HUMAN FACTORS, 1993, 35(3), 413–429

Experimental Evaluation of a Model of Mental Workload

P. A. HANCOCK[1] *and* J. K. CAIRD,[2] *University of Minnesota, Minneapolis, Minnesota*

This study was designed to test predictions from a model of mental workload. The model predicts that mental workload grows as perceived distance from a task goal increases and the effective time for action decreases. Decreases in workload can be achieved by actions that bring the task goal into the region of acceptable time/distance constraints for successful resolution. We reported an experiment that tested these assertions using the Timepools performance task. Timepools generates a spatial representation of a shrinking temporal target to assess the effects of path length (i.e., the number of sequential targets to be acquired) and shrink rate (i.e., elapsed time during which the circle is halved in area) on reaction time (RT), movement time (MT), error rate, and the subjective perception of workload. Data from the experiment indicated systematic effects for task-related factors across performance and workload measures. Path length and shrink rate had differential effects on both RT and MT, which were also reflected in the components of the individual workload scales. The results support a general form of the workload model which may help researchers and practitioners in the difficult task of workload prediction.

INTRODUCTION

The basic leitmotiv of human factors is a primary concern for the human in the design and operation of systems. Although it has been recognized that operator performance and general well-being are contingent on characteristics of the task and the capabilities of machines and are influenced by the complexities of the operator's environment, our mandate has been and remains person- or user-centered in orientation (Kantowitz and Sorkin, 1983). In order to ensure the safety, health, comfort, and long-term productive efficiency of the operator, a reasonable goal is to regulate task demands so that they neither underload nor overload an individual. Although the dangers of overload have long been recognized, many of our recent concerns are with the stresses of underload and boredom (Becker, Warm, Dember, and Hancock, 1991; Hancock and Warm, 1989), particularly as operations become the subject of progressively increased automation.

In order to regulate mental load, we must be able to measure it. For tasks composed principally of physical demand, this has posed soluble problems, and the measures to

---

[1] Requests for reprints should be sent to P. A. Hancock, Human Factors Research Laboratory, University of Minnesota, 164 Norris Hall, 172 Pillsbury Dr. SE, Minneapolis, MN 55455.

[2] Now at The University of Calgary, Alberta, Canada.

assess energy expenditure have been well established (Åstrand and Rodahl, 1970; Wilson and Corlett, 1990). Contemporary problems in load measurement have arisen because of two major developments. The first is the well-publicized transformation in the composition of many performance tasks from physical to cognitive demand (see Westrum, 1991). In and of itself, this is sufficient to generate some concern. However, it is the complexity of evolving systems that drives such an interest from a passing concern to a central human factors issue.

Complexity refers to the number of constraints on the spatial and temporal dimensions of operations and also the increase in interconnections among systems (Hancock, Chignell, and Kerr, 1988). Such complexity may act to increase the amplitude and frequency of workload variations placed on the human operator, such that modern systems in stable operational states require little in the way of active response, whereas such systems in transient or unstable operational states impose a heavy demand. One contemporary aspect of workload is the question of the load associated with apparently simple systems monitoring. Recent research (see Hancock and Warm, 1989; Hancock, Warm, and Dember, 1991) has indicated a consistently high level of workload associated with such conditions of performance. Consequently, variations in task-generated workload in response to unexpected demand may be imposing an additional burden on an already stressed operator, whose individual response is difficult to discern (Damos, 1988; Hancock, Meshkati, and Robertson, 1985).

The confluence of these trends leads to the present concern. Although a clear recognition of the importance of workload prediction exists (Chignell and Hancock, 1985), there are relatively few theoretically grounded models with which to attack the problem. The present experiment was designed to test predictions from one recently proposed model of mental workload (Hancock and Chignell, 1988). Although considerable work has focused on the issue of measurement per se, few models have been developed through which task-related workload may be predicted (for a summary of previous efforts, see Gopher and Donchin, 1986; Hancock and Meshkati, 1988; Kantowitz, 1987; Moray, 1979; and O'Donnell and Eggemeier, 1986). The traditional tie to resource theory (Wickens, 1980, 1984, 1987, 1988) has met with varied success, but this is an attention-based rather than a workload-based link.

The model evaluated here describes a workload surface derived from a three-dimensional representation that is contingent on the distance from the performance goal, the effective time for response, and the level of effort expended by this operator. The specific purpose is to test this dynamic model using a temporal performance task that permits the manipulation of both the time available to operators and the distance they have to travel to achieve their goals.

THEORETICAL FRAMEWORK

Mental workload is recognized as a multidimensional construct that is largely driven by the characteristics of local task demands. It has been linked in some fashion to many major theories of human cognition, such as the automatic-versus controlled-processing position (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977) and the aforementioned attentional resource constructs (Kahneman, 1973; Wickens, 1980), as well as other models of the human operator, such as those founded in control theory (Jex, 1988). However, many conceptualizations are static and prescriptive (i.e., focused on what the operator should do) rather than dynamic and interactive (i.e., evolved from operators' perception of and response to the task, or what they actually do).

The present conceptualization of workload is conceived in three dimensions: effective time for action, perceived distance from the desired goal state, and level of effort required to achieve the desired goal. Mental workload is assumed to increase as the distance from the desired goal state and time constraints is increased (see Hancock and Chignell, 1988). Successful performance depends on satisfying the demands of a task within the time available for action. At extremes of workload, perceived time may not be coincident with what is typically conceived of as "physical" or "real" time. Tasks always have a time limit. However, we can imagine many tasks that can be so extended in time that temporal restrictions become a negligible concern. Indeed, one such pursuit, chess, has had to include time limitations to retain its viability as a competitive pursuit.

Figure 1 illustrates our approach to the evaluation of mental workload that includes each of these facets. The abscissa scales the perceived distance between the current and the desired goal state. In the simplest case, perceived distance can represent a physical quantity that separates the performer from the goal (e.g., in running a race the athlete would be aware of the remaining portion of the distance). However, actual distance is not always coincident with perceived distance. The last few miles of a marathon, for example, may be perceived as being much longer in terms of needed effort than the first few.

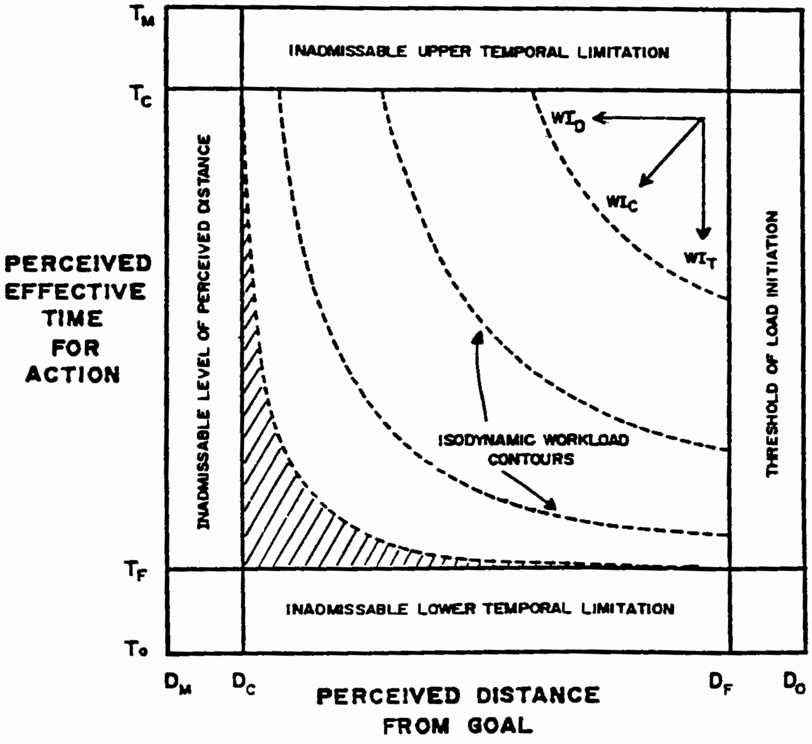

Figure 1. *Mental workload expressed as a function of perceived distance from goal state and effective time for action. Reproduced and modified from P. A. Hancock and M. H. Chignell (1988). Mental workload dynamics in adaptive interface design.* IEEE Transactions on Systems, Man, and Cybernetics, 18, *647–658. © IEEE 1988.*

Tasks emphasizing cognitive effort need to be scaled for distance in a similar manner. That is, how far do operators perceive themselves having to go to finish a task? Obviously such perception depends on task characteristics and previous experience; however, most operators can differentiate among the start, middle, and end phase of a job or task.

It is in this sense that the notion of distance is predominantly used in the present work. D_0 represents the goal state in which there is no task-generated load. The operator usually has multiple task goals, so achievement of D_0 is probably rare in practical operations. D_F is a minimal level of workload associated with the initiation of task-related response. As the perceived distance from the goal state increases, from D_F to D_C along the axis, workload grows accordingly.

D_C represents the ceiling level of perceived distance of an individual operator. A task exceeding this level may not be performed by that operator regardless of the time available. Consequently, one contingency of mental workload is the desire and belief of an individual operator that he or she can do the task within the time available. D_M represents the level of perceived distance that may be reconciled by experts or highest-level performers. The distance between D_M and D_C indicates that, even after extended practice, some levels of demand and associated load can exceed many individuals' best capability. As the task-related skill of the operator increases, the distance between D_C and D_M decreases and that between D_C and D_F decreases for any constant demand task. The threshold at D_C might be described by a discrete state change or may represent a portion of a continuum of graceful degradation in efficiency.

The ordinate represents effective time for action. T_0 represents the immediate present, often referred to in the time perception literature as the *specious* present (see James, 1890). This represents a time window in which the operator has essentially no time available for any task-related response. T_F is the effective floor for operator response time. The average for the minimal times for most human sensory and motor processes is well established (Card, Moran, and Newell, 1986). T_C is the upper limit on the time axis. It is taken to represent linked events designed to achieve a common goal. An example might be a flight mission, which may be defined by such temporal boundaries as take-off and landing. T_M is the maximum time available. This boundary can also be regarded as somewhat context specific, but in an absolute sense it represents the working lifetime of the operator.

In practical terms, T_C and T_M are functions of the task at hand and should be defined in terms of specific task goals. For example, the result of the introduction of time limits into competitive chess was a practical restriction on this temporal axis. Scaling on the time axis depends on the interplay of perceived time, which is contingent on endogenous (operator-specific) temporal information and so-called real time as reflected in the temporal invariants intrinsic to the task and/or environment.

With respect to spatial constraints alone and temporal constraints alone, admissible workload is bounded by the thresholds D_F, D_C, T_F, and T_C. However, within this area is a region of inadmissible workload that results from a combination of constraints on goal distance and effective time available for action. This region, shown as the shaded area in Figure 1, reflects the workload limitations on an operator when task demand is within his or her perceived range of capability but a time deadline precludes successful execution.

Assuming an arbitrary equivalence for the ranges represented on each axis, and if reduction of perceived distance per unit of effective time is considered constant, then contours appear as in the area of admissible load in

Figure 1. These contours possess an equivalence such that load resolution at any location along any one contour requires a consistent effort on behalf of the performer. In our previous work we refer to these functions as *isodynamic workload contours* (Hancock and Chignell, 1988). As illustrated in Figure 1, equal workload increases (WI) may be generated by increasing the perceived distance from the goal (workload increase resulting from distance, WI_D), by decreasing the effective time for action (workload increase resulting from time, WI_T), or by the two attributes in combination (WI_C).

A fuller representation of workload is given if the contours are developed in a third dimension, giving a surface that ascends toward the top right of the diagram. This is illustrated in Figure 2, in which the demands of the task and time constraints on performance drive the individual toward the lower left region while the actions of the performer oppose this impetus and serve to return the individual toward the region at the upper right. Within this framework, effort expenditure may be an adaptive strategy whereby the operator expends a constant rate of effort to attain the desired goal within the set time horizon set. This adaptive action on behalf of the operator might have little distortional effect in the center of the two-dimensional space in Figure 1, but it is likely to change contour shape at extremes, where threshold transition is imminent. The shape of the isodynamic workload contours shown in Figure 1 will change depending on the intrinsic constraints of the task under consideration. Such shapes will be empirically driven according to particular task demands and are explored in the present experiment.

The function for the change in load on each axis in Figure 2 is not arbitrary but adopts an ogival form. This function is used as an initial representation of the summed normal range

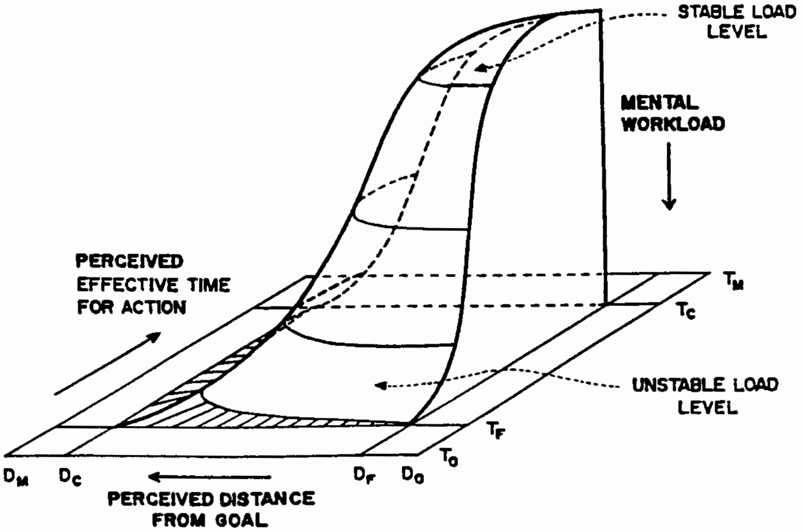

Figure 2. *Superimposed on the base described in Figure 1 is a third dimension representing degree of workload. Stable levels of load are presented at the top right, and workload increases with transition down the surface shown. Reproduced and modified from P. A. Hancock and M. H. Chignell (1988). Mental workload dynamics in adaptive interface design.* IEEE Transactions on Systems, Man, and Cybernetics, 18, *647–658. © IEEE 1988.*

of adaptive responses open to the operator (Hancock and Meshkati, 1988). In effect, the model assumes a parametric range of strategies by which an individual can solve the imposed problem. The nonlinearities in the topological structure imply that equal changes in either spatial or temporal constraint do not result in equal changes of mental workload. Rather, in the region of stability, at the top right, there are many strategies by which an operator can attain the goal state, and associated workload is low. However, toward the identified maximal thresholds, the number of strategies open to the operator are highly restricted, and associated workload is high.

These characteristics mean that the rate of progress toward the goal state is highly contingent on the operator's location on the surface of the topological structure. It can be envisaged that an individual's position on the workload surface is subject to constant change. Task demands—as represented, for example, by diminishing time for completion—act to push the individual from the stable regions at the top right of Figures 1 and 2 toward the thresholds of failure, principally at the lower left portions of these figures. Operator actions in resolving task demands serve to oppose this tendency. However, as Figure 2 indicates, the workload associated with these actions differs according to one's current position on the workload surface. The ogival curves assumption implies that there are few stereotypical responses at high extremes of demand and a plethora of choices as the individual approaches the goal state. This assertion provides a number of predictions, based upon which the model may be subject to falsification.

The model illustrated in Figure 2 does not include the operator's progress with respect to the changing loads of multifaceted goals but, rather, represents progress toward a unitary goal state. In any real-world task, the operator has a number of embedded hierarchically structured goals that need simultaneous reconciliation. For example, one goal may be related to effective system control in the face of external perturbing influences, whereas a simultaneous need to coordinate activity with other operators (e.g., formation flight) presents a second source of workload. These efforts to reconcile embedded goals summate to provide an overall level of workload that must be considered for any real-world implementations.

The model therefore suggests some hypotheses that are evaluated in the present experiment. Specifically, the change in task load function is tested by manipulating goal distance and time pressure, which reflects changes in the predicted workload surface as assessed by primary task performance and subjective estimates of mental workload.

METHOD

Experimental Task

To test predictions from this model, we used a performance task named Timepools (Johnson and Hart, 1987). Timepools is a software program that presents a matrix of circles on a computer screen, as shown in Figure 3. The actions of the subject in making discrete movements to a target location are similar to those used in traditional Fitts' law investigations (Fitts, 1954), except that targets shrink at rates that can be preset by the experimenter or made contingent on some part of the participant's performance.

Procedure

Six subjects volunteered to participate in this experiment. Each participant viewed a 7 × 5 matrix of white 1.9-cm-diameter circles against the blue background of a 35.6-cm-diagonal monitor (see Figure 3). Using an electro-optical mouse connected to a PC

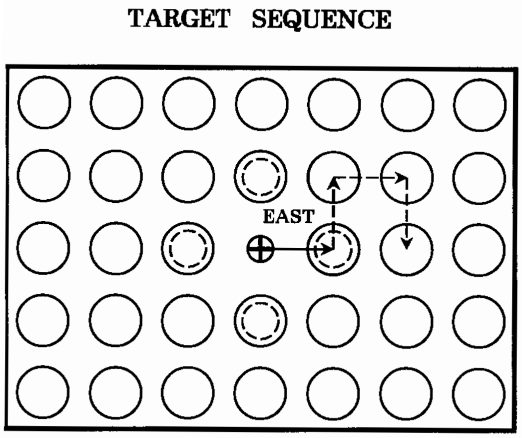

Figure 3. *A diagrammatic representation of Timepools. The 7 × 5 matrix of white circles appears on the computer monitor against a blue background. The subject stabilizes the cursor, represented by a red cross, in the start circle. The four adjacent circles begin to shrink, and the subject attempts to capture the circle indicated in the message presented, in this instance EAST. The dotted line represents a sequence of attempts to capture four targets or a path length of four.*

clone, the subject fixed a red cursor inside the start circle. The start circle was randomly located within the field of circles at the start of a trial, and its smaller size made it discriminable from the other circles. After 500 ms of stabilization, the four adjacent circles began to shrink at specified rates. At the same time that shrinkage began, a direction appeared above and to the right of the start circle in white letters specifying which of the shrinking circles (e.g., NORTH) was to be captured. Although it was possible to present more complex movement problems to operators, only the four primary compass directions of north, south, east, and west were used. The subject then attempted to capture the indicated circle by placing the cursor within that circle.

Whether the subject was successful or unsuccessful in his or her attempt to capture a target, the next target in the sequence of targets would be presented after a 500-ms period of stabilization. At this time the circle matrix would be reset, which returned all the circles to their original sizes except the circle in which the previous target had been, which was now the smaller start circle size. The next direction in a multiple path was then presented. The subject attempted to capture that circle, and so forth, until a whole path or trial was completed.

During each move within a path sequence, a time record was taken of the subject's *reaction time* (RT)—the elapsed time between the presentation of the stimulus and the initiation of movement—and *movement time* (MT), defined as the time to complete the movement once initiated. If a subject was unable to capture the target before it disappeared, or if a subject moved to the wrong target, the trial was classified as an error. After each performance condition, operator mental workload was assessed with the NASA-Task Load Index (TLX; Hart and Staveland, 1988) and Subjective Workload Assessment Technique (SWAT; Reid and Nygren, 1988) workload scales.

For each shrink rate, the area of the target circle was halved according to the four times chosen (i.e., 250, 500, 1000, and 2000 ms, respectively). The four shrink rates were completely crossed with four path lengths composed of 2, 4, 8, and 16 sequential target captures, respectively. This yielded 16 conditions per participant with 10 trials in each condition cell, giving a total of 160 trials per subject. The order of the presentation of combinatorial conditions was randomized across the six participating subjects (but see Poulton, 1982). Each participant was aware of which condition he or she was in; subjects did not know the respective path ahead of time, only the number of "steps" in that condition. It was this recognition of the number of steps in the condition that represented the manipulation of the distance to the goal state.

Model Predictions

Path length (i.e., the number of targets to be captured) and shrink rate of individual targets allowed for the generation of conditions that tested point predictions, which compose the theoretical workload surface as illustrated in Figure 2. A priori, it was expected that the most difficult condition would be the one in which the fastest shrink rate (i.e., 250 ms) was crossed with the longest path length (i.e., 16 steps), and, conversely, the easiest condition was expected to be the one in which the slowest shrink rate (i.e., 2000 ms) was crossed with the shortest path length (i.e., two steps). Absolute prediction of the relative difficulty of each of the other conditions was not made directly, as the relative importance of the change in shrink rate versus the change in path length could not be established prior to performance assessment.

RESULTS

Results from the present experiment indicated that there was a significant main effect of shrink rate on movement time (MT), $F(3,80) = 109.55, p < 0.0001$. In general, MT increased with the speed of shrink rate, such that shorter movement times accompanied longer shrinkage times and vice versa. On MT there was also a significant effect of path length, $F(3,80) = 5.68, p < 0.001$, in which MT tended to decrease as path length grew. In addition, there was a strong indication of an interaction between path length and shrink rate, $F(9,80) = 1.88, p = 0.051$ (see Figure 4). There were significant main effects of shrink rate on reaction time, $F(3,80) = 3.02, p < 0.029$. Again, there was a trend of RT reduction for increasing shrink rates. On RT there was also a significant effect of path length, $F(3,80) = 18.92, p = 0.0001$, which followed a similar pattern to that of MT (see Figure 5). For error rate, both path length and shrink rate again were significant, $F(3,80) = 15.39, p < 0.0001$, $F(3,80) = 178.91, p < 0.0001$, respectively, but no interaction was present (see Figure 6). On the sum of reaction time and movement time—here labeled *total response time* (TRT)—there were significant main effects of shrink rate and path length, $F(3,80) = 67.12, p < 0.0001$, $F(3,80) = 13.54, p < 0.0001$, but again no interaction (see Figure 7).

With respect to workload scores, there were significant main effects of both path length data and shrink rate, $F(3,80) = 3.2, p < 0.028$, $F(3,80) = 4.48, p < 0.005$, but no interactive effects between these factors on the overall workload score (OWL) of the NASA-TLX. This pattern was exactly replicated by the results of the overall SWAT assessment technique, $F(3,80) = 3.3, p < 0.026$, $F(3,80) = 7.9, p < 0.0001$, and these findings are illustrated in Figures 8 and 9.

However, analysis of the components dimensions of each scale provided a more detailed picture. On the time dimension of the SWAT procedure, there was a significant effect of shrink rate, $F(3,80) = 9.2, p < 0.0001$, but differences attributable either to path length or the interaction between these factors were not significant ($p > 0.05$). The converse pattern was seen in the effort dimension, in which a significant effect of path length, $F(3,80) = 5.8, p < 0.0015$, is associated with no effect of shrink rate. There were no significant interactive effects on any of the SWAT dimensions, which is consistent with the pattern for performance measures.

The SWAT stress dimension gave a pattern of results consistent with the SWAT time findings. That is, a significant effect of shrink rate was found, $F(3,80) = 3.8, p < 0.014$, but no effect of path length. For the TLX dimensions, both path length and shrink rate were significant for the mental demand, $F(3,80) =$

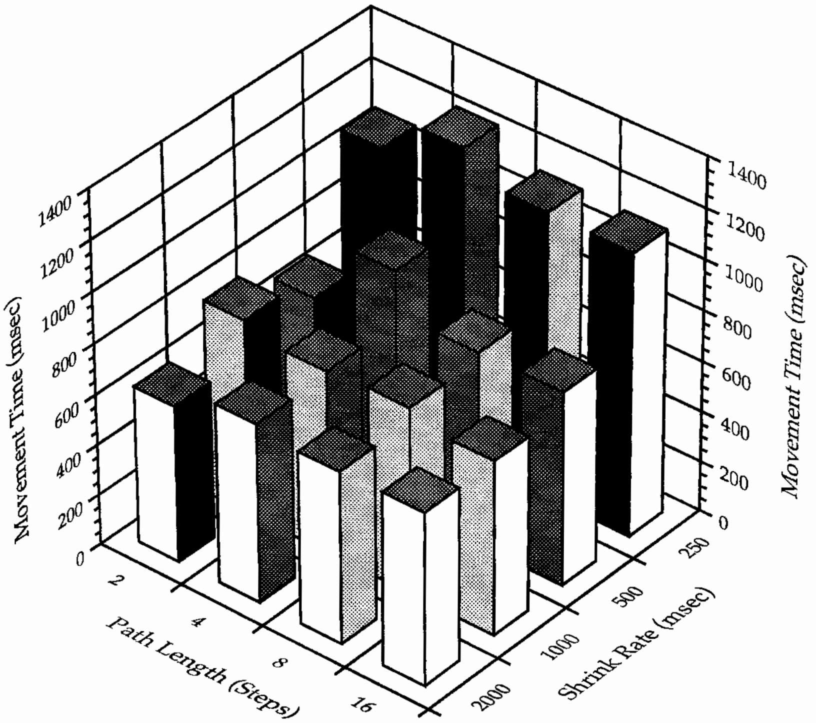

Figure 4. *Movement time plotted as a function of path length (steps) and shrink rate in milliseconds.*

4.5, $p < 0.006$, $F(3,80) = 3.1$, $p < 0.03$; physical demand, $F(3,80) = 3.4$, $p < 0.021$, $F(3,80) = 2.8$, $p < 0.05$; and effort, $F(3,80) = 3.4$, $p < 0.021$, $F(3,80) = 4.1$, $p < 0.009$, scales. However, shrink rate was significant only for the temporal demand, $F(3,80) = 6.9$, $p < 0.0001$, and performance, $F(3,80) = 4.2$, $p < 0.008$, dimensions, as were the time and stress dimensions of the SWAT procedure. Unlike the SWAT effort dimension, there were no TLX dimensions that were differentially sensitive to path length alone.

The present results are represented by 16 discrete points on a surface. There are no significant interactive effects, so the differences all reside exclusively between levels of the two independent variables. Post hoc analyses revealed consistent topological surfaces, in that no reversals occurred whereby a value decreased and then subsequently increased significantly across the range of each independent variable. Thus we examined the nature of the surfaces and their relationship within the predicted surfaces, rather than doing a point-by-point post hoc comparison. For total response time there is evidence of consistency with the predicted surface. That is, a significant decrease in TRT was observed across shrink rate and path length.

With respect to subjective measures of workload, both the summed TLX scores and the overall SWAT values also followed this predicted pattern. That is, there were significant reductions in overall workload for both scales across path length and shrink rate. For each scale, the highest loading condition oc-

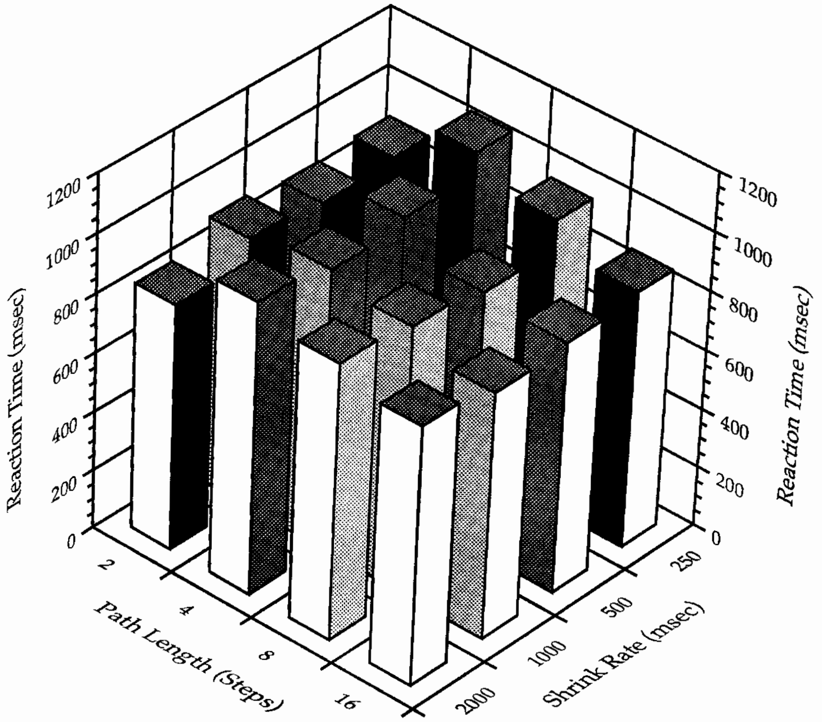

Figure 5. *Reaction time plotted as a function of path length (steps) and shrink rate in milliseconds.*

curred either near the combination of the 250-ms shrink rate and the 16-step path condition, which was specified a priori as being the most demanding. These results, from the overall primary task performance measures and the summed subjective workload scores, are consistent with and supportive of the predictions of the model as originally proposed. However, components of performance and workload reveal a more complex and potentially informative pattern.

On both movement time and reaction time, there was a significant effect of shrink rate and path length. These differential components of response proved sensitive to differences in the task manipulation. A comparison of these differences in measured performance with those recorded for components of the subjective workload scales proved interesting. For the three subsidiary SWAT scales, analyses indicated that the time and stress dimensions varied significantly with shrink rate but not at all with path length. This suggests either that the RT measure and the time and stress scales tapped some common source of workload or that response on the time and stress scales, taken after trial block completion, proved sensitive to reaction time performance. Whether driven by the independent variable directly or indirectly through the performance outcome, time is confirmed as a critical component of workload. The effort dimension of the SWAT scale showed significant differences according to path length but did not change as a function of shrink rate. Similar reasoning as that applied to the

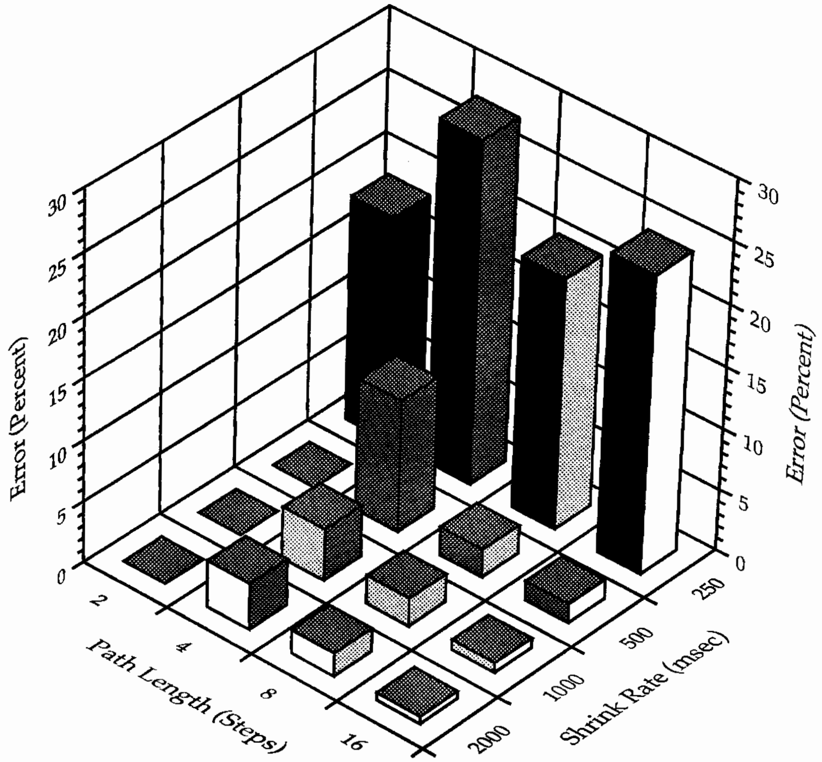

Figure 6. *Error percentage plotted as a function of path length (steps) and shrink rate in milliseconds.*

time and stress dimensions suggests that some additional discrete contribution to workload may be attributed to increasing path length and reflected in movement time and perceived effort.

The picture for the component dimensions of the TLX is similar in that the temporal demand and performance dimensions of the TLX were sensitive only to shrink rate and unaffected by path length. Again, the factor of time is raised as the common influence. However, unlike the effort dimension of the SWAT, no dimensions of the TLX were selectively influenced by path length alone; all other dimensions varied significantly as a function of both path length and shrink rate.

The dynamic model provides a symmetrical representation for change in workload as a function of distance from the goal state and available time for response. The generic representation of the model, given in Figure 2, illustrates an equivalence between these axes. That is, units of time, for the purpose of convenience here, are considered as equal to units of distance. Consequently, a change of two steps in path length is considered equivalent to a change in shrink rate of 250 ms.

Conditions extending from this baseline were chosen according to a doubling schedule. Of course, there is no a priori reason to suppose that two additional steps on a path would be equivalent in workload value to an additional rate of shrinkage of target of 250 ms. If participants interpret these manipula-

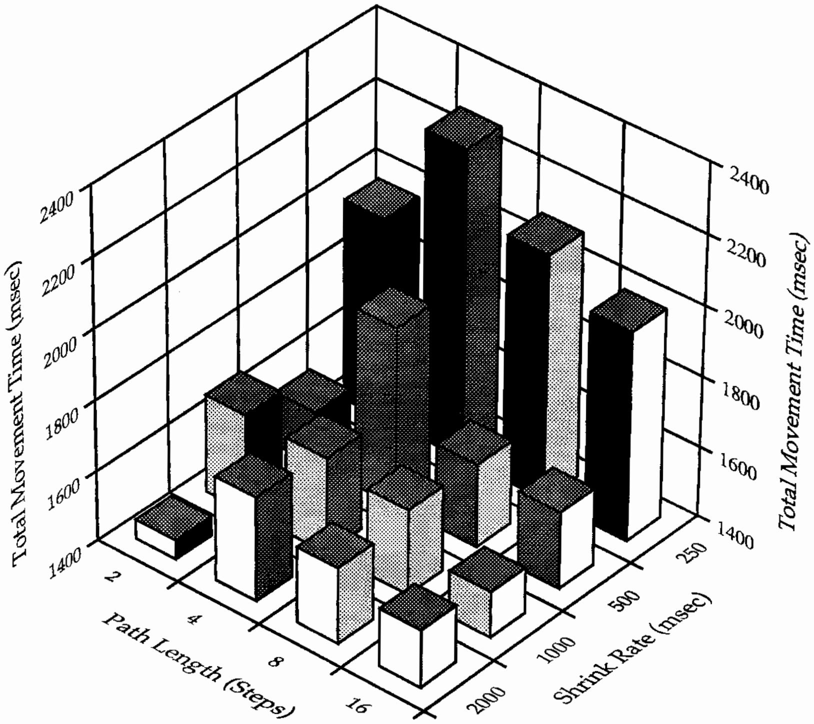

Figure 7. *Total movement time (reaction time + movement time) plotted as a function of path length (steps) and shrink rate in milliseconds.*

tions presented to them as equivalent, then symmetry might be expected in response. However, deviations from symmetry might reflect an intrinsic inequality of this scaling, depending on the particular task at hand, and such deviations should be expected. However, the fundamental characteristics of the workload surface should remain consistent across tasks.

In its simplest terms, then, the prediction of the model is for workload to decrease linearly and symmetrically across path length and shrink rate from 16 steps and 250 ms to 2 steps and 2000 ms. In terms of primary task performance, this would be reflected in significant reductions of reaction time, movement time, and total response time isomorphically mapped to reductions in overall levels and contributory subscales of the NASA-TLX and SWAT subjective workload scales (for the purposes of illustration, those overall trends that are consistent with predictions of the model shown in Figure 10).

DISCUSSION

The overall data provide confirmation of the model proposed by Hancock and Chignell (1988). Some differences in outcome can be accommodated by suggesting that individuals occupy different locations on the workload surface. Nevertheless, the model would be falsified if changes in workload level were to fluctuate irregularly across the time and distance axes or task demands, which generate local minima and maxima in the work-

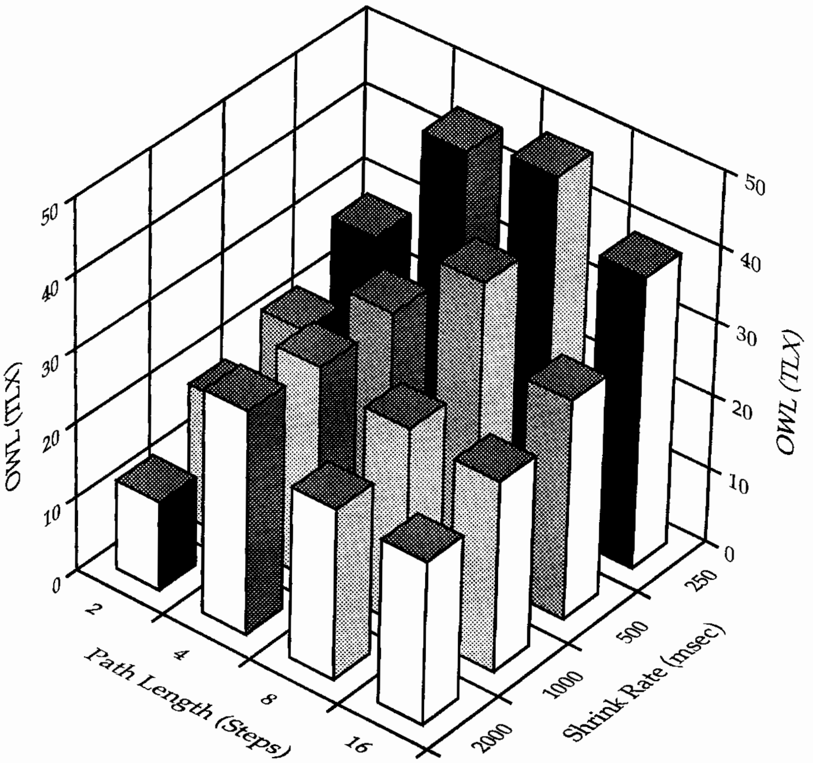

Figure 8. *Total operator workload (NASA-TLX) plotted as a function of path length (steps) and shrink rate in milliseconds.*

load surface. The present data support the general form of the model and indicate its utility in predicting global levels of task-related workload.

The topology of the surface of the model presented in Figure 2 generates a number of specific predictions that can be addressed by the results. In general, the model predicts that workload increases with greater distances from the goal state and reduced effective time for action. With respect to the present task, this implies highest subjective workload and poorest primary task performance at the fastest shrink rate and longest path length. There is certainly strong evidence that performance was poorest and workload highest at the fastest shrink rate. However, such a pattern cannot be affirmed with as great a certainty as for the effects of path length. Indeed, the longest response time and highest error rate was at the 4-step condition, although the 16-step condition was second with respect to error rate.

The converse of this hypothesis was supported. That is, the combination of the slowest shrink rate and shortest path length did prove to have the best performance in terms of both TRT and error, and it also showed the lowest workload for the TLX and second-lowest SWAT workload. With respect to these findings, the model does receive confirmatory support. The continuity of the model's surface requires that workload decrease and primary task performance improve systemati-

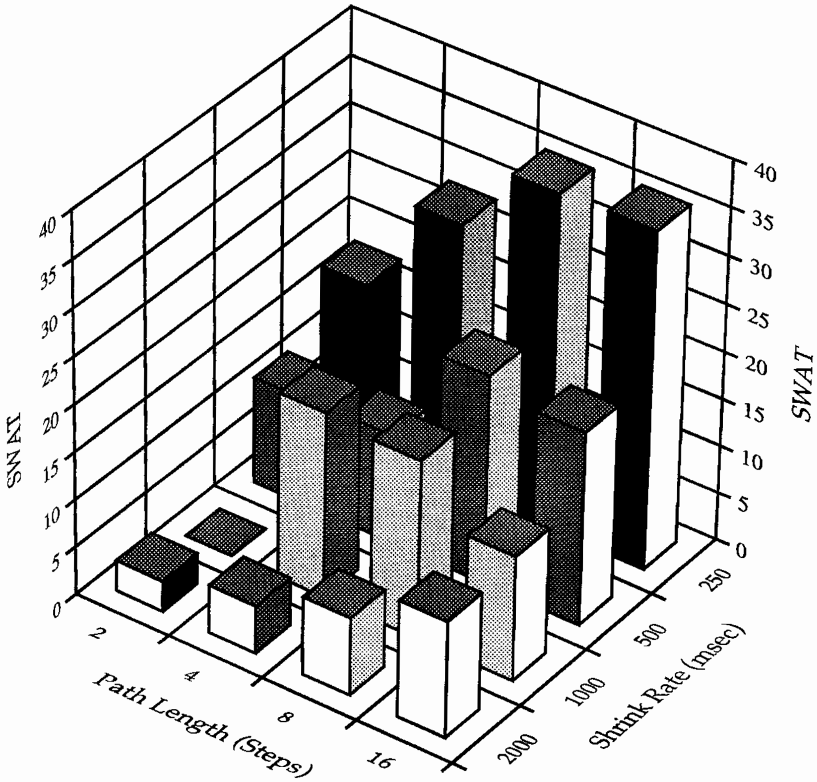

Figure 9. *SWAT plotted as a function of path length (steps) and shrink rate in milliseconds.*

cally as these constraints are sequentially released with shorter path lengths and slower shrink rates. Again, the overall pattern of the findings does support this assertion. Of particular note is the clear support for this trend in the overall findings of both subjective workload scales. Of course, the model also predicts an effect for skill that was not manipulated here but is an issue that should be addressed in future work.

Recent developments in workload research and application have led some researchers to question whether workload assessment is a declining issue. With respect to both practical and theoretical developments in human factors, other researchers would contend that this is not the case (Derrick, 1988; Hancock, 1989; Vidulich, 1988). Yet despite a large and growing body of literature, workload progress has stagnated somewhat for lack of a definitive theoretical direction. The traditional and continued link between workload and resource theories of attention, both being facets of what Freeman (1948) referred to as the *energetics of performance*, has failed to yield the hoped-for progress. The subsequent fractionation of multiple resource theory, together with periodic reports of workload dissociation (Yeh and Wickens, 1988), has therefore rendered context-independent workload prediction an arduous endeavor.

In the present theoretical approach, global

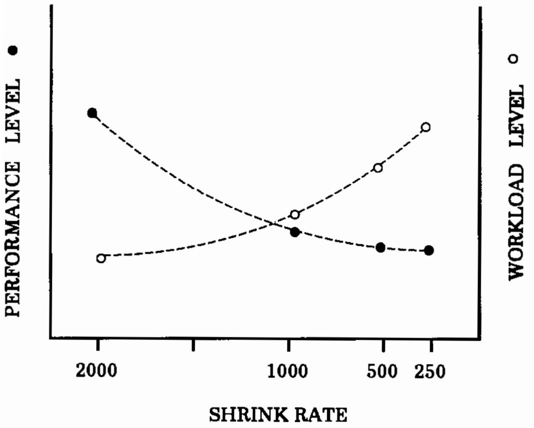

Figure 10. *Hypothetical workload level and performance as a function of shrink rate.*

metrics were generated for the axes of the model presented, those being effective time for action and distance from the goal state. The generality of these global metrics represents both a strength and a weakness in the model. The strength lies in the ability to supersede strict, context-specific task conditions so that the model may be applied to a spectrum of systems and their various operational conditions. The weakness lies in the necessity to translate or equate aspects of specific performance conditions with the global time/distance measures that we have proposed. Were these the sole properties of the model, then the potentially arbitrary nature of the time and distance representations and the fact that the workload surface and its local variants are essentially the outcome of empirical evaluation rather than actual given conditions would render this approach problematic.

In addition, the ability to move axes with specific data would render the model open to the same criticisms as those that have been directed at the inverted U explanation of performance under stress (Hancock, 1987; Hancock and Warm, 1989). However, this criticism is obviated by the fact that the model contains physical anchors representative of physical limitations and psychophysical anchors representative of performer limitations. Given the prediction that the morphology of the workload surface also transcends specific task conditions, the model generates sufficient constraints to provide testable propositions. Its further utility can be evaluated only by the degree to which it usefully predicts workload in extended regions of the performance space, given that experimental values are known for one specific region.

The value of the proposed model has been confirmed in the present experiment using typical performance measures such as RT, MT, and error and common subjective workload assessment techniques such as SWAT and the NASA-TLX. What remains to be determined is the value of this approach when more complex representations of overall system performance replace the somewhat arbitrary laboratory-based methods of evaluating operator capability (see Hancock and Chignell, 1987). Specific consideration of the facets of performance measured here, together with a consideration of workload subscales, suggests specific linkages between objective and subjective reflections of response that were unsuspected prior to the experimental inquiry. Although these links may themselves offer an intrinsically useful relationship to the workload theorist, the differentiation within performance and workload militates against an oversimplistic theoretical structure. Further empirical validation is always needed; however, the present approach offers a framework for practitioner and theorist alike from which to establish a unified study of human mental workload.

ACKNOWLEDGMENT

Part of this work was presented at the 33rd Annual Meeting of the Human Factors Society, (Hancock, Chignell, Vercruyssen, and Denhoff, 1989). This research was supported by the National Aeronautics and Space Administration through Grant NCC 2-379 from NASA Ames Research Center. Sandra Hart was the technical monitor for the grant. The views expressed are those of the authors and do not necessarily represent those of the named agency. We are grateful for data-collection assistance from Nancy Knabe and Marshall Denhoff. Max Vercruyssen and Mark Chignell provided valuable comments on earlier versions of this paper. We would like to thank two anonymous reviewers for their incisive and helpful comments on an earlier version of the present work.

REFERENCES

Åstrand, P. O., and Rodahl, K. (1970). *Textbook of work physiology.* New York: McGraw-Hill.

Becker, A. B., Warm, J. S., Dember, W. N., and Hancock, P. A. (1991). Effects of feedback on perceived workload in vigilance performance. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1491–1494). Santa Monica, CA: Human Factors and Ergonomics Society.

Card, S. K., Moran, T. P., and Newell, A. (1986). An engineering model of human performance. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 45.1–45.35). New York: Wiley.

Chignell, M. H., and Hancock, P. A. (1985). Knowledge-based load leveling and task allocation in human-machine systems. In *Proceedings of the 21st Annual Conference on Manual Control* (9.1–9.11). Columbus, OH: National Aeronautics and Space Administration.

Damos, D. L. (1988). Individual differences in subjective estimates of workload. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 231–237). Amsterdam: North-Holland.

Derrick, W. L. (1988). Dimensions of operator workload. *Human Factors, 30,* 95–110.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47,* 381–391.

Freeman, G. L. (1948). *The energetics of human behavior.* Ithaca, NY: Cornell University Press.

Gopher, D., and Donchin, E. (1986). Workload: An examination of the concept. In K. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 41.1–41.49). New York: Wiley.

Hancock, P. A. (1987). Arousal theory, stress and performance: Problems of incorporating energetic aspects of behavior into human-machine systems function. In L. S. Mark, J. S. Warm, and R. L. Huston (Eds.), *Ergonomics and human factors: Recent research* (pp. 170–179). New York: Springer-Verlag.

Hancock, P. A. (1989). The effect of performance failure and task demand on the perception of mental workload. *Applied Ergonomics, 20,* 197–205.

Hancock, P. A., and Chignell, M. H. (1987). Adaptive control in human-machine systems. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 305–345). Amsterdam: North-Holland.

Hancock, P. A., and Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man and Cybernetics, 18,* 647–658.

Hancock, P. A., Chignell, M. H., and Kerr, G. (1988, August). *Defining task complexity and task difficulty.* Presented at the XXIV International Congress of Psychology, Sydney, Australia.

Hancock, P. A., Chignell, M. H., Vercruyssen, M., and Denhoff, M. (1989). Experimental evaluations of a model of mental workload. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 1233–1237). Santa Monica, CA: Human Factors and Ergonomics Society.

Hancock, P. A., and Meshkati, N. (Eds.). (1988). *Human mental workload.* Amsterdam: North-Holland.

Hancock, P. A., Meshkati, N., and Robertson, M. M. (1985). Physiological reflections of mental workload. *Aviation, Space and Environmental Medicine, 56,* 1110–1114.

Hancock, P. A., and Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

Hancock, P. A., Warm, J. S.,, and Dember, W. M. (1991). *The workload of sustained attention.* (Tech. Report HFRL 91-02). Minneapolis: University of Minnesota, Human Factors Research Laboratory.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

James, W. (1890). *Principles of psychology.* New York: Holt.

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 5–39). Amsterdam: North-Holland.

Johnson, W. W., and Hart, S. G. (1987). Step tracking shrinking targets. In *Proceedings of the Human Factors Society 31st Annual Meeting* (pp. 248–252). Santa Monica, CA: Human Factors and Ergonomics Society.

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

Kantowitz, B. H. (1987). Mental workload. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 81–121). Amsterdam: North-Holland.

Kantowitz, B. H., and Sorkin, R. D. (1983). *Human factors: Understanding people-system relationships.* New York: Wiley.

Moray, N. (Ed.). (1979). *Mental workload: Its theory and measurement.* New York: Plenum.

O'Donnell, R. D., and Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 42.1–42.49). New York: Wiley.

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin, 92,* 673–690.

Reid, G. B., and Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: North-Holland.

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84,* 1–66.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Vidulich, M. A. (1988). The cognitive psychology of subjective mental workload. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 219–229). Amsterdam: North-Holland.

Westrum, R. (1991). *Technologies and society: The shaping of people and things*. Bellmont, CA: Wadsworth.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson and R. Pew (Eds.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of attention* (pp. 63–104). Orlando, FL: Academic.

Wickens, C. D. (1987). Attention. In P. A. Hancock (Ed.), *Human factors in psychology* (pp. 29–80). Amsterdam: North-Holland.

Wickens, C. D. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors, 30,* 599–616.

Wilson, J. R., and Corlett, E. N. (1990). *Evaluation of human work*. London: Taylor & Francis.

Yeh, Y. Y., and Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30,* 111–120.