Exploring two methods for estimating performance tradeoff

DAVID NAVON
University of Haifa, Haifa, Israel

Navon (in press) found that subjects could adjust their performance of a task to designated levels with no effect on a concurrent task, provided that the demands on the two tasks were not negatively correlated. In two experiments designed to examine the generalizability of that result under different conditions, subjects responded to task requirements less well than in Navon (in press), and performance tradeoff was not obtained even when the demands were negatively correlated.

One way to measure dual-task interference is to estimate performance tradeoff. Navon (in press) compared two methods of estimation, an *optimum–maximum* method, which calls for keeping performance of one of the tasks at a certain predesignated level while maximizing performance of the other task, and a more customary method, the *minimal-requirements* method, which assigns requirements for minimal performance for the two tasks. In that study, I used a digit-classification task paired with a letter-classification task and found that, although subjects under both methods adjusted fairly well their performance of a task to the level being required, there was no evidence of a necessary concomitant effect on the concurrent task. Under the method that required subjects to maximize the performance of one of the tasks, performance on that task was kept high regardless of how well the other task was performed. A linear performance tradeoff was obtained with the minimal requirements method. However, in the light of the former finding, tradeoff does not seem to be due to any limit on joint performance (such as scarcity of processing resources), but rather to compliance of the subject with what is figured to be the objectives of the experimental situation.

This paper reports two attempts to examine the generalizability of those results under slightly different conditions.

EXPERIMENT 1

In this experiment, I tried to replicate the findings reported in Navon (in press) under a different motivation set: The subjects participated in this experiment not as a part of course requirements, but rather for payment, the magnitude of which depended on their performance.

Method

Apparatus and Setting. Stimuli were presented on a VT-11 graphic terminal. Presentation and data acquisition were controlled by a PDP 11-34 computer. Viewing distance was about 60 cm.

Design and Procedure. The subjects were required to perform concurrently two tasks: (1) a digit-classification task in which the subject had to tell whether a visually presented digit (out of the digits 1–8) was odd or even, and (2) a letter-classification task in which the subject had to tell whether a visually presented Hebrew letter (out of eight possible ones) belonged to the first or to the second half of the Hebrew alphabet. A trial started with the onset of a digit and a Hebrew letter at two continually presented frames on the screen. The subject sat with his/her fingers on a computer keyboard. To respond to the digit, the subject pressed one of two response keys; to respond to the letter, the subject pressed one of two other keys. The stimulus disappeared when a correct response was given or when 4 sec with no correct response had elapsed. If an incorrect response was given, the stimulus remained until the subject made the correct response. The subsequent trial was initiated immediately following the offset of both stimuli.

All trials were administered in a single session that lasted about 60–80 min. The first 384 trials were devoted to practice and calibration. The mean and standard deviation of latencies until the first correct response in both tasks were calculated for the last 96 of those trials. These statistics were used during the rest of the session to give performance feedback and to set performance requirements. The subsequent 96 trials were devoted to practice with feedback indicators. A running average of the latency for each task during the 8 trials immediately preceding the current one was calculated and displayed on the screen as a vertical arrow above the frame within which the task-relevant stimulus was displayed. The length of the arrow was proportional to the difference in standard scores between the mean plus 2 standard deviations and the standardized score of the current average. In other words, the arrow extended higher, the greater the speed of responding relative to the typical speed of the subject measured previously. The experimental trials started after the practice with the feedback indicators was terminated. They were divided into two conditions, a minimal-requirements condition and an optimum–maximum condition. In the former condition, the subject was presented with two horizontal bars that crossed the path of the two arrows; the subject was told to operate so that the arrows would be at the height of the bars or higher, namely, *at least* as fast as indicated by the arrows. The subject's score, which was later used to calculate the payoff, was proportional to the difference between actual and required performance in *only* the task that was performed worse than the other, relative to the requirement. The requirements were set at the (1) mean minus .5 *SD*, (2) mean minus .25 *SD*, (3) mean, (4) mean plus .25 *SD*, and (5) mean plus .5 *SD*. In the minimal-requirements conditions, the requirements on each task were inversely related to the requirements on the other task—that is, they could be in-

---

terpreted as relative priorities on the two tasks. Each pair of requirements was administered in a block of 24 practice trials and 48 test trials. On the digits task, order of requirements was from Requirements 1 to 5 for half of the subjects, and from 5 to 1 for the other half. In the optimum–maximum condition, one of the above requirements was set in turn for one of the tasks, and the subject was to try to operate as much as possible *at* the level of performance required. He/she was told that scoring would be maximal on the task when performance was precisely at the required level and that it would decline when performance was *either* better *or* worse. The feedback indicator for the task was an arrowhead without an arrow line. There was no requirement on the other task, and the instruction was to maximize the speed on that task. No requirement bar was displayed for that task, only a feedback indicator was presented. The subject was told that scoring on the task would be simply proportional to the speed on the task. There were 10 requirement blocks in this condition, each consisting of 16 practice trials and 32 test trials. For half of the subjects, the digits task was the to-be-optimized task in the *first* five blocks; for the remaining subjects, the digits task was the to-be-optimized task in the *last* five blocks. The order of requirements for the to-be-optimized task was increasing Requirements 1 to 5 for half of the subjects, and decreasing 5 to 1 for the remaining half. The order of the two conditions was also counterbalanced across subjects. The subjects were given the specific instructions for the conditions only immediately before the administration.

Stimuli. The frames measured 2 × 3 cm on the screen. The vertical height of a digit or a letter was about 2 cm. The horizontal space between the frames measured 1 cm. The bases of the arrows were 2.5 cm above the top of the frames, and they corresponded to a latency that was 2 standard deviations above the mean. The length of the arrows was 2 cm per standard deviation below that latency. The digits were presented within the left frame; the letters were presented within the right one.

Subjects. Sixteen subjects were used. All were students of the University of Haifa who participated in the experiment for payment that depended linearly on their performance score.

Results and Discussion

Analyses were conducted on latencies to the responses that were correct on the first response attempt. Two types of within-subject ANOVAs were performed on the data: one in which the factors were condition and requirement on the letters task (henceforth called *letter-manipulation* analysis), and one in which the factors were condition and requirement on the digits task (henceforth called *digit-manipulation* analysis). That is, in each of the analyses all data from the minimal-requirements condition were analyzed, but only half of the data from the optimum–maximum condition. Analyses were performed on both the digits task and the letters task.

The condition factor was not found significant in any of the analyses, although in the two analyses of the digits-task performance it approached significance [$F(1,15) = 4.14, p > .05$, for digits, and $F(1,15) = 1.39, p > .10$, for letters, in the digit-manipulation analysis; and $F(1,15) = 4.25, p > .05$, for digits, and $F(1,15) = 1.72, p > .10$, for letters, in the letter-manipulation analysis]. The requirement factor was found significant only in the performance of the letters task [$F(4,60) = 1.2, p > .10$, for digits, and $F(4,60) = 3.67, p < .01$, for letters, in the digit-manipulation analysis; and $F(4,60) < 1$, for digits, and $F(4,60) = 2.81, p < .05$, for letters, in the letter-manipulation analysis]. The interaction was not found significant in either of the analyses [$F(4,60) < 1$, for both digits and letters, in the digit-manipulation anal-

ysis; and $F(4,60) = 2.13, p > .05$, for digits, and $F(4,60) = 1.60, p > .10$, for letters, in the letter-manipulation analysis].

Thus, the letters task, but not the digits task, was sensitive to the requirements, regardless of the condition. In separate analyses, it was found that, in the minimal-requirements condition, there was a significant effect of the requirement on the letters task [$F(4,60) = 2.83, p < .05$], but not on the digits task [$F(4,60) = 1.37, p > .10$]. In the optimum–maximum condition, no significant effect was found [$F(4,60) = 1.13, p > .10$, for digits, and $F(4,60) = 1.99, p > .10$, for letters, in the digit-manipulation analysis; and $F(4,60) < 1$, for digits, and $F(4,60) = 1.15, p > .10$, for letters, in the letter-manipulation analysis].

In sum, although the performance of the subjects in the minimal-requirements condition conformed partly to the requirements, the subjects in the optimum–maximum condition did not respond to the requirement, not even on the task on which the requirements were set. These results do not replicate those obtained by Navon (in press). The reason might be that a subject who works for payment adopts strategies that look more financially promising than the strategy suggested in the instructions. For example, the subjects might have reasoned that the marginal utility resulting from staying within the requirement limits in the optimum–maximum condition was not worth the effort. A similar motivation presumably did not exist in the experiment reported by Navon (in press), so that subjects there were more willing to follow the instructions. It is important to note that this experiment fails to replicate the one reported by Navon (in press), *not* in that performance tradeoff is more robust here, but rather in that it does not exist at all.

EXPERIMENT 2

The performance tradeoff that was obtained by Navon (in press) could be attributed to the subjects' need to perform the two tasks, or parts of them, in a serial fashion. Experiment 2 was undertaken to determine whether or not the findings could be replicated when one of the tasks imposed a continuous load on the subject. For this purpose, the digit-classification task was replaced by a task of memory for digits. Performance decrements due to concurrent memory load are documented in the literature (e.g., Logan, 1979).

Method

The method was similar to that of Experiment 1 except for the following points. At the beginning of each trial, four digits were presented sequentially to the subject, each for 200 msec, with intervals of 300 msec each. One second after the presentation of the last digit, a letter was presented and the subject was to respond to it as in Experiment 1. One second after the correct response, a digit was presented. On half of the trials, the digit had appeared in the memory set, and, on the other half, it had not. The subject had to indicate whether or not the digit belonged to the memory set. The numbers of trials were 192 in the practice and calibration block, 48 in the feedback-indicators practice block, 16 practice

trials and 32 test trials for each requirement in the minimal-requirements block, and 12 practice blocks and 24 test trials for each requirement in the optimum–maximum condition. The experiment lasted nearly 2 h. The subjects were 16 students of the University of Haifa who participated in the experiment as a part of their course requirements in the Department of Psychology.

Results and Discussion

As in Experiment 1, analyses were conducted on latencies to the responses that were correct on the first response attempt. Two types of within-subject ANOVAs were carried out on the data: a letter- and a digit-manipulation analysis.

The condition factor was not found significant in either of the analyses [$F(1,15) = 2.78, p > .10$, for digits, and $F(1,15) = 1.50, p > .10$, for letters, in the digit-manipulation analysis; and $F(1,15) = 2.57, p > .10$, for digits, and $F(1,15) = 1.11, p > .10$, for letters, in the letter-manipulation analysis]. The requirement factor was found significant in only one of the analyses [$F(4,60) = 1.58, p > .10$, for digits, and $F(4,60) = 1.74, p > .10$, for letters, in the digit-manipulation analysis; and $F(4,60) < 1$, for digits, and $F(4,60) = 4.48, p < .01$, for letters, in the letter-manipulation analysis]. The interaction was found significant in only one of the analyses [$F(4,60) = 1.42, p > .10$, for digits, and $F(4,60) = 3.18, p < .025$, for letters, in the digit-manipulation analysis; and $F(4,60) < 1$, for both letters and digits, in the letter-manipulation analysis].

Thus, the subjects' performance of the digits task is not adjusted to task requirements in a significant way. Perhaps this follows in part from the inability of the subjects to control sufficiently the latency to a single stimulus and in part from the limited sensitivity of the speed of responding at test time to what happens during storage time. On the other hand, the letters task was found more sensitive to requirements. To pinpoint the locus of the effect, separate analyses were performed for each of the conditions. It was found that in the minimal-requirements condition, the requirement had a significant effect on performance of the letters task [$F(4,60) = 3.02, p < .25$], but not on performance of the digits task [$F(4,60) < 1$]. In the optimum–maximum condition, the picture is more complicated. The requirement clearly does not affect the concurrent task [$F(4,60) = 1.38, p > .10$, for letters, in the digit-manipulation analysis; and $F(4,60) = 1.20$,

$p > .10$, for digits, in the letter-manipulation analysis]. As for the task on which the requirements are presented, although the requirement effect is not significant [$F(4,60) = 1.94, p = .11$, for digits, in the digit-manipulation analysis; and $F(4,60) = 1.87, p = .13$, for letters, in the letter-manipulation analysis], when latencies longer than 2,500 msec are trimmed, the effects become significant [$F(4,60) = 4.08, p < .01$; and $F(4,60) = 2.92, p < .05$, respectively] (and the effect on the letters task is ordered exactly as prescribed by the requirements). The results are not conclusive, yet they suggest that, although the subjects tried fairly successfully to adjust their performance to the requirements in the optimum–maximum task, that had no effect on the performance of the concurrent task.

In sum, performance tradeoff was not observed in either of the conditions. Navon (in press) reported absence of tradeoff only in the optimum–maximum condition. It is possible that the tradeoff he did obtain resulted from sequential processing of the tasks, or parts of them. In this experiment, in which the responses did not overlap temporally, there was no way that the subject could control the order of processing, so that tradeoff was not manifested.

GENERAL DISCUSSION

The hypothesis tested by these experiments is that performance tradeoff is an artifact of demand characteristics. In Experiment 1, no tradeoff was observed at all, and the subjects were not responding to the requirements. It is possible that when subjects are highly motivated to perform well, they do not conform to demand characteristics, an explanation that is consistent with the hypothesis. However, it is not less reasonable to attribute the discrepancy to factors, such as difference in subject population, that are irrelevant to the hypothesis. Because, in Experiment 2, the subjects did try to adjust their performance to the requirements, the lack of tradeoff there is more indicative of the absence of a real limit on joint performance.

REFERENCES

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception & Performance*, 5, 189-207.

Navon, D. (in press). Do people allocate limited processing resources among concurrent activities? In H. Rachlin and L. Green (Eds.), *Advances in behavioral economics* (Vol. 2). Norwood, NJ: Ablex.