human factors work. Some researchers in cognitive neuroscience are aware of the importance of ecological validity (e.g., see Kingstone et al., 2006), but typically tend to study mental processes in isolation independent of considerations of the artifacts and technologies of the world that require the use of those processes. Neuroergonomics goes one critical step further. It postulates that the human brain, which implements cognition and is itself shaped by the physical environment, must also be examined in interaction with the environment in order to understand fully the interrelationships of cognition, action, and the world of artifacts (Parasuraman, 2003). A recent review of progress in human factors describes the historical changes in the field from its beginnings in behaviorism, its adoption of the information processing view, and culminating in the neuroergonomic approach (Proctor & Vu, 2010).

In this chapter we discuss how our understanding of three areas in human factors research— mental workload, effects of stress on performance, and individual differences in cognition and human performance—can be enhanced by examining them from both cognitive and neuroergonomic perspectives.

3. MENTAL WORKLOAD

Mental workload is probably one of the most widely invoked concepts in human factors research and practice (Bailey & Iqbal, 2008; Loft et al., 2007; Moray, 1979, Parasuraman & Hancock, 2001; Tsang & Wilson, 2006; Wickens, 2008). System designers and managers raise the issue of mental workload when they ask questions such as: How busy is the operator? How complex are the tasks that the operator is required to perform? Can any additional tasks be handled above and beyond those that are already performed? Will the operator be able to respond to unexpected events? How does the operator feel about the tasks being performed? Each of these questions could be asked of the people in the surgical scenario described at the start of this chapter. Answers to the questions can be provided given that mental workload can be measured in an existing system or modeled for a system that is not yet built.

Mental workload characterizes the demands of tasks imposed on the limited information processing capacity of the brain in much the same way that physical workload characterizes the energy demands upon the muscles. In any resource-limited system, the most relevant measure of demand is specified relative to the supply of available resources, as discussed in Chapter 10. Thus a context for conceptualizing this supply-demand relationship associated with mental workload is provided by the two functions shown in figure 11.1. The X-axis depicts increasing resource demands of a task (or set of tasks) in a way that can encompass either the demands of a single task, or multitask demands (e.g., requirement to supervise more than a single unmanned vehicle or robot). We will distinguish between the single and multitask cases below.

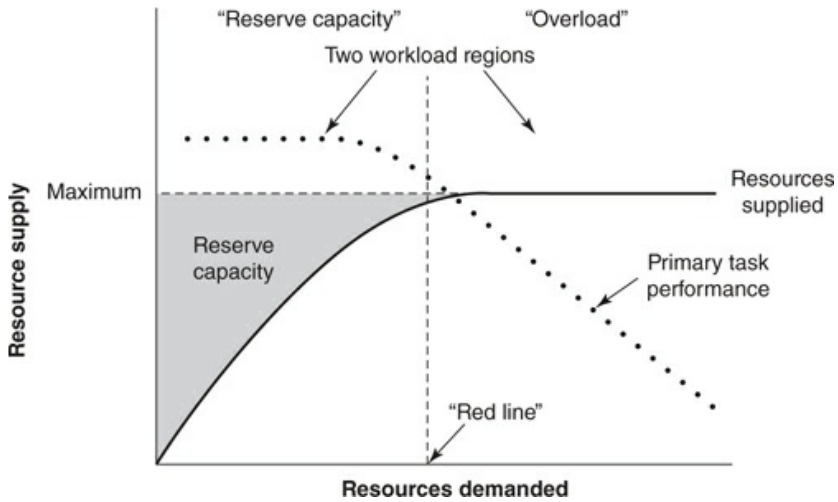

FIGURE 11.1 Schematic relationship among primary-task resource demand, resources supplied, and performance, indicating the "red line" of workload overload.

The Y-axis represents two functions. A "resource supply" function (solid line) reflects the fact that when demands are increased from 0 (doing nothing) to some level, the operator has ample supply to meet those demands. But as a limited capacity or limited resource system, when the demand exceeds the supply, no further resources can be supplied; the solid line flattens. Of course this level cannot be established precisely, and hence the leveling is gradual, not abrupt. The dashed line represents performance on the task(s) in question. Almost by definition, when supply exceeds demand, performance remains perfect, and is unchanged by differences in demands. Once demand equals supply, further demand increases will lead to further performance decrements. The discontinuity or "knee" on the two curves is sometimes referred to as the "red line" of workload (Hart & Wickens, 2010; Rennerman, 2009; Wickens, 2009); or given its fuzziness, a "red zone." Importantly, as we describe below, the red line divides two regions of the supply demand space. The region at the left can be called the "reserve capacity" region. That to the right can be labeled the "overload region." The two regions have different implications for workload theory, prediction, and assessment, as well as the kinds of concerns of engineering psychologists. We treat these in sequence below.

3.1 Workload Overload

Both engineering psychologists and designers are interested in predicting when demand exceeds supply and performance declines as a result, as well as in applying different remedies when this overload condition occurs. As we discussed in Chapter 10, when this performance decrement results because of **multitasking** overload, models such as the multiple resource model can offer a framework for design or task changes that will reduce the demand and resulting decrement in performance (see Figure 10.1 in Chapter 10). This may include using separate, rather than common resources or reducing the resource demands of the task. Examples of methods for reducing resource demands include reducing working memory load (see Chapter 7), automating parts of the task (as discussed in Chapter 12), reassigning some of the tasks to another operator or changing procedures in such a way that previously concurrent tasks can now be performed sequentially.

The multiple resource model is a useful tool for predicting what can be done to lower the multitask resource demand, and this reduction can be quantified by computational models (e.g., Horrey & Wickens, 2004b; Wickens, 2005). Hence, such models can be used to predict the **relative workload** (e.g., workload reduction) of different design alternatives. Multiple resource models can also predict the reduction in performance decrement achieved by operator training via developing **automaticity** of one or more of the component tasks (refer back to Figure 10.2), but such models cannot predict how much training is required to move demands below the red line. In the same way, the computational models of multiple resources are not yet able to predict the level of resource demand and resource competition that is at the red line (such that further demand increases will degrade performance and decreases will not improve it). That is, such models do not well predict the **absolute workload**.

Increasing demands can also be imposed by increasing the difficulty of a *single task* (rather than multitasking) as when the working memory load is increased (see Chapter 7), the relational complexity of a cognitive task is increased (Halford, Baker, et al., 2005; Halford, Wilson, & Phillips, 1998), the bandwidth of a tracking task is increased (driving along a winding road at faster and faster speeds, see Chapter 5) or the number of aircraft that a controller needs to supervise in his/her sector rises (Ayaz et al., 2012).

In these cases, where a particular variable can be counted (e.g., number of chunks, number of variable interactions, number of turns/second or number of aircraft, respectively), it is straightforward to predict relative workload (more is higher) and in many cases, data have provided a reasonable approximation to a red line. For example, we have noted that the red line for working memory at roughly seven chunks of information (see Chapter 7). For relational complexity it is roughly three (Halford et al., 2005). For tracking bandwidth, it is roughly one cycle per second (Wickens & Hollands, 2000).

Several variables can moderate these count "constants," effectively moving the red line to the left or right along the X-axis of Figure 11.1. In the case of the air traffic controller, for example, the degree of uncertainty in trajectory as well as the complexity of the airspace greatly affect the number of planes that can be adequately supervised (Hilburn, 2004). Similar modulating factors influence the number of unmanned vehicles that can be supervised (Cummings & Nehme, 2010).

One of the most important count variables, which can be employed in either single or multitask circumstances is time: simple timeline analysis computes the ratio of time required (TR) to time available (TA) (Parks & Boucek, 1989). We discuss timeline analysis further below in the context of reserve capacity. More specifically, timeline analysis will enable the system designer to "profile" the workload that operators encounter during a typical mission, such as landing an aircraft or starting up a power-generating plant (Kirwan & Ainsworth, 1992). In a simplified but readily usable version, it assumes that workload is proportional to the ratio of the time occupied performing tasks to total time available. If one is busy with some measurable task(s) for 100 percent of a time interval, workload is 100 percent during that interval. In a simple model, this may be defined as a "red line." Thus, the workload of a mission would be computed by drawing lines representing different activities, of length proportional to their duration. The total length of the lines would be summed and then divided by the total time (Parks & Boucek, 1989), as shown in Figure 11.2. In this way the workload encountered by or predicted for different members of a team (e.g., pilot and copilot) may be compared and tasks reallocated if there is a great imbalance. Furthermore, epochs of peak workload or work overload in which load is calculated as greater than 100 percent can be identified as potential bottlenecks.

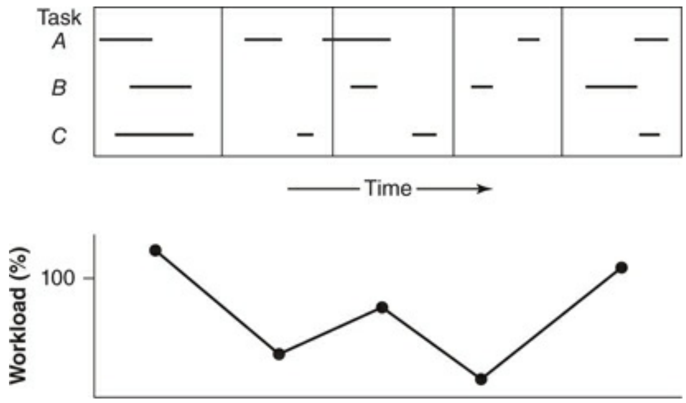

FIGURE 11.2 Time-line analysis. The percentage of workload at each point is computed as the average number of tasks per unit time within each time window.

Importantly, time line analysis is equally applicable to both the overload region (TR/TA >1) and the reserve capacity region (TR/TA <1), and in the latter it can be used equally well in workload **predictive models** (if tables are available to look up the time required to perform different tasks) and in **workload assessment**, as discussed below, if observers can carefully record operator activity (including non-observable cognitive tasks). While the 100 percent level may be initially set as the red line, observations by Parks and Boucek (1989) suggest instead that it is the 80 percent level where errors in performance begin to occur.

The important general point to be made here is that for both single and multitask demands in the overload region above the red line, simple measures of performance are adequate to measure “workload,” and models of multitask performance or single task count variables can predict workload increases (performance decreases) or *relative workload* above the red line. Count variables can be used to predict *absolute workload* values, both above and blow the red line, but multitask interference models can not easily do so at the current stage of their maturity.

3.2 Reserve Capacity Region

When demands are below the red line, both single task count variables and multiple resource models can continue to offer reliable **relative predictions** (for example, four chunks will have a higher workload than three chunks; and using separate modality resources will create more reserve capacity compared to common resources). However, as the dotted line curve in Figure 11.1 makes quite clear, *performance* on the task of interest is no longer adequate to assess differences in workload. (We refer to this task as the primary task.) Not only do primary task performance measures fail to reflect differences in resource demands below the red line, many such primary tasks are often very coarse, or nonexistent in their performance measures because a final output may not reflect vast differences in the complexity of cognition that supported it. For example, a decision-making task might have only one of two outcomes (correct or wrong, see Chapter 8). Yet reaching the decision may involve considerable working memory activity needed to entertain alternative hypotheses and evaluate possible outcomes. All of this cognitive activity is a large contributor to mental workload, and yet variation in it will not be reflected by the simple rightwrong measure of decision-making performance. As another example, the task of “maintaining situation awareness” can impose a high level of workload (and potential interference with other activities; Wickens, 2002c), but there may be no measure of performance at all associated with this task (unless situation awareness is periodically “probed” by potentially intrusive measures like SAGAT; see Chapter 7).

3.3 Measures of Mental Workload and Reserve Capacity

To cope with the inadequacy of primary task performance measures for workload assessment, engineering psychologists have developed four other categories of workload measurement tools, all designed, directly or indirectly, to assess the amount of reserve capacity (e.g., the distance to the left of the red line). We discuss three of these first: behavior, secondary task, and subjective measures, before describing physiological measures in some detail, given their anchoring in neuroergonomics.

3.3.1 BEHAVIORAL MEASURES Even as *performance* might not change with increasing resource demands in the reserve capacity region, behavior often will. If behavior is defined in its simplest form as "doing something," then, as described above, time line analysis serves as a very effective measure of workload. In many manual control tasks, control activity, mean control velocity, or high frequency power can serve as an effective behavioral measure of workload that may change with task demands even as tracking error (performance) remains constant.

3.3.2 SECONDARY TASKS Secondary tasks have the greatest fidelity. One simply asks: if the operator is performing the primary task at an adequate level (e.g., perfect performance), how well can she/he perform a concurrent secondary task? The better that task can be performed, we infer the more residual capacity there is available to it from the primary task, and lower the resource demands of the primary task. A variety of secondary tasks have been employed in a variety of circumstances, such as responding to an unexpected probe stimulus, estimating time passing, or doing mental arithmetic, and many of these have been reviewed in other papers (e.g., Gopher & Donchin, 1986; Hancock & Meshkati, 1988; Hart & Wickens, 2010; Hendy, Liao, & Milgram, 1997; Moray, 1979, 1988; O'Donnell & Eggemeier, 1986; Tsang & Wilson, 1997; Wierwille & Williges, 1978; Williges & Wierwille, 1979).

One major problem with secondary tasks is that researchers cannot always control the amount of attention given to them. For example, in some cases the secondary task may be intrusive on the primary task whose workload is measured, disrupting that level of performance. Ideally, this performance should be perfect and untouched, or at least primary task decrements caused by the secondary task should be the same across all versions of the primary task to be compared. Such disruption may bias the measurement itself (after all, here the secondary task is given more resources than it should receive) in much the same way that measuring the length of a worm with a ruler could cause the worm to contract if the ruler touches it in the measurement process. Alternatively, in order to avoid such intrusiveness, operators (particularly in high risk environments like driving or flying) might choose to ignore the secondary task altogether, providing no measure whatsoever.

In order to partially guard against these problems, researchers have recognized the importance of **embedded secondary tasks** (Raby & Wickens, 1994). These are in fact natural components of the total task scenario, but typically of the lower level of priority characteristic of a secondary task. Examples of embedded secondary tasks might be periodic glances to the rear or side view mirror (to measure driver workload) or offering periodic position reports to an air traffic controller (to measure pilot workload). As an embedded secondary task, Metzger and Parasuraman (2005) measured the latency with which air traffic controllers made a check mark on the flight strip of an aircraft when it had reached a navigational waypoint, and found that controllers either delayed or omitted such checking when they had more aircraft in their sector to control and their workload increased.

3.3.3 SUBJECTIVE MEASURES Subjective measures of workload experienced are widely used techniques for workload assessment (e.g., Hart & Wickens, 2010; Hill et al., 1992; O'Donnell & Eggemier, 1986; Tsang & Vidulich, 2006; Vidulich & Tsang, 1986; Wierwille & Casali, 1983). With this method workers using a unidimensional scale are simply asked to report on the scale what their experience of workload is (or was over some previous duration). A variety of scales abound, such as the Bedford scale or the Modified Cooper-Harper (Wierwille & Casali, 1983) scale (See Hart & Wickens, 2010 for a review). Many of these scales have the advantage that a verbal description of the red line can be actually placed at a given level. For example, a rating of 7 on a 10-point scale might be described as "no extra attention is available to give to any additional tasks." Subjective measures can therefore effectively serve the range of task demands both above and below the red line.

There are also multidimensional workload scales (Boles, Bursk, et al., 2007; Hill, Iavecchia, et al., 1992; Reid & Nygren, 1988; Vidulich & Tsang, 1986). These assume that mental workload has different components, just as physical workload may be imposed separately on, say, the arms, legs, or fingers. Perhaps the most widely used of these is the NASA TLX (Task Load Index) scale, which asks users to provide separate subjective ratings on subscales of mental demand, physical demand, time demand (time pressure), performance, effort, and frustration level (Hart & Staveland, 1988).

Multidimensional scales do not always provide sufficient added information relative to uni-dimensional scales, in order to justify the added time requirements for generating multiple ratings. But often when comparing qualitatively different systems ("apples and oranges") on a workload scale, such qualitative differences can be well revealed by differing differences along the different TLX subscales. For example, one procedure in programming a navigational device may impose a high time (temporal) load, but low mental load, whereas another, with which it is compared, may show the reverse. Descriptions of TLX provide guidance on how subscales can be combined to produce a single scale if desired (Hart & Staveland, 1988).

3.3.4 PURPOSE OF WORKLOAD ASSESSMENT Importantly, all measures of workload, whether performance based, subjective or physiological, can be assessed for two qualitatively different purposes. **Offline measures** are assessed during system evaluation, and designers can take the results of this assessment, diagnose a workload deficiency, and proceed to invoke some remedy, to move demands far enough below the redline to preserve a certain margin of residual capacity. In contrast, **online measures** of workload are assessed while the operator is performing the task outside the laboratory in the operational environmnent and can be used in **adaptive automation** to reduce the demands if workload is either over the red line, or perhaps is increasing toward it (see Chapter 12).

Online measures are only one source of evidence available for automation to infer workload, and other sources will be discussed in the next chapter. However, we note here that for online applications, there is heightened concern for the intrusiveness of the measure. Anything that might interfere with the primary task such as performing a secondary task or even responding to a probe in order to give a subjective measure could have serious possible consequences. For this reason, engineering psychologists have been particularly interested in "passive," less intrusive neurophysiological measures to index workload. We now turn to these in some detail.

3.4 Neuroergonomics of Workload

3.4.1 OVERVIEW Over a century ago, the famous neuroscientist Sir Charles Sherrington suggested that mental work was fundamentally *brain work*. He proposed that the movement of blood through the brain's arteries was in response to the demands placed on neurons by the need for cognitive processing. Such neuronal demands were met by the brain supplying increased oxygenated blood to the area of the active neurons. Sherrington therefore proposed that neuronal function was reflected in cerebral blood flow, which he was able to measure in animal models (Roy & Sherrington, 1890), but which he suggested could also be applied, in principle, to the study of human brain function. It took nearly a century before technological developments, first with the invention of positron emission tomography (PET) but later with functional magnetic resonance imaging (fMRI), allowed for noninvasive measurement of cerebral blood flow in humans, and thereby provided a basis for confirming Sherrington's hypothesis in humans. fMRI findings of increased cerebral blood flow in regions of the prefrontal cortex with increased task demand, e.g., higher working memory load, have pointed to neural correlates of resources (Parasuraman & Caggiano, 2005; Posner & Tudela, 1997). Furthermore, other fMRI findings (Just et al., 2003) have supported the distinction between perceptual/cognitive, verbal/spatial, and input and output modality-specific processing, which are components of the multiple resource model of Wickens (1984).

Cognitive neuroscience research using fMRI will continue to enhance our knowledge of the specific neural systems associated with attention and cognitive processing and will therefore contribute to a better theoretical understanding of the components of mental workload. But fMRI is an expensive, restrictive, and non-portable technology that is not suitable for routine use or for low-cost practical applications. However, a variety of other neuroergonomic techniques are available for the assessment of mental workload. These methods fall into three general classes: (1) electrophysiological; (2) hemodynamic; and (3) autonomic. We discuss examples of techniques in each of these classes in the following.

3.4.2 EEG The electroencephalogram (EEG) records the brain's electrical activity from electrodes placed on the scalp of a participant's head. Spectral power in different EEG frequency bands has been found to be sensitive to increased working memory (WM) load and demand for attentional resources. Given the important role that WM plays in comprehension, reasoning, and other cognitive tasks (Baddeley, 2003; see also Chapter 7), many studies have examined changes in the spectral structure of the EEG in tasks in which WM demands are varied.

A common WM task in which workload can be easily manipulated is the "N-back" task, in which participants are given a continuous stream of stimuli and must respond, not to the current stimulus, but that presented N stimuli back. This task is trivial when N = 0 and relatively easy when N = 1, but much more difficult when N >1. To be successful in such tasks when the WM demand is high (e.g., N = 3), participants must continuously apply mental effort and typically report high levels of workload and show increased neural activity in frontal and parietal regions of the brain (Owen et al., 2005). The spectral structure of the EEG shows systematic load-related modulation during such N-back task performance. Typically, when recorded from midline frontal electrode sites, EEG activity in the theta band (4–7 Hz) is increased in power for high WM load compared to low load (Gevins & Smith, 2003). Frontal midline theta increases have also often been reported for other difficult tasks requiring sustained concentration (Gevins et al., 1998). In contrast to the midline frontal theta, activity in the alpha band (8–12 Hz) shows an inverse relationship with task load, being reduced with high WM demand. The attenuation of EEG alpha with visual attention and with cognitive load has been shown in many studies since its initial demonstration by the discoverer of EEG, Hans Berger in 1929.

Frontal theta activity (4–7 Hz) increases while alpha power (8–12 Hz) decreases as more resources have to be allocated to the task and thus provide sensitive measures of mental workload (Gevins & Smith, 2003). Spectral power in these two frequency bands can be fairly easily computed from the raw EEG, including in near real-time (several seconds) using readily available software packages.

EEG measures have also been found to index operator mental workload in more complex tasks that are more representative of operational environments. These include tasks such as the Multiple Attribute Task Battery (Gevins & Smith, 2007), simulated process control (Hockey et al., 2009), and operational tasks such as flight, air traffic control (ATC), and road and rail transportation (Brookhuis & De Waard, 1993; Hankins & Wilson, 1998; Lei & Roetting, 2011; Wilson, 2001, 2002). For example, Brookings et al. (1996) recorded EEG from Air Force controllers while varying the difficulty of a simulated ATC task along two dimensions, the number or volume of aircraft to be controlled, and the aircraft mix (complexity). Right hemisphere frontal and temporal EEG theta band activity increased with workload. Midline central and parietal areas showed theta band activity to also increase with increased workload to both types of task manipulation. Alpha band activity decreased with increased task complexity but not with the number of aircraft being monitored. Thus these EEG components were differentially sensitive to different aspects of mental workload.

Can EEG be used to assess mental workload reliably in operational settings? Yes, but only with some difficulty. One problem is that EEG can be contaminated by eye movement and muscular artifacts in such environments. While it is relatively easy to remove these artifacts off line, after recordings have been made and stored, *on-line* artifact removal is more challenging. However, the recent development of mathematical techniques such as independent components analysis (ICA) has allowed for implementation of measurement of artifact-free EEG in an online manner (Jung et al., 2000). Real-time measurement of artifact-free EEG in operational settings is currently a topic of much research and development.

3.4.3 EVENT-RELATED POTENTIALS Event-related potentials (ERPs) represent the brain's neural response to specific sensory, motor, and cognitive events. ERPs are computed by recording the EEG and by averaging EEG epochs time-locked to a particular stimulus or response event. At the present time ERPs hold a somewhat unique position in the tool shed of cognitive neuroscientists because they provide the *only* neuroimaging technique that has high temporal resolution, of the order of milliseconds, compared to techniques such as PET and fMRI which are inherently sluggish (because they index cerebral hemodynamics). ERPs are often used whenever researchers need to examine the relative timing of neural mechanisms underlying cognitive processes with millisecond precision. For example, the timing information provided by ERPs provided critical evidence for the "early selection" view of attention because of findings showing attentional modulation of neural activity after about 100 ms post-stimulus (Hillyard et al., 1998).

The latency of one prominent ERP component, the P300, increases with the difficulty of identifying targets but not with increases in the difficulty of response choice, suggesting that P300 provides a relatively pure measure of perceptual processing/categorization time, independent of response selection/execution stages (Kutas et al., 1977; see Chapter 9). P300 amplitude is also proportional to the amount of attentional resources allocated to the target (Johnson, 1986; Polich, 2003). Thus, any diversion of resources away from target discrimination in a dual-task situation will lead to a reduction in P300 amplitude. Isreal, Chesney, et al. (1980) used this logic to examine the temporal locus of added workload demands in a dual-task situation. They showed that P300 amplitude decreased when a primary task, tone counting, was combined with a secondary task of visual tracking. However, increases in the difficulty of the tracking task did not lead to a further reduction in P300 amplitude. Thus, they argued that P300 reflects processing resources associated with perceptual processing and stimulus categorization, but not responserelated processes (see Chapter 10, Section 3.1). In a subsequent study, Wickens, Kramer, et al. (1983) showed reciprocal changes in P300 amplitude as resources were flexibly allocated between primary and secondary tasks.

Several studies have used the auditory P300 to assess the workload demands of different complex tasks. Isreal, Wickens, et al. (1980) showed the sensitivity of P300 to display complexity in an air-traffic monitoring type of task. Ullsperger et al. (2001) used secondary-task P300 amplitude changes to make inferences regarding the amount and type of resource demand of a gauge monitoring task. More recent studies have used P300 to assess the workload demands of learning to use different computer systems. For example, one of the problems associated with educational systems such as hypermedia is to assess how demanding they are for individual learners, and thereby to adapt them on a person-by-person basis, as discussed in Chapter 7. Schultheis and Jamieson (2004) found that P300 amplitude to auditory stimuli was sensitive to the difficulty of text presented in a hypermedia system. They concluded that auditory P300 amplitude and other measures, such as reading speed, may be combined to evaluate the relative ease of use of different hypermedia systems. For another example from the domain of driving assessment, Baldwin and Coyne (2005) found that P300 amplitude was sensitive to the increased difficulty of simulated driving in poor visibility due to fog, compared to driving in clear conditions. The unique value of this neuroergonomic measure was shown by the finding that performance-based and subjective indices were not affected by the visibility manipulation.

3.4.4 ULTRASOUND MEASURES OF CEREBRAL BLOOD FLOW EEG and ERP represent the class of electrophysiological measures. Two hemodynamic measures, in addition to PET and fMRI, are Transcranial Doppler Sonography (TCD) and near infrared spectroscopy. TCD is an ultrasound device that can be used as a noninvasive method to monitor cerebral blood flow. TCD therefore provides another technique that, like fMRI can be used to examine Sherrington's view that mental work is associated with brain work, as reflected in cerebral blood flow to the left or right cerebral hemispheres. TCD uses a small 2 MHz pulsed Doppler transducer to gauge arterial blood flow, typically of the middle cerebral artery (MCA), which can be isolated through the cranial "windows" in the temporal bone on each side of the head (Aaslid, 1986). The low weight and small size of the TCD transducer and the ability to embed it in a headband allow for measurement of cerebral blood flow while not limiting, or becoming hampered by, head and body motion (Tripp & Warm, 2007).

When a particular area of the brain becomes metabolically active due to cognitive processing, byproducts such as carbon dioxide increase, leading to a dilation of blood vessels serving that area. This, in turn, results in increased blood flow to that region. Several TCD studies have shown that changes in the difficulty of perceptual and cognitive tasks are accompanied by increases in cerebral blood flow in either the left or right hemisphere (see reviews by Duschek and Schandry, 2003; Stroobant & Vingerhoets, 2000). Shaw et al. (2010) examined dynamic changes in cerebral blood using TCD in a simulated air defense task in which participants had to protect a "no fly zone" by engaging enemy aircraft that approached the zone. They found that cerebral blood flow closely tracked changes in the number of enemy threats that lead to changes in mental workload.

3.4.5 NEAR INFRARED SPECTROSCOPY AND CEREBRAL OXYGENATION The TCD technique provides only an indirect index of oxygen utilization in the brain, as revealed by changes in blood flow. A more direct measure of cerebral oxygenation would be useful as another indicator of "brain work"—engagement of neurons recruited in the service of cognitive processing. Optical imaging, in particular near infrared spectroscopy (NIRS), provides such a measure. NIRS typically uses near-infrared light that is emitted by several sources embedded in a strap that is placed over the front of the head. The strap also contains several infrared detectors that detect the light after it has passed through the skull and brain. Changes in light absorption, typically measured at two wavelengths, are used to calculate relative changes of oxygenated and deoxygenated blood in the frontal cortex. NIRS has a precision advantage over TCD, given its ability to assess activation in several frontal brain regions, and not just in the left and right hemispheres as with TCD.

Previous research using the NIRS as well as fMRI has shown that tissue oxygenation increases with the information-processing demands of the task being performed (Toronov et al., 2001). More recently, Ayaz et al. (2012) used NIRS to examine cerebral oxygenation in experienced controllers monitoring air traffic in a high-fidelity simulator. Controller communications with pilots were via standard voice or visual text **data link** (see Chapter 7). Ayaz et al. (2012) found that there was a systematic increase in blood oxygenation as the number of aircraft that had to be controlled increased from 6 to 12 to 18. These neural changes were accompanied by similar changes in subjective workload, as measured by the NASA-TLX.

3.4.6 HEART-RATE VARIABILITY Autonomic measures constitute the third class of neuroergonomic measure. Of these, heart-rate variability has been the object of sustained study. Several investigators have examined different measures associated with the variability or regularity of heart rate as a measure of mental load. Variability is generally found to decrease as the load increases, particularly that variability which cycles with a period of around 10 seconds (0.1 Hz) (Mulder & Mulder, 1981). When this variability is associated specifically with the cycles resulting from respiration, the measure is termed *sinus arrhythmia* (Backs et al., 2003; Derrick, 1988; Mulder et al., 2003; Sirevaag et al., 1993; Vicente et al., 1987).

Heart rate variability is sensitive to a number of different difficulty manipulations and therefore appears to be more sensitive than diagnostic. Derrick (1988) investigated this measure with four quite different tasks performed in different combinations within the framework of the multiple-resource model. His data suggested that the variability measure reflected the total demand imposed on all resources within the processing system more than the amount of resource competition (and therefore dual-task decrement) between tasks. Backs et al. (2003) examined three different heart rate measures during simulated driving over easy or difficult curved courses and found that they were differentially affected by curve radius. They concluded that the differential effects indicated that the perceptual demands of driving could be distinguished from central and motor processing demands.

3.4.7 PUPIL DIAMETER Several investigators have observed that the diameter of the pupil correlates quite closely and accurately with the resource demands of a large number of diverse cognitive activities (Beatty, 1982). These include mental arithmetic (Kahneman et al., 1967), short-term memory load (Peavler, 1974), visual search (Porter et al., 2007), air traffic control monitoring load (Jorna, 1997), simulated driving (Recarte & Nunes, 2003), and on-the-road driving (Razael & Klette, 2011). This diversity of responsiveness suggests that the pupilometric measure may be highly sensitive, although as a result it is undiagnostic of the type of workload demand. It will reflect demands imposed anywhere within the information-processing system. However, changes in ambient illumination must be monitored since these also affect the pupil and because of its association with the autonomic nervous system, the measure will also be susceptible to variations in emotional arousal.

3.4.8 VISUAL SCANNING, ENTROPY, AND THE "NEAREST NEIGHBOR INDEX" While discussed as a measure of selective attention allocation in Chapter 3, visual scanning—the direction of pupil gaze—can also contribute extensively to workload modeling in two different ways. First, as we have noted, dwell time can serve as an index of the resources required for information extraction from a single source. In an aircraft simulation, Bellenkes et al. (1997) found that dwells were longest on the most information rich flight instrument (the artificial horizon or artificial horizon instrument; see Chapter 3) and that dwells were much longer for novice than expert pilots, reflecting the novices' greater workload in extracting the information. Second, scanning can be a diagnostic index of the source of workload within a multielement display environment. For example, Bellenkes et al. found that long novice dwells on the artificial horizon display were coupled with more frequent visits, and hence that instrument served as a major "sink" for visual attention. Little time was left for novices to monitor other instruments, and as a consequence their performance declined on tasks using those other instruments. Dinges et al. (1987) and Wikman et al. (1998) used scanning as a critical measure of the in vehicle head down time caused by the workload associated with different in-vehicle systems such as maps, radio buttons, etc.

Analyzing the degree of randomness of visual scanning or **entropy** can also be potentially informative regarding mental workload (Ephrath et al., 1980; Harris et al., 1986). One view is that as mental workload increases, a person's pattern of visual exploration of a region of interest in a display becomes more stereotyped and less random because they fixate on only the few regions of the display containing the relevant information so that entropy decreases. Conversely, a reduction in mental workload should increase entropy. Hilburn et al. (1997) confirmed this finding when examining the effects of automation on the mental workload and visual scanning patterns of experienced air traffic controllers. A challenge is that the entropy measure in this and other related studies typically ignores visual fixations outside a defined region of interest. Di Nocera et al. (2007), however, argued that all areas of visual fixation should be analyzed, and proposed a derived measure of mental workload called the Nearest Neighbor Index (NNI), defined as the ratio of the average of the observed minimum distances between fixation points and the mean distance that one would expect if the distribution of fixations was random. Di Nocera et al. (2007) found that the NNI index was significantly higher during the demanding take off and landing phases of flight operations than during cruise flight, pointing to the utility of NNI as an index of mental workload.

3.4.9 COSTS AND BENEFITS OF PHYSIOLOGICAL MEASURES OF WORKLOAD Neuroergonomic indexes have two advantages over behavioral and subjective measures of workload: (1) Such measures provide a relatively continuous record of data over time. (2) They are not obtrusive into primary-task performance. But they sometimes require that electrodes be attached, so a degree of physical constraint is imposed, and therefore they are not truly unobtrusive in a physical sense. However, the latest generation of eye tracking devices does not require any instrumentation of the participant, as the infrared sensors can be mounted on the desk or the side of the display being monitored. Other measures do require that the participant be fitted with the sensor in some manner, e.g., an EEG cap or a head strap for NIRS. These constraints will influence user acceptance.

Many physiological measures have a further potential cost in that they are, generally, one conceptual step removed from the inference that the system designers would like to make. That is, workload differences measured by physiological means must be used to *infer* that performance breakdowns would result or to *infer* how the operator would feel about the task. Secondary measures assess the former directly, whereas subjective measures assess the latter

There are many factors such as cost, ease of implementation, intrusiveness, etc., that must be taken into consideration when choosing a workload assessment technique for engineering psychology applications. Some of these factors (e.g., cost) may rule out the use of physiological measures in favor of simpler indexes such as subjective measures. Some individuals may also not wish to be "wired up" for physiological recording in work environments, so operator acceptance is another important factor to consider. With increasing miniaturization and development of "dry electrode," wireless wearable systems, some of these concerns are diminishing. At the same time, even if practical considerations rule out the use of physiological measurement, the neuroergonomic approach may nevertheless remain important for theory development, which in turn may lead to more sensitive assessment of mental workload (Kramer & Parasuraman, 2007).

3.5 Relationship Between Workload Measures

If all measures of workload demonstrated high correlation with one another and the residual disagreement was due to random error, there would be little need for further validation research in the area. The practitioner could adopt whichever technique was methodologically simplest and most reliable for the workload measurement problem at hand. Generally, high correlations between measures will be found if the measures are assessed across tasks of similar structure and widely varying degrees of difficulty. However, the correlations may not be high and may even be negative when quite different tasks are contrasted. For example, consider an experiment conducted by Herron (1980) in which an innovation designed to assist in a target-aiming task was subjectively preferred by users over the original prototype but generated reliably poorer performance than the original. Similar dissociations have been observed by Wierwille and Casali (1983) and by Childress et al. (1982) and, who measured pilot workload associated with cockpit-display innovations.

We use the term *dissociation* to describe these circumstances in which conditions that are compared have different effects on different workload measures. The understanding of attention and resource theory can be quite useful in interpreting why these dissociations occur. Yeh and Wickens (1988) suggested that subjective measures directly reflect two factors: the effort that must be invested into performance of a task and the number of tasks that must be performed concurrently. These two factors, however, do not always influence performance. To illustrate, consider the following situations:

- **A.** If two different tasks are in the underload region on the left of Figure 11.1, the greater resources invested on the more difficult task (and therefore that higher subjective workload) will not yield better performance.
- **B.** Subjective measures often fail to reflect differences due to data limits (see Chapter 10, figure 10.2), particularly if the lower level of performance caused by the lower level of the data limit is not immediately evident to the performer who is giving the rating. (Note however that this is an advantage of the NASA TLX measure, which allows the operator to separately rate "performance" and "mental effort.")
- **C.** In the context of the performance-resource function, if two systems are compared, one of which induces a greater investment of effort, this one will probably show higher subjective workload, even as its performance is improved (through the added effort investment). This dissociation is shown when effort investment is induced through monetary incentives (Vidulich & Wickens, 1986). However, it also appears that greater effort is invested when better (e.g., higher resolution) display information is available to achieve better performance. Thus in tracking tasks, features like an amplified error signal (achieved through magnification or prediction and inducing more precise corrections) will increase tracking performance but at the expense of higher subjective ratings of workload (Yeh & Wickens, 1988).

D. Yeh and Wickens (1988) concluded that a very strong influence on subjective workload is exerted by the number of tasks that must be performed at once. The subjective workload from time-sharing two (or more) tasks is almost always greater than that from a single task. We can see here the source of another dissociation with performance because a single task might be quite difficult (and result in poor performance as a result), whereas a dual-task combination, if the tasks are not difficult and use separate resources, may indeed produce a very good performance in spite of its higher level of subjective load.

The presence of dissociations often leaves the system designer in a quandary. Which system should be chosen when performance and workload measures do not agree on the relative merits between them? The previous discussion, and the chapter as a whole, does not provide a firm answer to this question. However, the explanation for the causes of dissociation and its basis on a theory of resources should at least help the designer to understand why the dissociation occurs, and thus why one measure or the other may offer a less reliable indicator of the true workload of the system in specific circumstances.

3.6 Consequences of Workload

Increases in workload do not inherently have "bad" consequences. Indeed, in many environments it is the low levels of workload that, when coupled with boredom, fatigue, or sleep loss can have negative implications for human performance (Chapter 2; Huey & Wickens, 1993). Adding task requirements can sometimes improve performance in low workload driving circumstances (Atchley & Chan, 2011). Given some flexibility, operators usually work homeostatically to achieve an "optimal level" of workload by seeking tasks when workload is low, and shedding them when workload is excessive (Hart & Wickens, 1990). This basis for strategic task management was discussed in Chapter 10.

In revisiting these task management issues, we must highlight the importance of understanding the strategy of task management that operators adapt when workload becomes excessive (i.e., crosses the red line from the underload to the overload region of Figure 11.1 as measured by the techniques described above). At a most general level, four types of adaptation are possible.

- People may allow performance of tasks to degrade, as a vehicle driver might allow lane position to wander as the workload of dealing with an in-vehicle automation system increases.
- People may perform the tasks in a more efficient, less resource consuming way. For example in decision making, they may shift from optimal algorithms to satisfactory heuristics.
- People may shed tasks altogether, in an "optimal" fashion, eliminating performance of those of lower priority. For example, under high workload, the air traffic controllers may cease to offer pilots weather information unless requested, while turning their full attention to traffic separation.
- People may shed tasks in a non-optimal fashion, abandoning those that should be performed, abandoning safe driving in favor of a cell phone conversation (see Chapter 10). Unfortunately, beyond the material covered in Chapter 10 on resource allocation, very little is known about general principles that can account for when people adopt one strategy or the other. However, as discussed there, training can certainly help (Orasanu, 1997).

4. STRESS, PHYSIOLOGICAL AROUSAL, AND HUMAN PERFORMANCE

We have all experienced stress at some point in our lives. Stress is typically seen as an emotional state of heightened arousal that can impair performance and, if severe enough, potentially disrupt behavior and have negative consequences for health. Stress is not always negative, however, for it may also serve as an energizing force that motivates people to perform well. Distinguishing the conditions under which stress impairs cognition and performance, and the mechanisms by which it does so, is one of the many challenges of stress research (Hancock & Desmond, 2001; Matthews et al., 2000).

The topic of stress has been studied from many different perspectives in the biological, psychological, and social sciences, with each discipline tending to define stress in different ways and examine different aspects of the phenomenon (Cohen et al., 1997). Within engineering psychology, the typical approach has been to adopt a stress-strain model in which an environmental stressor, such as noise, is compared to a condition without the stressor and effects on performance, physiology, and subjective feelings are assessed. The simple stress-strain model is shown in Figure 11.3. Stressors may include environmental influences such as noise, vibration, heat, dim lighting, and high acceleration, as well as such psychological factors as anxiety,