

EFFECTS OF SENSITIVITY, CRITERION SHIFTS, AND SUBJECTIVE CONFIDENCE ON THE DEVELOPMENT OF AUTOMATICITY IN AIRLINE LUGGAGE SCREENING

Poornima Madhavan and Cleotilde Gonzalez
Dynamic Decision Making Laboratory
Department of Social and Decision Sciences
Carnegie Mellon University

We examined the effect of cognitive factors on the development of automaticity in a complex task. Participants ($n = 24$) performed a luggage screening task where we manipulated stimulus mapping (consistent vs. varied), frame size (small vs. large), memory set size (1 vs. 4) and time constraint, and examined their effects on participants' sensitivities, criterion shifts and confidence. Results revealed that the highest cognitive advantage in terms of high sensitivities and minimal deviations from optimal beta was afforded by the combination of small memory sets and consistent mapping of targets, after extended practice. Varied mapping of stimuli under high memory loads exerted a negative effect on sensitivities and induced a greater shift from optimal beta. The concurrence of high memory loads with varied mapping also led to decreases in confidence that hindered automatic detection of targets. The results have implications for training individuals to develop appropriate decision-making strategies in complex vigilance tasks.

INTRODUCTION

Several complex tasks in the world today ranging across fields such as aviation, military and healthcare require operators to develop highly skilled and automated levels of performance in response to critical stimuli. The development of automaticity is important as these environments are characterized by multiple stimuli and distractors, as well as environmental variables such as time pressure and workload that challenge performance in the absence of a practiced skill set. The present paper is focused on examining cognitive factors that affect the development of automaticity in a complex environment characterized by high demands for vigilance, varying levels of task difficulty, multiple targets and time constraints.

A classic finding from the dual process theory of automaticity (Schneider & Shiffrin, 1977) is that automatic processing develops when stimuli are consistently-mapped to target and distractor categories over a period of extended experience. On the other hand, under varied-mapping conditions wherein stimuli may be targets in one instance but distractors in another, performance occurs under controlled processing, which is voluntary, serial, and requires attention. Several researchers have attempted to adapt

Schneider and Shiffrin's automaticity paradigm in different contexts. These researchers have demonstrated that consistency in stimulus mapping combined with small frame size and memory set size significantly improves response time and accuracy in contexts ranging from simple target detection tasks (Fisk & Schneider, 1984; Kramer et al., 1991) to complex tasks involving multiple dynamic stimuli (Gonzalez et al., in review). Despite the clear relevance of the dual-process theory to complex tasks, the cognitive factors that mediate the effect of the above contextual factors on automatic decision processes have seldom been examined in these contexts.

The primary goal of this research was to examine whether 'automatic' detection of targets, defined in the context of luggage screening as the efficient matching and retrieval of instances from memory without heavy dependence on focused attention, is a feasible goal for training screeners in the real world. First, we examined how Schneider and Shiffrin's 'traditional' automaticity variables - stimulus mapping, frame size, memory set size and practice - influence participants' target detection sensitivity (d') leading to better discriminability between signal and noise. Next, we examined how these task characteristics lead to shifts in decision criterion

settings (beta) or the propensity to generate ‘target present’ vs. ‘target absent’ responses. Research has demonstrated that subjective confidence in decisions influences shifts in decision bias, ultimately affecting performance accuracy (Madhavan et al., 2006). Therefore, our third goal was to examine the manner in which confidence is affected by the automaticity variables mentioned above. In addition, we incorporated time constraint into our paradigm in order to better capture the high levels of workload inherent in a luggage screening task.

Based on Schneider and Shiffrin’s original findings, we hypothesized that the workload created due to large frame size, large memory set size and varied mapping of targets, combined with high time constraints will lead to (a) low detection sensitivity, (b) greater deviations from optimal beta, and (c) lower levels of confidence, relative to the effect of lower workload. In addition, we expected practice to lead to an increment in sensitivity, which in turn, will lead to proportionate increases in confidence.

METHOD

Participants

24 students from Carnegie Mellon University completed the experiment on two consecutive days. The duration of the experiment was approximately 2.5 hours on each day. Participants were paid a total of \$50 for their participation.

Tasks and Procedures

Participants completed 800 trials of a luggage screening task. On each trial, participants observed a two-color x-ray luggage-image cluttered with a variety of everyday objects (e.g., clothes, hair dryers, pill bottles) moving from one end of the screen to the other. Each image moved onto the screen from the left, froze for a brief duration in the center of the screen and then continued exiting the screen towards the right. Participants were assigned to one of two time constraint conditions: (i) in the ‘low time constraint’ condition ($n = 12$), the luggage image took 4 seconds to enter the screen, froze for 4 seconds and took 4 seconds to exit the screen; (ii) in the ‘high time constraint’ condition ($n = 12$), each of the above actions lasted for a duration of 2 seconds instead of 4, thereby increasing the degree of constraint on participants’ image-viewing duration.

The images were presented in 16 blocks of 50 trials each, constituting a training and test phase. At the beginning of each trial block, participants were asked to memorize a set of targets. Participants were required to make a decision about stopping the bag or not, according to their belief that a member of the memory set was hidden in the bag. After the decision, the image disappeared and participants rated their confidence on a scale of 1 (not confident at all) to 5 (extremely confident), followed by feedback. The trials performed on the first day was designated the ‘training phase’ and those performed on the second day were designated the ‘test phase’. The target probability was 40% i.e., 20 bags in every block of 50 bags contained a digitally superimposed weapon.

In addition to time constraint and practice, and in keeping with Schneider and Shiffrin’s (1977) original methodology, three within-subject independent variables were manipulated:

1) *Frame size (FS)*: One half of the images were ‘high clutter’ images while the other half was ‘low clutter’ images. The latter contained approximately 50% of the clutter of objects in the former.

2) *Memory set size (MSS)*: The number of objects in the memory set was 1 or 4.

3) *Stimulus mapping (SM)*: On consistent mapping (CM) trials, the task was to look specifically for certain weapons among miscellaneous objects. The target weapons on one CM block never appeared as distractors during other CM blocks. On varied-mapping (VM) trials, a weapon that was a target in one VM block appeared as a distractor in another VM block.

The measures used to examine the development of automaticity were: (1) target detection sensitivity (d'), (2) decision criterion setting (beta), and (3) subjective confidence in decisions.

RESULTS

Multiple ANOVAs revealed that sensitivity and confidence were higher, and beta was closer to optimal when FS (clutter) was low rather high. FS did not interact significantly with the other variables. Since highly cluttered bags are more typical of the real world than less cluttered luggage and the combination of a small MSS, CM, practice and low time constraint presented a more significant cognitive advantage on highly cluttered images, we present results only for the highly cluttered images below.

Sensitivity (d'). A 2 (MSS) X 2 (SM) X 2 (phase) X 2 (time constraint) mixed ANOVA on participants' sensitivities revealed significant main effects for MSS, $F(1, 22) = 46.17, p < .01$, SM, $F(1, 22) = 12.17, p < .01$, and, time constraint, $F(1, 22) = 9.14, p < .01$. Contrary to initial expectations, phase did not have a significant main effect on sensitivities, $F(1, 22) = .54, p = .32$. Results also revealed the following interactions: SM and time constraint, $F(1, 22) = 5.68, p < .05$, MSS and SM, $F(1, 22) = 269.5, p < .01$, SM and phase, $F(1, 22) = 4.51, p < .05$, and, a three-way interaction between MSS, SM and phase, $F(1, 22) = 8.65, p < .01$.

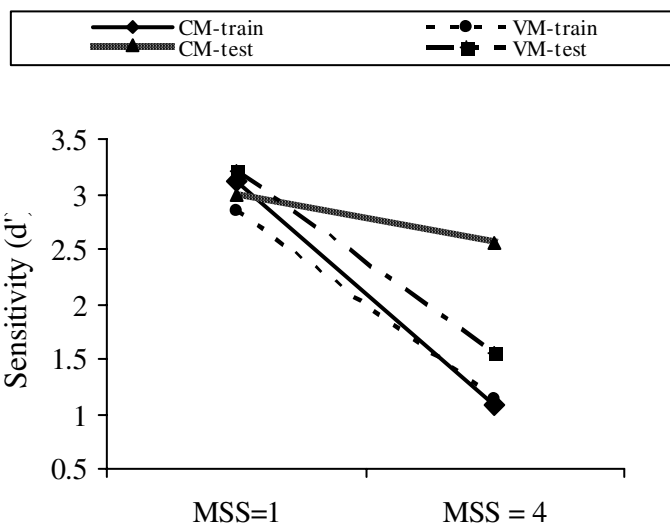


Figure 1. Sensitivities under high time constraint

When time constraint was low, sensitivity was higher on VM ($M = 2.66, SD = .51$) than on CM trials ($M = 2.35, SD = .40$), $t(11) = 2.63, p < .05$ across both MSSs and phases, which is contrary to expectations. An increase in time constraint led to a decrease in sensitivity across all trials, with a different pattern of sensitivity shifts emerging due to high levels of time constraint. Sensitivities of participants in the high time constraint condition are illustrated in Figure 1. As can be seen in the figure, during training, participants' demonstrated greater sensitivity on CM ($M = 3.11, SD = .67$) relative to VM trials ($M = 2.85, SD = .81$), $t(23) = 1.76, p < .05$, when MSS was small. However, an increase in MSS led to no significant differences between CM ($M = 1.09, SD = .66$) and VM trials ($M = 1.13, SD = .67$). Contrary to the pattern on training trials, on the test phase, sensitivity did not differ significantly on CM ($M = 3.0, SD = .95$) and VM trials ($M = 3.21, SD = 1.43$), $t(23) = 1.73, p = .11$, when MSS was small; when

MSS was large, sensitivity dropped on VM trials such that d' on VM ($M = 1.55, SD = .57$) was significantly lower than on CM trials ($M = 2.55, SD = .75$), $t(23) = 2.19, p < .05$.

Response criterion settings (beta). Given the signal to noise ratio, the optimal beta was 1.5. Most participants demonstrated an upward shift away from optimal beta, suggesting a trend toward conservative responding. A 2 (MSS) X 2 (SM) X 2 (phase) X 2 (time constraint) mixed ANOVA on response times revealed significant main effects for MSS, $F(1, 22) = 3.44, p < .05$, SM, $F(1, 22) = 18.26, p < .01$, and phase, $F(1, 22) = 10.92, p < .01$. Time constraint did not have a significant main effect on decision criterion settings, $F(1, 22) = 2.23, p = .56$. Results also revealed a significant interaction between MSS and phase, $F(1, 22) = 3.6, p < .05$, and a three-way interaction between MSS, SM and phase, $F(1, 22) = 8.51, p < .01$.

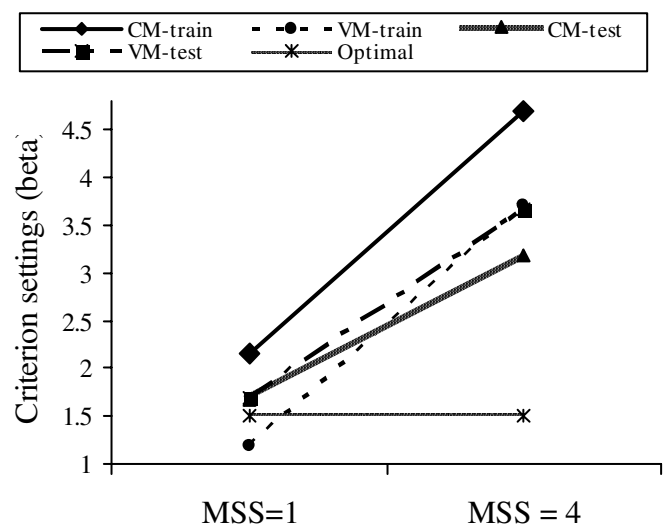


Figure 2. Response criterion settings

Since time constraint did not have a significant effect on response criteria, beta values averaged across both levels of time constraint are illustrated in Figure 2. As can be seen in the figure, all participants demonstrated significant upward deviations from optimal beta as a consequence of increases in workload (MSS). During training, criterion settings were liberal and closer to optimal beta, with no significant differences between CM ($M = 2.16, SD = 2.88$) and VM trials ($M = 1.18, SD = 1.42$), $t(23) = .50, p = .62$, when MSS was small; however, contrary to expectations, an increase in MSS led to a greater upward deviation from optimal beta on CM ($M = 4.70, SD = 3.25$) trials than on VM trials ($M = 3.70,$

SD = 3.57), $t(23) = 1.05, p < .05$. Similar to the training phase, response criteria on the test phase were closer to optimal beta, with no significant differences between CM ($M = 1.68, SD = 3.37$) and VM trials ($M = 1.7, SD = 3.34$), $t(23) = 1.06, p = .30$, when MSS was small. However, in keeping with expectations, participants demonstrated greater deviations from optimal beta on VM trials ($M = 3.66, SD = 4.95$) than on CM trials ($M = 3.19, SD = 3.3$), $t(23) = 1.87, p < .05$, when MSS was large.

Decision confidence. A 2 (MSS) X 2 (SM) X 2 (phase) X 2 (time constraint) mixed ANOVA on confidence estimates revealed significant main effects for MSS, $F(1, 22) = 27.25, p < .01$, SM, $F(1, 22) = 7.87, p < .05$, and time constraint, $F(1, 22) = 4.13, p < .05$. Contrary to expectations, practice (phase) did not have a significant main effect on decision confidence, $F(1, 22) = .57, p = .75$. Results also revealed significant interactions between MSS and SM, $F(1, 22) = 22.32, p < .01$, MSS and time constraint, $F(1, 22) = 4.00, p < .05$, and a three-way interaction between MSS, SM and phase, $F(1, 22) = 4.85, p < .05$.

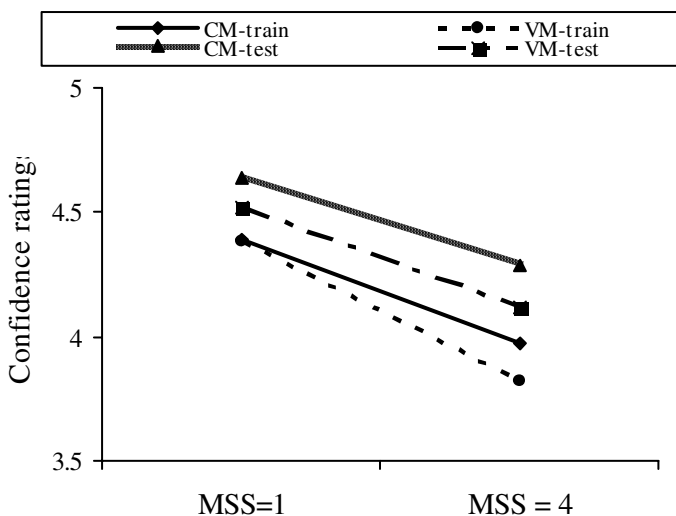


Figure 3. Confidence ratings

Similar to the analyses for criterion settings, confidence ratings averaged across both levels of time constraint are illustrated in Figure 3. As shown in the figure, confidence ratings on the training phase were not significantly different on CM ($M = 4.39, SD = .67$) and VM trials ($M = 4.38, SD = .52$), $t(23) = .07, p = .94$, when MSS was small. However, in keeping with expectations, an increase in MSS led to a greater drop in confidence on VM trials ($M = 3.82, SD = .86$)

relative to CM ($M = 3.97, SD = .87$) trials, $t(23) = 1.53, p = .06$. Contrary to the training phase, confidence estimates on the test phase conformed to the same pattern across both MSSs. Specifically, confidence was consistently higher on CM trials (small MSS: $M = 4.64, SD = .51$; large MSS: $M = 4.29, SD = .83$) than on VM trials (small MSS: $M = 4.52, SD = .57, t(23) = 2.50, p < .05$; large MSS: $M = 4.12, SD = .91, t(23) = 2.75, p < .05$).

DISCUSSION

We found the highest cognitive advantage in terms of high levels of sensitivity and minimal deviations from optimal beta when there was a small MSS, CM of targets and extended practice. VM exerted a negative effect on performance by reducing sensitivity and inducing a greater shift away from optimal beta under conditions of high memory load relative to CM conditions. Furthermore, the concurrence of low memory loads with CM led to increases in decision confidence that potentially influenced response patterns. This is consistent with previous research that varied-mapping of stimuli shows enormous effects of memory load leading to degradation in performance (Schneider & Shiffrin, 1977).

According to Schneider and Shiffrin (1977), automatic processing is stored in long-term memory and is triggered by the consistency of targets and responses. Following sufficient experience on a task, automatic detection of targets continues to occur independently of focused attention. While it has been established that automatic sequences of action do not require focused attention, the results of the present study demonstrate that the development of automaticity in this complex task occurred due to a combination of several factors. On the one hand, consistent mapping and low workload influenced operators' sensitivities or their abilities to distinguish the presence of a signal from background noise. In addition, contextual factors lead to psychological shifts in response bias (or the tendency to generate 'target present' vs. target absent' responses) and decision confidence. This suggests that the development of automaticity in complex tasks is the consequence of both perceptual shifts as well as cognitive changes which influence the efficacy of decision making during the course of a task.

Contrary to Schneider and Shiffrin's dual process theory that suggests the lack of susceptibility of CM

to workload, participants in the present study demonstrated a reduction in sensitivity, a shift away from optimal beta and a decrement in decision confidence even on trials in which stimuli were consistently mapped. One possible reason for this could be the lack of sufficient training provided to participants (50 trials per block) on each phase. Secondly, VM was conceptualized such that targets in one VM block appeared as distractors in another VM block, with targets and distractors not varying *within* the same block. Such a manipulation might have weakened the relative effects of CM vs. VM in the present context.

Conclusions and Implications

Despite the above design limitations, the results of the present study have implications for developing training solutions for security personnel in differentiating targets from non-targets within short time frames. As indicated by the results of the present study, methods of training that consistently map targets and responses, and impose low memory loads and low time constraints on the screener will likely lead to better acquisition of long-term detection skills than inconsistent mapping of targets and responses, large memory loads and high levels of time constraint. While the likelihood of objects in passenger luggage being consistently mapped in the real world is low, the results of the present study suggest that training interventions based on consistent mapping, with low workload are likely to be more effective in skill acquisition than methods based on varied mapping of stimuli with high levels of workload.

Whether the training interventions suggested above will indeed lead to better transfer of knowledge in the real world characterized by higher levels of workload and inconsistent mapping of stimuli remains to be examined. Furthermore, the development of automaticity in the context of luggage screening needs to be uni-directional, i.e., automatic detection is appropriate only in instances in which a target has been identified as potentially dangerous. It is possible

that less automatic (or controlled) processing might be more appropriate when there is ambiguity or uncertainty about the identity of a suspicious object. Therefore, further research is required before concrete suggestions can be made toward encouraging automaticity development in airline luggage screening.

ACKNOWLEDGEMENTS

This research was supported by the Multidisciplinary University Research Initiative Program (MURI), grant number N00014-01-1-0677. We thank Polina Vanyukov for helping in data collection, and Jack Lim for programming the tool used in this experiment.

REFERENCES

- Fisk, A. D., & Schneider, W. (1984). Consistent attending versus consistent responding in visual search: Task versus component consistency in automatic processing development. *Bulletin of the Psychonomic Society*, 22(4), 330-332.
- Gonzalez, C., Thomas, R.P., & Madhavan, P. (in review). The effects of stimulus and response mapping on the development of automaticity in complex environments. *Journal of Cognitive Engineering and Decision Making*.
- Kramer, A. F., Strayer, D. L., & Buckley, J. (1991). Task versus component consistency in the development of automatic processing: A psychophysiological assessment. *Psychophysiology*, 28(4), 425-437.
- Madhavan, P., Wiegmann, D.A., & Lacson, F.C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors*, 48 (2).
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review*, 84(1), 1-66.