Mental Workload and Situation Awareness

Chapter · February 2006

DOI: 10.1002/0470048204.ch9

Pamela S Tsang
Wright State University

CHAPTER 9

MENTAL WORKLOAD AND SITUATION AWARENESS

Pamela S. Tsang
Wright State University
Dayton, Ohio

Michael A. Vidulich
Air Force Research Laboratory
Wright-Patterson Air Force Base, Ohio

1 INTRODUCTION

Human engineering seeks to understand and improve human interactions with machines to perform tasks. This goal can be especially difficult to achieve in dynamic complex systems that characterize much of modern work. A good example of the problem can be seen in considering the human operator's reaction to automation. Many modern tasks, such as controlling the complex reactions in a process control plant (Wickens and Hollands, 2000), would not be possible without the assistance of automation. However, many researchers (e.g., Kessel and Wickens, 1982; Moray, 1986; Wiener and Curry, 1986; Bainbridge, 1987; Tsang and Vidulich, 1989; Adams et al., 1991) have pointed out that automated assistance can come at a high price. Moray (1986) expressed this issue well when he noted that as the computer automation did more, the operator would do less and therefore experience less mental workload, but: "Is there a price for the advantages? It could be said that the information processing demands become so alien to the operator that if called upon to reenter the control loop such reentry is no longer possible. ... The system will be poorly understood and the operator will lack practice in exercising control, so the possibility of human error in emergencies will increase" (Moray, 1986, p. 40–5).

Naturally, there is an issue of how well the requirements of using any machine matches the capabilities of the human operators. Some mismatches should be easy for an observer to see, especially physical mismatches. For example, Fitts and Jones (1961a) found that 3% of 460 "pilot errors" were due to the pilot being physically unable to reach the required control. But other aspects of how well the human operator can match the requirements of using a machine to accomplish a task may not be so obvious, especially mental mismatches. For example, Fitts and Jones (1961b) found that 18% of errors in reading and interpreting aircraft instruments were due to problems associated with integrating the information from multirevolution indicators (e.g., altitude displays with different pointers for one, tens, and hundreds of feet). An outside observer would not necessarily know by watching the pilot that such a misinterpretation had occurred.

Given the impossibility of seeing the mental processes of an operator performing a task, it is not surprising that human engineering specialists have developed concepts to relate the impact of various task demands on the human operator and on system performance. Vidulich (2003) identified two such concepts: mental workload and situation awareness (SA). Vidulich echoed the argument of other researchers that mental workload and SA had taken on the quality of meta-measures for system evaluation (Hardiman et al., 1996; Selcon et al., 1996). Selcon et al. (1996) pointed out that it is seldom possible to evaluate all of the mission-relevant uses of a display. Thus, any evaluation must generalize from the evaluation environment to the real world. It has been argued that rather than focusing on task-specific performance, it might be preferable to examine meta-measures that encapsulate the cognitive reaction to performance with a given interface. So, for example, an interface that allows a task to be performed with a more comfortable level of mental workload and better SA would be preferable to one that did not.

To appreciate the potential roles that workload and SA might play in supporting system development, several visions of future systems will be considered. Fallows (2001 a,b) presented an intriguing vision of the future of aviation. Examining the increasing bottlenecks and delays inherent in the existing airline industry, Fallows proposed that a simple scaling-up of the existing system with more planes and more runways at existing airports would not be a practical nor economically feasible approach to keep pace with the projected increases in airline travel. Fallows made a compelling argument that the increased reliability of aircraft mechanical systems combined with innovative research on cockpit interfaces will not only revitalize general aviation, but also lead to the emergence of a much more extensive small aircraft "taxi" service. Such an expansion would naturally lead to more pilots flying that lack the extensive training of current professional pilots. To maintain an acceptable level of safety, Fallows assumes that interface changes will occur to make the demands of piloting an aircraft tolerable for a wide range of pilots.

In the current context it is important to emphasize the role that human engineering is expected to assume in the development of the expanded air system. Fallows (2001 a,b) points out that current research, such as that performed by the National Aeronautics and Space Administration's (NASA) Advanced General Aviation Transport Experiments (AGATE) alliance, will decrease the difficulty of piloting and increase flight safety for a new generation of more user-friendly aircraft. NASA is aggressively pursuing the goal of making advanced cockpit technologies effective and affordable even to general aviation pilots. These technologies include highway-in-the-sky displays, head-mounted displays (Fiorino, 2001), and synthetic vision systems that use advanced sensors to present a view of the world to the pilot during degraded visual conditions (Wynbrandt, 2001). In other words, Fallows expects that future general aircraft cockpits will take advantage of advanced interface technologies to reduce the mental workload and increase the SA of the general aviation pilot. To this end, understanding the human pilot and building systems that best accommodate the human's cognitive strengths while supporting human frailty will remain a vital component of making those systems effective and safe.

In addition to changes to the cockpit, air traffic control (ATC) is also expected to face changing demands. Jha et al. (2003) discuss the proposed future architecture of the National Airspace System (NAS). Most proposed changes are expected to increase the pilot's autonomy in controlling the route of their aircraft, including the requirement to maintain required separation between aircraft. Flight time, delays, and fuel expenditures are all expected to profit from such changes. Despite the increase in pilot autonomy, controllers are still responsible for running the ATC system safely. In other words, the controller's role would be shifted from that of an active controller to one more like that of a monitor. Assuming that the equipment and the pilots all perform correctly, the controller's workload would be expected to decrease since they are interacting with fewer aircraft. It also seems plausible that the controller's SA of the airspace would decrease as well, since they would not be focusing as much attention as they previously did on many of the aircraft. However, as Metzger et al. (2003) point out, this may not be the case. For an unknown period of time, certain aircraft in the system would be properly equipped to participate according to the new rules, allowing more autonomy, but others would not be and would require the controller to manage their flight path. Metzger et al. (2003) explored the effect of the proportions of equipped and unequipped aircraft in the controller's airspace and the availability of a decision support system. The decision support was largely effective. Controllers were more able to control mixed types of aircraft, and rated their workload lower when decision support was available. Although Metzger et al. (2003) did not discuss their results in terms of SA, there were interesting findings of relevance. When the decision support tool was available, the controllers' attention (measured by eye fixations) to the radar display was partially drawn away to the decision support tools. This behavior could presumably reduce the controller's SA of information on the radar screen. The impact of such automated decision support tools will require careful analysis to ensure that important information from other sources is not lost due to their use. It is important to consider the impact of any technological support in the larger context of overall system performance and workload imposed on the human operator.

So technological changes constantly bring about a practical need to know about the cognitive processing of the operator. There seems to be a consensus that the concepts of SA and mental workload are useful for assessing the impact of such changes on the human operator. But although mental workload and SA are both concepts for understanding the human reaction to interacting with a system to achieve a goal, as we shall see, they are separate concepts. The utility of assessing SA therefore does not diminish that of mental workload. In fact, parallel studies of both are not only good for applied evaluations but could also help sharpen their respective definitions and stimulate new understanding (Endsley, 1993; Wickens, 2001; Vidulich, 2003).

In the next section we review a framework for understanding both mental workload and SA as parts of human information processing. Then select approaches for assessing mental workload and SA are reviewed. Finally, the roles of mental workload and SA in meeting the demands of future systems are considered.

2 THEORETICAL UNDERPINNINGS OF MENTAL WORKLOAD AND SITUATION AWARENESS

In 1994, Pew stated that *situation awareness* had replaced *workload* as the buzzword of the 1990s. But can the concept of situation awareness replace that of mental workload? Hendy (1995) and Wickens (1995) argued that the concepts of situation awareness and mental workload are clearly distinct but are also intricately related to each other. A decade later, there now seems to be a consensus that one concept does not replace, or encompass, the other, even though the two concepts are affected by many of the same human variables (such as limited processing capacity and the severe limit of working memory) and system variables (such as task demands and technological support).

Figure 1 provides a conceptual sketch of the relationship between mental workload and situation awareness and is not intended to be a complete representation of all the processes that are involved. There are two main components in this figure: the attention and mental workload loop and the memory and situation awareness loop. The ensuing portrayal will make clear that mental workload and situation awareness are intricately intertwined, as one affects and is affected by the other. Although convention would bias us in thinking that elements on top or on the left in the figure might have temporal precedence over those at the bottom or on the right, this is not necessarily the case with the dynamic interplay between workload and SA. For example, task demands could be initiated by an external task (such as the need to respond to an air traffic controller's request) as well as by an internal decision to engage in solving a nagging problem. Despite the seemingly discrete and linear depiction of the relations among the elements in the figure, the elements are actually thought of to be mutually interacting adaptively in response to both exogenous demands and endogenous states (e.g., Hockey, 1997).

2.1 Attention and Workload

Since the 1970s, much has been debated and written about the concept of mental workload (e.g., Welford, 1978; Moray, 1979; O'Donnell and Eggemeier, 1986; Adams et al., 1991; Huey and Wickens, 1993; Gopher, 1994; Kramer et al., 1996; Tsang and Wilson, 1997). Gopher and Donchin (1986) offered the following description of the role of mental workload:

> The term workload is used to describe aspects of the interaction between an operator and an assigned task. Tasks are specified in terms of their structural properties; a set of stimuli and responses are specified with a set of rules that map responses to stimuli. There are, in addition, expectations regarding the quality of the performance, which derive from knowledge of the relation between the structure of the task and the nature of human capacities and skills. . . . [W]orkload is invoked to account for those aspects of the interaction between a person and a task that cause task demands to exceed the person's capacity to deliver. . . . [M]ental workload is clearly an attribute of the information processing and control systems that mediate between stimuli, rules, and responses. (p. 41–3)

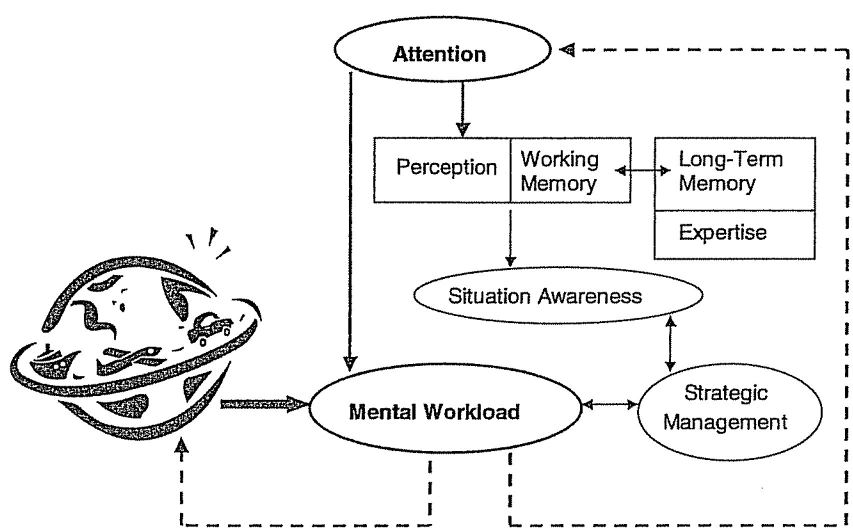

Figure 1 Theoretical framework illustrating the relationship between mental workload and situation awareness.

A commonly accepted notion is that mental workload is very much a function of the *supply and demand* of attentional or processing resources. The attention–workload loop in Figure 1 is minimally sketched but is described in more detail here. There are two main determinants of workload: the exogenous task demands as specified by factors such as task difficulty, task priority, and situational contingencies (represented by the globe in Figure 1); and the endogenous supply of attentional or processing resources to support information processing such as perceiving, updating memory, planning, decision making, and response processing. Further, this supply is modulated by individual differences such as one's skill level or expertise. The ultimate interest in measuring workload, of course, is how it might affect system performance represented via the feedback loop to the globe in Figure 1. Mental workload can be expressed in subjective experience, performance, and physiological manifestations. A host of assessment techniques have now been developed and are used in both laboratories and applied settings. They are reviewed in Section 3.

Although there are numerous theoretical accounts of attention, the one readily embraced and adopted in the workload literature is the energetics account (e.g., Hockey et al., 1986). Central to the present discussion is the notion that attentional resources are demanded for task processing, but they are of limited supply. Performance improves monotonically with increased investment of resources up to the limit of resource availability (Norman and Bobrow, 1975). An important implication of this relationship is that performance could be the basis of inference for the amount of resources used and remained. The latter, referred to as *spare capacity*, could serve as reserve fuel for emergencies and unexpected added demands. Further, attentional resources are subject to voluntary and strategic allocation. According to Kahneman (1973), attention is allocated via a closed feedback loop with continuous monitoring of the efficacy of the allocation policy that is governed by enduring dispositions (of lasting importance, such as one's own name and well-learned rules), momentary intentions (pertinent to the task at hand), and evaluation of the performance (involving self-monitoring of the adequacy of performance in relation to task demands). Because attention can be deployed flexibly, researchers advocate the need to examine the allocation policy in conjunction with the joint performance in a multitask situation in order to assess the workload and spare capacity involved (e.g., Gopher, 1994).

Among the most convincing support for the limited, energetic, and allocatable property of attentional resources are the reciprocity effects in performance and certain neuroindices observed between time-shared tasks. As the demand or priority of one task changes, the increase in performance, P300 amplitude, or PET-measured activity in one task has been observed to be accompanied by a decrease in the corresponding measures in the other task (Gopher et al., 1982; Wickens et al., 1983; Kramer et al., 1987; Sirevaag et al., 1989; Fowler, 1994; Tsang et al., 1996; Parasuraman and Caggiano, 2002; Just et al., 2003).

By the late 1970s, the notion of Kahneman's undifferentiated or all-purpose attentional resource was challenged by a body of data that suggested multiple specialized resources for different types of processing (see Allport et al., 1972; Kinsbourne and Hicks, 1978; Navon and Gopher, 1979; Friedman and Polson, 1981; Wickens, 1984). Based on an expansive systematic review of the interference pattern in the extant dual-task data, Wickens (1980, 1987) proposed a multiple resource model. According to this model, attentional resources are defined along three dichotomous dimensions: (1) stages of processing with perceptual/central processing requiring resources different from those used for response processing, (2) processing codes with spatial processing requiring resources different from those used for verbal processing, and (3) input/output modalities with visual and auditory processing requiring different processing resources and manual and speech responses also requiring different processing resources. An important application of this model is its prediction of multiple-task performance that is common in many modern complex work environment. The higher the similarity in the resource demands among the task components, the more severe the competition for similar resources, the less spare capacity, and the higher the level of workload that would result. The other side of the coin is that it would be less feasible to exchange and reallocate resources among task components that utilize highly dissimilar resources (see Wickens, 2002). That is, it would be more difficult to manage the workload levels dynamically between tasks that rely on dissimilar resources. According to the multiple resource model, the intensity aspect and the structural aspect are intrinsically intertwined when characterizing the processing demand of a task. Similarity in the resource demand among the time-shared tasks effectively determines resource availability as well as resource exchangeability. An increased degree of resource similarity promotes resource sharability but effectively reduces resource availability that leads to increased resource competition.

The energetic and specificity aspects of the attentional resources are receiving converging support from subjective (e.g., Tsang and Velazquez, 1996; Rubio et al., 2004), performance (e.g., Tsang et al., 1996; Wickens, 2002), and neurophysiological (e.g., Just et al., 2003; Parasuraman, 2003) measures. First, parametric manipulation of task demands have been found to produce systematic and graded changes in the level of subjective workload ratings, performances, and amount of neuronal activation. Second, all of these measures have been found to be sensitive to the competition for specific resource demands. Further, increased neuronal activation associated with different types of processing (e.g., spatial processing and verbal processing) are found to be localized in different cortical regions.

As mentioned above, the supply or availability of processing resources is subject to individual differences such as one's ability and skill level. Recently, Just et al. (2003) pointed out a set of neurophysiological data that appear to support the notion that a higher level of skill or ability effectively constitutes a larger resource supply. Parks et al. (1988) used a verbal fluency task that required subjects to generate as many words as possible that began with a given stimulus letter. Those who were more proficient at this task, and presumably had higher verbal ability, exhibited a lower level of positron emission tomography (PET) measures of brain activity. Just et al. (2003) proposed that the difference between the more and less proficient subjects lay in the proportion of resources they needed to perform the task. Since the same task was performed, the task demand objectively should be the same for all the subjects. The lower level of brain activity therefore would suggest that the more proficient subjects had a larger supply of resources. In another study, Haier et al. (1992) found that weeks of practice in the spatial computer game Tetris led to improved performance and a reduced amount of PET-measured activity. Just et al. proposed that practice improved the subjects' procedural knowledge, and the newly acquired, more efficient procedures entailed a lower level of resource use. In practice, a reduced level of resource requirement by one task would translate to increased spare resources for processing other tasks.

2.2 Memory and Situation Awareness

Like the concept of mental workload (and many other psychological concepts, such as intelligence), situation awareness (SA) is difficult to define precisely (e.g., Durso and Gironlund, 1999). Pew (1994) defines a situation as "a set of environmental conditions and system states with which the participant is interacting that can be characterized uniquely by a set of information, knowledge and resource options" (p. 18). A commonly referenced working definition for SA came from Endsley (1990, p. 1-3): situation awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." An on-line dictionary states that aware "implies knowledge gained through one's own perceptions or by means of information." These definitions connote both perception of the now and present and connection with knowledge gained in the past. As Figure 1 denotes, SA is most closely linked to the perceptual and the working memory processes. Certainly, it is not sufficient just for the information relevant to the situation to be available; the information needs to be perceived by the operator, and perception entails far more than the detection of signals or changes. For pattern recognition, object categorization, and comprehension of meaning to occur, contact with knowledge is necessary. But knowledge stored in long-term memory is accessible only through short-term or working memory. Baddeley (1990) introduced the term working memory to emphasize that short-term memory is far more than a temporary depository for information: It is an active process involved in maintaining information available in short-term memory. Working memory is an effortful process subject to capacity as well as attentional limits.

Adams et al. (1995) make a distinction between the process and product of SA: "*product* refers to the state of awareness with respect to information and knowledge, whereas *process* refers to the various perceptual and cognitive activities involved in constructing, updating, and revising the state of awareness" (p. 88). To elaborate, although only the perceptual and working memory processes are explicitly linked to SA in Figure 1, SA is supported by other processes that are subject to attentional limits. The product of SA is a distillation of the ongoing processing of the interchange between information perceived from the now and present (working memory) and knowledge and experience gained from the past (long-term memory). Both the process and product are influenced by one's experience. As will be made clear below, this distinction between the process and product of SA has profound implications on the interaction of mental workload and SA and on the appropriate assessment techniques.

Just as given the same objective task demand, mental workload could vary due to individual differences in resource supply as a result of skill and ability differences, given the same situation, SA could vary due to individual differences. Although long-term memory is not linked directly to SA in Figure 1, it plays a critical role since it stores the knowledge and experience associated with skill development. Ericsson and Simon (1993) pointed out: "Recognition and retrieval processes are determined in part by information in LTM [long-term memory], because the information in STM [short-term memory] is not sufficiently specific to determine a unique product of recognition and retrieval" (p. 197). The extant view of the nature of expertise further expounds on the role of memory in determining the content of SA.

Expertise is mostly learned, acquired through many hours of deliberate practice (e.g., Glaser, 1987; Chi et al., 1988; Druckman and Bjork, 1991; Adams and Ericsson, 1992; Ericsson, 1996). A fundamental difference between novices and experts is the amount of acquired domain-specific knowledge. In addition to having acquired declarative knowledge (facts), experts have a large body of procedural (how-to) knowledge. With practice, many procedural rules (productions) become concatenated into larger rules that can produce action sequences efficiently (Druckman and Bjork, 1991). However, the expertise advantage goes beyond a quantitative difference. The organization of knowledge is fundamentally different between experts and novices. An expert's knowledge is highly organized and well structured, so that retrieving information is much facilitated. The large body of organized knowledge enables experts readily, to see meaningful patterns, to make inferences from partial information, to constrain search, to frame the problem, to apprehend the situation, to update perception of the current situation continuously, and to anticipate future events, including eventual retrieval conditions (Glaser, 1987; Charness, 1995; Vidulich, 2003). An accurate account of the current situation allows an experienced operator to retrieve rapidly the appropriate course of action directly from memory, enabling swift suitable responses.

In addition, Ericsson and Kintsch (1995) proposed that a long-term working memory (LTWM) emerges as expertise develops and is a defining feature of an advanced level of skill (Ericsson and Delaney, 1998). Whereas working memory has severe capacity and temporal limits, LTWM is hypothesized to have a larger capacity that persists for a period of minutes (or even hours). Experts do not have a larger memory capacity, the critical aspect is *how* information is stored and indexed in long-term memory. With a meaningful system for organizing information that already would have been built in LTM, even very briefly seen, seemingly random, incoming information might be organized similarly. Retrieval cues can then be devised and used to access information in LTM quickly. One caveat is that skilled memory is highly domain specific (Ericsson and Charness, 1997). For example, Chase and Ericsson (1982) and Staszewski (1988) reported three people who, after extensive practice, developed a digit span in the neighborhood of 100 numbers. Being avid runners, the subjects associated the random digits to facts related to running that already existed in their LTM (e.g., date of the Boston Marathon). These subjects had a normal short-term memory span when the studies began, and after practice demonstrated the normal span when materials other than digits were tested. Charness (1995) pointed out that such escapes from normal limits also have been observed in perceptual processing. Reingold and Charness (1995) found that when chess symbols (as opposed to letters designating chess pieces) were used, highly skilled players could make their decision in some cases without moving their eye from the initial fixation point at the center of the display. In contrast, weaker players had to make direct fixations. When letter symbols instead of chess piece symbols were used, even the experts were forced to fixate on the pieces directly much more often. Charness (1995) also pointed out that these observations show that experts can both accurately encode a situation and prepare an appropriate response much more quickly than their less skilled counterparts, but only in the domain of their expertise.

To further illustrate the workings of LTWM, Ericsson and Kintsch (1995) described the medical diagnosis process that requires one to store numerous individual facts in working memory. Having developed a retrieval structure in LTM that would foster accurate encoding of patient information and effective reasoning, medical experts were found to be better able to recall important information at a higher conceptual level that subsumed specific facts and to produce more effective diagnosis. A very important function of LTWM appeared then to be providing working memory support for reasoning about and evaluating diagnostic alternatives (Norman et al., 1989; Patel and Groen, 1991). That is, expert problem solving is more than just quick retrieval of stored solutions to old problems. Expertise is also associated with effective application of a large amount of knowledge in reasoning to cope with novel problems (Charness, 1989; Horn and Masunaga, 2000).

Finally, experts show metacognitive capabilities that are not present in novices (Druckman and Bjork, 1991). These capabilities include knowing what one knows and does not know, planning ahead, efficiently apportioning one's time and attentional resources, and monitoring and editing one's efforts to solve a problem (Glaser, 1987).

2.3 Mental Workload and Situation Awareness

As several researchers have emphasized, the two concepts are intricately intertwined (e.g., Wickens, 2002; Vidulich, 2003). In this section we attempt to sharpen their distinction and to examine their interactions more closely. Wickens (2001, p. 446) contrasts the two concepts in the following way: "Mental workload is fundamentally an *energetic* construct, in which the quantitative properties ("how much") are dominant over the qualitative properties ("what kind"), as the most important element. In contrast, situation awareness is fundamentally a *cognitive* concept, in which the critical issue is the operator's accuracy of ongoing understanding of the situation (i.e., a qualitative property)." In practice, one assesses the amount and type of workload and the quality (scope, depth, and accuracy) of the content of SA (Vidulich, 2003).

Both the level of workload and the quality of SA are shaped by exogenous and endogenous factors. Exogenous factors are inherent in the situation (e.g., task demands and situation complexity and uncertainty). Endogenous factors are inherent in a person's ability and skill. The same level of task demands could impose different levels of workload on the operator, depending on her ability or skill level. As discussed above, a high skill level is functionally equivalent to having a larger processing resource supply. A moderate crosswind could be a challenge for a student pilot trying to land a plane but a rather routine task for a seasoned pilot. An overly sensitive warning alarm could be exceedingly disruptive to assessment of the situation by a new operator but could safely be ignored by an experienced operator with intimate knowledge of the workings of the system. Although calibrating the exogenous demands is not always straightforward, their influences on workload is obvious. Less apparent is the endogenous influences on the interplay between the level of workload and the quality of SA.

To the extent that workload is caused and SA supported by many of the same cognitive processes, they are enabled by, and subject to the limits of, many of the same processes. The more demanding the task, the more complex the situation and the more "work" is required to get the job done and the situation assessed. By our definition, the higher the level of workload, the more attention is needed for task performance and the less is left for keeping abreast of the situation. The SA process could actually compete with task performance for the limited resource supply, and therefore a high level of workload could lead to poor SA. On the other hand, SA could be improved by working harder (e.g., more frequent sampling and updating of information). That is, a high-level workload is sometimes necessary to maintain a good SA. Thus, a high level of workload could be associated with either a low or high degree of SA (Endsley, 1993). But poor SA may or may not impose more workload. One could simply not be doing the work necessary to attain and maintain SA, and if one is not aware of the dire situation that one is in and takes no action to correct the situation, no additional work would be initiated. Although a low degree of SA is never desirable, an awareness of one's lack of SA could start a course of action that could increase the level of workload in the process of attaining or restoring SA. The ideal scenario is one where a high degree of SA would support more efficient use of resources and thereby producing a low level of workload. In short, mental workload and SA could support each other as well as compete with each other.

Strategic management is proposed to be needed for the balancing act of maintaining adequate SA without incurring excessive workload. Strategic management is also referred to as executive control and is a much discussed topic in the literature. One point of contention is what exactly constitutes executive control, since a host of higher-level cognitive functions have been included under the rubric of executive control. The coordinating of multiple tasks (including the allocation of limited processing resources), planning, chunking or the reorganizing of information to increase the amount of materials that can be remembered, and the inhibiting of irrelevant information have all been labeled as part of the executive control. As Figure 1 indicates, strategic management is skills-based and is highly dependent on one's apprehension of the situation. For example, a beginner tennis player would be content to have made contact with the tennis ball and would not have the spare resources or the knowledge to ponder game strategies. After having mastered the basic strokes (which have become more automatic), however, the strategic component would take on more central importance. But strategic management is not attention free. Even though declarative and procedural knowledge develops as expertise develops and are used to support performance, there are components in many complex performances that are never automatic. High-performing athletes, chess players, musicians, and command and control officers expend considerable effort to perform at the level that they display.

Recent neurophysiological evidence provides some support for the notion that executive control is a distinct construct and consumes processing resources. Just et al. (2003) point out that the executive system is identified primarily with the prefrontal cortex which does not receive direct sensory input but has widespread connections with a number of cortical areas associated with various types of processing (e.g., spatial and verbal processing). Further, neuropsychological patients with lesions in the frontal lobe show impairments in planning and other higher-level cognitive functions (Shallice, 1988). Importantly, a number of functional magnetic resonance imaging studies show a higher level of activation in the prefrontal cortex in (1) a problem-solving task that requires more planning than one that requires less (Baker et al., 1996), (2) a working memory task that requires more updating of a larger amount of information (Braver et al., 1997), and (3) a dual-task (a semantic category judgment and a mental rotation task) than the single-task performance (D'Exposito et al., 1999). These results show that the activation in the prefrontal cortex vary systematically with the task demand.

Returning to Figure 1, strategic management competes directly with all the processes that generate mental workload for processing resources. But strategic management could optimize performance by planning and by smartly allocating the limited resources to the processes that need resources the most to meet system requirements. An efficacious strategic management would, of course, require a high-quality situation assessment. In the last section we discuss potential human factors support (such as display support, automation aids, training) that would improve the potential of attaining this ideal scenario of a high level of SA without an exceedingly high level of workload.

3 METRICS OF MENTAL WORKLOAD AND SITUATION AWARENESS

Measures of mental workload and SA have often been divided into three categories, based on the nature of the data collected: subjective ratings, operator performance, and psychophysiological measures. There are several properties that should be considered when selecting measures of cognitive activity: sensitivity, diagnosticity, intrusiveness, validity, reliability, ease of use, and operator acceptance. In addition to the foregoing concerns, Tenny et al. (1992) caution that there is an additional matter to consider in the case of SA measures. They distinguish between the *process* of building SA and the actual awareness that is the *product* of that process. SA measures should be designed and selected based on which aspect (i.e., process or product) the evaluator wishes to assess.

As outlined below, each group of measures has its strengths and weaknesses and a thoughtful combination of measures can lead to a more complete picture. Since the various workload and SA measures have different properties, one should have a good understanding of the properties of each measure so that the most appropriate choice(s) can be made. Readers are encouraged to consult more in-depth coverage of the metrics presented here (Gopher and Donchin, 1986; O'Donnell and Eggemeier, 1986; Lysaght et al., 1989; Vidulich et al., 1994a; Bryne, 1995; Tsang and Wilson, 1997; Gawron, 2000; Vidulich, 2003).

3.1 Performance Measures

System designers are typically most concerned with system performance. Some might say that the workload or SA experienced by an operator can be important only if it affects system performance. Consequently, performance-based measures might be the most valuable to system designers. There are two main categories of performance-based workload measures: primary task performance and secondary task performance. SA assessment also has made use of primary task performance. But instead of assessing secondary task performance, SA researchers have often employed recall-based memory probe performance or real-time performance. Although primary task performance is obviously the measure that is most strongly linked to the system designer's goal of optimizing system performance, Vidulich (2003) suggested that the secondary task method of workload assessment and the memory probe method of SA assessment are prototypical measures of the theoretical concepts behind workload and SA.

3.1.1 Primary Task Performance

Primary Task Workload Assessment The primary task method of workload assessment consists of monitoring the operator's performance and noting what changes occur as the task demands are varied. This methodology is grounded in the framework presented above. Since human operators have a finite capacity to deal with the demands of a task, as that task's demands continue to increase, task performance would be expected to deteriorate, and at some point the operator will no longer able to perform the task adequately. For example, an automobile driver might have more difficulty maintaining a proper course as the weather becomes more windy, and if the wind increases even more when the road is slippery, the driver may fail completely to keep the car in the proper lane.

It should be noted that mental workload is not the only thing that can influence operator performance. The operator's level of motivation might change, for example. Kahneman (1973) suggested that the human's capacity to perform mental work is related to the person's arousal level. In addition, Kahneman also pointed out that humans can monitor their own performance, and if the performance is found wanting and there are resources available (perhaps by increasing arousal), more resources can be allocated to the task to maintain or augment performance. Recently, Salvendy and his colleagues found that including a factor that reflected a person's skill, attitude, and personality contributed significantly to the predictive value of their projective modeling technique (Bi and Salvendy, 1994; Xie and Salvendy, 2000a,b).

In considering all of these issues, Gopher and Donchin (1986, p. 41–25) concluded: "In summary, direct measures of performance on the task of interest are usually a poor indicator of mental workload because they often do not reflect variation in resource investment due to difficulty changes, they do not diagnose the source of load, and they do not make possible a systematic conversion of performance units into measures of relative demands or load on the processing system." Thus, although the primary task performance is, clearly, very important to system evaluators as a test of whether design goals have been achieved, primary task performance by itself typically does not provide an adequate test of an operator's mental workload.

Primary Task SA Assessment Despite the problems in using primary task performance as a workload measure, it has become a common tool for assessing the impact of human–machine interface modifications intended to improve SA. For example, Vidulich (2000) found that it was common for researchers to propose an interface alteration that would improve SA and test it by determining if performance improved when the alteration was in place. The logic behind primary task performance based measures of SA is well illustrated by Andre et al. (1991). In a study of aircraft cockpit design, Andre et al. (1991) postulated that the pilot's ability to recover from disorienting events was a direct measure of how well the attitude information provided by the cockpit supported the pilots' SA of current and future attitudes of the aircraft. Following their stated logic, Andre et al. (1991) concluded that the display incorporating inside-out reference frames produced better performance and supported superior SA than the alternatives studied.

3.1.2 Secondary Task Measures of Workload

The secondary task measure of workload has been considered the prototypical measure of mental workload (e.g., Ogden et al., 1970; Gopher, 1994; Vidulich, 2003). A system evaluator would usually desire to assess primary task performance even if it was not being interpreted as a mental workload indicator. In contrast, a secondary task is usually only incorporated in a system assessment for assessing mental workload. More important, the secondary task technique is a procedure that is optimally suited to reflect the commonly accepted concept of mental workload (see the theoretical framework above). Workload is often assessed to determine whether the human operator is working within a tolerable information-processing capacity while performing the required task. It follows logically that if there is unused capacity, the operator could perform another task. For example, it is expected that spare capacity would be very valuable in emergencies or when under stress (Wickens, 2001; Hockey et al., 2003).

The secondary task measure of mental workload also offers some practical advantages for workload assessment in comparison to primary task performance assessment. Many secondary tasks have been developed and calibrated for use in different evaluations (Gawron, 2000). Many of these tasks vary the resources demanded as characterized by multiple resource theories. Presumably, this allows an evaluator to select secondary tasks to compete for specific resources of the primary task. Thus, a well selected secondary can be diagnostic of the primary task's resource demands.

The secondary task measure can be assessed in environments where primary task performance is difficult to obtain or is not available. This is often the case in many real-world systems, such as automobiles, ships, and airplanes that do not have performance-recording capability. Also, with highly automated systems in which the primary role of the operator is that of monitoring and supervising, little observable performance would be available for analysis. Finally, as noted above, primary task measures may not be sensitive to very low workload levels because operators could increase their efforts to maintain a stable level of performance (e.g., O'Donnell and Eggemeier, 1986). Adding a secondary task will increase the overall task demand to a level that performance measures may be more sensitive.

With the secondary task method, the operator is required to perform a second task concurrently with the primary task of interest. It is explained to the operators that the primary task is more important and the primary task performance must be performed to the best of their ability whether or not it is performed with the secondary task. Operators are to use only their spare capacity to perform the secondary task. Since the primary and secondary tasks would compete for the limited processing resources, changes in the primary task demand should result in changes in the secondary task performance as more or less resources become available for the secondary task.

These changes can be interpreted within a performance operating characteristic (POC) representation. As illustrated in Figure 2*a*, the performance trade-off between tasks can be examined by plotting them within a space representing their joint performance. In other words, the POC reflects the subject's allocation strategy for distributing attention between the time-shared tasks. Figure 2*a* shows possible trade-off between Tasks A and B performed as a dual-task combination. If the two tasks did not compete for any resources, perfect time-sharing could be observed. This is represented by the "X" on the figure, which would indicate that both tasks were performed at their respective single-task levels when performed together. Such perfect time-sharing is rare, but has been observed (e.g., Allport et al., 1972). The data lines in Figure 2*a*, being much closer to the origin of the graph than the perfect time-sharing point, indicate substantial interference in dual-task conditions. Such interference would typically show up in dual-task studies that manipulated the relative priorities of the two tasks. The dashed line shows a perfect trade-off pattern between the tasks. As one task's performance improves by a certain amount, a comparable degradation is observed in the other task. The dotted line shows the two task's joint performance being somewhat better than the perfect trade-off case. This would be expected to occur if the two tasks required at least some different types of information-processing resources.

The secondary task procedure differs from the standard laboratory dual-task study in that the priorities are not usually manipulated in the latter. The primary task's performance must be defended. Figure 2*b* illustrates what can be expected in this situation. In this example, the primary task performance (x-axis) of two possible interfaces is being evaluated. In this hypothetical example, the subjects have done a good job of following the secondary task (y-axis) instructions and are performing the primary task at very near the single-task level. Notice that both primary task interfaces (Y and Z) are maintaining the same level of primary task performance. However, interface Y's secondary task performance is substantially better than interface Z's secondary task performance. This result would be interpreted as interface Y inflicting less workload on the operator while performing the primary task than would interface Z.

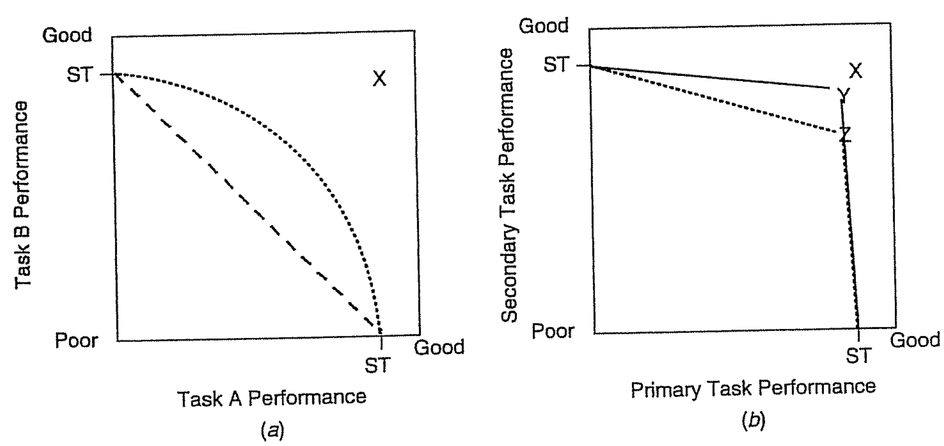

Figure 2 (*a*) Hypothetical performance operating characteristic (POC). Tasks A and B are two tasks that have been performed both independently and together. ST = level of single-task performance. The dashed and dotted lines illustrate possible joint performance when the two tasks are performed together. X, perfect time-sharing. (*b*) Hypothetical secondary task POC. The primary task was performed with two different interfaces. Y and Z, joint performances observed when the secondary task is performed with the two versions of the primary task.

An important consideration in the selection of a secondary task is the type of task demand of both the primary and secondary tasks, according to the logic of multiple-resource theory. Secondary task performance will be a sensitive workload measure of the primary task demand only if the two tasks compete for the same processing resources. The greater the dissimilarity of the resource demands of the time-shared tasks, the lower the degree of the interference there would be between the two tasks. Although a low degree of interference usually translates to a higher level of performance (which of course is desirable), this is not compatible with the goal of workload assessment. A fundamental assumption of the secondary task method is that the secondary task will compete with the primary task for limited processing resources. It is the degree of interference that is used for inferring the level of workload. Care must therefore be taken to assure that the secondary task selected demands resources similar to those of the primary task.

One drawback of the secondary task method is that the addition of an extraneous task to the operational environment may not only add to the workload, but may fundamentally change the processing of the primary task. The resulting workload metric would then be nothing more than an experimental artifact. The *embedded secondary task technique* was proposed to circumvent this difficulty (Shingledecker, 1984; Vidulich and Bortolussi, 1988). With this method, a normally occurring part of the overall task is used as the secondary task. In some situations, such as piloting a jet fighter aircraft, task shedding is an accepted and taught strategy that is used when primary task workload becomes excessive. Tasks that can be shed can perhaps serve as naturally lower-priority embedded secondary tasks in a less intense workload evaluation situation. However, a naturally lower-priority operational task may not always be available. Another drawback is that using the secondary task method requires considerable background knowledge and experience to properly conduct a secondary task evaluation and to interpret the results. For example, care must be taken to control for the operator's attention allocation strategy, so as to assure that the operator is treating the primary task as a high-priority task. The use of secondary tasks may also entail additional software and hardware development.

Despite the drawbacks and challenges of using the secondary task procedure, it is still used profitably for system assessment. For example, Ververs and Wickens (2000) used a set of secondary tasks to assess a simulated flight path following and taxiing performance with different sets of HUD symbology. One set of symbology presented a "tunnel in the sky" for the subjects to follow during landing approaches. The other display was a more traditional presentation of flight director information. The tunnel display reduced the subject's flight path error during landing. Subjects also responded more quickly and accurately to secondary task airspeed changes and were more accurate at detecting intruders on the runway. However, the other display was associated with faster detections of the runway intruder and quicker identification of the runway. The authors concluded that although the tunnel display produced a lower workload during the landing task, it also caused cognitive tunneling that reduced sensitivity to unexpected outside events. More recently, Leyman et al. (2004) used a secondary task to assess the workload associated with various simulated office tasks of varying complexity. The subjects in the experiment typed a practiced paragraph as the secondary task which was time-shared with a random word memory task of varying list lengths, a geographical reasoning task, or a scheduling task. The secondary task typing performance showed significant degradation. The secondary task results validated the workload assessments of a new electromyographic (EMG) office workload measure.

3.1.3 Memory Probe Measures of Situation Awareness

The first popular and standardized procedure for assessing SA was the memory probe technique. It can be considered the prototypical SA measurement tool (Vidulich, 2003). Any memory probe technique attempts to assess at least part of the contents of memory at a specific time during task performance, so it assesses the product of SA processes. As represented by the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1988, 1990), the typical memory probe procedure consists of unexpectedly stopping the subject's task, blanking the displays, and asking the subject to answer questions to assess his or her knowledge of the current situation. The questions asked are typically drawn from a large set of questions that correspond to experimenter's assessment of the SA requirements for task performance. The subject's answers are compared to the true situation to determine the SAGAT score. Vidulich (2000) found that this SAGAT-style approach with unpredictable measurement times and random selection of queries from large sets of possible questions was generally sensitive to interface manipulations designed to affect SA. In contrast, as memory probes were made more specific or predictable, the sensitivity to interface manipulations appeared to be diminished. For example, Vidulich et al. (1994b) used a memory probe procedure in which the memory probe, if it appeared, was at a predictable time and the same question was always used. This procedure failed to detect a beneficial effect of display augmentation to highlight targets in a simulated air-to-ground attack, even though there was a significant benefit in task performance (i.e., more targets destroyed) and a significant increase in SA ratings. In contrast, Vidulich et al. (1995) used a SAGAT-like approach with many different questions that were asked during unpredictable trial stoppages. In this case the memory probe data showed a significant SA benefit of the presence of a tactical situation display.

Although the memory probe procedure is attractive due to its assessing the information possessed by the subject at specific moment in time, it does have practical constraints that limit its applicability. First, it is clearly intrusive to stop task performance unexpectedly to ask questions. Endsley (1988) demonstrated that the performance of a simulated air-to-air combat task in trials that included SAGAT stoppages did not significantly differ from trials that did not. But even if the SAGAT stoppages are always unobtrusive to simulator performance, there are assessment environments where such stoppages are impossible (e.g., actual airplane flight tests). Also, the number of questions required that must be selected randomly and presented unpredictably can result in a large number of trials being needed to assess all questions.

The research of Strater et al. (2001) can be considered a typical SAGAT evaluation. They assessed U.S. Army platoon leaders in simulated Military Operations on Urbanized Terrain (MOUT) exercises. The platoon leaders varied from relatively inexperienced lieutenants to relatively experienced captains. SAGAT data were collected in a scenario that had the soldiers assaulting an enemy position and a scenario that involved defending a position. SAGAT probe questions were developed that could be used in either scenario. Results showed that the soldiers were more sensitive to different information depending on the scenario type. For example, in the assault scenarios the soldiers were more sensitive to the location of adjacent friendly forces than they were in the defend scenario. In the defend scenario, soldiers were more sensitive to the location of exposed friendly elements. Significant effects of soldier experience level were also detected. For example, experienced soldiers were more sensitive to enemy locations and strength than were inexperienced soldiers. The authors suggested that the data collected from such experimentation could be used to improve training programs, by helping to identify better information-seeking behaviors for the novices.

3.1.4 Situation Awareness Real-Time Performance Assessment

Real-time performance has been used as a potential indicator of SA (Durso et al., 1995; Pritchett and Hansman, 2000; Vidulich and McMillan, 2000). The logic of assessing real-time performance is based on the assumption that if an operator is aware of task demands and opportunities, she will react appropriately to them in a timely manner. This approach, if successful, would be unintrusive to task performance, diagnostic of operator success or failure, and potentially useful for guiding automated aiding. Since the continuous stream of operator performance is assessed, real-time performance should illuminate the SA processes of the operator.

The Global Implicit Measure (GIM) (Vidulich and McMillan, 2000) is an example of this approach. The GIM is based on the assumption that the operator of a human–machine system is attempting to accomplish known goals at various priority levels. Therefore, it is possible to consider the momentary progress toward accomplishing these goals as a performance-based measure of SA. Development of the GIM was an attempt to develop a real-time SA measurement that could effectively guide automated pilot aiding (Brickman et al., 1995, 1999; Vidulich, 1995; Shaw et al., 2004). In this approach, a detailed task analysis was used to link measurable behaviors to the accomplishment of mission goals. The goals will be varied depending on the mission phase. For example, during a combat air patrol, a pilot might be instructed to maintain a specific altitude and to use a specific mode of the on-board radar, but during an intercept the optimal altitude might be defined in relation to the aircraft being intercepted, and a different radar mode might be appropriate. For each phase, these measurable behaviors that logically affect goal accomplishment are identified and scored. The scoring was based on the contribution to goal accomplishment. The proportion of mission-specific goals being accomplished successfully according to the GIM algorithms indicated how well the pilot was accomplishing the goals of that mission phase. More important, the behavioral components scored as failing should identify the portions of the task that the pilot was either unaware of or unable to perform at the moment. Thus, GIM scores could potentially provide a real-time indication of the person's SA as reflected by of the quality of task performance and a diagnosis of the problem if task performance deviates from the ideal, as specified by the GIM task analysis and scoring algorithms.

Vidulich and McMillan (2000) tested the GIM metric in a simulated air-to-air combat task using two cockpit designs that were known from previous evaluations to produce different levels of mission performance, mental workload, and rated SA. The subjects were seven U.S. military pilots or weapons systems officers. The real-time GIM scores distinguished successfully between the two cockpits and the different phases of the mission. No attempt was made to guide adaptation on the basis of the GIM scores, but the results suggested that such an approach has promise.

3.2 Subjective Measures

Subjective measures consist primarily of using techniques that usually require subjects to quantify their experience of workload or SA. Many researchers are suspicious of subjective data, perhaps as a holdover from the behaviorists' rejection of introspection as an unscientific research method (Watson, 1913). However, Annett (2002a,b) argued that subjective ratings are maligned unfairly. In an in-depth discussion of the issues, he contended that the lack of precision associated with subjective measures was expected to prohibit their use in setting design standards. However, he also concluded that subjective ratings could be useful for evaluating the mechanism underlying performance or for the comparative evaluation of competing interface designs. Such a comparative process is how subjective ratings of workload and SA are typically used.

Vidulich and Tsang (1987) and Tsang and Vidulich (1994) found three variables that were useful for categorizing subjective rating techniques: dimensionality, evaluation style, and immediacy. *Dimensionality* refers to whether the metric required the subjects to rate their experiences along a single dimension or multiple dimensions. *Evaluation style* refers to whether the subjects were asked to provide an absolute rating of an experience or a relative rating comparing one experience to another. *Immediacy* distinguishes between subjective metrics that were designed to be used as soon as possible after the to-be-rated experience and those that were used at the end of a session or even at the end of an experiment.

Although it is theoretically possible to create a subjective technique that combines any level of the three variables, in practice two basic combinations have dominated. The most common techniques combine multidimensionality, the absolute evaluation style, and immediacy. The typical alternative to the multidimensional-absolute-immediate approach, are techniques that are usually unidimensional, use a relative comparison evaluation style, and are collected retrospectively rather than immediately.

3.2.1 Multidimensional Absolute Immediate Ratings

The subject's immediate assessment after trial completion should minimize the potentially damaging effects of any bias the subject may have regarding the task conditions and the likelihood that the ratings are based on between-condition comparisons (Tsang and Vidulich, 1994). Also, ratings after the trial should benefit from the freshest memory for the experience of performing the trial. The absolute scale design should also encourage the subjects to consider each trial condition individually. The multidimensional aspect supports diagnosticity, because the subjects can be more precise in describing how experimental conditions influence their experience.

Workload Ratings Although numerous scales have been developed, two popular multidimensional, absolute, and immediate ratings scales are the National Aeronautics and Space Administration's Task Load Index (NASA-TLX) (Hart and Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988). NASA-TLX is based on six scales (i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration level), and the ratings on the six scales are weighted according to the subject's evaluation of their relative importance. SWAT is based on three rating scales (i.e., time load, mental effort load, and psychological stress load). The relative roles of the three scales is determined by the subjects' rankings of the workload inflicted by each combination of the various levels of workload (1 to 3) in each of the three workload scales. A conjoint analysis is then conducted to produce a look-up table that translates the ordinal rankings to ratings with interval-scale properties. Both NASA-TLX and SWAT ultimately produce a workload rating from 0 to 100 for each trial rated.

NASA-TLX and SWAT have been compared to each other and to a number of other rating scales a number of times (e.g., Battiste and Bortolussi, 1988; Hill et al., 1992; Rubio et al., 2004). In reviewing the comparisons, Rubio et al. (2004) noted that SWAT and NASA-TLX both offer diagnosticity, due to their multiple scales, and have generally demonstrated good concurrent validity with performance. Rubio et al. (2004) also pointed out that both NASA-TLX and SWAT have demonstrated sensitivity to difficulty manipulations, although some researchers have found NASA-TLX to be slightly more sensitive, especially for low levels of workload (e.g., Battiste and Bortolussi, 1988; Hill et al., 1992).

NASA-TLX and SWAT have also been compared in terms of their ease of use. Each technique is composed of two major parts: the scales that the subjects fill out after each trial and a procedure for converting the raw scale ratings into the final workload scale. The actual scales used by SWAT are fewer than NASA-TLX (three vs. six) and only require the subject to choose one of three possible levels instead of rating on a 0-to-100 scale as NASA-TLX incorporates. That means that SWAT would be easier to collect in a prolonged task such as flying, while the task performance actually continues. On the other hand, NASA-TLX's paired comparison technique to generate scale weights is much easier than SWAT's card-sorting procedure for both the subject to complete and the experimenter to process. The NASA-TLX procedure only requires the subject to make 15 forced choices of importance between the individual scales. The raw count of the number of times that each scale was considered more important than another is then used to weigh the individual scale ratings provided by the subject. In contrast, the SWAT card sort requires each subject to consider and sort 27 cards (each representing a possible combination of rating scale selections), and then the experimenter must use specialized software to convert the card sort data into an overall workload scale.

Some researchers have investigated simpler methods of generating weights for SWAT. The simple sum of the three SWAT dimensions has been shown to exhibit the same pattern of significant findings as SWAT ratings using the conjoint analysis of card-sort data (Biers and Maseline, 1987; Biers and McInerney, 1988; Luximon and Goonetilleke, 2001). Additionally, Luximon and Goonetilleke (2001) found that SWAT sensitivity could be improved by using a continuous scale rather than a three-level discrete scale. As with SWAT, the weighting procedure of NASA-TLX has undergone testing. Both Nygren (1991) and Hendy et al. (1993) have argued that the NASA-TLX weighting procedure does not add to NASA-TLX's effectiveness.

Lee and Liu (2003) provide an example of the use of NASA-TLX to assessing the workload of China Airline pilots flying a Boeing 747 aircraft in a high-fidelity 747 simulator. Lee and Liu found that the overall NASA-TLX ratings discriminated successfully among four flight segments: takeoff, cruise, approach, and landing. As expected, takeoff, landing, and approach were all rated higher than cruise in mental workload. Lee and Liu also used the multidimensional scales of NASA-TLX to diagnose the causes of the higher workload. For example, they found that temporal demand was an important contributor to the takeoff and approach segments; effort was a more important contributor to landing. The authors concluded that training programs should be designed to help the pilot cope with the specific expected stresses of different flight segments.

SA Ratings Multidimensional, absolute, and immediate ratings have also been a popular approach for assessing SA. Probably the most commonly used subjective rating tool for SA has been the Situation Awareness Rating Technique (SART), developed by Taylor (1990). The SART technique characterizes SA as having three main dimensions: attentional demands (D), attentional supply (S), and understanding (U). The ratings on each of the three dimensions are combined into a single SART value according to a formula (Selcon et al., 1992): SA = U − (D − S).

Inasmuch as SART contains ratings of attentional supply and demand, it can be seen to incorporate elements of mental workload in its evaluation. However, in a direct comparison of NASA-TLX and SART, it was found that although both were sensitive to task demand level, SART was also sensitive to the experience level of the 12 Royal Air Force pilot subjects (Selcon et al., 1991).

3.2.2 Unidimensional Relative Retrospective Judgments

The unidimensional, relative, retrospective judgment approach is based on the assumption that the subject who has experienced all of the task conditions is considered a subject matter expert with knowledge about the subjective experience of performing the various task conditions under consideration. This approach attempts to extract and quantify subjects' opinions about the experiences associated with task performance.

Workload Judgments The use of unidimensional, relative, retrospective judgments was strongly supported by the work of Gopher and Braune (1984). Inspired by Stevens's (1957, 1966) psychophysical measurement theory, Gopher and Braune adapted it to the measurement of subjective workload. The procedure used one task as a reference task with an arbitrarily assigned workload value. All of the other tasks' subjective workload values were evaluated relative to that of the reference task. The resulting ratings were found to be highly sensitive in a number of studies (e.g., Tsang and Vidulich, 1994; Tsang and Shaner, 1998). In addition, high reliability of these ratings was revealed by split-half correlations of repeated ratings of the task conditions.

Another approach to collecting unidimensional, relative, retrospective judgments was developed by a mathematician, Thomas Saaty (1980). Saaty's technique was named the *Analytic Hierarchy Process* (AHP) and was developed to aid decision making. When applied to workload assessment, the AHP requires subjects to perform all pairwise comparisons of all task conditions. These comparisons fill a dominance matrix, which is then solved to provide the ratings for each task condition. Saaty's AHP was originally designed to evaluate all dimensions relevant to a decision and then combine the multiple dimensions to support selection of one option in a decision-making task. However, Lidderdale (1987) demonstrated that a unidimensional version of the AHP could be an effective workload assessment tool and inspired further investigations using the tool. Vidulich and Tsang (1987) compared the AHP to NASA-TLX and a unidimensional, absolute, immediate rating of overall workload in assessing the workload of selected laboratory tasks. The AHP was found to be both more sensitive and more reliable than the other techniques. Vidulich (1989) compared several methods for converting dominance matrices to the final ratings and used the results to create the Subjective Workload Dominance (SWORD) technique. In one application, Toms et al. (1997) used SWORD to evaluate a prototype decision aid for landing an aircraft. The participating pilots performed landings with and without the decision aid and in both low- and high-task-load conditions. Task load was varied by changing the information available to the pilot. Overall, the results showed that the decision aid improved landing performance while lowering mental workload.

Vidulich and Tsang performed a series of studies to examine the various approaches to subjective assessment. Although specific instruments were compared in these studies, the goal was not to determine which instrument was superior. Rather, the objective was to determine which assessment approach can elicit the most and accurate workload information. Tsang and Vidulich (1994) found that the unidimensional, relative, retrospective SWORD technique with the highly redundant pairwise comparisons superior to a procedure using relative comparisons to a single reference task. Tsang and Velazquez (1996) found that compared to an immediate absolute instrument, a relative, retrospective psychophysical scaling was more sensitive to task demand manipulation and had higher concurrent validity with performance. Tsang and Velazquez (1996) also found that a subjective multidimensional retrospective technique, the Workload Profile, provided diagnostic workload information that could be subjected to quantitative analysis. Rubio et al. (2004) confirmed the diagnostic power of the Workload Profile technique. They found the Workload Profile more diagnostic than either NASA-TLX or SWAT. Collectively, these studies suggested a relative-retrospective approach advantage.

SA Judgments Unidimensional, relative, retrospective judgments have also been applied to SA assessment. For example, the SWORD workload technique was adapted to measure SA (SA-SWORD; Vidulich and Hughes, 1991). Vidulich and Hughes used the SA-SWORD to evaluate the effect of data-linked information in an advanced fighter aircraft simulation. The technique demonstrated good sensitivity to the experimental manipulation and good reliability. Toms et al. (1997) used the SA-SWORD to assess SA along with SWORD to measure workload. Their results showed that a decision aid's benefits to landing performance and mental workload were also associated with improved SA. In this case, workload assessment and SA assessment both showed that the decision aiding was valuable.

3.3 Physiological Measures

A host of physiological measures have been used to assess mental workload with the assumption that there are physiological correlates to mental work. The most common measures include cardiovascular (e.g., heart rate and heart rate variability), ocular (e.g., pupil dilation, eye movement measures), and measures of brain activity. The present review focuses on the brain measures because (1) it would seem that brain activity could most directly reflect mental work; (2) in line with our framework that hypothesizes both an intensity aspect and a structural aspect to mental work, many of the brain measures have been demonstrated to be sensitive to parametric manipulation of task demands and to be diagnostic with regard to the types of cognitive demands involved in certain task performance; and (3) there already exist reviews of the nonbrain measures (e.g., Beatty, 1982; Stern et al., 1984; Wilson and Eggemeier, 1991; Jorna, 1992; Mulder, 1992; Backs and Boucsein, 2000; Kramer and Weber, 2000; Kramer and McCarley, 2003), but the brain-imaging workload studies are relatively new.

3.3.1 Electroencephalographic Measures

Electroencephalographic (EEG) measures are recorded from surface electrodes placed directly on the scalp and have been shown to be sensitive to momentary changes in task demands in laboratory studies (e.g., Glass, 1966), simulated environments (e.g., Fournier et al., 1999; Gevins and Smith, 2003), and real-world settings (e.g., Wilson, 2002b). Spectral power in two major frequency bands of the EEG have been identified as being sensitive to workload manipulations: the alpha (7 to 14 Hz) and theta (4 to 7 Hz) bands. Spectral power in the alpha band that arises in widespread cortical areas is inversely related to the attentional resources allocated to the task, whereas theta power recorded over the frontal cortex increases with increased task difficulty and higher memory load (Parasuraman and Caggiano, 2002). Sterman and Mann (1995) reported a series of EEG studies conducted in simulated and operational military flights. A systematic decrease in power in the alpha band of the EEG activity was observed with a degraded control responsiveness of a T4 aircraft. A graded decrease in the alpha band power was also observed as U.S. Air Force pilots flew more difficult in-flight refueling missions in a B2 aircraft simulator. Brookings et al. (1996) had Air Force air traffic controllers perform computer-based air traffic control simulation (TRACON). Task difficulty was manipulated by varying the traffic volume (number of aircraft to be handled), traffic complexity (arriving to departing flight ratios, pilot skill, and aircraft types), and time pressure. Brookings et al. found the alpha power to decrease with increases in traffic complexity and the theta power to increase with traffic volume.

Kramer and Weber (2000) point out that a distinct advantage of EEG measures is their sensitivity to variations of mental workload and their potential to track momentary fluctuations in mental workload associated with rapid changes in task demands. However, the sensitivity of EEG measures to numerous artifacts such as head and body movements poses special difficulties for extralaboratory applications. Kramer and Weber also point out that it is not yet entirely clear whether the EEG measures reflect changes in the level of general arousal or changes in more specific cognitive operations. That is, EEG measures may not be particularly diagnostic with regard to the specific types of demand incurred. More in-depth discussion can be found in Gevins et al. (1995) and Davidson et al. (2000).

3.3.2 Event-Related Potentials

Evoked potentials are embedded in the background of EEG and are responsive to discrete environmental events. There are several different positive and negative voltage peaks and troughs that occur 100 to 600 ms following stimulus presentation. The P300 component has been extensively studied as a mental workload measure (e.g., Gopher and Donchin, 1986; Parasuraman, 1990; Wickens, 1990; Kramer and Weber, 2000). The P300 is typically examined in a dual-task condition with either the oddball paradigm or the irrelevant probe paradigm. In the *oddball paradigm*, the P300 is elicited by the subject keeping track of an infrequent signal (e.g., counting infrequent tones among frequent tones). One drawback of the oddball paradigm is that the additional processing of the oddball and having to respond to it could inflate the true workload of interest artifactually. As an alternative, additional stimuli (e.g., tones) are presented but subjects are not required to keep track of them in the *irrelevant probe paradigm*. Needless to say, creating artifactual workload in the assessment process is of particular concern in applied settings.

With the oddball paradigm, the amplitude of the P300 has been found to decrease with increased task difficulty manipulated in a variety of laboratory tasks (e.g., Hoffman et al., 1985; Strayer and Kramer, 1990; Backs, 1997). Importantly, P300 is found to be selectively sensitive to perceptual and central processing demands. For example, Isreal et al. (1980) found that the amplitude of P300 elicited by a series of counted tones was not sensitive to response-related manipulations of tracking difficulty but was affected by manipulations of display perceptual load. Many of the laboratory-based findings have been replicated in simulator studies. For example, Kramer et al., (1987) had student pilots flew an instrument flight plan in a single-engine aircraft simulator. The P300s elicited by the secondary tone-counting task decreased in amplitude with increasing turbulence and subsystem failures (see also Fowler, 1994). Using the irrelevant probe paradigm, Sirevaag et al. (1993) had senior helicopter pilots fly low-level high speed flight in a high-fidelity helicopter simulator. The P300 amplitude was found to increase with increased difficulty in the primary tracking task. In addition, the P300 amplitude elicited by the secondary irrelevant probes decreased with increases in the communication load. Kramer and Weber (2000) point out further that the irrelevant probe paradigm would work only if the irrelevant probes are presented in a channel that would be monitored anyway.

In short, event-related potential (ERP) measures have been found to be sensitive to, and diagnostic of, changes in the perceptual and central processing task demands. One potential drawback is the possibility of artifactually augmenting the real workload of interest if the ERP measures are elicited from a secondary task. As with the performance, it would be ideal if a secondary task naturally embedded in the test environment could be used. But because ERP signals are relatively small and ensemble averaging across many stimuli is necessary for meaningful interpretation, there are not always sufficient stimuli available from the embedded secondary task or the primary task.

3.3.3 Brain Imaging Measures

Two measures of the brain's metabolic responses are considered here: positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) measures. These measures most notably have been used to localize cortical regions associated with various cognitive processing (e.g., D'Exposito et al., 1999; Posner and DiGirolamo, 2000). Equally important, recent research has shown that these measures exhibit systematic variations with parametric manipulation of task difficulty (Parasuraman and Caggiano, 2002). That is, they have been shown to be both diagnostic and sensitive measures.

As an example, Corbetta et al. (1990) had subjects determine whether two stimuli presented in two frames separated by a blank display were the same or different. Between the two frames, the stimuli could vary in one of three dimensions: shape, color, and velocity. In the selective-attention condition, one of the dimensions would be designated as the relevant dimension, and zero, one, or two irrelevant dimensions could covary with the relevant dimension. In the divided-attention condition, any one of the dimensions could vary. The behavioral data (d') indicated that the divided-attention condition was more difficult. In the selective-attention condition, increased blood flow was observed in the region of the visual cortex known to be related to the processing of the relevant dimension designated. Corbetta et al. proposed that the increased neural activity in the specialized regions for the different dimensions was a result of a top-down attentional control since the sensory information should be the same across selective- and divided-attention conditions.

Just et al. (2003) reviewed a series of PET studies and found lower brain metabolic rate to be associated with higher language proficiency (Parks et al., 1988) and increased practice with a spatial computer game (Haier et al. 1992). Just et al. interpreted these results to mean that high-ability, high-skill persons could process more efficiently, thereby requiring a smaller amount of their total amount of processing resources available. They effectively would have a larger supply of processing resources (for other processing). Just and Carpenter (1992) proposed that the computational work underlying thinking must be accompanied by resource utilization. In their 3CAPS model, a brain region is considered a resource pool. Computational activities are resource consuming in the sense that they all operate by consuming an entity called activation. The intensity and volume of brain activation in a given cortical area is expected to increase in a graded fashion with increased computational load. Indeed, Just et al. (1996) found that with increasing sentence complexity, the level of neuronal activation and the volume of neural tissue activated increased in four cortical areas associated with language processing (Wernicke's, Broca's, and their right hemisphere homologues). With a spatial mental rotation task, Carpenter et al. (1999) found a monotonic increase in signal intensity and volume activation in the parietal region as a function of increased angular disparity between the two stimuli whose similarity was to be judged.

Highlighting the results from a number of studies that use an array of behavioral and neurophysiological measures (ERPs, PET, and fMRI), Just et al. (2003) propose cognitive workload to be a function of resource consumption and availability. Several similarities between the 3CAPS model and Wickens' multiple resource model are apparent. According to both models, (1) mental workload is a function of supply and demand of processing resources, (2) resources can be modulated in a graded fashion, (3) specific resources are used for different types of cognitive processing (e.g., verbal and spatial task demands bring about activations in different cortical regions), and (4) supply or availability of resources can be modulated by individual differences in ability and skill or expertise.

The PET and fMRI studies described so far were all conducted in the laboratory. Although more applied studies would certainly be desirable, the methodologies involved with these newly available technologies are still being refined. Practical concerns aside, results of these laboratory studies are encouraging and at the same time point to the need for continual learning about interpreting these brain images and validating the interpretations. Notwithstanding, one simulated study on pilot performance can be presented. Pérès et al. (2000) had expert (with at least 3000 flight hours and flight instructor qualifications) and novice (with less than 50 flight hours) French Air Force pilots perform a continuous simulated flight control task at two speeds (100 and 200 knots) while fMRI measures were collected. The fMRI measures showed that neuronal activation was dominant in the right hemisphere, as would be expected for a visual spatial task. Further, novice pilots exhibited more intense and more extensive activation than expert pilots. At the high-speed condition, the expert pilots exhibited increased activation in the frontal and prefrontal cortical areas and reduced activity in visual and motor regions. This suggested to researchers that the expert pilots were better able to use their knowledge to focus their resources for the higher-level functions in working memory, planning, attention, and decision making. In contrast, novice pilots' increased activation in the high-speed condition was more widespread and extended across the frontal, parietal, and occipital areas, suggesting that they were engaged in nonspecific perceptual processing. Interestingly, when the expert pilots were asked to track at an even higher speed (400 knots), their pattern of activation resembled that of the novice pilots tracking at 200 knots.

Notably, there is paucity of physiological studies on SA included in this chapter. This is partly because compared to the concept of mental workload, the concept of SA is relatively new (Pew, 1994; Wickens, 2001) and both its theoretical and methodological development have not reached the level of maturity that the concept of mental workload has. It is also the case that the concept of SA does not refer to a specific process. Although complex performance generally entails multiple processes, it is often possible to identify many of the processes and hence the type of workload involved. However, whereas SA is supported by the many of the same processes, SA is an emergent property that has not been hypothesized to be associated with specific cortical regions or other physiological responses.

As a class of measures, many of the physiological measures have the distinct ability to serve as a continuous measure for on-line assessment. Some of them are diagnostic with regard to the type of cognitive demands entailed (e.g., ERPs, PET, fMRI). Most of them are not cognitively intrusive, as in having to perform additional work in order to provide a measure. The main drawback with most physiological measures is that they are equipment intensive, which makes real-world assessment impractical. But it is not impossible. Some physiological measures are or will become more feasible with the present rapid technological advances (e.g., see Gevins et al., 1995; Wilson, 2000; Wilson, 2002a; Kramer and McCarley, 2003; Parasuraman, 2003). Even for the costly fMRI studies, attempts have been made to assess the mental workload of simulated flight performance (Pèrés et al., 2000).

3.4 Multiple Measures of Workload and Situation Awareness

There are several facets to the undertaking of assessing workload and SA of a complex, dynamic human–machine system. First, there are a number of candidate measures to choose from, each with strengths and weaknesses. Measures that provide global information about the mental workload of these tasks may fail to provide more specific information about the nature of the demand. Measures that could provide more diagnostic information may be intrusive or insensitive to other aspects of interest, and certain sensitive measures may be collected only under restrictive conditions. Still, many workload measures often associate. For example, many subjective measures have been found to correlate with performance (e.g., Tsang and Vidulich, 1994; Hockey et al., 2003; Rubio et al., 2004). Just et al. (2003) present a convincing account of how the associations among a number of behavioral and neurophysiological measures support the extant understanding of many cognitive concepts relevant to mental workload (see also Wickens, 1990; Fournier et al., 1999; Lee and Liu, 2003). Importantly, when the measures do not associate, they do not do so in haphazard ways. The association and dissociation patterns among measures should therefore be evaluated carefully rather than treated as unreliable randomness. Below we discuss in greater detail the dissociation between the subjective and performance workload measures and the relation between the workload and SA measures.

3.4.1 Dissociations among Workload Measures

When different types of workload measures suggest different trends for the same workload situation, the workload measures are said to *dissociate*. Given that mental workload is a multidimensional concept and the various workload measures may be differentially sensitive to the different workload dimensions, dissociations among workload measures are to be expected. Measures having qualities of general sensitivity (such as certain unidimensional subjective estimates) respond to a wide range of task manipulations but may not provide diagnostic information about the individual contributors to workload. Measures having selective sensitivity (such as secondary task measures) respond only to specific manipulations. In fact, the nature of the dissociation should be particularly revealing with regard to the characteristics of the workload in the task under evaluation.

Several conditions for the dissociation of performance and subjective measures have been identified (Vidulich and Wickens, 1986; Vidulich, 1988; Yeh and Wickens, 1988):

1. Dissociation tends to occur under low-workload conditions (e.g., Eggemeier et al., 1982). Performance could already be optimal when the workload is low and thus would not change further with additional effort that would be reflected in the subjective measures.
2. Dissociation would occur when subjects are performing data-limited tasks (when performance is governed by the quality of the data rather than by the availability of resources). If subjects are already expending their maximum resources, increasing task demand would further degrade performance but would not affect the subjective ratings.
3. Greater effort would generally result in higher subjective ratings, however, greater effort could also improve performance (e.g., Vidulich and Wickens, 1986).
4. Subjective ratings are particularly sensitive to the number of tasks that subjects have to time-share. For example, performing an easy dual task (that results in good performance) tends to produce higher ratings than does performing a difficult single task (that results in poor performance) (e.g., Yeh and Wickens, 1988).
5. Performance measures are sensitive to the severity of the resource competition (or similarity

of resource demand) between the time-shared tasks, but subjective measures are less so (Yeh and Wickens, 1988).

6. Given that subjects only have access to information available in their consciousness (Ericsson and Simon, 1993), subjective ratings are more sensitive to central processing demand (such as working memory demand) than to demands that are not represented well consciously, such as response execution processing demand. Dissociation would therefore tend to occur when the main task demands lie in response execution processing (Vidulich, 1988). McCoy et al. (1983) provided an excellent list of realistic examples of how performance and subjective ratings may dissociate in system evaluations and discussed how the dissociations can be interpreted in meaningful ways.

Hockey (1997) offers a more general conceptual account for the relations among performance, subjective, and physiological measures. Hockey proposes a compensatory control mechanism that allocates resources dynamically through an internal monitor very much like the one proposed by Kahneman (1973). Performance may be protected (as in primary task performance) by recruiting further resources, but only at the expense of increased subjective effort and physiological costs and degraded secondary task performance or strategic management of the overall system performance. Alternatively, performance goals may be lowered. Although performance will then degrade, no additional effort or physiological cost will be incurred. Hockey emphasizes that the efficacy of the control mechanism hinges on the accuracy of the perception of the situation. For example, Sperandio (1978) found air traffic controllers to switch strategy when the traffic load increased. Beyond a certain number of aircraft that the controllers handled, controllers would switch to a uniform strategy across aircraft as opposed to paying more individual attention to the various aircraft. Although this strategy should reduce the cognitive resources needed for dynamic planning, it would also probably produce less optimal scheduling. That is, the primary task performance might have been preserved with the strategy switch, but some secondary goals would have suffered (Hockey, 1997).

3.4.2 Relations of Workload and Situation Awareness Measures

Wickens (2001) point out that due to the energetic properties of workload, many physiological and subjective rating measures are suited for capturing the quantitative aspects of workload. In contrast, physiological measures are likely to be poor candidates for assessing the quality or content of SA. Self-ratings of one's awareness are unlikely to be informative since one cannot be aware of what one is not aware. However, subjective SA ratings could still be useful if they are used for system evaluative purposes. As illustrated earlier, subjects often could indicate reliably which system design affords greater SA. Last, Wickens pointed out that explicit performance measures designed to examine what one is aware of (content of SA) have no parallel use for workload assessment. However, implicit performance measures such as those used to check for reaction to unexpected events can be used to assess both workload and SA.

As discussed earlier, there is not one fixed relationship between workload and SA. Although high SA and an acceptable level of workload is always desirable, workload and SA can correlate positively or negatively with each other, depending on a host of exogenous and endogenous factors. Two sample studies will be described to illustrate their potential relationships. Vidulich (2000) reviewed a set of studies that examined SA sensitivity to interface manipulations. Of the nine studies that manipulated the interface by providing additional information on the display, seven showed an increase in SA, four showed a concomitant reduction in workload, and three showed a concomitant increase in workload. In contrast, of another nine studies that manipulated the interface by reformatting the display, all nine showed an increase in SA, six showed a concomitant reduction in workload, and none showed an increase in workload. In short, although different patterns in the relationship between the workload and SA measures were observed, the various patterns were reasonably interpretable given the experimental manipulations. In another study, Alexander et al. (2000) examined the relationship between mental workload and situation awareness in a simulated air-to-air combat task. Seven pilots flew simulated air intercepts against four bombers supported by two fighters. The main manipulations were two cockpit designs (the conventional cockpit with independent gauges and a virtually augmented cockpit designed by a subject-matter expert) and four mission phases of various degrees of difficulty and complexity. A negative correlation between the workload and SA measures were observed for both the cockpit design and the mission complexity manipulations. The augmented cockpit improved SA while reduced workload, whereas increased mission complexity decreased SA and increased workload. These results underscore the value of assessing both the mental workload and SA involved in any test and evaluation.

3.4.3 Need for Multiple Measures

There are however several broad guiding principles that would be helpful in measures selection. Muckler and Seven (1992) hold that "the distinction between 'objective' and 'subjective' measurement is neither meaningful nor useful in human performance studies" (p. 441). They contend that all measurements contain a subjective element as long as the human is part of the assessment. Not only is there subjectivity in the data obtained from the human subject, the human experimenter also imparts his or her subjectivity in the data collection, analysis, and interpretation. Thus, performance measures are not all objective, nor are subjective measures entirely subjective (see also, Annett, 2002a,b; Salvendy, 2002). Muckler and Seven advocate that the selecting of a measure (or a set of measures) be guided by the information needs. Candidate measures can be evaluated by considering their relative strengths (such as diagnosticity) and weaknesses (such as intrusiveness). In addition, Kantowitz (1992) advocates using theory to select the measures. Kantowitz made an analogy between theory and the blueprint of a building. Trying to interpret data without the guidance of a theory is like assembling bricks randomly when constructing a building. To elaborate, Kantowitz points out that an understanding of both the substantive theory of human information processing and the psychometric theory of the measurements is helpful. The former dictates what one should measure, and the latter suggests ways of measuring them. Another useful (if not required) strategy is to use multiple measures as much as feasible. As discussed above, even seemingly dissociate measures are informative (and sometimes especially so) if one is cognizant of the idiosyncratic properties of the different measures. In fact, Wickens (2001) points out that converging evidence from multiple measures is needed to ensure an accurate assessment of the level of workload incurred and the quality of SA attained.

To emphasize the value of assessing multiple measures, Parasuraman (1990) reported a study that examined the effectiveness of safety monitoring devices in high-speed electric trains in Europe (Fruhstorfer et al., 1977). Drivers were required to perform a secondary task by responding to the occurrence of a target light in a cab within 2.5 seconds. If no response was made, a loud buzzer would be activated. If the buzzer was not responded to within an additional 2.5 seconds, the train's braking system was activated automatically. Over a number of train journeys, onset of the warning buzzer was rare, and the automatic brake was activated only once. However, the EEG spectra showed that the secondary task performance could remain normal even when the drivers were transiently in stage 1 sleep.

4 DESIGN FOR MENTAL WORKLOAD AND SITUATION AWARENESS: INTEGRATED APPROACH TO OPTIMIZING SYSTEM PERFORMANCE

It is probably fair to say that after a decade of debate, there is now a general agreement that mental workload and SA are distinct concepts and yet are intricately intertwined. Both can be affected by very many of the same exogenous and endogenous factors and have a significant impact on each other and on system performance. One implication is that fairly well understood psychological principles can be applied to both concepts. For example, in the framework presented above, both workload and SA are subject to attentional and memory limits, and both can be supported by expertise. There exists an established body of knowledge about the effects of these limits and the enabling power of expertise to allow fairly reliable performance predictions. But the fact that the two concepts are distinct also means that they each contribute uniquely to the functioning of a human–machine system. Below we review three research areas that could be exploited for developing support that would manage workload and SA cooperatively to optimize system performance.

4.1 Adaptive Automation

Automation is often introduced to alleviate the heavy demand on an operator or to augment system performance and reduce error. Many modern complex systems simply cannot be operated by humans alone without some form of automation aids. However, it is now recognized that automation often redistributes, rather than reduces, the workload within a system (e.g., Wiener, 1988; Lee and Moray, 1992). Further, an increasing level of automation could distance the operator from the control system (e.g., Adams et al., 1991; Billings, 1997). The upshot of this is that even if automation reduces mental workload successfully, it could reduce SA and diminish an operator's ability to recover from unusual events. The idea of adaptive automation was introduced as a means of achieving the delicate balance of a manageable workload level and an adequate SA level. This idea has been around for some time (e.g., Rouse, 1977, 1988) and is receiving much attention in recent research (e.g., Rothrock et al., 2002; Parasuraman and Bryne, 2003). Proponents of adaptive automation argue that static automation that entails predetermined fixed task allocation will not serve complex dynamic systems well. Workloads can change dynamically due to environmental and individual factors (e.g., skill level and effectiveness of strategies used). It has been proposed that a major environmental determinant of workload is rapid (Huey and Wickens, 1993) and unexpected (Hockey et al., 2003) changes in task load. So ideally, more or fewer tasks should be delegated to automation dynamically. More automation would be introduced during moments of high workload, but as the level of workload eases more tasks would be returned to the operator, thereby keeping the operator in the loop without overloading the person. The key issue is the development of an implementation algorithm that could efficaciously adapt the level of automation to the operator's state of workload and situation awareness.

Parasuraman and Bryne (2003) describe several adaptation techniques that rely on different inputs to trigger an increase or decrease in the extent of automation in the system. One technique relies on physiological measures and another relies on performance measures. One obvious advantage of physiological measures is their continuous availability and noninvasive nature. A number of physiological measures have been evaluated for their potential to provide real-time assessment of workload. These include heart rate variability (e.g., Jorna, 1999), EEG (e.g., Gevins et al., 1998; Prinzel et al., 2000; Gevins and Smith, 2003), ERPs (Parasuraman, 1990; Kramer et al., 1996) and eye movement measures (Hilburn et al., 1997; Kramer and McCarley, 2003).

Although performance-based measures are not used as much as physiological-based measures, recent studies have demonstrated their potential promise as well. In one study, Kaber and Riley (1999) used a secondary monitoring task along with a target acquisition task. Adaptive computer aiding based on secondary-task performance was found to enhance primary-task performance. Notice that the tasks used here afford fairly continual performance measures, a property that not many performance-based measures possess. Also, as discussed above, for the secondary-task methodology to provide useful workload information, the time-shared tasks would need to be competing for some common resources, which of course could add to the workload. In another set of studies, Vidulich and colleagues propose that the Global Implicit Measure (GIM, described above) could be developed as a real-time situation awareness measurement that could guide effective automated pilot aiding based on real-time scoring of both continuous and discrete tasks.

4.2 Display Design

To the extent that excessive workload could reduce SA, any display that supports performance without incurring excessive workload would at least indirectly support SA as well (see, e.g., Previc, 2000). Wickens (1995) propose that displays that do not overtax working memory and selective attention are particularly attractive because SA depends heavily on these processes. Wickens (1995, 2002, 2003) discusses various display principles (e.g., proximity compatibility principle, visual momentum) that have been shown to support various types of performance (e.g., flight control as opposed to navigation) and display features (e.g., frame of reference) that would lend support to SA.

While display formats that facilitate information-processing support performance and thereby free up resources for SA maintenance, Wickens (2002) shows that display formats could also affect the product (type) of the SA. For example, a display with an egocentric frame of reference (an inside-out view with a fixed aircraft and a moving environment) provides better support for flight control, whereas an egocentric frame of reference (an outside-in view with a moving aircraft and fixed environment) provides better support for noticing hazards and general awareness of one's location. Wickens points out further that there are often trade-offs between alternative display formats. For example, whereas an integrated, ecological display generally provides better information about three-dimensional motion flow, a three-dimensional representation on a two-dimensional viewing surface tends to create ambiguity in locating objects in the environment. Such ambiguity is less of a problem in a two-dimensional display format. But it would take more than one two-dimensional display to present the same information in a three-dimensional display. It has been shown that it can be more cognitively demanding in trying to integrate information from two separate two-dimensional displays. The trade-off between promoting SA for objects in the environment and accomplishing other tasks at a lower workload level could only be resolved with regard to the specific goals or the priorities of competing goals of the system.

4.3 Training

Given the role that expertise plays in one's workload and situation awareness, there is great potential in training to support SA and to permit tasks to be accomplished with less resources at a lower level of workload. The issue is: What does one train for? That expertise is based largely on a large body of domain-specific knowledge suggests that a thorough understanding of the workings of the system would be helpful, particularly in nonroutine situations. Although expertise speeds up performance and experts generally perform at a high level under normal situations, their expertise is particularly useful in unexpected circumstances because of their ability to use their acquired knowledge to recognize and solve problems. The concern that operators trained on automated systems (which would be especially helpful for novices because of the presumed lower level of workload involved) might never acquire the needed knowledge and experience to build up their expertise is certainly a valid one until there exist automated systems with total reliability that would never operate outside a perfectly orchestrated environment. One possibility might be to provide some initial and refresher training in a nonautomated or less automated simulated system.

Although there exists in the literature a large body of training research that aims at accelerating the learning process and there is much evidence to support the advantages of not subjecting a trainee to an excessive level of workload, there are additional considerations when the goal is to build SA as well. One, it would be most useful to have some ideas about the knowledge structure that experts have so that the training program can build upon reinforcing this structure. After all, it is the structure and organization of information that support fast and accurate pattern recognition and information retrieval. Two, experts do not merely possess more knowledge, they are better at using it. This would suggest that training should extend to strategic training. Given the growing body of evidence to support that strategic task management (or executive control) is a higher-level generalizable skill, much of the strategic training could be accomplished with low-cost low-physical-fidelity simulated systems such as a complex computer game (see Haier et al., 1992; Gopher, 1993). The strategic training can be at odds with the goal of keeping the level of workload down while the operators are in training. However, research has shown that the eventual benefits outweigh the initial cost in mental workload. As desirable as it is to train to develop automatic processing that is characterized as fast, accurate, and attention free, this training strategy may have only limited utility in training operators who have to function within a dynamic complex system. This is because there would be relatively few task components in these systems that would have an invariant stimulus–response mapping (a requirement for automatic processing to be developed and applied).

All three research areas underscore the interdependence of the concepts of workload and SA. The design of any efficacious technical support or training program would need to take into account the interplay of the two. Any evaluation of the effectiveness of these supports would need to assess both the operator's workload and situation awareness in order to have a clear picture of their impact on system performance.

5 CONCLUSIONS

So the years of research into mental workload and situation awareness have been profitable. The research has developed a multitude of metric techniques, and although the results of different mental workload or SA assessment techniques sometimes show dissociations, they seem to fit within the theoretical constructs behind the measures. Workload is primarily a result of the limited attentional resources of humans, whereas SA is a cognitive phenomenon emerging from perception, memory, and expertise. The concepts of workload and SA have been studied extensively in the laboratory and have been transitioned successfully to real-world system evaluation. Indeed, workload and SA have been useful tools of system evaluators for years, and now they are providing vital guidance for shaping future automation, display, and training programs. In short, these concepts have been, and should continue to be, essential tools for human factors researchers and practitioners.

REFERENCES

Adams, R. J., and Ericsson, K. A. (1992), "Introduction to Cognitive Processes of Expert Pilots," DOT/FAA/RD-92/12, U.S. Department of Transportation, Federal Aviation Administration, Washington, DC.

Adams, M. J., Tenny, Y. J., and Pew, R. W. (1991), "Strategic Workload and the Cognitive Management of Advanced Multi-task System," State of the Art Report CSERIAC 91-6, Crew System Ergonomics Information Evaluation Center, Wright-Patterson Air Force Base, OH.

Adams, M. J., Tenny, Y. J., and Pew, R. W. (1995), "Situation Awareness and the Cognitive Management of Complex Systems," *Human Factors*, Vol. 37, pp. 85–104.

Alexander, A. L., Nygren, T. E., and Vidulich, M. A. (2000), "Examining the Relationship Between Mental Workload and Situation Awareness in a Simulated Air Combat Task," Technical Report AFRL-HE-WP-TR-2000-0094, Air Force Research Laboratory, Wright-Patterson Air Force Base, OH.

Allport, D. A., Antonis, B., and Reynolds, P. (1972), "On the Division of Attention: A Disproof of the Single Channel Hypothesis," *Quarterly Journal of Experimental Psychology*, Vol. 24, pp. 255–265.

Andre, A. D., Wickens, C. D., Moorman, L., and Boschelli, M. M. (1991), "Display Formatting Techniques for Improving Situation Awareness in the Aircraft Cockpit," *International Journal of Aviation Psychology*, Vol. 1, pp. 205–218.

Annett, J. (2002a), "Subjective Rating Scales: Science or Art?" *Ergonomics*, Vol. 45, pp. 966–987.

Annett, J. (2002b), "Subjective Rating Scales in Ergonomics: A Reply," *Ergonomics*, Vol. 45, pp. 1042–1046.

Backs, R. W. (1997), "Psychophysiological Aspects of Selective and Divided Attention during Continuous Manual Tracking," *Acta Psychologica*, Vol. 96, pp. 167–191.

Backs, R. W., and Boucsein, W. (2000), *Engineering Psychophysiology: Issues and Applications*, Lawrence Erlbaum Associates, Mahwah, NJ.

Baddeley, A. (1990), *Human Memory: Theory and Practice*, Allyn, & Bacon, Boston.

Bainbridge, L. (1987), "Ironies of Automation," in *New Technology and Human Error*, J. Rassmusen, K. Duncan, and J. Leplat, Eds., Wiley, New York, pp. 271–283.

Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S., et al. (1996), "Neural Systems Engaged by Planning: A PET Study of the Tower of London Task," *Neuropsychologia*, Vol. 34, pp. 515–526.

Battiste, V., and Bortolussi, M. (1988), "Transport Pilot Workload: A Comparison of Two Subjective Techniques," in *Proceedings of the Human Factors Society 32nd Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 150–154.

Beatty, J. (1982), "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources," *Psychological Bulletin*, Vol. 91, pp. 276–292.

Bi, S., and Salvendy, G. (1994), "Analytical Modeling and Experiment Study of Human Workload in Scheduling of Advanced Manufacturing Systems," *International Journal of Human Factors in Manufacturing*, Vol. 4, pp. 205–235.

Biers, D. W., and Maseline, P. J. (1987), "Alternative Approaches to Analyzing SWAT Data," in *Proceedings of the Human Factors Society 31st Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 63–66.

Biers, D. W., and McInerney, P. (1988), "An Alternative to Measuring Subjective Workload: Use of SWAT Without the Card Sort," in *Proceedings of the Human Factors Society 32nd Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 1136–1139.

Billings, C. E. (1997), *Aviation Automation: The Search for a Human-Centered Approach*, Lawrence Erlbaum Associates, Mahwah, NJ.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997), "A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory," *NeuroImage*, Vol. 5, pp. 49–62.

Brickman, B. J., Hettinger, L. J., Roe, M. M., Stautberg, D., Vidulich, M. A., Haas, M. W., and Shaw, R. L. (1995), "An Assessment of Situation Awareness in an Air Combat Task: The Global Implicit Measure Approach," in *Experimental Analysis and Measurement of Situation Awareness*, D. J. Garland and M. R. Endsley, Eds., Embry-Riddle Aeronautical University Press, Daytona Beach, FL, pp. 339–344.

Brickman, B. J., Hettinger, L. J., Stautberg, D., Haas, M. W., Vidulich, M. A., and Shaw, R. L. (1999), "The Global Implicit Measurement of Situation Awareness: Implications for Design and Adaptive Interface Technologies," in *Automation Technology and Human Performance: Current Research and Trends*, M. W. Scerbo and M. Mouloua, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 160–164.

Brookings, J., Wilson, G. F., and Swain, C. (1996), "Psychophysiological Responses to Changes in Workload During Simulated Air Traffic Control," *Biological Psychology*, Vol. 42, pp. 361–378.

Byrne, E. A. (1995), "Role of Volitional Effort in the Application of Psychophysiological Measures to Situation Awareness," in *Experimental Analysis and Measurement of Situation Awareness*, D. J. Garland and M. R. Endsley, Eds., Embry-Riddle Aeronautical University Press, Daytona Beach, FL, pp. 147–153.

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1999), "Graded Functional Activation in the Visuospatial System with the Amount of Task Demand," *Journal of Cognitive Neuroscience*, Vol. 11, pp. 9–24.

Charness, N. (1989), "Age and Expertise: Responding to Talland's Challenge," in *Everyday Cognition in Adulthood and Late Life*, L. W. Poon, D. C. Rubin, and B. A. Wilson, Eds., Cambridge University Press, Cambridge. pp. 437–456.

Charness, N. (1995), "Expert Performance and Situation Awareness," in *Experimental Analysis and Measurement of Situation Awareness*, D. J. Garland and M. R. Endsley, Eds., Embry-Riddle Aeronautical University Press, Daytona Beach, FL, pp. 35–42.

Chase, W. G., and Ericsson, K. A. (1982), "Skill and Working Memory," in *The Psychology of Learning and Motivation*, Vol. 16, G. H. Bower, Ed., Academic Press, New York.

Chi, M., Glaser, R., and Farr, M., Eds. (1988), *On the Nature of Expertise*, Lawrence Erlbaum Associates, Mahwah, NJ.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1990), "Attentional Modulation of Neural Processing of Shape, Color, and Velocity in Humans," *Science*, Vol. 248, pp. 1556–1559.

Davidson, R. J., Jackson, D. C., and Larson, C. L. (2000), "Human Electroencephalography," in *Handbook of Psychophysiology*, J. T. Cacioppo, L. G., Tassinary, and G. G., Berntson, Eds., Cambridge University Press, Cambridge, pp. 27–52.

D'Esposito, M., Zarahn, E., and Aguire, G. K. (1999), "Event-Related Functional MRI: Implications for Cognitive Psychology," *Psychological Bulletin*, Vol. 125, pp. 155–164.

Druckman, D., and Bjork, R. A., Eds. (1991), *In the Mind's Eye: Enhancing Human Performance*, National Academy Press, Washington, DC, Chap. 4.

Durso, F. T., and Gironlund, S. D. (1999). "Situation Awareness," in F. T. Durso, Ed., *Handbook of Applied Cognition*, Wiley, New York, pp. 283–314.

Durso, F. T., Truitt, T. R., Hackworth, C. A., Ohrt, D., Hamic, J. M., Crutchfield, J. M., and Manning, C. A. (1995), "Expertise and Chess: A Pilot Study Comparing Situation Awareness Methodologies," in *Experimental Analysis and Measurement of Situation Awareness*, D. J. Garland and M. R. Endsley, Eds., Embry-Riddle Aeronautical University Press, Daytona Beach, FL, pp. 189–195.

Eggemeier, F. T., Crabtree, M. S., Zingg, J. J., Reid, G. B., and Shingledecker, C. A. (1982), "Subjective Workload Assessment in a Memory Update Task," in *Proceedings of the 26th Human Factors Society Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 643–647.

Endsley, M. R. (1988), "Situation Awareness Global Assessment Technique (SAGAT)," in *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, NAECON 1988*, Vol. 3, Institute of Electrical and Electronics Engineers, New York, pp. 789–795.

Endsley, M. (1990), "A Methodology for the Objective Measurement of Pilot Situation Awareness," in *Situation Awareness in Aerospace Operations*, AGARD-CP-478, AGARD, Neuilly-sur-Seine, France, pp. 1–1 to 1–9.

Endsley, M. (1993), "Situation Awareness and Workload: Flip Sides of the Same Coin," in *Proceedings of the 7th International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH, pp. 906–911.

Ericsson, K. A. (1996), "The Acquisition of Expert Performance: An Introduction to Some of the Issues," in *The Road to Excellence*, K. A. Ericsson, Ed., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 1–50.

Ericsson, K. A., and Charness, N. (1997), "Cognitive and Developmental Factors in Expert Performance," in *Expertise in Context*, P. J. Feltovich, K. M. Ford, and R. R. Hoffman, Eds., MIT Press, Cambridge, MA, pp. 3–41.

Ericsson, K. A., and Delaney, P. F. (1998), "Working Memory and Expert Performance," in *Working Memory and Thinking*, R. H. Logie and K. J. Gilhooly, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 93–114.

Ericsson, K. A., and Kintsch, W. (1995), "Long-Term Working Memory," *Psychological Review*, Vol. 105, pp. 211–245.

Ericsson, K. A., and Simon, H. A. (1993), *Protocol Analysis: Verbal Reports As Data*, rev. ed., MIT Press, Cambridge, MA.

Fallows, J. (2001a), *Free Flight: From Airline Hell to a New Age of Travel*, Public Affairs, New York.

Fallows, J. (2001b), "Freedom of the Skies," *The Atlantic*, Vol. 287, No. 6, June, pp. 37–49.

Fiorino, F. (2001), "Tech Trickle-Down Enhances GA Safety," *Aviation Week and Space Technology*, Vol. 155, No. 7, August 13, p. 52.

Fitts, P. M., and Jones, R. E. (1961a), "Analysis of Factors Contributing to 460 'Pilot-Error' Experiences in Operating Aircraft Controls," in *Selected Papers on Human Factors in the Design and Use of Control Systems*, H. W. Sinaiko, Ed., Dover, New York, pp. 332–358; reprinted from "Analysis of Factors Contributing to 460 'Pilot-Error' Experiences in Operating Aircraft Controls," Memorandum Report TSEAA-694-12, Aero Medical Laboratory, Air Materiel Command, Wright-Patterson Air Force Base, OH.

Fitts, P. M., and Jones, R. E. (1961b), "Psychological Aspects of Instrument Display," in *Selected Papers on Human Factors in the Design and Use of Control Systems*, H. W. Sinaiko, Ed., Dover, New York, pp. 359–396; reprinted from "Psychological Aspects of Instrument Display," Memorandum Report TSEAA-694-12A, Aero Medical Laboratory, Air Materiel Command, Wright-Patterson Air Force Base, OH.

Fournier, L. R., Wilson, G. F., and Swain, C. R. (1999), "Electrophysiological, Behavioral, and Subjective Indexes of Workload When Performing Multiple Tasks: Manipulations of Task Difficulty and Training," *International Journal of Psychophysiology*, Vol. 31, pp. 129–145.

Fowler, B. (1994), "P300 as a Measure of Workload During a Simulated Aircraft Landing Task," *Human Factors*, Vol. 36, pp. 670–683.

Friedman, A., and Polson, M. C. (1981), "Hemispheres as Independent Resources Systems: Limited-Capacity Processing and Cerebral Specialization," *Journal of Experimental Psychology*, Vol. 7, pp. 1031–1058.

Fruhstorfer, H., Langanke, P., Meinzer, K., Peter, J., and Plaff, U. (1977), "Neurophysiological Vigilance Indicators and Operational Analysis of a Train Vigilance Monitoring Device: A Laboratory and Field Study," in *Vigilance: Theory, Physiological Correlates, and Operational Performance*, R. R. Mackie, Ed., Plenum Press, New York, pp. 147–162.

Gawron, V. J. (2000), *Human Performance Measures Handbook*, Lawrence Erlbaum Associates, Mahwah, NJ.

Gevins, A., and Smith, M. E. (2003), "Neurophysiological Measures of Cognitive Workload During Human–Computer Interaction," *Theoretical Issues in Ergonomics Science*, Vol. 4, pp. 113–131.

Gevins, A., Leong, H., Du R., Smith, M. E., Le, J., DuRousseau, D., et al. (1995), "Towards Measurement of Brain Function in Operational Environments," *Biological Psychology*, Vol. 40, pp. 169–186.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998), "Monitoring Working Memory During Computer-Based Tasks with EEG Pattern Recognition," *Human Factors*, Vol. 40, pp. 79–91.

Glaser, R. (1987), "Thoughts on Expertise," in *Cognitive Functioning and Social Structure over the Life Course*, C. Schooler and K. W. Schaie, Eds., Ablex, Norwood, NJ, pp. 81–94.

Glass, A. (1966), "Comparison of the Effect of Hard and Easy Mental Arithmetic upon Blocking of the Occipital Alpha Rhythm," *Quarterly Journal of Experimental Psychology*, Vol. 8, pp. 142–152.

Gopher, D. (1993), "The Skill of Attention Control: Acquisition and Execution of Attention Strategies," in *Attention and Performance XIV*, D. Meyer and S. Kornblum, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 299–322.

Gopher, D. (1994), "Analysis and Measurement of Mental Load," in *International Perspectives on Psychological Science*, Vol. 2, *The State of the Art*, G. d'Ydewalle, P. Eelen, and P. Bertelson, Eds., Lawrence Erlbaum Associates, Hove, East Sussex, England, pp. 265–291.

Gopher, D., and Braune, R. (1984), "On the Psychophysics of Workload: Why Bother with Subjective Ratings?" *Human Factors*, Vol. 26, pp. 519–532.

Gopher, D., and Donchin, E. (1986), "Workload: An Examination of the Concept," in *Handbook of Perception and Human Performance*, Vol. II, *Cognitive Processes and Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., Wiley, New York, pp. 41–1 to 41–49.

Gopher, D., Brickner, M., and Navon, D. (1982), "Different Difficulty Manipulation Interacts Differently with Task Emphasis: Evidence for Multiple Resources," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 8, pp. 146–157.

Haier, R. J., Siegel, B., Tang, C., Abel, L., and Buchsbaum, M. (1992), "Intelligence and Changes in Regional Cerebral Glucose Metabolic Rate Following Learning," *Intelligence*, Vol. 16, pp. 415–426.

Hardiman, T. D., Dudfield, H. J., Selcon, S. J., and Smith, F. J. (1996), "Designing Novel Head-Up Displays to Promote Situational Awareness," in *Situation Awareness: Limitations and Enhancement in the Aviation Environment*, AGARD-CP-575, Advisory Group for Aerospace Research and Development, Neuilly-sur-Seine, France, pp. 15–1 to 15–7.

Hart, S. G., and Staveland, L. E. (1988), "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., Elsevier, Amsterdam, The Netherlands, pp. 139–183.

Hendy, K. D. (1995), "Situation Awareness and Workload: Birds of a Feather?" in *AGARD Conference Proceedings 575: Situation Awareness: Limitations and Enhancements in the Aviation Environment*, *Advisory Group for Aerospace Research and Development*, Neuilly-sur-Seine, France, pp. 21–1 to 21-7.

Hendy, K. C., Hamilton, K. M., and Landry, L. N. (1993), "Measuring Subjective Workload: When Is One Scale Better Than Many?" *Human Factors*, Vol. 35, pp. 579–601.

Hilburn, B., Jorna, P. G., Byrne, E. A., and Parasuraman, R. (1997). "The Effect of Adaptive Air Traffic Control (ATC) Decision Aiding on Controller Mental Workload," in M. Mouloua and J. Koonce, Eds., *Human Automation Interaction*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 84–91.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., and Christ, R. E. (1992), "Comparison of Four Subjective Workload Rating Scales," *Human Factors*, Vol. 34, pp. 429–439.

Hockey, G. R. J. (1997), "Compensatory Control in the Regulation of Human Performance Under Stress and High Workload: A Cognitive–Energetic Framework," *Biological Psychology*, Vol. 45, pp. 73–93.

Hockey, R., Gaillard, A., and Coles, M. (1986), *Energetics and Human Information Processing*, Matinus Nijhoff, Dordrecht, The Netherlands.

Hockey, R. J., Healey, A., Crawshaw, M., Wastell, D. G., and Sauer, J. (2003), "Cognitive Demands of Collision Avoidance in Simulated Ship Control," *Human Factors*, Vol. 45, pp. 252–265.

Hoffman, J., Houck, M., MacMilliam, F., Simons, R., and Oatman, L. (1985), "Event Related Potentials Elicited by Automatic Targets: A Dual-Task Analysis," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 11, pp. 50–61.

Horn, J. L., and Masunaga, H. (2000), "New Directions for Research into Aging and Intelligence: The Development of Expertise," in *Models of Cognitive Aging*, T. J. Perfect and E. A. Maylor, Eds., Oxford University Press, New York, pp. 125–159.

Huey, B. M., and Wickens, C. D., Eds. (1993), *Workload Transition*, National Academy Press, Washington, DC.

Isreal, J. B., Chesney, G. L., Wickens, C. D., and Donchin, E. (1980), "P300 and Tracking Difficulty: Evidence for Multiple Resources in Dual Task Performance," *Psychobiology*, Vol. 17, pp. 57–70.

Jha, P. D., Bisantz, A. M., and Parasuraman, R. (2003), "Lens Model Analysis of Pilot and Controller Decision Making Under Free Flight," in *Proceedings of the 12th International Symposium on Aviation Psychology*, Wright State University, Dayton, OH, pp. 619–624.

Jorna, P. G. A. M. (1992), "Spectral Analysis of Heart Rate and Psychological State: A Review of Its Validity as a Workload Index," *Biological Psychology*, Vol. [illegible], pp. 237–258.

Jorna, P. G. A. M. (1999), "Automation and Free Flight: Exploring the Unexpected, in *Automation Technology and Human Performance*, M. W. Scerbo and M. Mouloua, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 107–111.

Just, M. A., and Carpenter, P. A. (1992), "A Capacity Theory of Comprehension: Individual Differences

Working Memory," *Psychological Review*, Vol. 99, pp. 122–149.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1996), "Brain Activation Modulated by Sentence Comprehension," *Science*, Vol. 274, pp. 114–116.

Just, M. A., Carpenter, P. A., and Miyake, A. (2003), "Neuroindices of Cognitive Workload: Neuroimaging, Pupillometric and Event-Related Potential Studies of Brain Work," *Theoretical Issues in Ergonomics Science*, Vol. 4, pp. 56–88.

Kaber, D. B., and Riley, J. M. (1999), "Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement," *International Journal of Cognitive Ergonomics*, Vol. 3, pp. 169–187.

Kahneman, D. (1973), *Attention and Effort*, Prentice-Hall, Englewood Cliffs, NJ.

Kantowitz, B. H. (1992), "Selecting Measures for Human Factors Research," *Human Factors*, Vol. 34, pp. 387–398.

Kessel, C. J. and Wickens, C. D. (1982), "The Transfer of Failure Detection Skills Between Monitoring and Controlling Dynamic Systems," *Human Factors*, Vol. 24, pp. 49–60.

Kinsbourne, M., and Hicks, R. (1978), "Functional Cerebral Space," in *Attention and Performance VII*, J. Requin, Ed., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 345–362.

Kramer, A. F., and McCarley, J. S. (2003), "Oculomotor Behaviour as a Reflection of Attention and Memory Processes: Neural Mechanisms and Applications to Human Factors," *Theoretical Issues in Ergonomics Science*, Vol. 4, pp. 21–55.

Kramer, A. F., and Weber, T. (2000), "Applications of Psychophysiology to Human Factors," in *Handbook of Psychophysiology*, 2nd ed., J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds., Cambridge University Press, New York, pp. 794–814.

Kramer, A. F., Sirevaag, E., and Braune, R. (1987), "A Psychophysiological Assessment of Operator Workload During Simulated Flight Missions," *Human Factors*, Vol. 29, pp. 145–160.

Kramer, A. F., Trejo, L. J., and Humphrey, D. G. (1996), "Psychophysiological Measures of Workload: Potential Applications to Adaptively Automated Systems," in *Automation and Human Performance: Theory and Application*, R. Parasuraman and M. Mouloua, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 137–161.

Lee, Y., and Liu, B. (2003), "Inflight Workload Assessment: Comparison of Subjective and Physiological Measurements," *Aviation, Space, and Environmental Medicine*, Vol. 74, pp. 1078–1084.

Lee, J. D., and Moray, N. (1994), "Trust, Self-Confidence, and Operators' Adaptation to Automation," *International Journal of Human–Computer Studies*, Vol. 40, pp. 153–184.

Leyman, E. L. C., Mirka, G. A., Kaber, D. B., and Sommerich, C. M. (2004), "Cervicobrachial Muscle Response to Cognitive Load in a Dual-Task Scenario," *Ergonomics*, Vol. 47, pp. 625–645.

Lidderdale, I. G. (1987), "Measurement of Aircrew Workload in Low-Level Flight," in *The Practical Assessment of Pilot Workload, AGARDograph* 282, A. Roscoe, Ed., *Advisory Group for Aerospace Research and Development*, Neuilly-sur-Seine, France, pp. 67–77.

Luximon, A., and Goonetilleke, R. S. (2001), "Simplified Subjective Workload Assessment Technique," *Ergonomics*, Vol. 44, pp. 229–243.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., Jr., and Wherry, R. J. Jr. (1989), "Operator Workload: Comprehensive Review and Evaluation of Workload Methodologies," ARI Technical Report 851, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

McCoy, T. M., Derrick, W. L., and Wickens, C. D. (1983). "Workload Assessment Metrics: What Happens When They Dissociate?" in *Proceedings of the 2nd Aerospace Behavioral Engineering Technology Conference, P-132*, Society of Automotive Engineers, Warrendale, PA, pp. 37–42.

Metzger, U., Rovira, E., and Parasuraman, R. (2003), "Controller Performance, Workload and Attention Allocation in Distributed Air–Ground Traffic Management: Effects of Mixed Equipage and Decision Support," in *Proceedings of the 12th International Symposium on Aviation Psychology*, Wright State University, Dayton, OH, pp. 803–809.

Moray, N., Ed. (1979), *Mental Workload: Theory and Measurement*, Plenum Press, New York.

Moray, N. (1986), "Monitoring Behavior and Supervisory Control," in *Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., Wiley, New York, pp. 40–1 to 40-51.

Muckler, F. A., and Seven, S. A. (1992), "Selecting Performance Measures: 'Objective' Versus 'Subjective' Measurement," *Human Factors*, Vol. 34, pp. 441–456.

Mulder, L. J. M. (1992), "Measurement and Analysis Methods of Heart Rate and Respiration for Use in Applied Environments," *Biological Psychology*, Vol. 34, pp. 205–236.

Navon, D., and Gopher, D. (1979), "On the Economy of the Human-Processing System," *Psychology Review*, Vol. 86, pp. 214–255.

Norman, D. A., and Bobrow, D. G. (1975), "On Data-Limited and Resource-Limited Processes," *Cognitive Psychology*, Vol. 7, pp. 44–64.

Norman, G. R., Brooks, L. R., and Allen, S. W. (1989), "Recall by Expert Medical Practitioners and Novices as a Record of Processing Attention," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 15, pp. 1166–1174.

Nygren, T. E. (1991), "Psychometric Properties of Subjective Workload Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload," *Human Factors*, Vol. 33, pp. 17–33.

O'Donnell, R., and Eggemeier, F. T. (1986), "Workload Assessment Methodology," in *Handbook of Perception and Human Performance*, Vol. II, *Cognitive Processes and Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., Wiley, New York, pp. 42–1 to 42-49.

Ogden, G. D., Levine, J. M., and Eisner, E. J. (1970), "Measurement of Workload by Secondary Tasks," *Human Factors*, Vol. 21, pp. 529–548.

Parasuraman, R. (1990), "Event-Related Brain Potentials and Human Factors Research," in *Event-Related Brain Potentials: Basic Issues and Applications*, J. W. Rohrbaugh, R. Parasuraman, and R. Johnson, Jr., Eds., Oxford University Press, New York, pp. 279–299.

Parasuraman, R. (2003), "Neuroergonomics: Research and Practice," *Theoretical Issues in Ergonomics Science*, Vol. 4, pp. 5–20.

Parasuraman, R., and Bryne, E. A. (2003), "Automation and Human Performance in Aviation," in *Principles and Practice of Aviation Psychology*, P. S. Tsang and M. A. Vidulich, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 311–356.

Parasuraman, R., and Caggiano, D. (2002), "Mental Workload," in *Encyclopedia of the Human Brain*, V. S. Ramachandran, Ed., Academic Press, San Diego, CA, Vol. 3, pp. 17–27.

Parks, R. W., Lowenstein, D. A., Dodrill, K. L., Barker, W. W., Yoshi, F., Chang, J. Y., et al. (1988), "Cerebral Metabolic Effects of a Verbal Fluency Test: A PET Scan Study," *Journal of Clinical and Experimental Neurophysiology*, Vol. 10, pp. 565–575.

Patel, V. L., and Groen, G. J. (1991), "The General and Specific Nature of Medical Expertise: A Critical Look," in K. A. Ericsson and J. Smith, Eds., *Toward a General Theory of Expertise* Cambridge University Press, Cambridge, MA, pp. 93–125.

Pérès, M., Van De Moortele, P. F., Pierard, C., Lehericy, S., Satabin, P., Le Bihan, D. et al. (2000), "Functional Magnetic Resonance Imaging of Mental Strategy in a Simulated Aviation Performance Task," *Aviation, Space, and Environmental Medicine*, Vol. 71, pp. 1218–1231.

Pew, R. A. (1994), "An Introduction to the Concept of Situation Awareness," in *Situational Awareness in Complex Systems*, R. D. Gilson, D. J. Garland, and J. M. Koonce, Eds., Embry-Riddle Aeronautical University Press, Daytona Beach, FL, pp. 17–23.

Posner, M. I., and DiGirolamo, G. J. (2000), "Cognitive Neuroscience: Origin and Promise," *Psychological Bulletin*, Vol. 126, pp. 873–889.

Previc, F. H. (2000), "Neuropsychological Guidelines for Aircraft Control Stations," *IEEE Engineering in Medicine and Biology*, March–April, pp. 81–88.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2000), "A Closed-Loop System for Examining Psychophysiological Measures for Adaptive Automation, *International Journal of Aviation Psychology*, Vol. 10, pp. 393–410.

Pritchett, A. R., and Hansman, R. J. (2000), "Use of Testable Responses for Performance-Based Measurement of Situation Awareness," in *Situation Awareness Analysis and Measurement*, M. R. Endsley and D. J. Garland, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 189–209.

Reid, G. B., and Nygren, T. E. (1988), "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., Elsevier, Amsterdam, The Netherlands, pp. 185–218.

Reingold, E., and Charness, N. (1995), "Perceptual Automaticity in Chess Skill: Evidence from Eye Movements," poster presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles.

Rothrock, L., Koubek, R., Fuchs, F., Haas, M., and Salvendy, G. (2002), "Review and Reappraisal of Adaptive Interfaces: Toward Biologically Inspired Paradigms," *Theoretical Issues in Ergonomics Science*, Vol. 3, pp. 47–84.

Rouse, W. B. (1977), "Human–Computer Interaction in Multitask Situations," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 7, pp. 293–300.

Rouse, W. B. (1988), "Adaptive Aiding for Human/Computer Control," *Human Factors*, Vol. 30, pp. 431–443.

Rubio, S., Díaz, E., Martín, J., and Puente, J. M. (2004), "Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods," *Applied Psychology: An International Review*, Vol. 53, pp. 61–86.

Saaty, T. L. (1980), *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill, New York.

Salvendy, G. (2002), "Use of Subjective Rating Scores in Ergonomics Research and Practice," *Ergonomics*, Vol. 45, pp. 1005–1007.

Selcon, S. J., Taylor, R. M., and Koritsas, E. (1991), "Workload or Situational Awareness? TLX vs. SART for Aerospace System Design Evaluation," in *Proceedings of the Human Factors Society 35th Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 62–66.

Selcon, S. J., Taylor, R. M., and Shadrake, R. M. (1992), "Multi-modal Cockpit Warnings: Pictures, Words, or Both?" in *Proceedings of the Human Factors Society 36th Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 57–61.

Selcon, S. J., Hardiman, T. D., Croft, D. G., and Endsley, M. R. (1996), "A Test-Battery Approach to Cognitive Engineering: To Meta-measure or Not to Meta-measure, That Is the Question!" in *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, Vol. 1, Human Factors and Ergonomics Society, Santa Monica, CA, pp. 228–232.

Shallice, T. (1988), *From Neuropsychology to Mental Structure*, Cambridge University Press, New York.

Shaw, R. L., Brickman, B. J., and Hettinger, L. J. (2004), "The Global Implicit Measure (GIM): Concept and Experience," in *Human Performance, Situation Awareness and Automation*, D. A. Vincenzi, M. Mouloua, and P. A. Hancock, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 97–102.

Shingledecker, C. A. (1984), "A Task Battery for Applied Human Performance Assessment Research," Technical Report AFAMRL-TR-84-071, Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, OH.

Sirevaag, E. J., Kramer, A. F., Coles, M. G. H., and Donchin, E. (1989), "Resource Reciprocity: An Event-Related Brain Potentials Analysis," *Acta Psychologica*, Vol. 70, pp. 77–97.

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., and Grenell, J. (1993), "Assessment of Pilot Performance and Mental Workload in Rotary Wing Aircraft," *Ergonomics*, Vol. 36, pp. 1121–1140.

Sperandio, A. (1978), "The Regulation of Working Methods as a Function of Workload Among Air Traffic Controllers," *Ergonomics*, Vol. 21, pp. 367–390.

Staszewski, J. J. (1988), "Skilled Memory and Expert Mental Calculation," in *The Nature of Expertise*, M. T. H. Chi and R. Glaser, Eds., Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 71–128.

Sterman, B., and Mann, C. (1995), "Concepts and Applications of EEG Analysis in Aviation Performance Evaluation," *Biological Psychology*, Vol. pp. 115–130.

Stern, J. A., Walrath, L. C., and Goldstein, R. (1984), "The Endogenous Eyeblink," *Psychophysiology*, Vol. pp. 22–33.

Stevens, S. S. (1957), "On the Psychophysical Law," *Psychological Review*, Vol. 64, pp. 153–181.

Stevens, S. S. (1966), "On the Operation Known as Judgment," *American Scientist*, Vol. 54, pp. 385–401.

Strater, L. D., Endsley, M. R., Pleban, R. J., and Matthews, M. D. (2001), "Measures of Platoon Leader Situation Awareness in Virtual Decision-Making Exercises," Technical Report 1770, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Strayer, D., and Kramer, A. F. (1990), "Attentional Requirments of Automatic and Controlled Processes," *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 16, pp. 67–82.

Taylor, R. M. (1990), "Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design," in *Situational Awareness in Aerospace Operations, AGARD-CP-478*, Advisory Group for Aerospace Research and Development, Neuilly-sur-Seine, France, pp. 3–1 to 3–17.

Tenny, Y. J., Adams, M. J., Pew, R. W. Huggins, A. W. F., and Rodgers, W. H. (1992), "A Principled Approach to the Measurement of Situation Awareness in Commercial Aviation," NASA Contractor Report 4451, Langley Research Center, National Aeronautics and Space Administration, Hampton, VA.

Toms, M. L., Cavallo, J. J., Cone, S. M., Moore, F. W., and Gonzalez-Garcia, A. (1997), "Pilot-in-the-Loop Evaluation of the Approach Procedures Expert System (APES)," Technical Report WL-TR-97-3102, Wright Laboratory, Wright-Patterson Air Force Base, OH.

Tsang, P. S., and Shaner, T. L. (1998), "Age, Attention, Expertise, and Time-Sharing Performance," *Psychology and Aging*, Vol. 13, pp. 323–347.

Tsang, P. S., and Velazquez, V. L. (1996), "Diagnosticity and Multidimensional Workload Ratings," *Ergonomics*, Vol. 39, pp. 358–381.

Tsang, P. S., and Vidulich, M. A. (1989), "Cognitive Demands of Automation in Aviation," in *Aviation Psychology*, R. S. Jensen, Ed., Gower Publishing, Aldershot, Hampshire, England, pp. 66–95.

Tsang, P. S., and Vidulich, M. A. (1994), "The Roles of Immediacy and Redundancy in Relative Subjective Workload Assessment," *Human Factors*, Vol. 36, pp. 503–513.

Tsang, P. S., and Wilson, G. (1997), "Mental Workload," in *Handbook of Human Factors and Ergonomics*, 2nd ed., G. Salvendy, Ed., Wiley, New York, pp. 417–449.

Tsang, P. S., Velazquez, V. L., and Vidulich, M. A. (1996), "The Viability of Resource Theories in Explaining Time-Sharing Performance," *Acta Psychologica*, Vol. 91, pp. 175–206.

Ververs, P. M., and Wickens, C. D. (2000), "Designing Head-Up Displays (HUDs) to Support Flight Path Guidance While Minimizing Effects of Cognitive Tunneling," in *Proceedings of the IEA 2000/HFES 2000 Congress*, Vol. 3, Human Factors and Ergonomics Society, Santa Monica, CA, pp. 3–45 to 3–48.

Vidulich, M. A. (1988), "The Cognitive Psychology of Subjective Mental Workload," in *Human Mental Workload*, P. Hancock and N. Meshkati, Eds., North-Holland, Amsterdam, The Netherlands, pp. 219–229.

Vidulich, M. (1989), "The Use of Judgment Matrices in Subjective Workload Assessment: The Subjective WORkload Dominance (SWORD) Technique," in *Proceedings of the Human Factors Society 33rd Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 1406–1410.

Vidulich, M. A. (1995), "The Role of Scope as a Feature of Situation Awareness Metrics," in *Experimental Analysis and Measurement of Situation Awareness*, D. J. Garland and M. R. Endsley, Eds., Embry-Riddle Aeronautical University Press, Daytona Beach, FL, pp. 69–74.

Vidulich, M. A. (2000), "Testing the Sensitivity of Situation Awareness Metrics in Interface Evaluations," in *Situation Awareness Analysis and Measurement*, M. R. Endsley and D. J. Garland, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 227–246.

Vidulich, M. A. (2003), "Mental Workload and Situation Awareness: Essential Concepts for Aviation Psychology Practice," in *Principles and Practice of Aviation Psychology*, P. S. Tsang and M. A. Vidulich, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 115–146.

Vidulich, M. A., and Bortolussi, M. R., (1988), "Speech Recognition in Advanced Rotorcraft: Using Speech Controls to Reduce Manual Control Overload," in *Proceedings of the American Helicopter Society National Specialist's Meeting: Automation Applications for Rotorcraft*, American Helicopter Society, Atlanta Southeast Region, Atlanta, GA, pp. 1–10.

Vidulich, M. A., and Hughes, E. R. (1991), "Testing a Subjective Metric of Situation Awareness," in *Proceedings of the Human Factors Society 35th Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 1307–1311.

Vidulich, M. A., and McMillan, G. (2000), "The Global Implicit Measure: Evaluation of Metrics for Cockpit Adaptation," in *Contemporary Ergonomics 2000*, P. T. McCabe, M. A. Hanson, and S. A. Robertson, Eds., Taylor & Francis, London, pp. 75–80.

Vidulich, M. A., and Tsang, P. S. (1987), "Absolute Magnitude Estimation and Relative Judgment Approaches to Subjective Workload Assessment," in *Proceedings of the Human Factors Society 31st Annual Meeting*, Human Factors Society, Santa Monica, CA, pp. 1057–1061.

Vidulich, M. A., and Wickens, C. D. (1986), "Causes of Dissociation Between Subjective Workload Measures and Performance: Caveats for the Use of Subjective Assessments," *Applied Ergonomics*, Vol. 17, pp. 291–296.

Vidulich, M., Dominguez, C., Vogel, E., and McMillan, G. (1994a), "Situation Awareness: Papers and Annotated Bibliography," Technical Report AL/CF-TR-1994-0085, Armstrong Laboratory, Wright-Patterson Air Force Base, OH.

Vidulich, M. A., Stratton, M., Crabtree, M., and Wilson, G. (1994b), "Performanced-Based and Physiological Measures of Situational Awareness," *Aviation, Space, and Environmental Medicine*, Vol. 65 (Suppl.), pp. A7–A12.

Vidulich, M. A., McCoy, A. L., and Crabtree, M. S. (1995), "The Effect of a Situation Display on Memory Probe and Subjective Situational Awareness Metrics," in *Proceedings of the 8th International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH, pp. 765–768.

Watson, J. B. (1913), "Psychology as the Behaviorist Views It," *Psychological Review*, Vol. 20, pp. 158–177.

Welford, A. T. (1978), "Mental Workload as a Function of Demand, Capacity, Strategy and Skill," *Ergonomics*, Vol. 21, pp. 151–167.

Wickens, C. D. (1980), "The Structure of Attentional Resources," in *Attention and Performance VIII*, R. S. Nickerson, Ed., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 239–257.

Wickens, C. D. (1984), "Processing Resources in Attention," in *Varieties of Attention*, R. Parasuraman and D. R. Davies, Eds., Academic Press, San Diego, CA, pp. 63–102.

Wickens, C. D. (1987), "Attention," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., North-Holland, New York, pp. 29–80.

Wickens, C. D. (1990), "Applications of Event-Related Potential Research to Problems in Human Factors," in *Event-Related Brain Potentials: Basic Issues and Applications*, J. W. Rohrbaugh, R. Parasuraman, and R. Johnson, Jr., Eds., Oxford University Press, New York, pp. 301–309.

Wickens, C. D. (1995), "Situation Awareness: Impact of Automation and Display Technology," in *AGARD Conference Proceedings 575: Situation Awareness: Limitations and Enhancements in the Aviation Environment, Advisory Group for Aerospace Research and Development*, Neuilly-sur-Seine, France, pp. K2-1 to K2-13.

Wickens, C. (2001), "Workload and Situation Awareness," in *Stress, Workload, and Fatigue*, P. A. Hancock and P. A. Desmond, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 443–450.

Wickens, C. D. (2002), "Multiple Resources and Performance Prediction," *Theoretical Issues in Ergonomics Science*, Vol. 3, pp. 159–177.

Wickens, C. D. (2003), "Aviation Displays," in *Principles and Practice of Aviation Psychology*, P. S. Tsang and M. A. Vidulich, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 147–200.

Wickens, C. D. and Hollands, J. G. (2000), *Engineering Psychology and Human Performance*, 3rd ed., Charles E. Merrill, Columbus, OH.

Wickens, C. D., Kramer, A. F., Vanasse, L., and Donchin, E. (1983), "The Performance of Concurrent Tasks: A Psychophysiological Analysis of the Reciprocity of Information Processing Resources," *Science*, Vol. 221, pp. 1080–1082.

Wiener, E. L. (1988), "Cockpit Automation," in *Human Factors in Aviation*, E. L. Wiener and D. C. Nagel, Eds., Academic Press, San Diego, CA, pp. 433–461.

Wiener, E. L., and Curry, R. E. (1980), "Flight-Deck Automation: Promises and Problems," *Ergonomics*, Vol. 23, pp. 955–1011.

Wilson, G. F. (2000), "Strategies for Psychophysiological Assessment of Situation Awareness," in *Situation Awareness: Analysis and Measurement*, M. R. Endsley and D. J. Garland, Eds., Lawrence Erlbaum Associates, Mahwah, NJ, pp. 175–188.

Wilson, G. F. (2002a), "A Comparison of Three Cardiac Ambulatory Recorders Using Flight Data," *International Journal of Aviation Psychology*, Vol. 12, pp. 111–119.

Wilson, G. F. (2002b), "An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures," *International Journal of Aviation Psychology*, Vol. 12, pp. 3–18.

Wilson, G. F., and Eggemeier, F. T. (1991), "Physiological Measures of Workload in Multi-task Environments," in *Multiple-Task Performance*, D. Damos, Ed., Taylor & Francis, London, pp. 329–360.

Wynbrandt, J. (2001), "Seeing the Flightpath to the Future: Synthetic Vision and Heads-Up Displays Are on the Way to a Cockpit near You," *Pilot Journal*, Vol. 2, No. 2, Spring, pp. 30–34, 94.

Xie, B., and Salvendy, G. (2000a), "Prediction of Mental Workload in Single and Multiple Tasks Environments," *International Journal of Cognitive Ergonomics*, Vol. 4, pp. 213–242.

Xie, B., and Salvendy, G. (2000b), "Review and Reappraisal of Modelling and Predicting Mental Workload in Single and Multi-task Environments," *Work and Stress*, Vol. 14, pp. 74–99.

Yeh, Y. Y., and Wickens, C. D. (1988), "Dissociation of Performance and Subjective Measures of Workload," *Human Factors*, Vol. 30, pp. 111–120.

HANDBOOK OF HUMAN FACTORS AND ERGONOMICS

Third Edition

Edited by
Gavriel Salvendy
Purdue University
West Lafayette, Indiana
and
Tsinghua University
Beijing, People's Republic of China

WILEY
JOHN WILEY & SONS, INC.