METACOGNITIVE JUDGMENTS IN A SIMULATED LUGGAGE SCREENING TASK

Jason S. McCarley & Jessica Gosney
University of Illinois at Urbana-Champaign
Urbana, IL

A pair of experiments examined the accuracy and potential role of predictive metacognitive judgments in a simulated baggage screening task. Procedure was modeled after the ease-of-learning task (Underwood, 1966). Subjects searched for knives hidden in x-ray images of passenger bags. Before performing the search task, subjects viewed each stimulus without the embedded target and rated the likelihood of finding the target if it was hidden in that image. Experiment 1 used a 2IFC search task. Experiment 2 used a speeded yes-no search procedure. Results suggest that predictive metacognitive judgments are only modestly accurate, but that the information on which such judgments are based is nonetheless used to regulate search behavior.

INTRODUCTION

Successful performance in complex mental tasks often requires effective *metacognition*, awareness and understanding of the performer's own cognitive processes (Flavell, 1979; Nelson & Leonesio, 1988; Perfect & Schwarz, 2002). Theorists distinguish two classes of metacognitive function, *monitoring* and *control* (Nelson & Narens, 1990). The role of metacognitive monitoring is to assess the efficacy of one's own cognitive processes and strategies. The role of metacognitive control, in turn, is to modify the information processing strategies, where appropriate, in response to the output of monitoring functions. Studies have found both that metacognitive judgments can be predictive of task difficulty, and that such judgments can be used to guide effort allocation (e.g., Nelson & Leonesio, 1988).

Although metacognition is often studied using relatively simple tasks such as paired-associate (e.g., Nelson, et al, 1994) and consonant trigram learning (e.g., Nelson & Leonesio, 1988), metacognitive processes can also contribute to human performance in more complex tasks such as aviating (Valot, 2002) and driving (Lesch & Hancock, 2004). The present study examined the accuracy and potential influence of metacognitive judgments in another real-world task, baggage x-ray screening. Data indicate that baggage screening is a difficult task, such that even under near-ideal conditions (low time stress, low noise levels, low target uncertainty) performance is imperfect (McCarley, et al, 2004). The potential role of metacognition in screening, then, would be in allowing an operator to judge whether a given bag can be adequately processed based on a visual inspection alone, and if so, how much time/effort should allotted to inspecting the bag (i.e., setting the criterion for terminating search; Chun & Wolfe, 1996). The experiments reported here thus had two goals. The first was to assess the accuracy of subjects' predictive metacognitive judgments in a simulated luggage screening task. The second was to assess the relationship between such predictive judgments and the time allocated to searching various bags. The design of both experiments was modeled after the *ease-of-learning* (EOL) judgment task used for the study of metacognitive monitoring and control in memory tasks (Nelson & Leonesio, 1988; Underwood, 1966).

EXPERIMENT 1

Experiment 1 measured the correlation between subjects' predictive metacognitive judgments of search difficulty and observed target detection rates in a 2-interval forced choice (2IFC) search task. The 2IFC task was used because it provides a measure of sensitivity independent of yes-no response bias (Wickens, 2002).

Method

Subjects were 20 young adults. All were naïve to the purpose of the study. Stimuli were 69 pairs of sparsely to densely cluttered baggage x-rays. The images within each pair were identical except that one contained a hidden knife and the other did not. After a short introductory task, the experiment comprised two phases. In the first, subjects viewed 69 target absent images and were asked to predict the likelihood of finding a target knife if one were hidden in each bag. Each trial, a single stimulus image was presented for a duration of 2 seconds. Following stimulus presentation, the subject rated the likelihood of being able to find a target in that bag. Ratings were made on a 1-5 Likert scale using a response box.

In the second phase of the experiment, subjects searched for target knives hidden in the same bags that had been presented in the rating task. A 2IFC procedure was used. Each trial, a pair of matched target present and target absent images was presented sequentially. Images were presented for 2 seconds each, and separated by a 2 second ISI. After viewing the stimulus sequence, the subject provided a key press to indicate whether the target present image had appeared in the first or second stimulus interval, then provided a 1-5 rating of confidence. The first five trials were treated as practice, leaving 64 trials for analysis.

Results

Mean accuracy in the 2IFC search task was .769, *SE* = .017. This value was reliably greater than .5, the value corresponding to chance performance [$t(19) = 15.82$, $p < .001$ by one-sample t-test).

Goodman-Kruskal gamma correlation (*G*) values (Nelson, 1984) were used to assess the relationship between A) predictive judgments and observed target detection rates, and B) post-trial confidence ratings and observed target detection rates. Across subjects, mean *G* for predictive ratings was .161, *SE* = .054. This value, although statistically larger than 0 [$t(19) = 2.968$, $p = .008$], was small, indicating only a modest ability for participants to predict which bags would allow easy target detection and which would not. Mean *G* for post-trial confidence ratings was substantially higher *M* = .771, *SE* = .055 [$t(19) = 9.208$, $p < .001$ by paired-samples test], suggesting that the relatively low value of *G* for predictive ratings was not an artifact of the design (e.g., the number of trials or choice of rating scale).

An additional analysis was conducted to determine if subjects with high predictive metacognitive accuracy also performed better on the search task itself. Across subjects, Pearson product-moment correlation between predictive *G* and search accuracy was .37, a value that was not statistically significant, $p = .11$. There was thus no reliable relationship between metacognitive accuracy and performance on the search task.

EXPERIMENT 2

Experiment 2 measured the correlation between subjects' predictive judgments of search difficulty and observed response times (RTs) in a free-viewing visual search task. Of particular interest is the possibility that RTs for target present and target absent decisions will be related differently to judgments of search difficulty. Note that a correct target present response is triggered by a perceptual event—detection of the target. RT for such a response is therefore determined by the amount of time required to localize and recognize the target. A correct target absent response, in contrast, occurs only once the observer has searched the image to a criterion level of confidence. Setting of this criterion may be related to an assessment of task difficulty. The following two predictions are therefore possible. First, if subjects are poor at predicting the difficulty of target detection, as the data of Experiment 1 suggest, then RTs for target present responses should be uncorrelated with predictive ratings. Second, if the criterion for executing a target absent response is related to the subject's metacognitive assessment of search difficulty, then RTs for target absent responses should increase when predicted difficulty is high.

Method

Subjects were 28 young adults. All were naïve to the purpose of the study. Stimuli and the rating task were identical to those of Experiment 1. The search task differed from that of Experiment 1 in the following ways. Each trial, the subject saw a single image and was asked to judge whether or not it contained the target. The stimulus image remained visible until the subject's response. Subjects were asked to make responses as quickly as possible without sacrificing accuracy. A target was present on ½ of all trials. Stimulus items were divided into two sets of 32 target present and 32 target absent images each (excluding a set of 5 practice stimuli), with presentation counterbalanced such that bags which served as target present stimuli for half of the subjects served as target absent stimuli for the rest.

Results

With the small number of trials in each of the target present and target absent conditions (32 each) divided across 5 levels of the rating scale, several subjects were left with empty data cells. The 5-point rating scale used for predictive judgments was therefore reduced to a 3-point scale by collapsing across the two lowest and two highest rating categories. Following this, 24 of the 28 subjects had no empty data cells. Data from the remaining 4 subjects were omitted from the ANOVA used for the primary analysis. A follow-up analysis, reported below, indicated that the omission of these data did not meaningfully alter the pattern of results. Data from all 28 subjects are presented in Figure 1.

RTs for correct responses were analyzed using a within-subjects ANOVA with response (target present vs. target absent) and predictive rating (1-3) as factors. Unsurprisingly, analysis revealed a reliable main effect of response, $F(1, 23) = 40.90$, $p < .001$, confirming that RTs were longer for target absent than for target present responses. Data produced no main effect of rating [$F < 1$], but did show a reliable interaction of response by rating, $F(2, 46) = 4.50$, $p = .016$. The basis of the effect is evident in Figure 1; RTs for target absent responses became shorter as the predicted likelihood of target detection increased, while RTs for target present responses were independent of ease of search ratings. One-way ANOVAs using rating as a within-subjects variable, conducted separately for target absent and target present responses, confirmed this. Target absent RTs showed a reliable effect of rating, $F(2, 52) = 5.84$, $p = .005$. Target present RTs did not, $F < 1$.

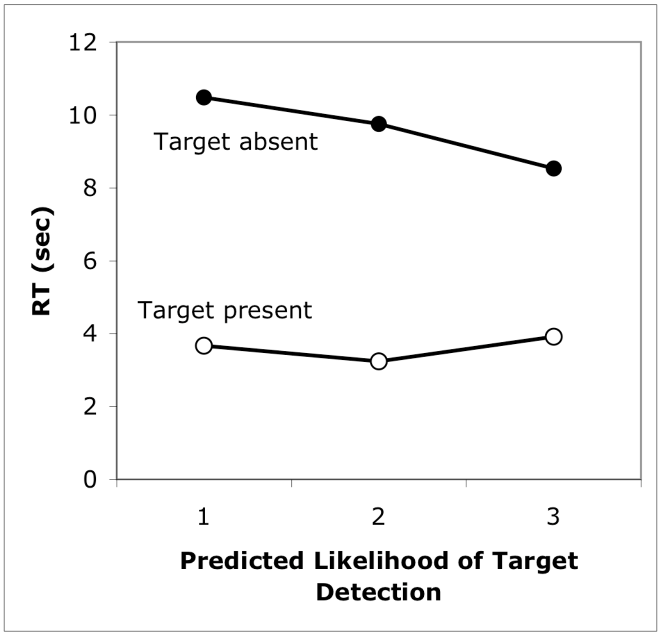

Figure 1. RTs for Experiment 2

As noted above, an additional analysis was conducted to ensure that results were not distorted by the omission of data from 4 subjects. Here, we calculated subject-by-subject correlations between predicted ease of target detection on the 3-point rating scale and mean RTs for target present and target absent responses. Mean correlation for target absent responses was negative, $M = -.31$ [$t(27) = -2.06$, $p = .049$ by one-sample t-test vs. a value of 0], indicating that target absent RTs declined as the predicted ease of target detection increased. Mean correlation for target present responses, in contrast, was not statistically different from 0 [$t(27) = .25$, $p = .80$]. A paired samples t-test directly comparing the mean correlations for target absent and target present responses produced a reliable effect [$t(27) = -2.38$, $p = .025$].

Mean hit rate was .70. Mean correct rejection rate was .08. Neither the hit rate nor the correct rejection rate varied reliably as a function of predicted likelihood of target detection.

DISCUSSION

A pair of experiments examined the potential role of metacognitive monitoring and control in a simulated baggage x-ray screening task. Experiment 1 found that subjects had only a modest ability to discriminate bags that would allow target detection from those that would not in a forced-choice task. Experiment 2 found that predictive judgments of the likelihood of detection were uncorrelated with target present RTs in a free-viewing visual search task, but were inversely correlated with target absent RTs. In total, the data imply that subjects' predictive metacognitive assessments were poor, but nonetheless were correlated with placement of the criterion by which target absent search was terminated.

It is important to note that these findings do not demonstrate that metacognitive judgments were made explicitly during search, or that the relationship between metacognitive assessments and criterion placement was causal. Rather, the data indicate only that predictive metacognitive assessments and criterion for terminating search are based in part on the same information. Future research will be necessary to examine the causal nature of the relationship between metacognitive judgments and search behavior in more detail.

Additional study will likewise be required to explore potential applications of the present results. Of particular interest is the possibility that an understanding of metacognitive processes might improve screener selection and/or training. As noted, predictive metacognitive judgments in the current studies were poor, both in the criterion-free search task of Experiment 1 and in the target present trials of the speeded task used in Experiment 2. Moreover, the data of experiment 1 found no reliable correlation between sensitivity and metacognitive accuracy. Participants in these studies, however, were inexperienced in the baggage screening task. It is possible that more experienced screeners might have produced better predictive judgments, or that among more skilled screeners metacognitive accuracy may have been more predictive of search performance. More interestingly, is possible that metacognitive training might help to inculcate effective monitoring and control skills, as has been demonstrated in other task domains (Dunlosky, Kubat-Silman, & Hertzog, 2003). In such case, metacognitive training efforts could be used to supplement training of the first-order cognitive and perceptual skills that mediate screener performance. Future experiments will be needed to test this possibility.

Footnote

The views expressed here are those of the authors and do not necessarily reflect those of the organizations with which they are affiliated or their sponsoring agencies. The research reported in this paper did not involve Transportation Security Administration screener personnel, screener training materials, and did not take place at and does not represent a true airport screening checkpoint.

REFERENCES

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology, 30*, 39-78.

Dunlosky, J., Kubat-Silman, A. K., & Hertzog, C. (2003). Training monitoring skills improves older adults' self-paced associative learning. *Psychology and Aging, 18*, 340-345.

Flavell, J. H. (1979). Metacognitive and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist, 34*, 906-911.

Lesch, M., & Hancock, P. A. (2004). Driving performance during concurrent cell-phone use: Are drivers aware of their performance decrements? *Accident Analysis and Prevention, 36*, 471-480.

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport security inspection. *Psychological Science, 15*, 302-306.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Review, 95*, 109-133.

Nelson, T. O., Dunlosky, J., Graf, P., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study time during multitrial learning. *Psychological Science, 5*, 207-213.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676-686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-141). New York: Academic Press.

Perfect, T. J., & Schwartz, B. L. (Eds.). (2002). *Applied metacognition*. Cambridge: Cambridge University Press.

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free-recall learning. *Journal of Experimental Psychology, 71*, 673-679.

Valot, C. (2002). An ecological approach to metacognitive regulation in the adult. In P. Chambres, M. Izaute & P. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 135-151). Boston: Kluwer.

Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford.