

Effects of Training with Added Difficulties on RADAR Detection

MICHAEL D. YOUNG¹, ALICE F. HEALY^{1*}, CLEOTILDE GONZALEZ²,
VARUN DUTT² and LYLE E. BOURNE Jr.¹

¹Department of Psychology and Neuroscience, University of Colorado, USA

²Dynamic Decision Making Laboratory, Social and Decision Sciences Department, Carnegie Mellon University, USA

Summary: Three experiments simulating military RADAR detection addressed a training difficulty hypothesis (training with difficulty promotes superior later testing performance) and a procedural reinstatement hypothesis (test performance improves when training conditions match test conditions). Training and testing were separated by 1 week. Participants detected targets (either alphanumeric characters or vehicle pictures) occurring among distractors. Two secondary tasks were used to increase difficulty (a concurrent, irrelevant tone-counting task and a sequential, relevant action-firing response). In Experiment 1, involving alphanumeric targets with rapid displays, tone counting during training degraded test performance. In Experiment 2, involving vehicle targets with both sources of difficulty and slower presentation times, training under relevant difficulty aided test accuracy. In Experiment 3, involving vehicle targets and action firing with slow presentation times, test accuracy tended to be worst when neither training nor testing involved difficult conditions. These results show boundary conditions for the training difficulty and procedural reinstatement hypotheses. Copyright © 2010 John Wiley & Sons, Ltd.

The goal of the present research is to examine the effects on visual search and decision-making performance of increasing difficulty by adding secondary tasks during training and testing. With practice on a visual search paradigm (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), automatic (rapid, often involuntary, requiring little attention) responses develop when items are consistently targets or distractors, called a consistent mapping (CM) condition. In contrast, when a given item might be a target on one trial and a distractor on another, a condition called varied mapping (VM), processing slows down, and task performance might diminish, even with extended practice. If another task is added, to be performed in parallel to visual search as would often occur in everyday work situations (e.g. in operating aircraft when pilots search for landmarks while answering questions from controllers) or other skilful operations (e.g. playing two-handed piano pieces while searching sheet music for changes in dynamics), the development of accurate detection responses is likely to require even more practice than it would in the VM condition alone.

Adding a secondary task to the primary visual search paradigm during training could have at least four different consequences at test: (1) The difficulty created by a secondary task could siphon cognitive resources from the primary visual search task. Under this assumption, adding a secondary task at training might prevent participants from fully learning the primary task during a fixed number of training trials so that test performance would suffer accordingly. This possibility would seem to be especially likely when the secondary task is irrelevant to the primary task and the two tasks occur concurrently. (2) The difficulty created by a secondary task could force the participants to engage in more effortful, elaborate or extensive processing of the primary task. Under this assumption, adding a secondary task

might actually enhance the learning of the primary task during training so that test performance would benefit accordingly. (3) The difficulty created by a secondary task coupled with a primary task at training could allow participants to learn how to cope with that same difficulty when the secondary and primary tasks are re-encountered together subsequently. Under this assumption, adding a secondary task at training might enhance test performance but only when the test involves the same secondary task. (4) The difficulty created by a secondary task could allow participants to overcome the boredom or task disengagement that might accrue during routine primary task training, especially when there are long periods of inactivity. Under this assumption, adding a secondary task at training might enhance test performance but only when the test conditions are not captivating.

These alternative possibilities for the effect of adding a secondary task to the primary task during training on primary task performance at test can be related to previous theoretical accounts in the literature. The first alternative corresponds to the logic originally outlined by Posner and Boies (1971) for primary and secondary task interrelations, according to which the two tasks compete for cognitive resources (see Pashler & Johnston, 1998, for a review of divided attention theory). The second alternative corresponds to the *training difficulty* hypothesis (Schneider, Healy, & Bourne, 2002; Young, Healy, & Bourne, 2005), by which test performance should be optimal when training is conducted under difficult conditions, regardless of the match between training and test, but only when the difficulties are 'desirable' (Bjork, 1994; McDaniel & Einstein, 2005). The third alternative corresponds to the *procedural reinstatement* hypothesis (Healy, Wohldmann, & Bourne, 2005), by which performance should be best at test when the procedures learned during training are repeated during testing. By this hypothesis, test performance should be optimal when the secondary task conditions at test match those at training. These last two hypotheses partially conflict because training under difficult conditions (i.e. with a secondary task) should aid subsequent

*Correspondence to: Dr. Alice F. Healy, Department of Psychology and Neuroscience, University of Colorado, Boulder, CO 80309-0345, USA.
E-mail: Alice.Healy@Colorado.edu

test performance under all conditions by the difficulty of training hypothesis but only under difficult testing conditions by the procedural reinstatement hypothesis. The type of difficulty (e.g. irrelevant concurrent or relevant sequential) might also have an important impact on performance, if an irrelevant task during training is less helpful to eventual test performance than a relevant task during training, and if a concurrent task is more distracting than a sequential task, especially when it employs the same cognitive resources (e.g. Navon & Gopher, 1979; Wickens & Kessel, 1980). That is, all types of difficulties are not necessarily equivalent in their effectiveness or their desirability for training, and an irrelevant concurrent task is more likely to be undesirable than a relevant sequential task. The fourth alternative corresponds to the *cognitive antidote* hypothesis (Kole, Healy, & Bourne, 2008), by which adding cognitive complications to a routine task serves to overcome a decline in accuracy that occurs with prolonged work on a simple primary task as participants become bored, fatigued or disengaged (see Hockey & Earle, 2006, for a discussion of the effects of effort and mental fatigue on performance).

To discriminate among these alternative potential effects of adding a secondary task to the primary task during training on primary task performance at test, we used the RADAR visual search and decision making paradigm developed by Gonzalez and Thomas (2008). This task requires dynamic detection and decision making under complex conditions that approximate those occurring outside the laboratory in military settings. Specifically, we examined the effects of adding a secondary task to the primary task of searching for targets during both CM and VM trials. In addition, we varied the processing load, namely whether 1 or 4 possible targets had to be held in memory (i.e. were in the memory set) coupled with whether there were 1 or 4 filled-in locations in the search space. In Experiment 1, the effects of a concurrent irrelevant secondary task (counting the number of tones that deviate from a standard tone) were examined in all four mapping and processing load combinations during both training and test, using digits and letters as targets and distractors. In Experiments 2 and 3, to make the task more realistic, digits and letters were replaced with nine unique coloured drawings of vehicles. Only the most difficult mapping and processing load combination, VM high load, was employed.

In addition to the concurrent, irrelevant tone-counting secondary task, a sequential, relevant action-firing secondary task was varied across both training and test phases in Experiments 2 and 3. Specifically after detecting a target, half of the participants in each tone-counting condition made an immediate, multi-step firing decision. The response was chosen on the basis of examining two pairs of threat level gauges. Participants were required to complete both target response and firing response within a fixed time period. Experiment 2 examined both irrelevant and relevant secondary tasks, whereas Experiment 3 examined only the relevant task.

EXPERIMENT 1

The aim of this experiment was to examine the development of target identification and decision making skill in a

complex RADAR task, with the additional subtask of counting tones (Gonzalez & Thomas, 2008). According to the training difficulty hypothesis formulated to account for training effects in foreign vocabulary acquisition (Schneider *et al.*, 2002; Young *et al.*, 2005), increasing the amount of effort during training (by adding an additional parallel task) should initially impede performance and acquisition of the primary skill of detecting targets. However, these more strenuous processing demands should eventually lead to a more durable skill than training without a parallel task, and thus participants should show better retention after a 1-week delay. Moreover, because of the clear distinction between targets and distractors provided on CM trials (targets were always letters and distractors were always digits or targets were always digits and distractors were always letters), with practice, participants should be less affected by an increase in processing load (due to both the number of targets in memory and the number of filled-in locations) compared to VM trials (e.g. Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The gap in performance for high processing load should be evident in added response time (RT) costs as well as in depressed accuracy of detection. With the superimposed concurrent task of tone counting, these costs of processing load might be exaggerated at both training and testing.

Method

Participants

Forty-eight undergraduate students participated in this experiment, divided into four equal groups with respect to their tone-counting condition according to a fixed rotation based on their time of arrival at training: 12 participants performed the tone-counting task in both training and testing sessions; 12 participants performed tone-counting in the first session only; 12 participants performed tone-counting in the second session only; and 12 participants never performed the tone-counting task. Additionally, half of the participants in each tone-counting condition had digits as their targets on CM trials, and the other half had letters as targets on CM trials. Introductory psychology course credit was awarded at the completion of two experimental sessions (training and test), each lasting approximately 2.5 hours, occurring between 6–8 days apart.

Stimuli and materials

We used the RADAR target detection and memory task and a tone task developed in past research. The full version of these tasks was described in an earlier article (Gonzalez & Thomas, 2008), and thus, here we focus on only those parts that are relevant for the present study.

Nine consonant letters (C, D, F, G, H, J, K, L, M, always in upper case) and nine single digits (1, 2, 3, 4, 5, 6, 7, 8, 9) were used as targets and distractors. When digits were targets in the CM condition, all other stimuli were letters (i.e. distractors in the CM condition and both targets and distractors in the VM condition). The opposite was the case when letters were targets in the CM condition. Thus, for a given participant, CM targets never occurred as distractors even on VM trials, so that the participant could develop automatic detection responses for the CM targets (e.g. Shiffrin & Schneider, 1977). In the VM high

load condition, four items (either letters or digits) were used as targets in each trial out of the total of nine possible items (either letters or digits) and the remaining five items served as distractors on such trials (employing all nine available items from the set). In both low load (one target) and high load (four targets) trials, a maximum of five distractors was presented. Targets comprising the memory set were presented at the beginning of the trial in a random order.

The tone-counting task participants monitored three types of tones varying in pitch—low (950 Hz), middle (1000 Hz) and high (1050 Hz)—and were asked to count the low- and high-pitched tones. Tones were played through computer speakers at irregular intervals (between 500–1500 milliseconds apart) during the trial on which the participant's main responsibility was target detection. A middle-pitched tone that could be ignored played while the memory set was initially displayed. Once the spacebar was pressed and until the end of the trial, 0–8 deviant (i.e. high- or low-pitched) tones were played in the background, along with the more frequent middle-pitched tone. A final count of deviant tones, the sum of both the high- and low-pitched tones, was prompted at the end of the trial. To insure that participants could discriminate among the three tone levels, all participants performing the tone-counting task were exposed pre-experimentally to the three tone types *via* a calibration procedure that lasted 1 minute, during which the spacebar was pressed each time a deviant tone sounded. During the 60 seconds of the calibration exercise, the participants heard tones, of which 16% were deviant (high or low) tones. Participants were allowed to proceed to the main task with an error rate on the calibration procedure under 20%. (No participants in this experiment or in Experiment 2 were omitted for failure to pass the calibration procedure, showing all had adequate discrimination abilities.)

Procedure and design

Both the training and test sessions consisted of 8 blocks of 20 trials. Thus, there was a total of 160 trials. Each trial was divisible into seven frames, or different successive screen presentations. The middle 5 frames contained (at most) one item from the memory set and the 5 characters used as distractors within the 4 RADAR locations (blips) that appeared in each frame. The first and last frame contained empty blips. In low processing load trials, only 1 of the 4 blips was filled per frame, and there was only 1 target in the memory set; in high processing load trials, all 4 blips were filled, and there were 4 targets in the memory set. For both processing loads, a target randomly appeared in at most 1 of the middle 5 frames (i.e. Frames 2–6). The block ordering for both experimental sessions was constant: (a) 20 trials of low processing load (1 target in the memory set with 1 target or distractor shown per frame) with CM (CM 1 + 1), (b) 20 trials of high processing load (4 targets in the memory set with 3–4 distractors plus 0–1 target(s) per frame) with CM (CM 4 + 4), (c) 20 trials of low processing load with VM (VM 1 + 1), (d) 20 trials of high processing load with VM (VM 4 + 4). These blocks were repeated in reverse order for Blocks 5–8, with a 5-minute intermediate break during which participants were allowed to leave the laboratory.

The memory set for a given trial was randomly selected from the appropriate target type (digits or letters). Each trial started with a presentation of the memory set displayed horizontally. The targets could be studied until the participant pressed the spacebar (occurring on average after 2676 milliseconds). At that time, the tone-counting task began, with deviant tones appearing during the following 7 frames of the trial. Participants were instructed that only one target should be expected on each trial for the duration of the experiment and to press the spacebar once to detect the target while the target was still on the screen. All trials lasted 7 frames, regardless of when the target was detected. The interframe interval was 1500 milliseconds. In each frame, four blips originated from the four principal corners of the RADAR screen (NE, NW, SE, SW), and converged uniformly to the centre of the RADAR screen (see Figure 1, top panel), then dissolved after 2062 milliseconds of being displayed (see Gonzalez & Thomas, 2008). After the seventh frame, the deviant tones stopped and the middle-pitched tone continued. Next to the RADAR screen, a window prompted the participant to make a quiet airspace report by pressing the spacebar to indicate that no targets were detected, which occurred in 5 out of the 20 trials per block. If a target was detected, no report was submitted, and the window vanished after a 2-second countdown. The number of deviant tones was reported in another window, appearing after the quiet airspace report. Points were awarded and deducted based on the speed and accuracy of target detection and on the accuracy of the tone count. A false alarm or a missed target resulted in a deduction of 1000 points, as did an incorrect deviant tone count. A correct quiet airspace report and perfect deviant tone count were worth 1000 points apiece. Based on the speed of a correct target detection, a maximum of roughly 700 points was awarded for responding at the start of a frame, with faster detections resulting in more points and with the points decreasing linearly as latency increased. Points were displayed as awarded during the trial for each component of the scoring and as total for the trial, as well as separate scores for the block (20 trials), and an overall score from the start until the finish of the experimental session (160 trials). However, the purpose of the scoring was to motivate participants and to provide them with performance awareness. None of the scores from any of the components were used in the analyses.

In summary this experiment used a mixed factorial design including the between-participants variables of training condition (silence, tone), testing condition (silence, tone), and type of target during CM trials (letters, digits) and the within-participants variables of mapping (CM, VM), processing load (1 + 1, 4 + 4) and session (training, test). Two dependent variables were examined here and in the subsequent experiments: Hit rate and correct RT.

Results

Mixed factorial analyses of variance (ANOVAs) were conducted to examine the effects of mapping (CM, VM), processing load (1 + 1, 4 + 4), and their interaction, separately for training and test as well as for training and test combined. The factor of training condition (silence, tone) was included in all three analyses, and the factor of

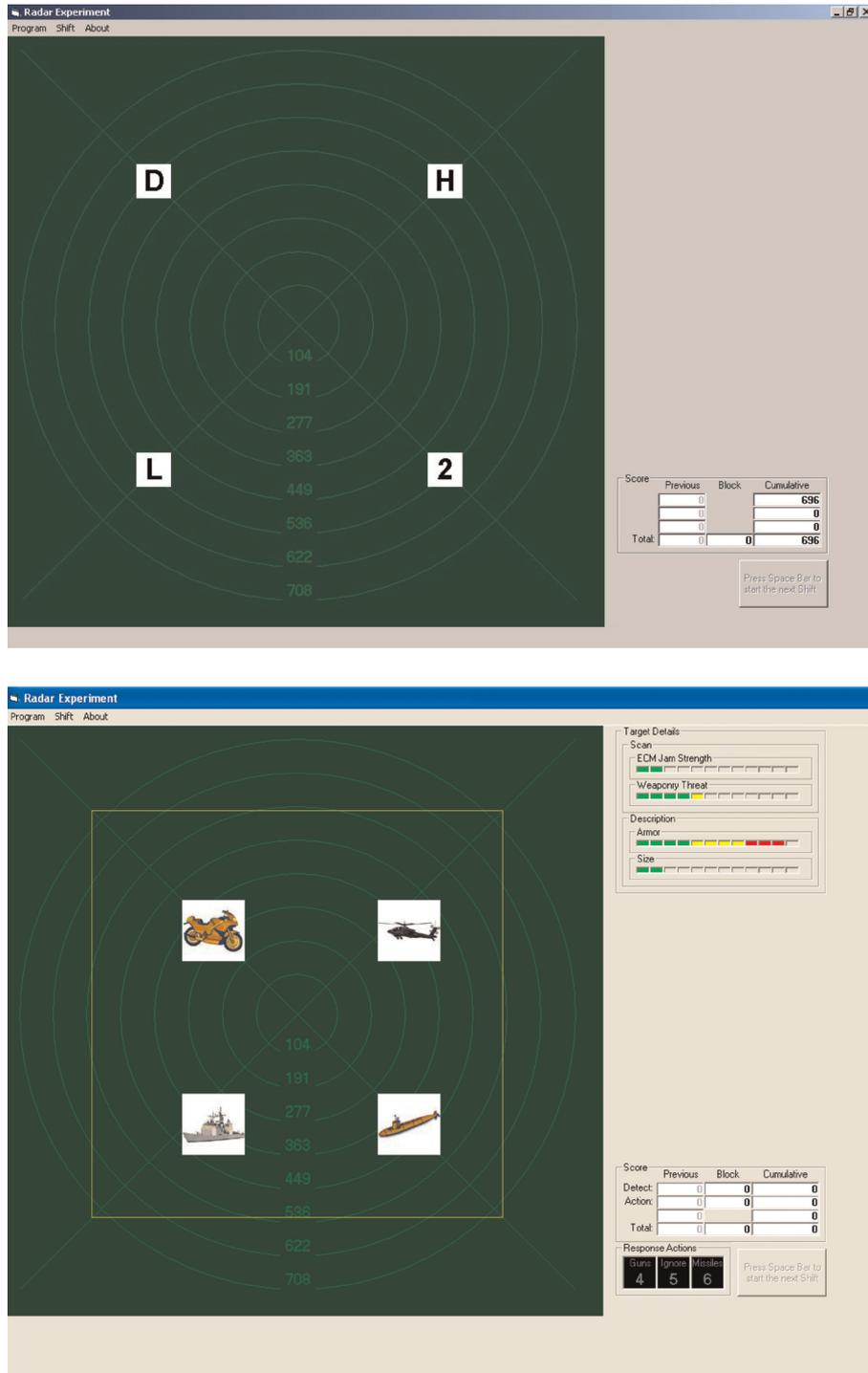


Figure 1. Sample display showing four filled-in blips in Experiment 1 (top panel) and in Experiments 2 and 3 (bottom panel)

testing condition (silence, tone) was also included in the analysis of test and in the combined analysis of training and test. In addition, the factor of type of target during CM trials (letters or digits) was included in all three analyses although in all cases throughout this experiment, the differences observed were not influenced by that factor. Finally, the factor of session (training, test) was included in the analysis combining the two sessions.

Training

A planned ANOVA on correct response time (RT) at training was conducted with the factors of mapping, processing load,

training condition, and target type. During the training session, mapping and processing load significantly impacted correct RTs for detection. CM trials ($M = 849.84$ milliseconds) were faster than VM trials ($M = 933.42$ milliseconds) ($F(1, 46) = 68.435, MSE = 9798.109, p < .0001, \eta^2 = .598$). Detecting targets in the low one-item processing load (1 + 1) was much faster ($M = 690.97$ milliseconds) than detecting targets in the high four-item processing load (4 + 4) ($M = 1092.29$ milliseconds) ($F(1, 46) = 1027.748, MSE = 15044.542, p < .0001, \eta^2 = .957$). The interaction of these two factors was also significant ($F(1, 46) = 148.224, MSE = 7335.373, p < .0001, \eta^2 = .763$). VM high processing

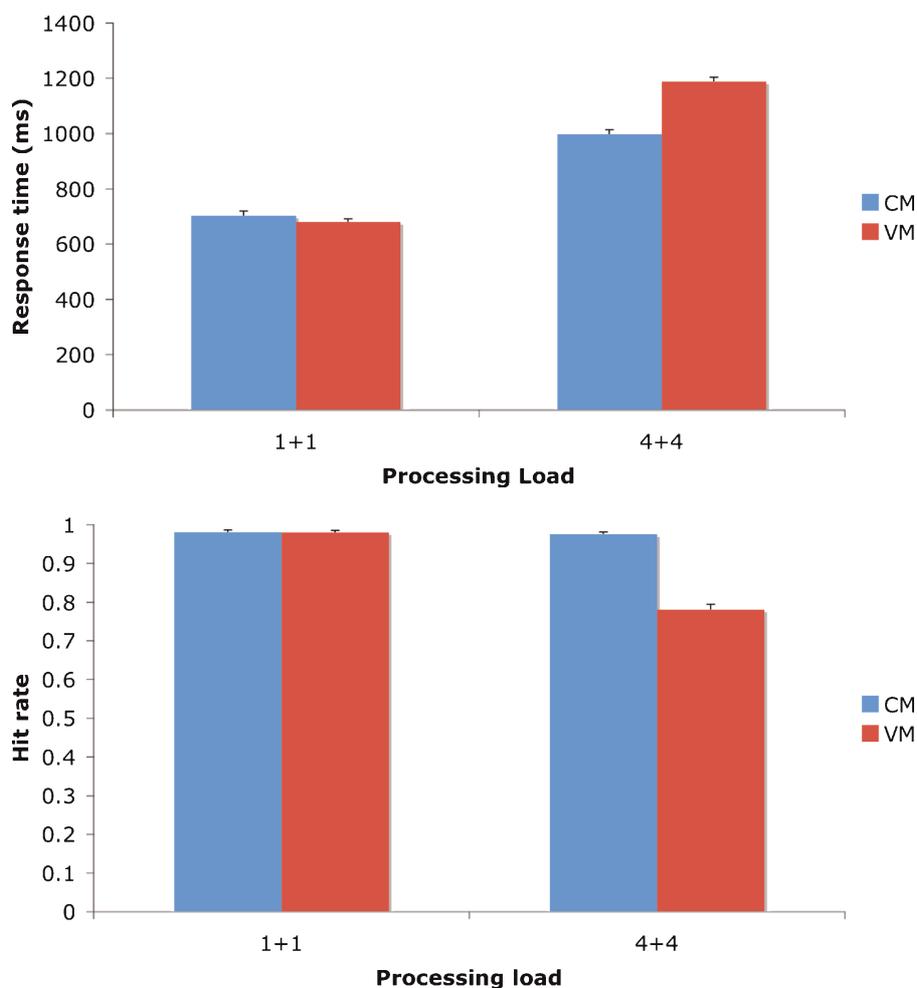


Figure 2. Correct response time (in milliseconds) (top panel) and hit rate (bottom panel) during training as a function of mapping and processing load in Experiment 1. 1 + 1 = low load; 4 + 4 = high load. Bars represent standard errors of the mean

load (4 + 4) trials were much slower than CM 4 + 4 trials, whereas CM 1 + 1 and VM 1 + 1 trials produced very similar correct RTs (see Figure 2, top panel). Also, in accordance with predictions based on the development of automaticity in CM but not VM trials (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), the effect of processing load was larger for VM than CM. Participants trained without the tone-counting task were faster ($M = 827.38$ milliseconds) than participants with the additional subtask ($M = 955.87$ milliseconds) ($F(1, 46) = 23.165$, $MSE = 68422.014$, $p < .0001$, $\eta^2 = .335$).

The same planned ANOVA was conducted on hit rate at training, the correct identification of a target from the memory set during the frame in which the target was presented.¹ Mapping and processing load effects, as well as the interaction of those two factors, occurred for hit rate. Participants hit more CM targets ($M = .978$) than VM targets ($M = .880$) ($F(1, 46) = 151.563$, $MSE = .006$, $p < .0001$, $\eta^2 = .767$), and also more targets in the 1 + 1 load ($M = .981$) compared to the 4 + 4 load ($M = .877$) ($F(1, 46) = 132.323$, $MSE = .008$, $p < .0001$, $\eta^2 = .742$). The interaction of processing load and mapping showed a very reduced hit rate for VM 4 + 4 trials

compared to near-ceiling performance for the other three trial types (see Figure 2, bottom panel) ($F(1, 46) = 155.850$, $MSE = .006$, $p < .0001$, $\eta^2 = .772$). Thus, as with correct RT, and consistent with the development of automaticity in CM but not VM, the effect of processing load was evident for VM but not for CM. Tone counting during training had no significant effect on hit rate in this session.

Test

A planned ANOVA with the factors of mapping, processing load, training condition, testing condition and target type was conducted on correct RT at test. In the test session, 1 week later, the processing load and mapping effects observed with correct RTs during training were replicated, as well as the interaction (see Figure 3, top left panel) (all $ps < .0001$). Participating in the tone-counting task during the test ($M = 927.61$ milliseconds) also slowed down RTs in this session compared to testing without tone counting ($M = 834.23$ milliseconds) ($F(1,44) = 6.493$, $MSE = 128910.978$, $p = .0144$, $\eta^2 = .129$). Tone counting during training had no overall effect on RT performance during test and did not interact with tone counting during test; thus, contrary to the difficulty of training and procedural reinstatement hypotheses, training with tone counting did not facilitate performance at test even when the test

¹We do not present the data for false alarm rates here or in the result sections of the other experiments because the multiple frames within each trial make this measure difficult to interpret.

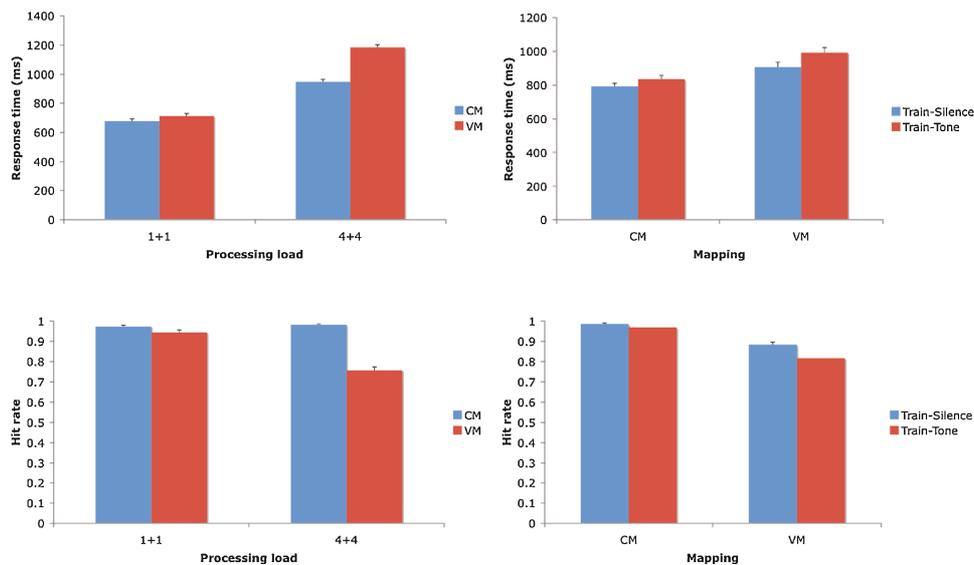


Figure 3. Correct response time (in milliseconds) (top panels) and hit rate (bottom panels) at test as a function of mapping and processing load (left panels) and as a function of mapping and training condition (right panels) in Experiment 1. 1 + 1 = low load; 4 + 4 = high load. Bars represent standard errors of the mean

also involved tone counting (training with tone and testing with tone $M = 962.71$ milliseconds, tone-silence $M = 862.68$ milliseconds, silence-tone $M = 892.50$ milliseconds, silence-silence $M = 805.78$ milliseconds). Exposure to the tone-counting task during training did, however, interact with mapping during test, because training under silence led to better performance than training with tone counting for VM trials, but the advantage for training under silence was smaller for CM trials ($F(1, 44) = 6.350$, $MSE = 6614.864$, $p = .0154$, $\eta^2 = .126$) (see Figure 3, top right panel).

The same planned ANOVA was conducted on hit rate at test. The mapping and processing load effects (including the interaction) for hit rate during training were also found at test (see Figure 3, bottom left panel) (all $ps < .0001$). Participants who experienced the tone-counting task during training were negatively impacted in their hit rate at test (training with tone counting: $M = .893$, training with no tone counting (in silence): $M = .935$) ($F(1, 44) = 4.618$, $MSE = .036$, $p = .0372$, $\eta^2 = .095$). Stimulus mapping nearly interacted with training condition (see Figure 3, bottom right panel) ($F(1, 44) = 3.775$, $MSE = .016$, $p = .0584$, $\eta^2 = .079$). As with RT, there was a larger decrement in hit rate for training with tone counting on VM trials than on CM trials. The tone-counting task condition at test did not have an overall effect on hit rate.

Training and test

An overall ANOVA was conducted on correct RT at training and test with the factors of mapping, processing load, training condition, testing condition, target type and session. Across training and test, both the mapping and processing load effects varied significantly. The difference between CM and VM trials in correct RT was more pronounced during test (CM $M = 812.88$ milliseconds, VM $M = 948.96$ milliseconds) than during training (CM $M = 849.84$ milliseconds, VM $M = 933.42$ milliseconds) ($F(1, 44) = 20.849$, $MSE = 6348.258$, $p < .0001$, $\eta^2 = .322$). In contrast, the difference between 1 + 1 and 4 + 4 processing loads was smaller during test

(1 + 1 $M = 695.87$ milliseconds, 4 + 4 $M = 1065.97$ milliseconds) than during training (1 + 1 $M = 690.97$ ms, 4 + 4 $M = 1092.29$ milliseconds) ($F(1, 44) = 4.072$, $MSE = 11492.463$, $p = .0497$, $\eta^2 = .085$). The three-way interaction between tone counting at training, mapping and experimental session (training and test) was also significant ($F(1, 44) = 7.483$, $MSE = 6348.258$, $p = .0089$, $\eta^2 = .145$). Tone counting at training naturally increased correct RTs more during the training session itself than during the test session, with comparable deleterious effects in the training session for the VM (tone $M = 992.39$ milliseconds, silence $M = 874.44$ milliseconds) and CM (tone $M = 919.36$ milliseconds, silence $M = 780.32$ milliseconds) trials, but in the test session the deleterious effect of training with the tone was larger for the more difficult VM trials (tone $M = 991.20$ milliseconds, silence $M = 906.73$ milliseconds) than for the easier CM trials (tone $M = 834.19$ milliseconds, silence $M = 791.56$ milliseconds).

The same overall ANOVA was conducted on hit rate at training and test. The hit rate for stimulus mapping also varied across training and test. Hit rate for CM trials remained the same across sessions (both $M = .978$), but VM trials had a lower hit rate during test ($M = .850$) than at training ($M = .880$) ($F(1, 44) = 5.593$, $MSE = .007$, $p = .0225$, $\eta^2 = .113$).

Discussion

Acquisition of the primary task of target detection was accomplished in the predicted manner: Responses were faster when targets and distractors came from different alphanumeric categories (CM, so there was a consistency in the definition of the targets across trials) than when they came from the same category (VM, so there was inconsistency in the definition of the targets across trials), and were faster at the lower processing load (1 + 1) than at the higher processing load (4 + 4). The disadvantage for the high processing load was larger for the VM condition than

for the CM condition, and this interaction was evident at both training and testing. Accuracy in terms of hit rate showed the same interaction; hit rate was lowest for the VM condition with the high processing load. This pattern of effects is consistent with the pattern predicted on the basis of the assumption that CM conditions can lead to automatic responding, which should not be affected by processing load. In contrast, only controlled, not automatic, responding is possible under VM conditions, so that processing load should diminish performance in that case (e.g. Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). In the present experiment, training was presumably not sufficient to allow complete automaticity to develop, and complete automaticity might never be possible, at least with the stimuli presented so far apart when the frame begins that they cannot all be seen together in the center of the visual field. However, participants should have at least become more automatic (i.e. more highly skilled) under CM than under VM conditions, and thus the effect of processing load should have been smaller under CM than under VM, as was in fact observed.

According to the training difficulty hypothesis (e.g. Bjork, 1994; Schneider et al., 2002), the secondary task should interfere with performance during training but promote high performance at test. In fact, adding a parallel secondary task during training was not an effective means of encouraging fast and accurate performance of the primary task at test. All participants who experienced the secondary task were negatively impacted by it in both sessions, as would be expected on the basis of the logic originally put forth by Posner and Boies (1971), by which a secondary task siphons cognitive resources from the primary task. Not only would cognitive resources be overworked by an additional task, but dividing attention between two tasks might increase the overall complexity of the combined tasks, making the difficulty of parallel tasks more than just the sum of the difficulties of the two component tasks (Salthouse, 1982). Indeed, there was no benefit at test for training in the difficult tone-counting condition, especially on VM trials, even when testing also involved that condition, contrary to the procedural reinstatement principle (e.g. Healy et al., 2005). This finding suggests that it might be important to master the primary task before beneficial effects from adding a secondary task can be observed. Even though the secondary task's cognitive demands were focused on auditory resources, whereas those of the primary task were focused on visual resources, this separation of modalities was not sufficient to elicit the kind of concentration and effort that would benefit eventual performance of the primary task. Rather, the addition of the secondary task seems to have essentially diverted attention from the primary task, thereby preventing adequate acquisition of that task. Perhaps if the secondary task were made less likely to divert attention from the primary task, there would be a different pattern of results. That possibility is explored in Experiment 2.

EXPERIMENT 2

In Experiment 1, when the tone-counting task was performed concurrently with visual search, it initially slowed down

visual search performance, and this difficulty during training impeded performance at test, even under tone counting and especially on the VM trials. These findings are in direct opposition to the training difficulty hypothesis (e.g. Schneider et al., 2002). However, it is possible that there were no benefits of training with difficulty because the secondary task was both concurrent with (thereby distracting the participants) and irrelevant to the primary task. Hence, in the present experiment the irrelevant concurrent secondary task of tone counting was compared to a relevant secondary task that was sequential (thus less likely to divert attention from the primary task), namely the action-firing task after target detection. Under this secondary task, after detecting a target, the participants made an immediate, multi-step firing response, which was determined first by examining two 'scan gauges' and then by examining two 'description gauges'. Both the target response and the firing action response were limited to a fixed time limit of 4.124 seconds, twice the limit used in Experiment 1.²

The two secondary tasks, tone counting and action firing, were crossed both in training and at test. The design ensured that for each secondary task, half of the participants in each training condition were exposed to the same condition at test as at training (e.g. tone in both sessions), whereas the other half were in different conditions at training and test (e.g. tone at training and silent at test). With this design, we can determine whether the two types of difficulty (concurrent, irrelevant difficulty created by the tone task and sequential, relevant difficulty created by the action-firing task) have differing impacts on training, retention and transfer of the primary visual search task.

Because the largest effects of the tone task in Experiment 1 occurred under the VM condition, Experiment 2 used only that condition. Experiment 2 was also restricted to the more difficult high processing load. In addition, to make the search task more realistic with respect to situations occurring naturally outside the laboratory (i.e. more ecologically valid), we changed the target stimuli from alphanumeric characters to vehicle images. During each trial, participants were presented with a set of four vehicle targets to detect among vehicle distractors, with four filled-in blips in every display (see Figure 1, bottom panel).

Method

Participants

Data were collected from 32 undergraduate students during two 40-minute experimental sessions, training and test. Training and testing with the tone-counting task and the action-firing response was counterbalanced during both sessions using a fixed rotation based on time of arrival at training, yielding 16 (i.e. 2⁴) possible participant conditions. There were two participants in each of these 16 conditions. A 6–8 day retention interval occurred between training and test. Most participants completed the second session exactly

²We doubled the frame time on the basis of a pilot experiment that required the action-firing response within the original frame time. Participants were unable to perform both detection and action-firing tasks within the time allotted; indeed about half of the participants in the pilot experiment were unable to provide a single correct detection response with these time constraints.

1 week following the first session, but some leeway was allowed in this and the other experiments to accommodate variation in personal schedules. Following the experiment, participants were granted introductory psychology course credit.

Stimuli and materials

Participants responded to pictures of vehicles during this experiment, rather than letters and digits. There was one small colour picture (74×73 pixels) for each of the following nine vehicle types: airplane, helicopter, humvee (military 4-wheeler), jeep, motorcycle, ship, submarine, tank and train car. Each vehicle appeared as both target and distractor across trials (the mapping was VM only). As each trial featured four enemy target vehicles, the remaining five vehicles were included as distractors. When presented in the memory set, the vehicles appeared in random order.

The stimuli for the tone-counting task in Experiment 1 were reused in this experiment.

Procedure and design

There were two blocks of 20 trials in training and test. Each block consisted of high load (4 + 4) VM trials, with 15 of the 20 trials from each block containing a single target. Each trial contained 7 frames (frame time 4124 milliseconds—twice the frame time of Experiment 1), with the middle 5 frames (Frames 2–6) containing the target and distractors. The first and last frames were blank.

A set of four randomized vehicle targets comprising the memory set preceded each trial. Pressing the spacebar initiated the frame sequence as well as the deviant tones for participants also responsible with providing a deviant tone count at the end of the trial. All participants pressed the spacebar when 1 of the 4 targets appeared on the screen, or waited until the end of the trial to submit a quiet airspace report. The deviant tone count was also reported at the end of the trial. For half of the participants during training and test (counterbalanced with the tone counting task), an additional subtask was performed while the target was still on the screen, the action-firing response, which was relevant to the primary detection task. Immediately upon pressing the spacebar, two pairs of threat level gauges appeared to the right of the RADAR screen in a space labelled 'target details'. The four target details were subdivided into two scan gauges (labelled 'ECM jam strength' and 'weaponry threat') on the top and two description gauges (labelled 'armor' and 'size') on the bottom. The four individual gauges each had readings from 1–12 units (coloured green at low levels, yellow at middle levels, and red at high levels). The action-firing response choices were ignore (5 key), guns (4 key) or missiles (6 key), in increasing order of severity. Gauge readings were randomized for each trial and type of target (e.g. the tank could not be expected to have the same gauge reading for each of its appearances). To make the correct action-firing response, the top scan gauges were consulted first. If less than half of the available gauge units were filled in across the two scan gauges, the ignore response should be selected, regardless of the description gauge readings. If more than half of the scan gauges were filled in, the bottom description gauges had to be examined.

High description gauge readings, along with sufficiently high scan readings, prompted a missile response, whereas low description gauge readings required a guns response. Points were also awarded (for participant motivation) for the speed (roughly 0–700 points) and accuracy (1000/–1000 points for correct/incorrect response) of the action-firing response.

In summary this experiment used a mixed factorial design including the between-participants variables of tone training condition (silence, tone), firing training condition (fire, no fire), tone testing condition (silence, tone) and firing testing condition (fire, no fire) and the single within-participants variable of session (training, test).

Results

Training

A planned ANOVA on correct RT at training was conducted with the factors of tone training condition and firing training condition. Presence of the task-relevant (action-firing response) and the task-irrelevant (tone counting) additional subtasks negatively impacted participants' ability to quickly detect 1 of the 4 items from the memory set. Counting tones during training was more harmful on correct RT for detection ($M = 2159.16$ milliseconds) than training in silence ($M = 1914.39$ milliseconds) ($F(1, 28) = 4.762$, $MSE = 201294.812$, $p = .0376$, $\eta^2 = .145$). Making an additional action-firing response ($M = 2184.37$ milliseconds) also added to correct detection RT compared to only performing the detection task ($M = 1889.18$ milliseconds) ($F(1, 28) = 6.926$, $MSE = 201294.812$, $p = .0137$, $\eta^2 = .198$).

The same planned ANOVA was conducted on hit rate at training. Hit rates were also significantly lower during training when the action-firing response was added ($F(1, 28) = 14.889$, $MSE = .031$, $p = .0006$, $\eta^2 = .347$). Without the action-firing response, the hit rate was .89, compared to .72 when the subtask was required.

Test

A planned ANOVA on correct RT at test was conducted with the factors of tone training condition, firing training condition, tone testing condition and firing testing condition. For the second session, only the task-irrelevant additional subtask (tone counting) had an overall effect on correct detection RT. Testing in silence was faster ($M = 1828.12$ milliseconds) than was testing with the tone-counting task ($M = 2156.59$ milliseconds) ($F(1, 16) = 4.966$, $MSE = 347599.386$, $p = .0405$, $\eta^2 = .237$). Exposure to the task-relevant action-firing subtask during either training or test did not have an overall effect on RT at test. There was, however, a three-way interaction of firing condition at training, tone counting condition at training, and tone counting condition at test ($F(1, 16) = 5.346$, $MSE = 347599.386$, $p = .0344$, $\eta^2 = .250$). Training with firing and without the tone-counting task then switching to tone counting at test resulted in much longer responses compared with when the same training conditions were tested in silence. In general, testing with tone counting led to longer RTs than testing in silence for all training condition combinations, except training with both tone counting and firing (see Figure 4). This interaction should be interpreted with some caution because of

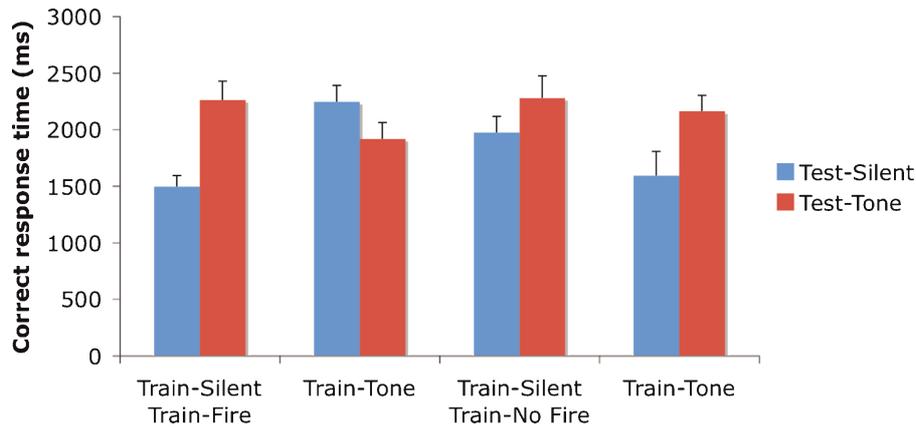


Figure 4. Correct response time (in milliseconds) at test as a function of firing condition at training, tone condition at training and tone condition at test in Experiment 2. Bars represent standard errors of the mean

the relatively small number of participants in each combination of conditions.

The same planned ANOVA was conducted on hit rate at test. Making action-firing responses at test did significantly lower hit rate ($F(1, 16) = 31.696, MSE = .034, p < .0001, \eta^2 = .665$). Without the action-firing response, the hit rate at test was .90, whereas with the action-firing response, the hit rate was .64.

Training and test

An overall mixed factorial ANOVA on hit rate at training and test was conducted with the factors of tone training condition, firing training condition and session. Session interacted with training with the action-firing response ($F(1, 28) = 14.299, MSE = .029, p = .0008, \eta^2 = .338$). After an advantage during training without the action-firing response (hit rate without action-firing response $M = .89$; hit rate with action-firing response $M = .72$), this training effect reversed during test, as training with the action-firing response resulted in a higher hit rate ($M = .80$) than without it ($M = .74$). This advantage for training with the action-firing response on test hit rate was due exclusively to testing with the action-firing response (see Figure 5), although the interaction of training and testing action-firing condition was not significant on hit rate at test. Also, the tone-counting task

at training yielded no significant interactions on the hit rate during training and test.

Discussion

Tone counting led to a significant increase in correct RT during both training and testing in Experiment 2, as it had in Experiment 1. In Experiment 1, however, tone counting during training led to worse test performance in terms of both correct RT and hit rate than did silence during training, especially in the VM condition. In contrast, in Experiment 2, tone counting when coupled with the action-firing response during training removed the increase in correct RT at test due to tone counting. Thus, tone counting during training hurt test performance in Experiment 1 but helped test performance in Experiment 2 when it was paired with the action-firing response during training, with respect to the deleterious effects on RT of having the tone counting task at test. We speculate that tone counting during training had a benefit in Experiment 2 not seen in Experiment 1 because of the increased frame time in Experiment 2 (in order to accommodate the sequential action-firing response; see Footnote 2). Another relevant difference between Experiments 1 and 2 is that the targets in Experiment 2 were of a different type and from a different category (i.e. pictures of vehicles) than those used in Experiment 1 (i.e.



Figure 5. Hit rate as a function of session, firing condition at training, and firing condition at test in Experiments 2 and 3. Bars represent standard errors of the mean

alphanumeric characters). Tone counting in Experiment 1 during training might have been especially disruptive because the numerical stimuli used might have been confused in working memory with the numbers in the running tone count.

The relevant secondary task (action firing) used in Experiment 2, like tone counting, yielded significant interference during training, in terms of both hit rate and correct RT. At test, there was no significant overall decrement due to action firing on correct RT, but there was a large overall decrement in hit rate. Most importantly, training with the new, relevant secondary task hurt performance during acquisition but had a long-term benefit on testing, where the hit rate was higher for those trained with the action-firing response than for those trained without it, apparently though only when testing also involved the action-firing response. This result is consistent with both the training difficulty hypothesis (e.g. Bjork, 1994; Schneider *et al.*, 2002) and the procedural reinstatement hypothesis (Healy *et al.*, 2005) because by both accounts training with difficulty should enhance long-term retention and transfer at least when the same difficulty occurs during testing as during training. Because of the advantage at test for training with the relevant sequential difficulty in Experiment 2, Experiment 3 focuses exclusively on the difficulty created by the action-firing response to provide further tests of the training difficulty and procedural reinstatement hypotheses.

EXPERIMENT 3

To gain an understanding of the scanning processes that apply to both the RADAR display and the four-item memory set, we manipulated orthogonally the location of the target on the screen and its placement in the memory set. These locations had not been controlled in the previous experiments. In Experiment 3, each of the four positions in the memory set was crossed with each of the four quadrants of the display so that each of the 16 combinations of memory set and display locations was used once in each of the 20 trials within a block, with the remaining 4 trials containing no targets. If participants scan the memory set in a fixed order from first to last (i.e. in serial self-terminating order; e.g. Sternberg, 1966; Taylor, 1976, for classic discussion of alternative scanning strategies), then we should find a significant effect on correct RT of memory set position, with the responses fastest to the first position and slowest to the fourth position. Likewise, if participants search the quadrants in a fixed order, then we should find a significant effect on correct RT of location on the RADAR screen. In particular, if participants scan the RADAR screen in the same direction as they read text, we would expect correct RTs to slow down from top to bottom and from left to right. Alternatively, participants might be biased to the right side of the RADAR screen because that side is closer to the threat level gauges used in the action-firing response.

Training with the action-firing task benefited hit rate at test in Experiment 2, at least when testing also involved action firing, supporting both the training difficulty and procedural reinstatement hypotheses. To provide additional verification

of these hypotheses, Experiment 3 included again the secondary task of action firing, which was manipulated in the same way as in Experiment 2. The tone counting task was not included in Experiment 3.

Method

Participants

Thirty-two participants, undergraduate students, were included, divided into four equal groups using a fixed rotation based on the time of arrival at training: Eight participants performed the action-firing response during training and test; eight participants performed the action firing response during training only; eight participants performed the action firing response during test only and eight participants were never exposed to the action firing response. The training and test experimental sessions (40 minutes each) were spaced 6–8 days apart. Introductory psychology course credit was given following the completion of the test session.

Stimuli and materials

The nine vehicle stimuli from Experiment 2 were reused. The position of the target was varied along two dimensions within each block, unlike the previous experiments. The first dimension was the target's position within the 4-vehicle memory set. Within a block, targets appeared equally often from the (proceeding from left to right) first, second, third and fourth positions in the memory set in each block. The second dimension varied in this experiment was the target's position on the RADAR screen. Within a block, targets appeared equally often in the four principle corners of the screen: NE, NW, SE and SW. These two constraints required that 16 of the 20 trials per block contain a target, rather than the 15 employed in the previous experiments. The tone counting task was discontinued for all participants.

Procedure and design

The procedure followed the one used in Experiment 2: There were two blocks of 20 high load VM trials per session, with each trial beginning with a randomized memory set of four vehicles. The frame length remained 4124 milliseconds.

In summary this experiment used a mixed factorial design including the between-participants variables of training condition (fire, no fire) and testing condition (fire, no fire) and the within-participants variables of position in memory set (first, second, third, fourth), direction (NE, NW, SE, SW) and session (training, test).

Results

Training

Mixed factorial ANOVAs were conducted on both correct RT and hit rate at training with the factors of training condition, position in memory set and direction. Location on the RADAR screen and position within the memory set influenced correct RTs. Targets originating in the upper-left (NW) corner of the screen were detected the fastest ($M = 1599.95$ milliseconds) among the four directions (NE $M = 1756.81$ milliseconds; SE $M = 1926.81$ milliseconds; SW $M = 1911.59$ milliseconds) ($F(3, 90) = 6.305$, $MSE = 238238.414$, $p = .0006$, $\eta^2 = .174$).

The first position in the memory set (left-most in presentation at the start of the trial, $M = 1668.43$ milliseconds) was also the position with the fastest average correct RT ($F(3, 90) = 3.475$, $MSE = 137165.728$, $p = .0193$, $\eta^2 = .104$) (2nd position: 1852.72 milliseconds, 3rd position: 1806.88 milliseconds, 4th position: 1848.45 milliseconds). The action-firing response condition did not affect correct RT and did not interact with either the location on the RADAR screen or position within the memory set. Hit rate was not affected by the action-firing response or by either position within the memory set or target location.

Test

Mixed factorial ANOVAs were conducted on hit rate and correct RT at test with the factors of training condition, testing condition, position in memory set and direction. The interaction between training condition and testing condition approached significance on hit rate at test ($F(1, 28) = 2.949$, $MSE = .019$, $p = .0970$, $\eta^2 = .095$) (see Figure 5). The interaction reflected a marked disadvantage for not training or testing with the action-firing response (i.e. the least difficult training and testing conditions).

Only the direction effect on correct RT found at training was replicated during the test ($F(3, 84) = 15.738$, $MSE = 163632.065$, $p < .0001$, $\eta^2 = .360$). The NW direction was the fastest, as scanning occurred, on average, left to right, then top to bottom, starting from this direction (see Figure 6, top panel). The effect of position within the memory set for correct RT only approached significance ($F(3, 84) = 2.380$, $MSE = 155331.109$, $p = .0754$, $\eta^2 = .078$)

(see Figure 6, bottom panel). There was however a significant memory set position effect on hit rate during test ($F(3, 84) = 3.585$, $MSE = .041$, $p = .0171$, $\eta^2 = .114$). The item in the second memory position had the highest hit rate, whereas the item in the fourth position had the lowest hit rate (1st position: .91, 2nd position: .95, 3rd position: .87, 4th position: .83). Hit rate was also influenced by direction during test ($F(3, 84) = 4.051$, $MSE = .030$, $p = .0097$, $\eta^2 = .126$). The NW direction not only had the fastest overall correct RT during training and during the test, but also the highest hit rate at test (NW: .95, NE: .92, SW: .86, SE: .86).

Discussion

The results of Experiments 2 and 3 are inconsistent with respect to their implications for the two training hypotheses. In accordance with the training difficulty hypothesis (e.g. Bjork, 1994; Schneider et al., 2002), difficult training (with the firing decision) did promote better test performance in terms of increased hits in Experiment 2. Experiment 2 also provided some support for the procedural reinstatement hypothesis (e.g. Healy et al., 2005), because at test under the firing condition, hit rate was especially high when training also involved firing. There was an interaction between training and testing condition that approached significance in Experiment 3 on hit rate at test. That interaction shows some benefit for difficult training. However, it is not fully consistent with either of the training hypotheses because performance was worst when participants both trained and

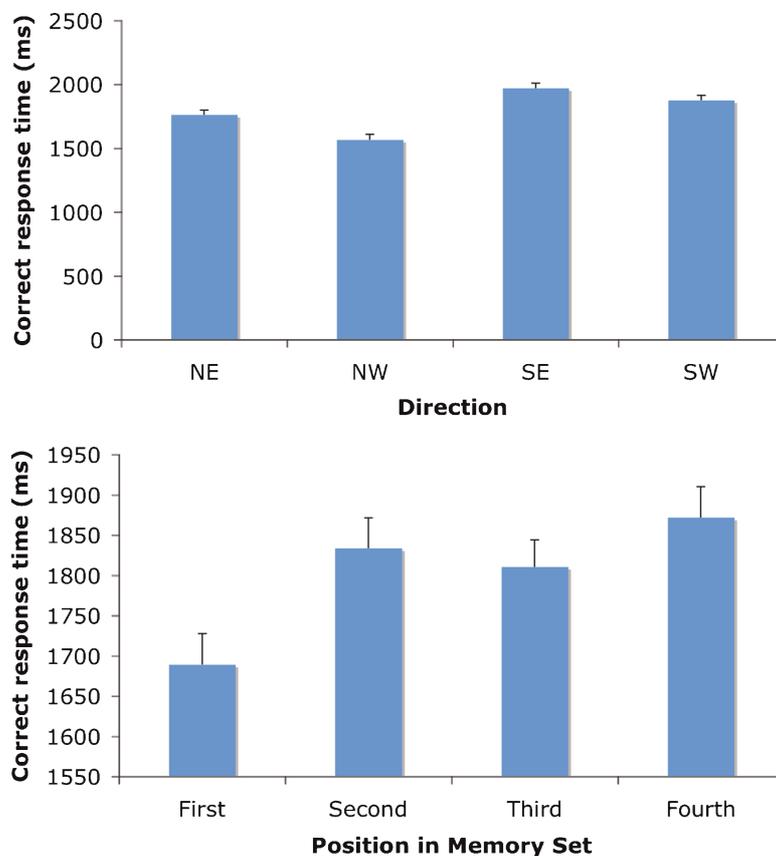


Figure 6. Correct response time at test as a function of direction (top panel) and as a function of position in the memory set (bottom panel) in Experiment 3. Bars represent standard errors of the mean

tested with no firing response. Possibly, as outlined in the Introduction, the explanation for this finding involves task disengagement on the part of the participants, none of whom were required to do tone counting in this experiment, which involved a longer frame time than in Experiment 1. The firing response, which is cognitively demanding and relevant to the primary task, presumably served to overcome any task disengagement, in accordance with the cognitive antidote hypothesis by Kole *et al.* (2008) for the simple data entry task. Participants who were not required to make an action-firing response (or count deviant tones) might have become bored or disengaged from the task during the course of the experimental session because they only responded with a simple key press to targets on one of the 7 frames within 16 of the 20 trials of each of the two experimental blocks, and they should have had sufficient time to see each of the four stimuli in each frame.

In Experiment 3, the pattern of correct detection RT as a function of direction implies that searching the RADAR screen proceeds in a left-to-right and top-to-bottom fashion. Moreover, the correct detection RT pattern as a function of memory set position implies that the memory set is searched starting with the first position. Thus, the RT results suggest that participants engage both visual and memory serial scanning in a consistent, systematic order (*cf.* Sternberg, 1966; Taylor, 1976).

GENERAL DISCUSSION

Two somewhat conflicting training hypotheses were tested in three experiments using the RADAR detection and decision task as primary and adding two types of secondary tasks, one a concurrent irrelevant task (tone counting) and the other a sequential relevant task (action firing). According to the procedural reinstatement hypothesis (Healy *et al.*, 2005), performance at test should be best when the testing conditions match the training conditions. In contrast, according to the training difficulty hypothesis (Schneider *et al.*, 2002), performance at test should be best when training conditions are difficult. The support for these hypotheses was mixed. In Experiment 1, difficult training led to worst performance at test, rather than improved performance, even when testing involved the same difficulty, which is contrary to both hypotheses. The difficulty in Experiment 1 involved the concurrent irrelevant tone-counting task, which is, thus, clearly not a 'desirable' difficulty with respect to eventual test performance (Bjork, 1994; McDaniel & Einstein, 2005). In Experiment 2, training with the same secondary tone-counting task also led to longer correct RTs at test. However, when that irrelevant concurrent secondary task was paired with the relevant sequential secondary task of action firing at training, test performance with the tone counting task was improved, in agreement with both hypotheses. Specifically, testing with tone counting resulted in longer RTs than testing in silence in all cases except when training involved both tone counting and action firing. In addition, training with the sequential relevant action-firing response hurt both accuracy and correct RT during acquisition but, again in agreement with the two training hypotheses, led to improved hit rate at

test, at least when the test also involved the action-firing response. In contrast to both previous experiments, Experiment 3 found that hit rate at test tended to be lowest when training and testing both involved no secondary task (in this case, action-firing). This finding for the testing conditions involving no action firing is consistent with the training difficulty hypothesis, under the assumption that action firing is a 'desirable' form of difficulty, but not with the procedural reinstatement hypothesis. However, a simpler explanation might invoke task disengagement at test because there are relatively long inter-response delays that required no cognitive processing so that participants might have become bored by the undemanding nature of the task. In particular, participants might have used the frequent interruptions with no required processing to daydream or focus on non-task goals or events outside the laboratory.

These experiments, thus, provide boundary conditions for the operation of the difficulty of training and procedural reinstatement hypotheses. Desirable difficulties at training not only depend on the difficulty's relevance to the task, but also on which cognitive resources are required to handle them. In Experiment 1, in which digits and letters were searched, the tone-counting task introduced undesirable difficulty because subvocal rehearsal of the memory set (1 or 4 letters or digits) presumably required the same cognitive resource (*i.e.* the phonological loop of working memory, *e.g.* Baddeley, 1992) that was employed with the updating of the tone count. Switching between rehearsing the memory set and updating the tone count, particularly during VM trials, can be exceedingly difficult because of the potential confusion between the digits appearing as stimuli and those that are sub-vocalized during the counting process.

In Experiments 2 and 3, the targets were changed from letters and digits to pictures of vehicles, which could be remembered as images, rather than exclusively by the names of the images. In other words, when separate cognitive resources can keep the memory set and tone count on-line simultaneously (*e.g.* when the memory set uses resources devoted to visual processing and the tone count uses resources devoted to auditory processing), positive effects of training with the tone-counting task are more likely than when cognitive resources are shared (*e.g.* Sanders & Schroots, 1969, with respect to the separation of digits and tones into different cognitive categories in a short-term recall task). This interpretation is more consistent with a 'capacity sharing' rather than a 'bottleneck' view of divided attention (Pashler & Johnston, 1998). Although both tasks created demands on the same memory pool, these demands could be efficiently divided (especially on modality grounds), but at a cost to the speed of processing for each task. The running tone count did not need to be inhibited in order to keep the memory set on-line (bottlenecking), but instead the memories for both tasks were maintained in parallel (capacity sharing). This logic implies that one of the boundary conditions for the operation of the difficulty of training hypothesis (*i.e.* for determining whether a specific difficulty is desirable to add at training) concerns whether the added task shares or competes with the primary task for the same resources.

In any event, the combined results of the three experiments make it clear that whether adding a secondary task to training

can offer a desirable difficulty is not dependent on having extensive training. In Experiment 1, where the added difficulty did not benefit performance, training included considerably more trials than in Experiment 3, where the added difficulty did benefit performance. In fact, there were four times as many trials during training in Experiment 1 (160 trials) than in Experiment 3 (40 trials).

In especially simple tasks, adding sources of difficulty during training has been shown to be useful for keeping primary task performance high at test (Kole et al., 2008). In Experiment 2, adding the tone-counting task to the firing response during training offset the increase in correct RT attributed to the tone-counting task at test, a benefit not found in Experiment 1. Furthermore, requiring the firing response during training led to an increase in hit rate at test. Frame time was increased from Experiment 1 to Experiments 2 and 3, which should allow for both additional subtasks to be performed with better concentration. By optimally filling up the frame with attention-demanding tasks, the amount of inactivity was reduced, and thus participants were more likely to stay properly engaged on the primary task. Without an attention-demanding task, participants might lower their overall effort towards the primary task (Hockey & Earle, 2006). With another related task added to the primary task, the level of effort might be raised and, thus, be useful for keeping performance levels higher for both tasks than for the primary task alone. For simple tasks in which there is a substantial risk of task disengagement, an additional cognitively demanding subtask can be effective in maintaining higher levels of attention for the primary task, a finding consistent with the cognitive antidote hypothesis (Kole et al., 2008).

In any event, these experiments make it clear that in designing an effective training programme, the actual primary and secondary tasks to be trained need to be carefully chosen. The specific principles of training that apply depend on properties of the tasks and their interrelationships, especially whether the primary and secondary tasks use the same pool of cognitive resources (e.g. Navon & Gopher, 1979; Wickens & Kessel, 1980). Although the present experiments involved a specific militarily relevant task (RADAR detection), there are many other non-military real-life tasks that have similar requirements. For example, as mentioned in the Introduction, in operating aircraft pilots often search for landmarks while answering questions from controllers, and when playing two-handed piano pieces pianists often read notes on sheet music while searching for changes in dynamics beneath the notes. The principles of training illuminated in the present study presumably apply to these other real-life tasks as well as to the RADAR task studied here.

ACKNOWLEDGEMENTS

This work was supported by Army Research Institute Contract DASW01-03-K-0002 and Army Research Office Grant W911NF-05-1-0153 to the University of Colorado. The authors are grateful to Scott Holmstrom for help testing participants in Experiment 1 and Carolyn Buck-Gengler

for help in data analysis. Earlier versions of this work were reported at meetings by Young, Healy, Gonzalez and Bourne (2007) (Experiment 1) and by Young, Healy, Gonzalez, Dutt and Bourne (2008) (Experiments 2 and 3).

REFERENCES

- Baddeley, A. (1992). Working memory. *Science*, 255, 556–559.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Gonzalez, C., & Thomas, R. P. (2008). Effects of automatic detection on dynamic decision making. *Journal of Cognitive Engineering and Decision Making*, 2, 328–348.
- Healy, A. F., Wohldmann, E. L., & Bourne, L. E. Jr. (2005). The procedural reinstatement principle: Studies on training, retention, and transfer. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 59–71). Washington, DC: American Psychological Association.
- Hockey, G. R. J., & Earle, F. (2006). Control over the scheduling of simulated office work reduces the impact of workload on mental fatigue and task performance. *Journal of Experimental Psychology: Applied*, 12, 50–65.
- Kole, J. A., Healy, A. F., & Bourne, L. E. Jr. (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. *Applied Cognitive Psychology*, 22, 917–937.
- McDaniel, M. A., & Einstein, G. O. (2005). Material appropriate difficulty: A framework for determining when difficulty is desirable for improving learning. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 73–85). Washington, DC: APA.
- Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86, 214–255.
- Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention* (pp. 155–189). Hove, England: Erlbaum (UK).
- Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, 78, 391–408.
- Salthouse, T. A. (1982). *Adult cognition: An experimental psychology of human aging*. New York: Springer-Verlag.
- Sanders, A. F., & Schroots, J. J. (1969). Cognitive categories and memory span: III. Effects of similarity on recall. *Quarterly Journal of Experimental Psychology*, 21, 21–28.
- Schneider, V. I., Healy, A. F., & Bourne, L. E. Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, 46, 419–440.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review*, 84, 1–66.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127–190.
- Sternberg, S. (1966). High-speed scanning and human memory. *Science*, 153, 652–654.
- Taylor, D. A. (1976). Stage analysis of reaction time. *Psychological Bulletin*, 83, 161–191.
- Wickens, C. D., & Kessel, C. (1980). Processing resource demands of failure detection in dynamic systems. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 564–577.
- Young, M. D., Healy, A. F., & Bourne, L. E. Jr. (2005). Learning French pronoun-verb pairs with contextual interference. *Paper presented at the 75th Annual Convention of the Rocky Mountain Psychological Association*, Phoenix, AZ.
- Young, M. D., Healy, A. F., Gonzalez, C., & Bourne, L. E. Jr. (2007). The effects of training difficulty on RADAR detection. *Poster presented at the joint meeting of the Experimental Psychology Society and the Psychonomic Society*, Edinburgh, Scotland.
- Young, M. D., Healy, A. F., Gonzalez, C., Dutt, V., & Bourne, L. E. Jr. (2008). Effects of training with added relevant responses on RADAR detection. *Poster presented at the 49th Annual Meeting of the Psychonomic Society*, Chicago, IL.