between actors (agents or humans) and "executable" actions, though the actors must also be aware that they can do these things, i.e. have sufficient "know-how" (Moore, 1985; Singh, 1991; van der Hoek et al., 1994).

These computational models allow agent-based systems to be designed where the agents can reason about each others' capabilities, and even perform task distribution and load balancing. However, these models generally assume task completion is binary (success or failure) and do not take into account graded senses of capability, which are more meaningful to human performance. Humans can often achieve better results by working "harder" (applying more effort or attention), they can dynamically reduce their effort on one task to accomodate performing other tasks in parallel, and they are often limited by pragmatic upper-bounds on performance (e.g. due to finite skills or attention). What is needed is a formal system that will enable an agent to understand when a human is too busy doing certain activities (e.g. flying an aircraft in combat or engaging an enemy) to do other things (e.g. monitor for new visual contacts, listen to background radio traffic). The agent needs to be able to compute the relative impact of new tasks on the accuracy of performing existing tasks, and the potential for delay in completion of individual tasks by their deadlines. This is different from just asking whether an operator is capable of doing the additional tasks "in principle."

Humans are capable of performing multiple tasks in parallel, and there is a great deal of literature on analyzing time-sharing performance (Wickens and Holland, 2000). Yet humans ultimately have limits on their processing capacity, exemplified by the notion of finite limits on attention, which has been rigorously documented. Furthermore, there is clear evidence that some task combinations are time-shared more efficiently than others, such as the difference between drawing a sketch while listening to the radio versus reading while listening to the radio. Some models, such as the multiple-resource model (e.g., Wickens, 1984) have postulated distinct and separate cognitive resources for different types of cognitive processing to explain the wide range of observed task interactions. Another important issue that makes human capabilities difficult to reason about is that performance is not a binary quantity, but rather a graded value (e.g. accuracy, reaction time), and humans can intentionally adjust task performance in a number of controllable ways, such as increasing quality by focusing attention and applying more cognitive "effort," or by reducing effort by spreading the task processing out over a longer interval of time (Hendy et al., 1997), such as multiplying multi-digit numbers together in one's head more slowly for greater accuracy. Therefore, whether or not a human member of a team is "capable" of doing something depends on a great many things, including what other tasks he or she is doing (their current workload), the degree to which the new task might interfere with them, the individual's skill level(s), attention management skill (Gopher, 1993), and the adaptability of the task performance with respect to the tightness of the constraints on completion (e.g. deadlines, quality criteria). This is a more situation-based or context-dependent perspective on capability.

Reasoning about capabilities at this quantitative level is important for modeling and understanding teamwork. To date, very little research has addressed the relationship between individual cognition and quality of teamwork, though the connection is discussed in (Huey and Wickens, 1993). An understanding of individuals' capabilities and workload are clearly important to the efficient operation of a team, such as for distributing tasks to the most appropriate/skillful members, balancing the load (to maintain flexibility), and proactively assisting or backing each other up. With regard to training, we hypothesize that each team member him/herself must develop enough reserve capacity on top of their individual taskwork to devote some attention to participating in the teamwork, such as tracking status or progress of team goals, sharing information relevant to others, or building distributed situation awareness.

In this paper, we present a formal, mathematical model for reasoning about capabilities, especially for agents to reason about and interact with their human teammates. The approach sythesizes ideas from a number of previous descriptions of workload, attention, and performance into a computational model that can be concretely implemented as a decision-making procedure in a multi-agent system. After establishing some terms and assumptions, a definition of capability will be presented in terms of whether a human could adapt his or her performance (i.e. to find a schedule and select effort levels) that would accomodate a given set of tasks with a set of specified constraints. We conclude by discussing the implications of this computational model of capabilities for modeling and understanding team performance, and for designing new approaches to team training.

A Formal Model

In this section, we present a quantitative model for reasoning about capabilities. For simplicity, we start with description of a single-resource model as a basis. Later we extend it to show how it can accomodate the assumption of multiple cognitive resources.

Single-Resource Foundation

Preliminarily, assume there is a single common cognitive resource for which tasks compete. Perhaps it might be labelled generically as "attentional resources." At any given time, a person might be using some amount of this resource, $u(t)$, but the resource is bounded, $\forall t\ 0 \leq u(t) \leq u_{max}$. Since the scale is arbitrary, we normalize resource utilization so that $u_{max} = 1.0$, putting it on a uniform scale of 0 to 1.

We assume this common resource can be allocated to, or divided among, several concurrent tasks. The amount of resource being applied to a given task i at a given moment t is referred to as "effort," and is denoted $e_i(t)$. We view the sum of resources being applied to all tasks at a given moment as a reflection or internal measure of workload. Let the set of tasks be called $\tau_1 ... \tau_i$. The "workload" is defined as:

$$w(t) \equiv \Sigma_i\ e_i(t)$$

and it is constrained not to exceed the limit, $0 \leq w(t) \leq u_{max}$.

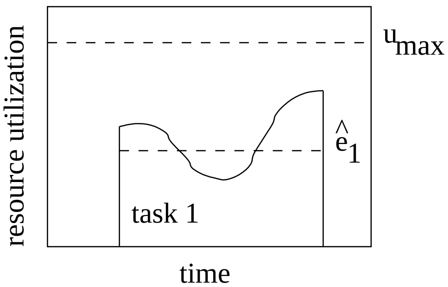

Figure 1: Resource utilization over time (with limit).

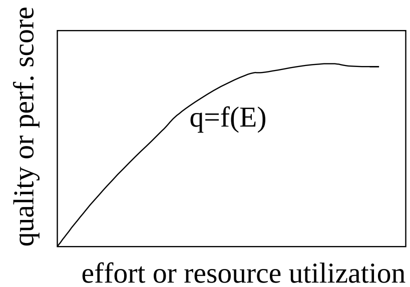

Figure 2: Illustration of a Performance Resource Function.

Whereas this notion of effort is defined on a moment-by-moment basis, the total effort expended on a task, or total resources applied E_i, is the sum of the effort allocated to that task over the duration of its performance (illustrated in Figure 1). In other words, it is the sum, or integral, of the moment-by-moment resources utilized in performing that task:

$$E_i \equiv \int_{t=start(i)}^{t=end(i)} e_i(t)\ dt$$

While the amount of resources applied to a task is not necessarily constant, we assume there is an average effort value $\overline{e}_i$, and our model is based on this approximation.

The amount of resources required for an individual to perform a given task depends on a number of internal and external determinants. Externally, the difficulty of the task, as well as constraints on accuracy or speed (i.e. deadlines), can influence the processing resources required (e.g. it is harder to do a task better or faster, and some tasks have parameters related to difficulty, such as number of items to remember, and so on). Internally, a specific individual's response to task demand can be affected by their innate ability and executive management skill, prior training, degree of automation, etc. We quantify the relationship between amount of resources applied to a task and quality of performance using a function for quality-effort tradeoff: $q_i = f(E_i)$ (also known as a Performance Resource Function (Norman and Bobrow, 1976); see Figure 2).[1] Quality can represent any number of performance measures specific to the task, such as accuracy, inverse of reaction time, etc. We do not place many restrictions on the form of this function, but typically, we assume it is monotonic: increasing effort on most tasks increases quality (Wickens, 1984). (Often, they reach a plateau where greater effort does not improve quality, in which case they are said to be "data-limited.")

[1]Quality of performance might also depend on some measure of task difficulty, which can be treated either by adding an argument for whatever variable parameterizes the degree of difficulty, or by simply viewing them as separate tasks.

Humans can apply the same amount of total effort to a task in a range of different ways. In particular, they might choose to work as hard as possible on the task, completing it in a short amount of time, or they might decide to spread the processing out over time, reducing the moment-by-moment effort, for example to have some reserve capacity left over to apply to other tasks in parallel. Given that we model effort applied as the integral of resource utilization over time, and we assume there is an average level of effort dedicated to a task, the relationship among momentary effort, total effort applied, and duration may be expressed as a simple formula:

$$E_i = \overline{e}_i \cdot d_i$$

where $d_i = end(i) - start(i)$ is the duration of the task. Therefore, the effort-duration tradeoff may be represented as a (presumably) hyperbolic function, and different levels of total effort appear as iso-curves (see Figure 3). Each point on a given curve bounds a box of constant area, representing the common degree of total effort. Harder versions of the same task correspond to curves farther out (dashed line in Figure 3), and improvements in ability, e.g. due to training, appear as curves closer to the origin.

We assume there are range bounds on both duration and effort. Of course, a task can utilize no more than 100% of a resource, and this puts a bound on the minimum execution time (speed), as a result of the hyperbolic function. Similarly, we assume there is a minimum amount of resource required, and a corresponding limit on the slowest effective rate of performance. We represent these ranges as $(d_{i,min}, d_{i,max})$ and $(u_{i,min}, u_{i,max})$.

The performance-resource function is not only a function of task, but also of the individual. We model the difference among individuals by assuming the form of the equation is the same, and applying a multiplier that represents their degree of skill s of individual j (relative to the average of the population, for which we set $\overline{s} = 1$):

$$q_{j,i} = f(E_i) \cdot s_{j,i}$$

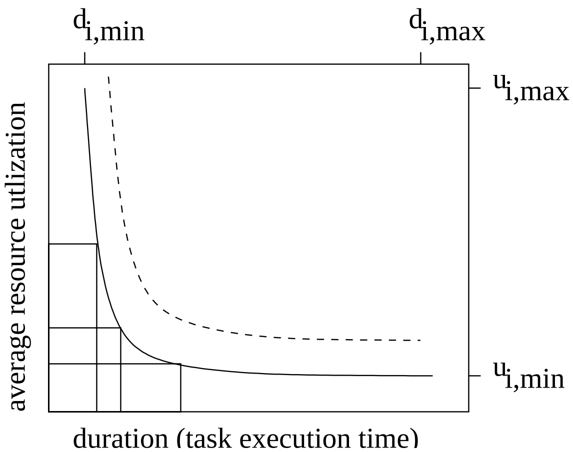

Figure 3: Effort iso-curves.

Hence the greater the skill, the greater the quality of performance for a fixed amount of effort (can be visualized as higher curves in Figure 2). This allows us to model the differences between novices and experts in a simple way.

Capability Assessment as Scheduling

Given this quantitative model of task performance, we can now offer an initial formal definition of "capability." Recall that we are interested not just in whether an individual is capable of doing something "in principle," but also whether it can be carried out effectively in the time allotted and to the level of quality or accuracy required, all within the context of other on-going activities. We view this as a kind of "scheduling" problem, where capability is determined by whether or not the individual can find an arrangement of processing so that all the tasks can be completed without violating any internal capacity limitations.

> **Definition 1:** A *schedule* for a set of tasks $\tau_1 ... \tau_n$ being processed or executed by an individual is a set of parameter vectors $\{\langle start(i), end(i), \overline{e}_i \rangle\}$ defining the start and end times of each task, along with planned average resource utilization to be applied to each.

> **Definition 2:** An individual j who is currently performing a set of tasks $\tau_1 ... \tau_n$, with quality constraints $q_i ... q_n$ and deadlines $dl_1 ... dl_n$ is said to be *capable* of performing a new task τ_{n+1} (with constraints q_{n+1} and d_{n+1}), if there exists a schedule $\mathcal{S}$ over $\tau_1 ... \tau_{n+1}$ defining the start and end times along with average resource utilization of each task $\langle start(i), end(i), \overline{e}_i \rangle$, such that all constraints remain satisfied. Specifically:

1. $q_i \leq f_j(E_i) = f(\overline{e}_i \cdot (end(i) - start(i)))$,
2. $end(i) < dl_i$, and
3. $\forall t \ w(t) = \Sigma_i \ \overline{e}_i \leq u_{max}$, where the sum runs over all tasks i s.t. $start(i) \leq t \leq end(i)$.

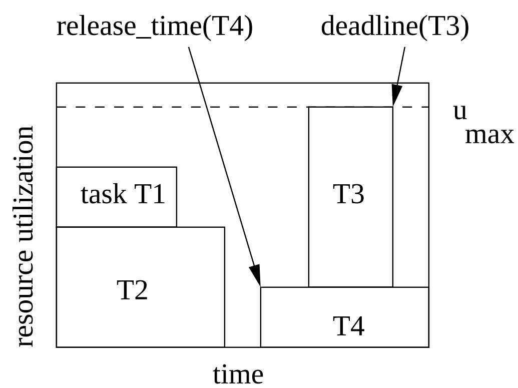

Figure 4: Example of task schedule.

We note that, among the existing tasks, some may currently be being processed, while others may be scheduled to start after some delay (i.e. pending tasks). If processing of certain currently executing tasks is considered uninterruptible, then we require $start(i) = t_{now}$ for those tasks in the revised schedule to maintain continuity (though the effort level may be modified). Figure 4 shows an example of a task schedule, with selected effort levels and durations, and some representative constraints.

The point of this definition of capabilities is that determination of capability in context must be done flexibly, since there are a wide variety of ways in which performance of tasks can be rearranged to accomodate multiple on-going activities. One primary mechanism is delaying processing of tasks that are not as time-critical. This naturally leads to a scheduling metaphor (Tulga and Sheridan, 1980). Various scheduling algorithms can be drawn from other fields, such as real-time systems. While exact solutions to these problems are often provably intractable, reasonably efficient approximation algorithms often exist (e.g. greedy, earliest deadline first, most-difficult task first, etc.). A major open question is: which approximations seem to correspond to the kinds of heuristics humans use in deciding how to carry out multiple tasks in complex environments?

One unique characteristic of this application of scheduling is that, in addition to manipulating start and end times, another dimension taken into account is the level of effort. In other words, individuals have the option of reducing or increasing their resources allocated to a given task, which can result in a correponding increase or decrease in duration required to produce equivalent performance. Hence, one may decide to defer processing of a new task until the existing ones are complete, or, if there is insufficient time, may decide to begin processing the new task right away by shifting some of their emphasis or attention away from the current tasks, as long as it will not threaten their successful completion.

Using our scheduling-based definition of task performance under resource constraints (both internal and external), we can implement a concrete, computational method for agents to estimate workload of humans and use it to simulate decision-making about when they are likely to accept or reject new tasks in a dynamic environment. Specifically, the model would predict task acceptance if and only if a feasible schedule can be found (at least by a reasonably plausible heuristic method) that would accomodate the new task along with all existing ones, where they would all be completed in time to meet their respective deadlines, and the effort requirements (workload) would not exceed the limits (maximum capacity) of the internal cognitive resource.

A pragmatic issue in developing such a computational method is that the performance characteristics for each individual would need to be derived. We believe that these parameters can be inferred from empirical observations under various controlled conditions by using data-mining techniques, but a detailed description of the methodology is outside the scope of this paper.

Extension to Multiple Resources

The problem with the model as we have presented up to this point is that it is based on a single-resource assumption; thus it cannot account for variable degrees of interaction among tasks of different types. To extend our model to incorporate multiple resouces, we start by assuming that there is a fixed set of resource pools, $r_1 ... r_n$. For example, these might represent the eight components in Wickens' (2000) model, with resources for: auditory input processing, visual input processing, perceptual/central processing, response processing, spatial processing, verbal processing, manual response processing, and speech response processing. Each of these is postulated to be used to different degrees (possibly zero) by any given task.

Thus, instead of a univariate curve for the performace-resource function, we have in principle a function of n dimensions, representing allocation of each resource independently (Tsang and Velasquez, 1996). However, to keep the model manageable (and for parsimony), we instead use a "profile" for each task to represent, under single-task, full-attention conditions, the relative amounts of each resource required: $\langle u(r_1), ..., u(r_n) \rangle$ (as illustrated in Figure 5). Then each resource level is modified proportionally based on what fraction of 100% attention, $att(\tau)$, is allocated to a given task τ in a specific situation, effectively parameterizing the resource demands. Hence the task demand, distributed over the various resource components, becomes:

$$\langle att(\tau) \cdot u(r_1, \tau), ..., att(\tau) \cdot u(r_n, \tau) \rangle$$

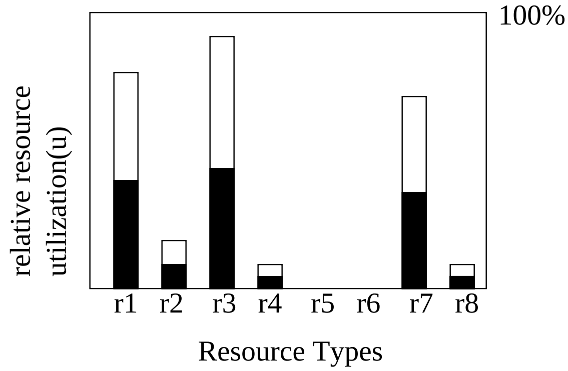

Figure 5: Resource-demand profile for a hypothetical task. The open bars represent the relative amount of demand by a task on each type of internal (cognitive) resource when it is being given full attention. The solid bars show resource allocation levels scaled down proportionally for a case where an individual is only able to focus half of the requisite attention on this particular task.

This approach allows us to treat task performance as intrinsically multi-dimensional. Now the rest of the model (including scheduling) can be applied as before, with the condition that, regardless of how the tasks combine or overlap, no individual resource may exceed its capacity at any point in time:

$$\forall t \; w_k(t) = \Sigma_i \; att(\tau_i, t) \cdot u(r_k, \tau_i) \leq u_{max}(r_k)$$

for all resource pools k, where i is summed over all tasks being executed at time t. The task difficulty and quality (i.e. accuracy) requirements set the level of effort required for the individual, the individual chooses a suitable duration and corresponding level of average emphasis to apply, and then this is used to compute utilization of each resource based on a scaled version of the single-task utilization profile. To determine whether an individual is capable of accepting a new task in the context of existing ones, a schedule must be sought that allows all tasks to complete within the time and quality constraints, while violating no limits on internal resources.

The primary benefit of this multi-dimensional model is that it can be used to simulate different degrees of interference among tasks depending on their type. For example, even though tasks A, B, and C are considered equally demanding, it might be more efficient to process A and C in parallel than A and B. This effect could be captured by saying that the profiles for A and B both share high demand for the same underlying resource, while the components utilized by A and C are relatively distinct. The phenomenon of differential interference based on task type has been called "structural similarity" in the literature (Wickens and Holland, 2000). Our work is intended to form a preliminary basis for theoretical and empirical modeling of this effect.

Discussion

Capabilities and workload are one part of the "shared mental model" that must be computed, along with others' beliefs, goals, situations, etc., to generate believable simulations of teamwork. This model could be applied to enhancing the simulation and generation of teamwork by influencing role selection, delegation, negotiation, and pro-active behavior. For example, responsibilities could be re-defined to take into account the degree to which one is capable of doing something, delegation policies and task allocation strategies could be modified to reflect an awareness of individuals' workload (i.e. to select a member for whom it would least interfere), and agents could adjust their initiative in offering to help team members with tasks based on an assessment of how over-loaded they are versus how much of a distraction it would be.

An important application of this computational model of capabilities could be for designing intelligent agents for use in team-training systems. Specifically, this model would allow agents to monitor, exercise, and evaluate individuals' ability on human teams to appropriately and effectively participate in the teamwork, as a function of their own skills, workload response, and attention management strategies. The goal would be the development of novel training interventions that could promote the balance of the cognitive demands of taskwork versus teamwork (i.e., spending time reasoning about each other).

Acknowledgments

This work was supported in part by MURI grant #F49620-00-1-0326 from DoD and AFOSR.

References

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In Meyer, D. and Kornblum, S., editors, *Attention and Performance XIV*. Cambridge, MA: MIT Press.

Hendy, K., Liao, K., and Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors*, 39:30–47.

Huey, M. and Wickens, C. (1993). *Workload transition: Implications for individual and team performance*. Washington, D.C.: National Academy Press.

Kleinman, D., Luh, P., Pattipati, K., and Serfaty, D. (1992). Mathematical models of team performance: A distributed decision-making approach. In Sweezy, R. and Salas, E., editors, *Teams: Their Training and Performance*. New York: Ablex.

Kozlowski, S. (1997). Training and development of adaptive teams. In Cannon-Bowers, J. and Salas, E., editors, *Making Decisions Under Stress*, pages 115–153. Washington, D.C.: American Psychological Association.

Moore, R. (1985). A formal theory of knowledge and action. In Hobbs, J. and Moore, R., editors, *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex.

Norman, D. and Bobrow, D. (1976). On the analysis of performance operating characteristics. *Psychological Review*, 83:508–510.

Rickel, J. and Johnson, W. (1997). Intelligent tutoring in virtual reality: A preliminary report. In *Proc. of the International Conference on Artificial Intelligence in Education*, pages 294–301.

Serfaty, D., Entin, E., and Johnston, J. (1997). Team coordination training. In Cannon-Bowers, J. and Salas, E., editors, *Making Decisions Under Stress*, pages 221–245. Washington, D.C.: American Psychological Association.

Singh, M. (1991). A logic of situated know-how. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 343–348.

Tsang, P. and Velasquez, V. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39:358–381.

Tulga, M. and Sheridan, T. (1980). Dynamic decisions and workload in multitask supervisory control. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(5):217–232.

van der Hoek, W., van Linder, B., and Meyer, J. C. (1994). A logic of capabilities. In *Lecture Notes in Computer Science*, volume 813, pages 366–378. Springer-Verlag.

Wickens, C. (1984). Processing resources in attention. In Parasuraman, R. and Davies, D., editors, *Varieties of Attention*. New York: Academic Press.

Wickens, C. and Holland, J. (2000). Attention, time-sharing, and workload. In Wickens, C. and Holland, J., editors, *Engineering Psychology and Human Performance*, chapter 11. Upper Saddle Reiver, NJ: Prentice Hall, 3rd edition.