THE EFFECT OF SPATIAL AND TEMPORAL TASK CHARACTERISTICS ON PERFORMANCE, WORKLOAD, AND STRESS

Grace W.L. Teo and James L. Szalma
University of Central Florida

The present study examined the Maximal Adaptability Model of Stress (Hancock & Warm, 1989) by investigating how the task characteristics of information rate (event rate) and information structure (display uncertainty) affect performance on a cognitively demanding signal detection task. Performance as well as perceived workload and stress were measured. Results supported a performance-workload association rather than performance insensitivity, but the pattern of decline in adaptation to task-induced stress generally conformed to the maximal adaptability model: At lower levels of demand the change in accuracy and workload was smaller, but at higher demand these changes increased in magnitude.

INTRODUCTION

The maximal adaptability model of stress and performance (Hancock & Warm, 1989) asserts that the most proximal source of stress that confronts an individual is the task he/she is to perform. The model delineates the task into two dimensions, these being information structure (the spatial organization of a task) and information rate (the temporal properties of a task). The model predicts that across a wide range of task demands individuals can maintain a stable adaptive state (see Figure 1). However, as demands are increased or decreased to extremes, adaptation failures occur. As can be seen in Figure 1, the failures in adaptation are progressive, such that subjective comfort (e.g., perceptions of stress and workload) begins to decline at levels of stress/task demand that are lower than the levels at which performance decrements are observed. This translates into regions in which there are performance insensitivities (performance remains stable but perceived workload increases) and at greater extremes of demand there are areas in which performance declines and perceived workload remains high (performance-workload associations; see Hancock, 1996). Such a pattern has been observed in a field study of police officers engaged in a firearms training task (Oron-Gilad, Szalma, Stafford, & Hancock, 2007). However, the task dimensions identified in the maximal adaptability model could not be manipulated. The current study was designed to evaluate this model using a cognitively demanding detection task in which the two dimensions were manipulated.

METHOD

Participants

Three hundred and nineteen psychology undergraduates (131 males, 188 females) at a large southeastern U.S. university were recruited for the study. Participants received course credit in exchange for participation. Participant ages ranged from 17 to 25 years.

Task

The study used an adaptation of the Bakan (1959) cognitive signal detection task developed by Warm, Howe, Fishbein, Dember, & Sprague (1984). The task consisted of a visual presentation of 2-digit numbers ranging from 01 to 99. Participants were instructed to respond when a critical signal appeared on the screen. Critical signals were 2-digit numbers whose digits differed by 0 or 1 (e.g. 01, 54, 99), and all other numbers were defined as neutral events. Throughout the task, two 2x2 grids appeared side-by-side on the screen where each of the 8 cells was a “display” in which the numbers appeared.

Manipulations

There were 2 independent variables in the study: Information Uncertainty (number of displays to be monitored) was the within-subjects variable, while Information Rate (event rate) was the between-subjects variable. Information uncertainty was manipulated by varying the number of the displays (i.e. 1, 2, 4 or all 8) participants had to monitor for the critical signals. The order in which these four levels were presented was balanced via a Latin Square. Before each condition, there was a ‘notification screen’, where a red outline around the display(s) notified participants of the number and location of the displays to be monitored for that condition. For the 1, 2 and 4 display conditions, the locations of the monitored displays were randomly selected, with the restriction that in the 2 or 4 display condition the monitored displays were always within the same 2x2 grid. Task uncertainty was lowest when participants had to monitor only 1 display and highest in the 8 display condition.

Information rate was manipulated by varying the speed of presentation of the numbers (event rate). The four event rates were 8, 12, 16, and 20 events/minute. The lowest difficulty level was at an event rate of 8 events/min, followed by 12 events/min, then 16 events/min, and the greatest difficulty entailed an event rate of 20 events/min. Each participant was randomly assigned to one of these four levels. Regardless of the event rate, all events (stimuli) were presented for 2500msec for each trial. The stimulus duration of 2500msec was deter-
mined via a pilot study after a pilot study which showed that it was just long enough to allow scanning of all 8 displays. The number of signal trials was set at 10 signals per 3-minute block of trials across all conditions. The number neutral events differed for the different levels of event rate.

Figure 1. The maximal adaptability model. (after Hancock & Warm, 1989).

Measures

**Workload.** The NASA-Task Load Index (TLX; Hart & Staveland, 1988) served as the measure of workload. It assesses six sources of workload, namely, **Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration**, as well as an aggregate measure of global workload.

**Stress.** Perceived stress was measured using a short version of the Dundee Subjective State Questionnaire (DSSQ) (Helton, 2004; Matthews et al., 2002) which consisted of the subscales of **Task Engagement, Distress, and Worry**. Participants rated the 20 items of the short DSSQ on a 5-point rating scale.

Procedure

The data were collected from groups of up to nine participants seated in cubicles that prevented visual contact. Each participant wore noise canceling headphones (Audio-Technica QuietPoint ATH-ANC7) to reduce the impact of any distracting sounds. During the session, if any participant needed to contact the experimenter he/she communicated via a chat window that restricted contact to the experimenter (i.e., participants could not contact one another or anyone outside the laboratory). Participants were asked to surrender their wristwatches, cell phones, or other portable devices for the duration of the experiment. Participants completed the pre-task stress state (DSSQ) questionnaire prior to engaging in the tasks. Each condition was 12 minutes in duration, with the first 3 minutes serving as practice trials. After each condition participants completed the NASA TLX and post-task version of the DSSQ, the order of which was counterbalanced. The duration of the entire study was approximately 2 hours.

RESULTS

Due to computer-based technical problems some participants did not complete several items on the workload and stress measures and were therefore not included in the analyses for those variables. Hence, the degrees of freedom are not constant across dependent measures.

Data were analyzed via 4 (Event Rate) x 4 (Display) Mixed ANOVAs for performance, workload, and stress. Due to computer-based technical problems some participants did not complete several items on different subjective measures and so were not included in the analyses for those variables. Hence, the degrees of freedom are not constant across dependent measures.

Proportion of Signals detected

Results showed a significant main effect of information uncertainty (Display), *F*(3,842) = 921.60, *p* < .001. As expected, the highest proportion of signals detected was in the 1-Display condition (*M* = 0.91, *SD* = 0.19), followed by the 2-Display condition (*M* = 0.85, *SD* = 0.19), then the 4-Display condition (*M* = 0.75, *SD* = 0.19), and the worst performance was in the 8-Display condition (*M* = 0.46, *SD* = 0.18). Pos-hoc comparisons using the Bonferroni correction indicated that differences between all possible comparisons of Display conditions were statistically significant.

There was also a significant main effect of information rate (Event Rate), *F*(3,314) = 23.39, *p* < .001. Detection rate was significantly better in the 8-events/min condition (*M* = 0.816, *SE* = 0.016) than both the 16-events/min (*M* = 0.748, *SE* = 0.016) and 20-events/min (*M* = 0.637, *SE* = 0.016) conditions, while the 12-events/min condition (*M* = 0.766, *SE* = 0.016) yielded significantly higher correct detections than the 20-events/min condition. Performance in the 20-events/min condition was significantly worse than in all other conditions.
The Display by Event Rate interaction was statistically significant, $F(8, 842)=11.684, p<.001$, suggesting that pattern of improvement in detections across the 4-Display conditions differed as a function of event rate (see Figure 2). As expected, the best detection rate was in the 1-Display, 8-events/min condition ($M=0.947, SD=0.137$), while the lowest level of performance was observed in the 8-Display, 20-events/min condition ($M=0.278, SD=0.121$).

**Proportion of False alarms**

The main effect of Display was significant, $F(3,551)=99.72, p<.001$. There was a significantly lower false alarm rate in the 1-Display condition ($M=0.01, SD=0.04$) than in the 4-Display ($M=0.02, SD=0.65$) and 8-Display ($M=0.07, SD=0.11$) conditions, and significantly smaller proportion of false alarms in the 2-Display condition ($M=0.02, SD=0.06$) compared to the 8-Display condition. Performance in the 8-Display condition was significantly worse than the other conditions.

Although the main effect of Event Rate was not significant, there was a significant Display by Event Rate interaction, $F(5, 551)=5.26, p<.001$ (see Figure 3). Specifically, among all the Display conditions, only the 8-Display condition showed significant differences in false alarm rates across the different levels of Event Rate. Post-hoc comparisons using the Bonferroni correction indicated that the proportion of false alarms was significantly higher in the 8-Display, 8-events/min condition ($M=0.10, SD=0.09$) compared to the 8-Display, 20-events/min ($M=0.06, SD=0.15$) as well as the 8-Display, 12-events/min ($M=0.05, SD=0.04$) conditions.

**Response time on Correct (Signal detected) trials**

There was a significant main effect for Display, $F(3, 780)=397.42, p<.001$. Post-hoc tests indicated that the differences between all possible pair wise comparisons of Display levels were significant. The fastest responses were made in the 1-Display condition ($M=1.01, SD=0.35$), followed by the 2-Display condition ($M=1.28, SD=0.29$), then the 4-Display condition ($M=1.56, SD=0.32$), and the slowest responses were observed in the 8-Display condition ($M=1.83, SD=0.46$).

The main effect of Event Rate was also significant, $F(3,315)=20.59, p<.001$. All pair wise comparisons of event rates yielded significant differences except for that between the 16-events/min and 20-events/min levels. However, it was the 20-events/min condition ($M=1.32, SE=0.02$) that yielded the fastest response time, followed by the 16-events/min condition ($M=1.35, SE=0.02$), then the 12-events/min condition ($M=1.45, SE=0.02$), and the slowest responses were in the 8-events/min condition ($M=1.56, SE=0.02$). This may reflect the greater time pressure experienced at higher event rates.

In addition, the Display by Event Rate interaction effect was significant, $F(7, 780)=397.42, p<.001$, indicating that the pattern of response times across the levels of Event Rate differed for the various Display conditions (see Figure 4). For the 1-Display condition, the response time at the 8-events/min level ($M=1.14, SD=0.49$) was significantly slower than that of all other levels of Event Rate, but no significant differences were found across levels of Event Rate in the 2-Display condition. In the 4-Display condition however, the two slowest Event Rate levels differed significantly from the two highest Event Rate levels, i.e. it was the 8-events/min ($M=1.71, SD=0.39$) and 12-events/min ($M=1.63, SD=0.31$) levels that were significantly different from the 16-events/min ($M=1.46, SD=0.25$) and 20-events/min ($M=1.46, SD=0.25$) levels. This same pattern was found in the 8-Display condition, the response time for correct trials for the 8-events/min ($M=2.05, SD=0.57$) and 12-events/min ($M=1.93, SD=0.31$) levels differed significantly from the 16-events/min ($M=1.72, SD=0.33$) and 20-events/min ($M=1.62, SD=0.46$) levels.
Figure 3: Proportion of false alarms as a function of display uncertainty and event rate. Note. Error bars are standard errors.

Figure 4: Response Time on Correct Trials as a function of the number of displays to be monitored and event rate. Note. Error bars are standard errors.

Workload

Global workload. Technical problems resulted in a number of missing values for the weights of the six components of the TLX. Thus, weighted ratings and a weighted average could not be computed and the unweighted ratings and unweighted average were analyzed instead. Results showed a significant main effect of Display, $F(2, 633)=227.278$, $p<.001$, with the differences between all levels of Display being significant. Post-hoc tests indicated that, as expected, global workload was highest in the 8-Display condition ($M=52.83$, $SD=16.40$), followed by the 4-Display condition ($M=39.04$, $SD=18.08$), then the 2-Display condition ($M=31.48$, $SD=17.62$), and lastly, the 1-Display condition ($M=26.25$, $SD=18.40$) yielded the lowest scores on global workload.

There was also a significant main effect of Event Rate, $F(3,274)=2.99$, $p=.031$, but the only significant difference was found between the 8-events/min ($M=35.07$, $SE=1.60$) and 20-events/min ($M=41.44$, $SE=1.64$) levels. The Display by Event Rate interaction was not statistically significant (see Figure 5).

Subscales. Analyses of the six components (i.e. Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration) of the workload measure showed that for all components, there was a significant main effect of Display, (Mental Demand: $F(3, 846)=186.98$, $p<.001$, Physical Demand: $F(3, 855)=17.37$, $p<.001$, Temporal Demand: $F(3, 852)=195.62$, $p<.001$, Performance: $F(3, 861)=48.51$, $p<.001$, Effort: $F(3, 858)=46.044$, $p<.001$, and for Frustration: $F(3, 843)=85.581$, $p<.001$). In addition, there were significant Display by Event Rate interactions, with the exception of Physical Demand, $F(9,855)=2.039$, $p=.041$. There was also no significant main effect of Event Rate for all components except Temporal Demand, $F(3, 284)=3.00$, $p=.031$ and Frustration, $F(3, 281)=4.70$, $p=.003$. When effects were observed, they generally conformed to that of global workload: as task demand increased workload increased.

Stress

The DSSQ consists of three scales that reflect the components of stress: Engagement, Distress and Worry. Analyses showed that there were no significant differences in pre-task state as a function of Event Rate ($p>.31$ in each case).

Analyses of the three stress components indicated that there were significant changes in all components between pre-task and post-task, i.e. there was significant main effect of Display for Task engagement, $F(4, 1048)=42.26$, $p<.001$, Distress, $F(4, 974)=74.95$, $p<.001$, and Worry, $F(4, 1076)=21.86$, $p<.001$. However, there was no significant main effect for Event Rate, but there was a statistically significant Display by Event Rate interaction (see Figure 6) for Task Engagement, $F(11, 1049)=1.86$, $p=.041$ Post-hoc analyses did not indicate any significant differences among the pairwise comparisons.
Figure 6: Task Engagement as a function of the number of displays to be monitored and event rate. Note. Error bars are standard errors.

DISCUSSION

The purpose for the present study was to test elements of the maximal adaptability model of Hancock and Warm (1989). The progressive decline in subjective and behavioral adaptation predicted by the model was observed: as task demands increased performance declined and workload increased, with steeper slopes at the higher levels of information demand. However, in this study the increase in perceived workload was associated with declining performance. Performance insensitivity was not observed, indicating that the functions describing the progressive failure may not have different thresholds of adaptive failure for subjective state and task performance as implied in the model shown in Figure 1. For both measures the threshold occurred at the same level of demand (4-display condition; see Figures 2 and 4). One could argue that the range of demand in this study was restricted to that range of the model in which both forms of adaptation failed. However, this is unlikely, as the least demanding condition (one display at 8 events/minute) was associated with high performance and low workload, suggesting this condition lay within the comfort zone of figure 1. By contrast, the most demanding condition (8 displays at 20 events per minute) was associated with very poor performance and high workload. Hence, the current results are not likely an artifact of range effects. Instead, it appears that the progressive failures in adaptation, in which subjective state fails before performance (see Figure 1), may not be ubiquitously true as implied by the maximal adaptability model. The current data do not preclude the possibility that there may be tasks in which such progressive functions occur, but this study does indicate that relative positions of the nested functions may depend on the nature of the task itself. In this case, it appears that the performance and subjective state functions roughly coincide such that the failures occur at similar levels of demand. Future research should vary the specific nature of the task in order to address the potential generality of the current results.

REFERENCES