

## THE IMPACT OF PERFORMANCE INCENTIVES DURING TRAINING ON TRANSFER OF LEARNING

Patricia C. Brennan<sup>1</sup>, Poornima Madhavan<sup>1</sup>, Cleotilde Gonzalez<sup>2</sup> and Frank C. Lacson<sup>3</sup>

<sup>1</sup>Old Dominion University, Norfolk, VA

<sup>2</sup>Carnegie Mellon University, Pittsburgh, PA

<sup>3</sup>Pacific Science & Engineering, San Diego, CA

We examined the effect of incentives on transfer of training in a visual search task. The incentives were designed to influence participants' detection of dangerous targets across two phases of airline luggage screening – training (familiar targets) and transfer (novel targets). Participants were assigned to one of five groups - hit-sensitive (points awarded for hits), miss-sensitive (points deducted for misses), equal-costs (equal points awarded and deducted for hits and misses), no-incentives (no points) and control (self-training). The goal was to examine which incentive structure was most effective in transfer of training. Results revealed that rewarding points for hits and deducting points for misses led to best transfer performance, although punishment had a stronger impact on performance than rewards. Rewarding hits implicitly primed participants to say 'yes' more often inflating their false alarm rate. Allowing participants to self-train also significantly benefited performance by minimizing the constraints imposed by fixed incentives-based training.

### INTRODUCTION

A visual search task consists of scanning for a target item in which the exact location or presence of the target is unknown. Fisk and Schneider (1981), suggest that vigilance decrements are expected for any task where the signal (i.e., weapon) to noise (i.e., hair curlers, blow dryers) ratio is unpredictable and automatic processing cannot take place. The purpose of this research was to examine whether providing incentives for achieving specific performance goals will be beneficial in reducing vigilance decrements and helping performance in such a complex task.

Incentives can also be framed qualitatively through goals. Goal framing is structuring information in the form of positive or negative consequences for a behavior or strategy, such as the cost of false alarms versus the benefits of a hit (Levin et al., 1998). In a study on goal framing and automation reliance and compliance, Lacson, Wiegmann and Madhavan (2005) investigated how positive and negative goal frames affected the participants' trust towards an automated aid in a simulated process control task. They found that negatively framed instructions (minimize misses and false alarms) increased reliance (i.e., agreeing when the aid says "no" or remains silent) whereas positively framed instructions (maximize hits and correct rejections) decreased

reliance (i.e., disagreeing with the aid when it said "no"). Additionally, providing information about the automation aid's reliability increased compliance (i.e., agreeing with the aid when it said "yes" or sounded an alert). They concluded that giving both positive (maximize hits and correct rejections) and negative task information (minimize false alarms and misses) as well as additional information about the automation aid (i.e., accuracy in targeting weapons) influenced operators' behavior toward fewer false alarms and higher hits in various ways. This is possibly because it provided a more realistic picture of the automation aid's error and reliability rates leading participants to utilize automation appropriately. In addition, it provided participants a clearer idea of their own goals.

Although the study by Lacson et al. answered a very interesting question regarding framing effects in real world decision making, there are some problems with the paradigm. First, goal framing was not operationalized in a very specific manner; participants were not provided real gains and losses (e.g., points, rewards, punishment) for correct and incorrect decisions. Second, the design was not completely crossed in that the study did not compare the performance of participants with framed incentives with those that received neutral (equal) frames or no frames at all. Lastly, the task by Lacson et al., (2005) was opaque in that the task was impossible for

participants to perform on their own without the assistance of an automated aid, which, we believe, was confounded with the goal frames themselves.

### Purpose of the Present Study

The purpose of this study is to address the flaws in earlier paradigms and to investigate the incorporation of incentives during training on transfer of training in a visual search task of inspecting luggage for weapons. Incentives, such as giving or taking away points for a hit, miss, correct rejection and false alarm are expected to affect decisions in detection behavior. In this study, we will extend the work by Lacson et al (2005), by examining performance as a result of framing effects when goals are related to incentives (rewarding points versus deducting points) for performing the task correctly or incorrectly. Specifically we are interested in how incentives (or rewards) versus punishment versus no incentives might affect detection behavior and specifically transfer of training.

## METHOD

### Participants

One hundred and thirty undergraduate students, graduate students and some community members performed a 2-day experiment and were given \$15 plus a performance bonus between \$0 and \$7.

### Tasks and Procedures

Participants were randomly assigned to either one of the four experimental groups or the control group. The experimental groups were: Hit-Sensitive (H-S), Miss-Sensitive (M-S), Equal-Costs, and No-Incentives. Participants who were assigned to the fifth control group were trained differently from the experimental groups (as described in the procedure section below). Points were given or taken away for hits and/or misses and served as incentives depending on the incentive structure. The number of points gained or lost determined how much performance bonus money was given at the end of the experiment. Refer to Table 1 below for the number of points given or taken away for the four outcomes of Hit, Miss, False Alarm or Correct Rejection. For each incentive structure, the costs were derived based on signal detection theory (Wickens & Hollands, 2000).

Table 1. Incentive structures

Incentive Structure	Outcome (Points)			
	Hit	Miss	FA	CR
Hit-Sensitive (Gain)	+350	-50	-24	+1
Miss-Sensitive (Loss)	+50	-350	-24	+1
Equal-Costs	+1	-1	-1	+1
No-Incentives	0	0	0	0
Control (Self-Training)	0	0	0	0

The visual search task consisted of 12 blocks with 30 trials (x-ray images of luggage) in each, totaling 360 images for the training session. For the four experimental groups, a unique four-item target set was shown to the participant before each block and all weapons were drawn from the four-item target set. The images were presented for three seconds. During each presentation, participants made a decision to search the bag (by clicking on a weapon if they detected it) or pass the bag (if they decided that the bag did not contain a weapon). Participants received textual feedback at the end of each trial and their point structure changed in real time according to the incentive structures illustrated in Table 1. The target base rate was 20%. This was higher than what is seen in the real world for purely statistical reasons. They were also asked to rate their confidence after every decision on a 5 point Likert scale. The control group only differed from the other groups in that they did not receive the four-item target set.

The transfer phase on Day 2 differed from the training phase by including novel or unique signal items that participants had not seen on Day 1. These items consisted of dangerous items that were not used during the training phase but were also from non-traditional categories such as sharp glass or metal objects. Furthermore, participants were not shown the target set before the beginning of the trial block as in Day 1. Instead they were instructed to use their knowledge and memory of the weapons seen on Day 1 to estimate what the new targets would be on Day 2. Participants then performed the visual screening task for 6 blocks of 30 trials each (180 luggage images). Again, 20% of the trials (6 out of 30) contained an actual target.

The Control group was different from the other four groups in that they were not explicitly trained with target exposures on Day 1. Instead, they trained themselves ("self-training") without the target exposures explicitly shown to them before each block during training.

## RESULTS

Multiple repeated measures ANOVAs revealed that in general performance improved from training to transfer and the control group remained fairly consistent in performance across training and transfer. During training, the hit-sensitive and miss-sensitive groups produced the most hits and the control group produced the least amount of hits. However, the hit-sensitive group had a higher rate of false alarms than all other groups. Our hypotheses for transfer are as follows 1) the hit-sensitive group will generate more hits but have more false alarms, 2) the miss-sensitive group will generate more hits but have more correct rejections, 3) the equal-costs group will attempt to balance hits and misses, and 4) the no-incentives group will be strongly influenced by context rather than incentives and try to maximize hits.

We used one-way ANOVAs to analyze transfer data followed by Tukey post-hoc analyses for between group comparisons (Maxwell & Delaney, 2004). In Figures 1-5, Group 1 represents the hit-sensitive group, Group 2 represents the miss-sensitive group, Group 3 represents the equal-costs group, Group 4 represents the no-incentives group, and Group 5 represents the control group

**Hit rate.** A one-way ANOVA on hit rates during transfer revealed a significant difference between groups,  $F(4, 73) = 6.91, p < .001$ , partial  $\eta^2 = .029$ . The results for hit rates are illustrated in Figure 1.

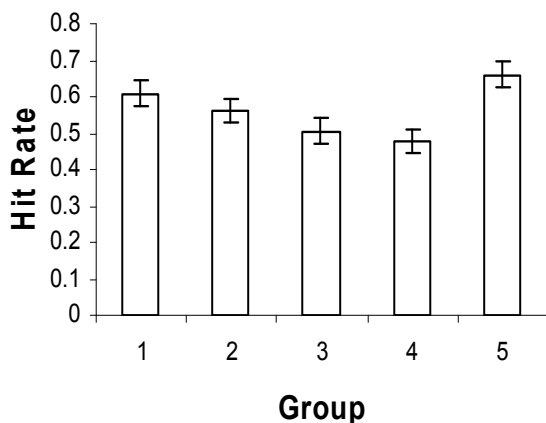


Figure 1. Hit rates.

The hit-sensitive group ( $M = .61, SD = .06$ ),  $p < .756$ , miss-sensitive ( $M = .56, SD = .13$ ),  $p < .756$ , and the control group ( $M = .66, SD = .13$ ),  $p < .699$ , were not significantly different from one another. However, the equal-costs ( $M = .51, SD = .10$ ),  $F(4, 73) = 12.16, p < .05$ , partial  $\eta^2 = .531$  and the no-incentives groups

( $M = .48, SD = .11$ ),  $F(4, 73) = 14.37, p < .001$ , partial  $\eta^2 = .589$  had significantly lower hit rates than the control group.

**False alarm rate.** A one-way ANOVA on false alarm rates during transfer revealed a significant difference between groups,  $F(4, 73) = 9.40, p < .001$ , partial  $\eta^2 = .353$  illustrated in Figure 2. As predicted, the hit-sensitive group ( $M = .60, SD = .33$ ) had significantly higher false alarm rates than all other groups including miss-sensitive ( $M = .27, SD = .20$ ),  $F(4, 73) = 12.03, p < .05$ , partial  $\eta^2 = .512$ , equal-costs ( $M = .20, SD = .17$ ),  $F(4, 73) = 14.55, p < .05$ , partial  $\eta^2 = .603$ , no-incentives ( $M = .23, SD = .25$ ),  $F(4, 73) = 13.63, p < .05$ , partial  $\eta^2 = .536$ , and the control group ( $M = .12, SD = .17$ ),  $F(4, 73) = 17.74, p < .05$ , partial  $\eta^2 = .675$ .

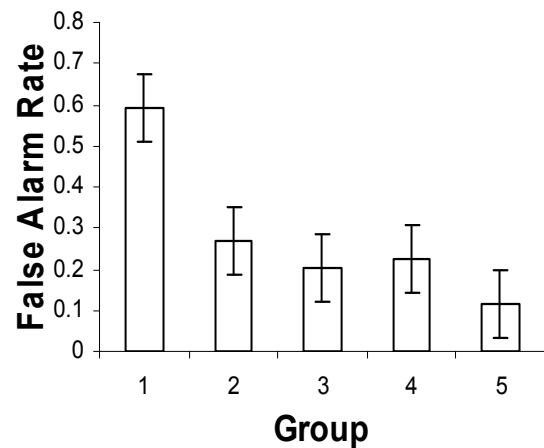


Figure 2. False Alarm rates.

**Response time for hits.** A one-way ANOVA on response time for hits revealed no significant differences between groups,  $F(4, 73) = .522, p < .720$ , partial  $\eta^2 = .029$ .

**Response time for false alarms.** A one-way ANOVA on response times during transfer revealed a significant difference between groups,  $F(4, 73) = 9.278, p < .001$ , partial  $\eta^2 = .350$ . The results in Figure 3 suggest that the hit-sensitive group ( $M = 3603.91\text{ms}, SD = 602.11$ ) had significantly longer response times for false alarms than the equal-costs ( $M = 2866.83\text{ms}, SD = 1232.07$ ),  $F(4, 73) = 6.48, p < .05$ , partial  $\eta^2 = .411$ , no-incentives ( $M = 2840.57\text{ms}, SD = 924.37$ ),  $F(4, 73) = 6.71, p < .05$ , partial  $\eta^2 = .439$ , and control groups ( $M = 1505.83\text{ms}, SD = 908.37$ ),  $F(4, 73) = 18.45, p < .001$ , partial  $\eta^2 = .806$ .

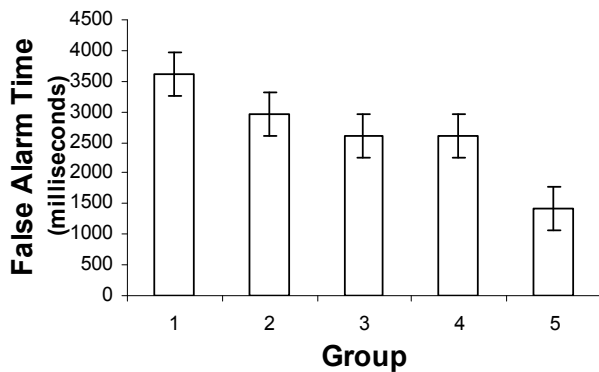


Figure 3. Response time for false alarms.

The control group had significantly shorter response times than all other groups including the hit-sensitive, miss-sensitive ( $M = 2946.8\text{ms}$ ,  $SD = 1096.889$ ),  $F(4, 73) = 12.67$ ,  $p < .001$ , partial  $\eta^2 = .582$ , equal-costs ( $M = .2866.83\text{ms}$ ,  $SD = 1232.07$ ),  $F(4, 73) = 11.96$ ,  $p < .001$ , partial  $\eta^2 = .532$ , and no-incentives groups ( $M = .2840.5\text{ms}$ ,  $SD = .924.377$ ),  $F(4, 73) = 11.74$ ,  $p < .001$ , partial  $\eta^2 = .589$ .

*Subjective confidence when generating hits.* A one-way ANOVA on confidence for hits during transfer revealed a significant difference between groups,  $F(4, 73) = 4.61$ ,  $p < .05$ , partial  $\eta^2 = .211$ , illustrated in Figure 4.

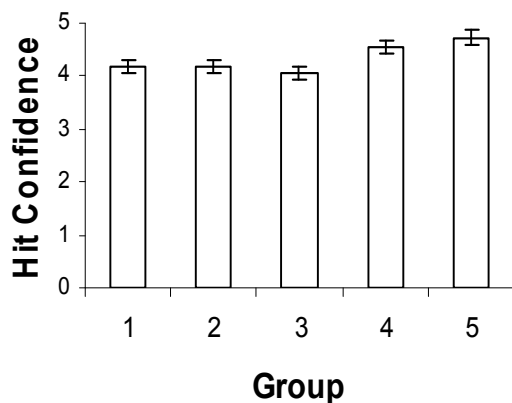


Figure 4. Subjective confidence when generating hits.

The hit-sensitive ( $M = 4.17$ ,  $SD = .52$ ), miss-sensitive ( $M = 4.20$ ,  $SD = .54$ ),  $p < .905$  and equal-costs ( $M = 4.07$ ,  $SD = .70$ ),  $p < .585$  were not significantly different from each other. However the control group ( $M = 4.73$ ,  $SD = .33$ ) had significantly higher confidence than the hit-sensitive  $F(4, 73) = 19.26$ ,  $p < .05$ , partial  $\eta^2 = .543$ , miss-sensitive  $F(4, 73) = 13.49$ ,  $p < .05$ , partial  $\eta^2 = .513$ , and equal-costs group  $F(4, 73) = 10.45$ ,  $p < .001$ , partial  $\eta^2 = .516$ .

*Subjective confidence when generating false alarms.* Overall, false alarm confidence was significantly lower than confidence on hits. A one-way ANOVA on confidence for false alarms during transfer revealed a significant difference between groups,  $F(4, 73) = 3.40$ ,  $p < .05$ , partial  $\eta^2 = .165$ , illustrated in Figure 5. The primary difference was between the no-incentives ( $M = 2.06$ ,  $SD = 1.4$ ) and control group ( $M = .93$ ,  $SD = .68$ ),  $p < .05$ , partial  $\eta^2 = .926$  with the control group having significantly lower confidence in false alarms than the no-incentives group.

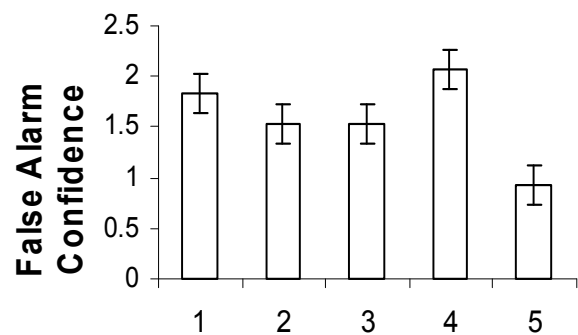


Figure 5. Subjective confidence when generating false alarms.

## DISCUSSION

The purpose of this study was to examine the role of incentives in improving transfer for learning in a complex visual search task. Interestingly, participants who ‘self-trained’ (the control group) showed some evidence of better transfer performance than participants who received training. This pattern of results suggest that participants who were not specifically shown targets during training was able to effectively create their own visualizations of threat objects over time and were able to connect targets during training to new targets at transfer. It appears that when participants did not have the pressures created by performance incentives and were not primed by specific targets during training it helped rather than hurt transfer of training. Their confidence could be a further indication that the absence of an incentive or ‘pressure’ to behave in a particular manner helped them develop a more realistic assessment of their own performance. Contrary to the pattern for the control group, for participants in the experimental groups who were trained with specific target exposures the introduction of novel targets

hindered performance by decreasing hit rates across training to transfer.

Despite the superior performance of the Control group, the results of the present study do provide some support for the hypothesis that providing incentives during training benefits performance at transfer when comparing the trained groups alone. The hit-sensitive and miss-sensitive groups generated more hits than the equal-costs and no-incentives groups illustrating that incentive structures help in motivating participants to detect targets; on the contrary, providing no training incentives or equal incentives for different outcomes creates cognitive confusion in terms of performance expectations and hampers transfer of training. Apparently, the incentive structures for equal-costs and no-incentives inhibited transfer of training.

The performance of the equal-costs group in this study suggests a pattern of confusion as a consequence of incentive structure. Specifically, participants were unaware of which performance outcome was more beneficial - generating more hits and risking false alarms or curtailing false alarms by allowing misses. The no-incentives group performed as expected; they simply performed the task without focusing on outcome.

Results suggest that incorporating incentive structures which reward hits or penalize misses lead participants to consciously maximize their hits. However, rewarding hits also implicitly primed participants to say 'yes' more often, resulting in a higher number of false alarms, which can be costly in terms of time and resource wastage. The hit-sensitive group also spent a significantly longer time searching for weapons that were not present, as was indicated by their lengthy response times for false alarms. Therefore, although the hit-sensitive incentive structure was the most effective in detecting targets, it also damaged performance by leading participants to invest inordinate amounts of time searching for targets when they were absent.

Interestingly, participants in the miss-sensitive group were able to curtail their 'yes' responses in order to optimize their false alarm rates. Additionally, they did not waste as much time looking for weapons when they were absent. Clearly, giving directional incentives (for hits or misses) does improve transfer performance in comparison to giving equal incentives or no incentives. However, surprisingly, punishment in the form of point deduction for misses had a more

powerful effect on balancing hits and false alarms than providing rewards for hits.

For all participants, confidence was significantly lower when generating false alarms than hits reflecting an implicit level of performance awareness. In keeping with the results for hits and false alarms, the control group appeared to have had the highest level of such performance awareness as was indicated by their significantly high confidence when generating hits and proportionately low confidence for false alarms.

## CONCLUSIONS

Overall, the results of this study indicate that rewarding hits and penalizing misses led participants to maximize their hits, with punishment exerting a more powerful motivating effect on performance than rewards. However, allowing participants to self-train was as effective as providing training with incentives. However, it is possible that if the task were longer in duration, the effects of self-training may not have been as powerful as training with hit- or miss-based incentives because there is no external motivation to compensate for vigilance decrements.

The results of this study suggest that incentives during training do have beneficial effects on transfer. In addition, allowing participants to construct their own mental images of target stimuli (self-training) appeared to be beneficial in transfer of learning in which novel targets are introduced. However, in order for incentive-based training to be effective in the real world, care must be taken to minimize adverse impacts of incentives such as excessive false alarms and slowing down of decision times as observed in this study.

## REFERENCES

- Fisk, A.D., & Schneider, W. (1981). Automatic and control processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors*, 25, 391-399.
- Lacson, F.C., Wiegmann, D.W., & Madhavan, P. (2005). Effects of attribute and goal framing on automation compliance and reliance. *Proceedings of the 49th Annual Conference of the Human Factors and Ergonomics Society*, pp 482-486. Santa Monica, CA.
- Levin, I.P., Schneider, S.L., & Gaeth, G.J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149-188.
- Maxwell, S.E., & Delaney, H.D. (2004). *Designing Experiments and Analyzing Data*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Wickens, Christopher D. and Hollands, Justin G. (2000): *Engineering Psychology and Human Performance*. Prentice Hall