Applied Cognitive Psychology, *Appl. Cognit. Psychol.* **26**: 194–206 (2012)
Published online 31 May 2011 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.1807

The Role of Incentive Framing on Training and Transfer of Learning in a Visual Threat Detection Task

POORNIMA MADHAVAN[1]*, FRANK C. LACSON[2], CLEOTILDE GONZALEZ[3] and PATRICIA C. BRENNAN[1]

[1]*Old Dominion University, Nortfolk, USA*
[2]*Pacific Science & Engineering, San Diego, USA*
[3]*Carnegie Mellon University, Pittsburg, USA*

Summary: We examined the effects of different incentives on skill acquisition and transfer during threat detection in airline luggage screening. The incentives were presented within positive (gains) or negative (losses) frames, and points were given or taken away accordingly during training (with familiar targets) and transfer (to novel targets). During training, incentives exerted a more beneficial effect on skill acquisition than training without incentives. During transfer, incentives benefitted performance largely when presented as losses or penalties. Incentives framed as gains primed participants to say 'yes' more often leading to a high ratio of false positives; however, incentives framed as losses lead participants to become more selective in their 'yes' responses leading to a lower number of false positives but a comparable probability of correct detections. Interestingly, participants that received no training outperformed participants that received incentive-based training, suggesting that incentives actually constrained rather than helped transfer of learning in this study. Copyright © 2011 John Wiley & Sons, Ltd.

When presented with a mentally challenging visual search task, how can incentives enhance performance and lead individuals to turn extrinsic motivation (e.g. monetary rewards) to intrinsic motivation (or, internally stimulating reasons for personal effort)? Visual search tasks are a type of perceptual task requiring attention from the individual that can become mentally challenging over time because of their complexity and cognitive demands. One example of such a complex visual search task is airport luggage screening that requires the human operator to scan for rarely occurring threat objects among a large variety of other objects that are typically distractors, or constitute 'perceptual noise' (i.e. pill boxes, clothes, blow dryers). The task of searching through background noise and deciding whether certain features of an object categorize it as a weapon or not, can be cognitively very challenging. In order to reduce the potential for errors and improve human performance in these domains, there is renewed emphasis on improving the quality of personnel training in these tasks, one of which involves designing and providing varied forms of incentives for desired performance.

Providing incentives during training has been shown to positively influence performance (Fridrici, Lohaus, & Glab, 2009; Scott & Goldwater, 1998) although there have been documented occasions when incentives do not succeed in achieving the desired goal immediately (Bregman & McAllister, 1983). In the context of luggage screening, some forms of incentive have been designed to improve screener performance. For example, the Threat Image Projection technique (designed by TSA; http://www.tsa.gov/approach/tech/tip.shtm) is implemented as a form of on-the-job training; this technique involves projecting false threat images into random X-ray images of luggage so that security officers remain alert. Occasionally, 'live' threats are often carried out by inspection teams as a further check. Incentives to perform include the possibility of screener promotion to management positions with penalties for detection failures. However, it must be noted that incentives in this context are generally challenging to implement because of the overwhelming emphasis on throughput (i.e. just getting passengers on their planes). In this article, we address the effects of an alternative incentive measure—positive and negatively framed incentives during training. First, we examined whether incentives framed as gains (giving points) or losses (taking away points) or some combination of the two enhance performance by serving as effective priming mechanisms from training to transfer. Second, we examined whether training using such extraneous incentives provides an advantage (or, disadvantage) over self-training wherein the decision maker acquires skills and learns the task on his/her own over time without assistance. In the following sections, we describe the visual search issues relevant to our paradigm, theoretical foundations of framing and its relationship with the incentive-based training method implemented in our research study and practical implications.

VISUAL SEARCH AND GOAL-DIRECTED PROCESSING

Visual search describes the difficult process of finding a target item among distractor items in often cluttered visual environments. The difficulty of visual search arises from physical and cognitive processing limitations that can prevent us from instantly recognizing the presence of a target item in a single glance (Boot, Becic, & Kramer, 2009). Attention is required to focus limited processing resources on specific regions of a scene in order to find the people and objects for which we are searching. Researchers have proposed a number of factors, both stimulus-related and cognitive, that influence attention allocation during visual search.

The manner in which attention is allocated during search appears to be determined, at least in part, by the properties of the scene being searched and the type of training given to the searchers. Evidence suggests that even when observers

*Correspondence to: Poornima Madhavan, Department of Psychology, 346-B Mills Godwin Building, Old Dominion University, Norfolk, VA 23529, USA. E-mail: pmadhava@odu.edu

know exactly where the target will appear, highly salient features known never to be associated with the target item can still capture attention in a seemingly stimulus-driven manner (Christ & Abrams, 2006). These findings suggest that in certain situations, visual salience plays a dominant role in controlling the direction of attention (e.g. Itti, 2006; Itti & Koch, 2000). However, cognitive or top-down factors also appear to play an important role.

In top-down driven visual search, the task goal or 'attention set' of the observer plays a significant role in determining theallocation of attention (e.g. Folk, Remington, & Johnston, 1992; Folk, Remington, & Wright, 1994). In other words, observers can modify their attentional focus depending on their task goals (Boot, McCarley, Kramer, & Peterson, 2004; Brockmole & Henderson, 2006; Peterson & Kramer, 2001). If goal-directed top-down processing does indeed play such a critical role in visual search, then 'framing' the goals of the search task to draw the searcher's attention toward certain aspects (versus away from other aspects) of the task will lead to qualitatively different search strategies. Previous research has focused extensively on the role of visual search in threat detection. Our approach is different from these earlier approaches in that we infuse the idea of conscious decision making and framing effects into the existing target search paradigm. In the succeeding section, we examine further the genesis of such framing effects in the decision making literature and the manner in which they can be shaped into performance incentives for a typical visual search paradigm.

FRAMING EFFECTS

Tversky and Kahneman (1981) defined framing effects as transparently and objectively identical decision options, which generate different outcomes depending on how the information is presented and consequently perceived as 'losses' or 'gains'. Specifically, framing effects are a representation of task information in either a positive (gains) frame (e.g. glass is half full) or negative (losses) frame (e.g. glass is half empty); such frames typically lead individuals to engage in risk seeking behaviors in negative (loss) frame situations and risk aversive behaviors in positive (gain) frame situations.

In their classic experiment, Tversky and Kahneman (1981) examined framing effects in the context of an 'Asian disease problem' that required participants to make decisions regarding an epidemic that was proposed to kill 600 people. Participants were given two alternatives to choose from within either a positive frame (or, gains frame), which focused on '*lives saved*' alone, or a negative frame (or, losses frame), which presented the same information framed as '*lives lost*' alone. Results revealed that individuals in the positively framed condition selected the riskless, or risk averse option of saving 200 people for sure over a 1/3 chance of saving all 600 people. On the other hand, participants in the negatively framed condition, chose the risky option of allowing a 2/3 probability of the death of all 600 people over the sure death of 400 people. The authors concluded that positively framed terms elicit risk aversion because people prefer the guarantee of saving 200 lives than a one in three chance of saving 600 lives. The negatively framed terms elicited risk-seeking behavior because the definite death of 400 people is less desirable than the two in three chance of the death of 600 people.

Meyerowitz and Chaiken (1987) found similar effects for positive versus negative framing in the context of breast self-examinations. They found that women were more compliant when information was presented in negative terms (i.e. the decreased chance of finding a tumor in a treatable stage) versus positive terms (i.e. the increased chance of finding a tumor in a treatable stage). These studies are just two examples among several studies that have revealed that under circumstances that are possibly life-threatening, either to oneself or to others, framing goals in negative terms elicits stronger incentive to perform risk-taking behaviors and a greater mitigation of negative decision choices.

Research suggests that a primary reason why positivity or negativity of framed information strongly impacts behavior is because of the role played by affect or emotion in the decision process (Gelder, Vries, & Pligt, 2009). Reportedly, people make decisions using two modes of thought—analytical and intuitive (Chaiken & Trope, 1999; Epstein, 1994; Sloman, 1996). The analytical technique involves the cognitive mode, which entails objective weighting of outcomes and probabilities. The affective mode relies primarily on motivational factors thereby favoring emotionality in decision making (Slovic, Peters, Finucane, & MacGregor, 2005). The *cognitive–affective tradeoff* theory (Gonzalez, Dana, Koshino, & Just, 2005) proposes that decision makers typically attempt to make emotionally acceptable choices with minimal computational effort. Sure gains are easy to process and lead to an obviously acceptable outcome. Sure losses are also easy to process; however, the negative affect associated with a sure loss causes people to engage in a more thorough consideration of the risky option. This theory implies that people can be influenced or trained to rely on their rational thought processes or their emotions by simply adding affective or analytical information to the description of the decision problem (Gelder et al., 2009).

All the research described previously suggests the following underlying message: framing effects represent an important method to influence decision making by exerting an influence on affective or analytical thought processes. Despite this general finding, few studies have explored whether framing can be used productively as a means to design incentives for improving performance on a task. Additionally, it is yet to be determined whether such incentives can be generalized into long term effective training strategies for personnel in complex visual search tasks.

INCENTIVE-BASED FRAMING VERSUS 'TRADITIONAL' FRAMING EFFECTS THEORY

Incentive framing refers to the process of structuring information in the form of positive or negative consequences for a behavior or strategy, such as costs or penalties for false alarms versus benefits or rewards for hits (Levin, Schneider, & Gaeth, 1998). Such incentive-based training will purportedly lead operators to focus on certain desired outcomes when performing a task (e.g. accuracy versus speed, hits versus correct rejections). This emphasis on one particular outcome over another during training can be considered a form of 'framing' as it involves presenting a decision situation in alternative ways to elicit the desired performance.

Framing in incentive-based training differs from the Asian disease problem studied by Tversky and Kahneman (1981) wherein specific predetermined response patterns were elicited by forcing participants to choose between two options, which were based entirely on the emotionality (i.e. the affective component) associated with decision outcomes. Incentive framing goes beyond traditional framing in that incentives reflect actual differences in quantities of gains versus losses. For example, an incentive gains frame would be 'you will receive 100 points for finding a target', whereas an incentive loss frame would be 'you will lose 100 points for missing a target'. This kind of framing incorporates both affective (emotion-based) and analytical (cognitive or rational) components by emphasizing both concrete cognitive goals (points awarded or deducted) as well as the emotional importance of attaining these goals (win more money and become richer or lose money and become poorer). In incentive framing, decision makers have the ultimate freedom to decide their own best strategy or behavior in order to meet these goals. An example would be airline luggage screening wherein each screener decides on a different strategy to detect a potential threat object (and creating the balance between hits and false alarms) although the overarching goal is the same—to maximize security within a reasonable scan time.

Very few studies have attempted to combine the concepts of incentive-based training and framing effects. One recent study attempted to examine this in the context of human interaction with automated decision aids (Lacson, Wiegmann, & Madhavan, 2005). The authors used a visual search task to investigate how positive incentive frames (i.e. maximize hits and correct rejections) and negative incentive frames (i.e. minimize misses and false alarms) affected participants' trust in an automated aid. Additionally, they also provided selective information about the automation aid's reliability. They found that providing both positive and negative incentives concurrently in addition to information about the automation aid's reliability influenced operator's behavior towards fewest false alarms and maximum hits. This is possibly because this combination of instructions provided a realistic picture of the automation aid's reliability rates in conjunction with the clearest picture of the participants' own goals. Although the study of Lacson et al. answered a very interesting question regarding framing of incentives in real world decision making, there are several problems with the paradigm. First, incentive framing was not operationalized in a very effective manner; participants were not provided real gains and losses (e.g. points, rewards, or punishment) for correct and incorrect decisions. Second, the design did not compare the performance of participants with framed incentives with those that received neutral frames or no frames at all. Lastly, the task described in Lacson et al. 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 incentive frames themselves. Furthermore, the studies by Lacson et al. and Meyerowitz and Chaiken (1987) have conflicting results suggesting the need for more research in order to understand the role of incentives in learning and transfer of learning.

PURPOSE OF THE PRESENT STUDY

The purpose of this study was to investigate the effects of incentive framing on performance and decision making in a visual search task of inspecting luggage for weapons. In this study, we extended the work of Lacson et al. (2005) by examining screening performance as a result of framing effects when incentives (positive versus negative versus none) are provided for performing the task correctly. In a recently published preliminary version of the present study (Lacson, Gonzalez, & Madhavan, 2008), we compared the efficacy of positive, negative and equal incentives frames across two different contexts: the context of threat detection (i.e. detection of weapons in airline passenger luggage), which is typically accompanied by highly critical consequences and the context of produce inspection (i.e. detection of fruit among vegetables), which is accompanied by less critical consequences. Results revealed that penalties for missing targets had the most powerful effect on eliciting the desired performance strategies; this effect was much stronger in the emotionally charged context of weapon detection. Although this study provided some insight into the impact of incentive structures on performance, there was no mechanism to draw conclusions regarding how performance trained with incentives compared with untrained performance or trained performance with no incentives.

The first goal of this study, therefore, was to compare incentive-based training with self-training wherein participants learned to detect targets on their own over time without explicit instructions or strategies (i.e. target exposures or incentives). The second goal was to determine which incentive structure(s) during training (hit-sensitive, miss-sensitive, equal-incentives or no-incentives) will provide the best incentives for superior performance both during skill acquisition and transfer of learning (i.e. the ultimate goal being the maximization of hits and minimization of false alarms). The overarching purpose of this study was to examine how incentive structures impact decision making and transfer of learning and how this can be used to train operators in visual search tasks. We were particularly interested in situations wherein the targets at transfer are novel (i.e. when participants encounter new targets that they have not encountered during training) as is often characteristic of real world luggage screening contexts.

METHOD

Pretest

The purpose of the pretest was to ensure that the stimuli that would be used to train and transfer learning in the actual luggage screening experiment were of comparable difficulty. We created a laboratory version of the task of screening carry-on luggage. The Department of Homeland Security's Transportation Security Administration supplied jpeg X-ray images of empty bags and a wide array of isolated objects such as laptops, pillboxes, toys, containers, and clothing. Colors in these images were coded according to the atomic density of the material X-rayed, with blue indicating metal, orange indicating organic material (plastics, clothing, food), and green indicating materials of intermediate density. Using Adobe Photoshop, we generated artificial packed bags. For each trial, we took an image of an empty bag and added objects into it that could overlap in a transparent manner. Although it must be noted that the final overlap of images created via Adobe Photoshop was not 100% identical to the overlap of objects in actual luggage at screening stations, our algorithm for superimposing objects attempted to approximate an actual packed bag as closely as possible. Objects were assigned and placed at random within the bags. The bags varied in size but only to a small extent. A target 'threat object' was digitally superimposed on select bags as described below. The orientations of all targets were held constant throughout the experiment.

Scaling luggage images for clutter

The purpose of this procedure was to ensure that the images were comparable in clutter. 'Clutter' was assessed along three parameters: the number of objects in each bag, the density of objects (or, the amount of overlap among objects) in any specific area in the bag, and the amount of empty (white) space ('uncluttered areas') in the entire bag. A total of 500 images of bags were generated overall; 30 participants observed all 500 images and rated their perceived clutter of each image on a scale of 1 (not cluttered at all) to 5 (extremely cluttered) using the parameters previously mentioned. There was no time limit for this task, but most participants on average finished rating the bags in less than 2 hours. Bags that received clutter ratings of less than '4' were digitally modified to increase the clutter. Out of these, a total of 100 bags with comparable ratings of 4 and above were selected for the final task.

Difficulty scaling and categorization of targets

Participants ($n = 10$) were presented with one specific X-ray image of luggage from the set selected previously, forty times in succession on a computer screen. A neutral gray screen appeared between trials to minimize carryover effects. On each of the 40 trials, a new threat object (described below) was embedded in the same luggage image. The participants' task was to click on the target (in the luggage image) as soon as they detected it. In addition, participants rated their perceived difficulty of detecting each target on a scale of 1 (cannot detect the target at all) to 5 (can detect the target very easily). Participants were shown digital images of each of the targets before they began the task and were allowed to refer to the target images at any time during the task. Although there was no time limit for this task, participants were instructed to spend a reasonable amount of time scanning each image and to ensure that the scan times across images were comparable. Most participants on average completed the task in less than an hour and a half. In a separate test, participants ($n = 10$) also categorized each of the 40 threat objects into five possible categories based on the dimensions of color and shape—guns, knives, sharp glass objects, scissors and metal tools.

For the final selection of targets, we excluded 'traditional' weapon categories such as guns and knives from the array and instead focused on two unconventional object categories—'sharp glass objects' and 'metal tools' from the above pretest. An example of an x-ray image of luggage and the selected targets are presented in Figure 1. Within these two categories, we then selected a subset of 10 targets that received difficulty ratings between '3' and '4' ($M = 3.55$, $SD = 0.12$), thereby ensuring that the selected targets were neither overly easy nor impossible to detect. The selected targets also had average detection times between 1.2 seconds and 3 seconds ($M = 2.$ 26 seconds, $SD = 0.89$), thereby ensuring that the targets could be detected within the maximum exposure time of 4 seconds per bag in the actual timed luggage screening task.

The luggage screening task

Participants

Participants ($n = 75$) comprised undergraduate and graduate students from Carnegie Mellon University and community members from the city of Pittsburgh (Pennsylvania). This sample size is based on a power analysis conducted using a power of.80, with a medium effect size, at an alpha level of.05 (Keppel & Wickens, 2007). Participants ranged between the ages of 18 to 45 and were both men and women with normal or corrected-to-normal vision. The task was a two-day experiment wherein participants were asked to play the role of airport security screeners and identify threats within X-ray images of passenger luggage. After completing the experiment, participants were given \$15 plus a bonus between \$0 and \$6 depending on their performance. The bonus was calculated as \$0.50 for every trial block where they generated correct diagnoses on at least 50% of the trials (across a total of 12 trial blocks on day 1). Participants were not informed about how the bonus was calculated and did not receive any performance bonus on day 2 of the experiment.

Experimental design

The luggage screening simulation was presented on a 19-in. monitor with a resolution of 800×600 pixels and Dell Pentium PC. Participants were randomly assigned to one of four experimental groups, each corresponding to a different incentive frame: hit-sensitive (HIT), miss-sensitive (MISS), equal-incentives (EQUAL), and no-incentives (NONE). In addition, participants were assigned to a control group that learned the task differently from the experimental groups (as described in the procedure in the succeeding section). Points were given or taken away for hits and/or misses and served as incentives depending on the framing 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 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

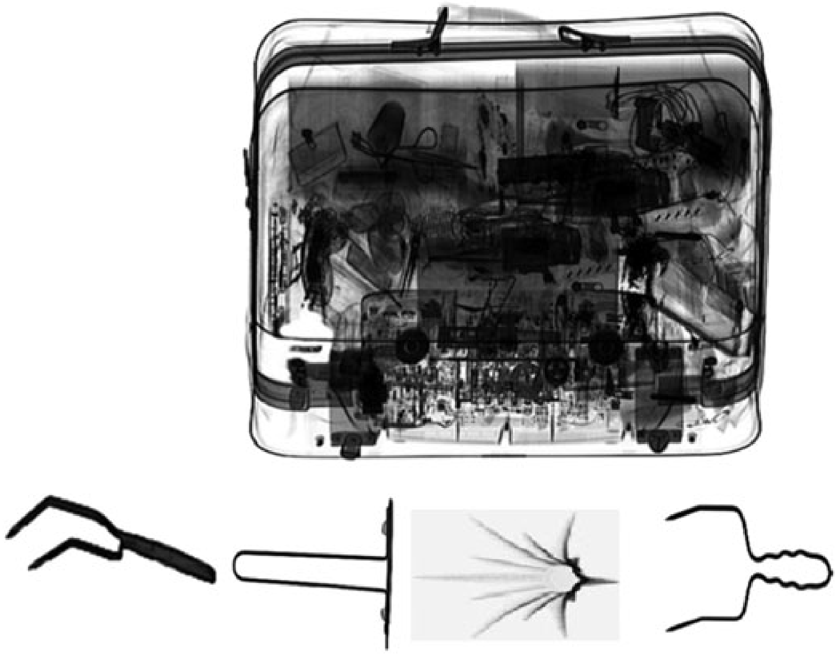

Figure 1. Sample luggage image and some targets used in the training phase

were derived based on signal detection theory as described below (Wickens & Hollands, 2000).

Incentive frames

As stated above, the incentive frame was defined by the number of points gained or lost, which in turn, determined how much performance bonus money was given at the end of the experiment. The logic was based on the signal detection parameter of response criterion setting or the ratio of 'target present' responses to 'target absent' responses. Optimal response criterion setting refers to the 'ideal' balance between 'present' and 'absent' responses that must be attained in order to minimize the tradeoff between hits and false alarms in a target detection context.

For each incentive frame, optimal response criterion setting (β_{opt}) was determined from Equation (1) (Wickens & Hollands, 2000). In Equation (1) , the quantities $p(N)$ and $p(S)$ refer to the noise and signal base rate, respectively. Values of correct rejection outcomes and hit outcomes are referred to as $V(CR)$ and $V(H)$, respectively. Costs of false alarms and misses are referred to as $C(FA)$ and $C(M)$, respectively.

$$\beta_{\text{opt}} = [p(N)/p(S)] * [\{V(CR) + C(FA)\}/\{V(H) + C(M)\}] \quad (1)$$

Table 1. Incentive frames

Incentive frame	Outcome (points)			
Group	Hit	Miss	False alarm	Correct rejection
Hit-sensitive (*Gain*)	+350	−50	−24	+1
Miss-sensitive (*Loss*)	+50	−350	−24	+1
Equal-incentives (*Neutral*)	+1	−1	−1	+1
No-incentives	0	0	0	0
Control (*self-training*)	0	0	0	0

In Equation (1) , if the probability of weapon presence (in a hypothetical luggage screening context) is .5, the monetary value associated with each hit and correct rejection is $10 and $5 respectively, and the monetary cost associated with each miss and false alarm is $10 and $5, respectively, β_{opt} would be calculated as follows:

$$\beta_{\text{opt}} = [1\text{–}.5/.5] * [5 + 5/10 + 10] = .50$$

The aforementioned equation was used to derive the optimal criterion setting and incentive distribution for each experimental group. A natural logarithmic transformation was then performed on β_{opt} to account for floor and ceiling effects.

1. The HIT group: Hit outcomes received the greatest point change (+350) in the HIT group and thus, this was thought of as the 'gain' or positive frame in which participants gained seven times more points for hits than they lost for misses.
2. The MISS group: Miss outcomes received the greatest point change (–350) in the MISS group, and thus, this was thought of the 'loss' or negative frame, because participants lost points with misses seven times more than they gained points for hits.
3. The EQUAL group: All outcomes in the EQUAL group received either +1 (for hits and correct rejections) or −1 (for misses and false alarms). Thus, losses and gains were balanced.
4. The NONE group: The NONE group received no points at all.
5. The control group received no training at all.

The HIT and MISS groups had an identical optimal response bias measure (ln $\beta_{\text{opt}} = -1.69$), meaning that the problems were structurally identical and only differed in the framing of the problem. The EQUAL group's optimal response bias (+1.69) corresponded to a strategy to maximize decision accuracy, regardless of signal presence.

The incentive frame in Table 1 was created using a number of assumptions related to the task having a signal base rate below chance (20%) and the representation of the task towards airline luggage screening. First, hits and misses were considered more valuable (higher absolute value) than false alarms and correct rejections. Second, false alarms costs were considered more valuable than correct rejections. Because non-signal trials occurred four times as often as signal trials, any changes in correct rejection and false alarm outcome values would have four times the effect on optimal beta than the same value change for hits and misses. With a target base rate of 20%, there were four times as many chances of generating correct rejections and false alarms because there were more luggage bags with no weapons as is characteristic of the actual luggage screening context. As a result, in this study, it was not feasible to create 'false-alarm-sensitive' and 'correct-rejection-sensitive' incentive frames that fit the previously mentioned assumptions and contained an identical optimal response bias with the HIT and MISS structures.

Dependent variables consisted of hit rate, which is the probability of correctly identified weapons out of a 20% base rate (six in every block of 30 bags) and false alarm rate, which is the probability of incorrectly designating a non-weapon as a weapon. Other dependent variables included response time for hits and false alarms measured in seconds, sensitivity and response criterion setting, and decision confidence measured on a scale from 0 to 5.

HYPOTHESES

Given the complexity of the luggage screening task, we hypothesized that participants who received training would perform significantly better than participants who attempted to self-train. In luggage screening, the ultimate goal is to localize and detect a target (i.e. find a weapon) when it is present. Therefore, ideally, both rewards (giving points) for finding targets and punishment (deducting points) for missing targets should have similar effects on maximizing hits and curtailing misses. However, we hypothesized that rewards and punishment will nevertheless have different effects on decision-making behavior because rewards represent a 'gain frame' and punishment represents a 'loss frame'. Specific hypotheses were as follows:

Hypothesis 1

Because the emphasis for the HIT group is on maximizing hits, participants in this group will say 'yes' more often to weapon detection and generate more hits than the EQUAL and NONE groups; this strategy will eventually lead the HIT group to generate several hits but also several false alarms.

Hypothesis 2

The MISS group, with the emphasis on minimizing misses, will behave similarly to the HIT group and say 'yes' more often to weapon detection and generate more hits than the EQUAL and NONE groups. However, the MISS group will also make an effort to increase correct rejections relative to the HIT group because of the overall 'losses' frame.

Hypothesis 3

The EQUAL group will attempt to balance their 'yes' and 'no' responses and generate equal hits and correct rejections.

Hypothesis 4

Regarding the NONE group, there are two possibilities: (i) Participants will balance their 'yes' and 'no' responses such that they generate an equal number of hits and correct rejections; and (ii) Alternatively, participants will weigh context more heavily (hits being more important than correct rejections for weapon detection in luggage screening) and behave similarly to the HIT group, thereby generating several 'yes' responses and several hits.

PROCEDURE

Day 1: Training

Participants were randomly assigned to one of the four experimental groups: HIT, MISS, EQUAL and NONE. Participants were given written instructions specifying a role as an airport screener with the task description of finding weapons. This instruction set also described the assigned outcome or incentive structure (HIT, MISS, EQUAL, NONE). Participants were only told how many points they would lose or gain and were instructed to maximize their score in order to receive a higher performance bonus ranging from $0 to $6.

Participants then performed the visual search task consisting of 12 trial blocks with 30 trial images (bags) in each, summing up to a total of 360 images for the training session. For the four experimental groups, a unique four-item target set (randomly drawn from the set selected in the pretest) was shown to the participant before each block, and all weapons were drawn from the four-item target set. An example of a luggage image and some examples of target objects are presented in Figure 1. Only 20% of the images in each block (six out of 30) contained a signal item from the four-item target set.

At the beginning of each trial, a luggage image was presented for 4 seconds. During the 4 seconds, if participants detected a target, they responded by physically moving the cursor to the target location inside the bag image and clicking on it. An overt response was required only on trials in which participants detected a target. If they did not respond within 4 seconds, the trial timed out. A click on the target was automatically recorded as a hit; a click on any other part of the X-ray image was recorded as a miss (on target-present trials) or as a false alarm (on target-absent trials).After each trial, participants' degree of confidence was collected on a 6-point Likert scale with 0 representing 'not confident at all' and 5 representing 'extremely confident'. After the confidence rating, feedback for the outcome of their decision was provided by showing the luggage bag with a red rectangular box highlighting the target when present. They were also informed via text message on the computer screen whether their decision was a hit, miss, false alarm or correct rejection. Participants were then asked to proceed to the next trial. A summary screen was shown after every two blocks (60 images) containing the current performance bonus and the cumulative rates for hits, misses, false alarms and correct rejections. The training phase lasted 60 minutes on average.

In addition to the four experimental groups, the control group detected the same targets as the experimental groups but without training. Specifically, control participants were not shown the four-item target set at the beginning of each trial block. Instead, they were required to independently extrapolate what the targets were and detect them. At the end of each trial, they received the same form of feedback as the experimental groups. Therefore, they could potentially use this feedback as a means to train themselves (or 'self-train') on the task. The objective was to create a baseline for participants' ability to learn the task without training with specific target-set exposures and to establish whether they have the ability to self-train in the absence of training stimuli.

Day 2: Transfer

Participants returned for the transfer phase on the following day at the same time set for the training phase. These objects were also drawn from the original set of 10 targets from the pretest. It was ensured that these 'novel' targets had never been presented to participants during training, and participants were not shown any memory sets at the beginning of the transfer block. Although the transfer targets were novel, they belonged to the same categories of 'metal tools' and 'sharp glass objects' as the training targets (see Figure 1). The difficulty of detection and category membership of all targets (training and transfer) was pretested as described earlier. It is important to note, however, that participants were not provided any information regarding stimulus categories or any clues regarding the physical identity of the targets during transfer. The novel targets are illustrated in Figure 2.

Additionally, the incentive structure was not given to participants on day 2; instructions simply stated that they should 'use previous training to maximize their perceived point total.' Participants then performed a shorter version of the visual screening task for six blocks of 30 trials (180 images). Again, 20% of the trials (six out of 30) contained a signal item. Outcome feedback emulated the training phase on day 1 by displaying the luggage bag with a red rectangular box highlighting the target when present and informing the participant of a hit, miss, false alarm or correct rejection. The difference between the two methods of outcome feedback (training versus transfer) was that there was no score or accumulated performance bonus given for the transfer day. On average, the transfer phase lasted 30 minutes.

Figure 2. Novel target items

The control group performed the same task as the experimental groups on day 2. It is important to note that on day 2, the experience of the screening task for the control group was very similar to their experience of the task on day 1, that is, a detection task with no target-set at the beginning of each block.

RESULTS

Analyses revealed that the block to block shifts in performance on each of the 2 days (training and transfer) were gradual and did not differ significantly across consecutive blocks for most participants. Therefore, the results below are averaged across blocks and are presented as a consolidated 'training phase' and 'transfer phase'. We used a two-way analysis of variance (ANOVA) to analyze performance across training and transfer followed by independent and paired-sample t-tests for *post hoc* comparisons between and within groups, respectively (Maxwell & Delaney, 2003). All results with alpha values below .05 are discussed as statistically significant, and alpha values between.051 and.090 are discussed as marginally significant. In Figures 3–8, group 1 represents the HIT group, group 2 represents the MISS group, group 3 represents the EQUAL group, group 4 represents the NONE group, and group 5 represents the control group.

Hit rate

The results for hit rates are illustrated in Figure 3. A 2 (Phase: training versus transfer) × 5 (Incentive frame: HIT, MISS, EQUAL, NONE, control) ANOVA on hit rates revealed significant main effects for phase, $F(1,69) = 166.62$, $p < .001$, partial $\eta^2 = 0.707$, and for incentive frame, $F(4,69) = 2.59$, $p < .05$, partial $\eta^2 = 0.131$. These main effects formed a significant interaction between phase and incentive frame, $F(4,69) = 21.32$, $p < .001$, partial $\eta^2 = 0.503$. *Post hoc* t-tests revealed that, as illustrated in Figure 3, all four groups that received training demonstrated a reduction in hit rates from training to transfer following the inclusion of novel targets (HIT group: $t(14) = 8.86$, $p < .001$; MISS group: $t(14) = 5.31$, $p < .001$; EQUAL group: $t(14 = 7.63$, $p < .001$; NONE group: $t(14) = 10.86$, $p < .001$). However, control participants who

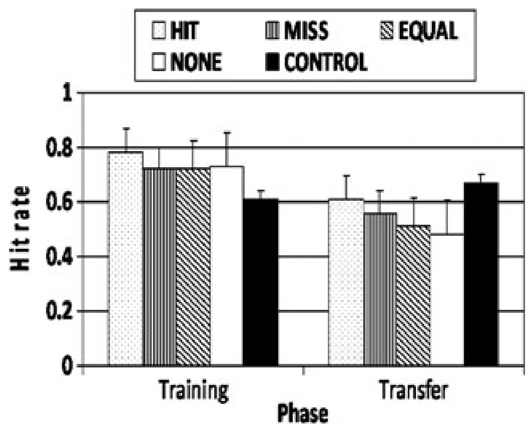

Figure 3. Hit rates

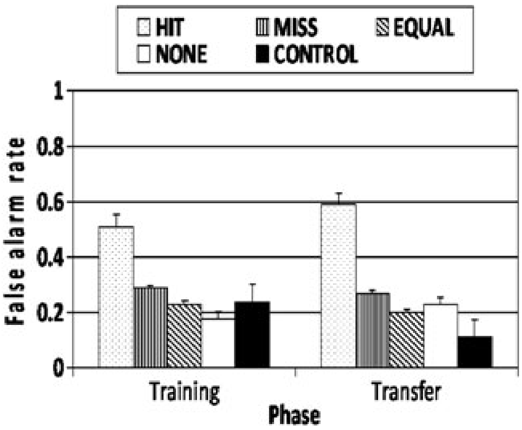

Figure 4. False alarm rates

self-trained demonstrated a significant increase in hit rate from training ($M = 0.61$, $SD = 0.087$) to transfer ($M = 0.68$, $SD = 0.13$), $t(14) = 2.436$, $p < .05$.

False alarm rate

The results for false alarm rates are illustrated in Figure 4. A similar 2×5 ANOVA on false alarm rates did not reveal a significant main effect for phase, $F(1,69) = 0.286$, $p = .82$, partial $\eta^2 = 0.004$. However, there was a significant main effect for incentive frame, $F(4,69) = 6.546$, $p < .001$, partial $\eta^2 = 0.275$. There was also a significant interaction between phase and incentive frame, $F(4,69) = 2.87$, $p < .05$, partial $\eta^2 = 0.143$. As depicted in Figure 4 and contrary to the pattern for hit rates, participants who received training either exhibited a marginally significant increase in false alarm rates at transfer (NONE group: $t(14) = 1.90$, $p = .07$) or did not vary significantly from training to transfer (HIT, MISS and EQUAL groups). However, control participants who self-trained demonstrated a significant reduction in false alarm rates from training ($M = 0.241$, $SD = 0.15$) to transfer ($M = 0.115$, $SD = 0.17$), $t(14) = 3.51$, $p < .01$.

Sensitivity (d′) and response criterion setting (c)

The data for hits and false alarms are represented within the signal detection framework of sensitivity and response criterion setting in Tables 2a and 2b, respectively. Because the hypotheses were conceptualized primarily as hits and false alarms, we do not present an in-depth discussion of sensitivity and criterion settings. However, as presented in Tables 2a and 2b, in keeping with the patterns for hits and false alarms described previously, control participants demonstrated the highest levels of sensitivity ($M = 1.67$, $SD = 0.15$) accompanied by liberal criterion settings ($M = 1.48$, $SD = 0.61$) at transfer, whereas the HIT group demonstrated the lowest level of sensitivity ($M = 0.051$, $SD = 0.16$) accompanied by relatively conservative response criteria ($M = 0.78$, $SD = 0.22$) at transfer.

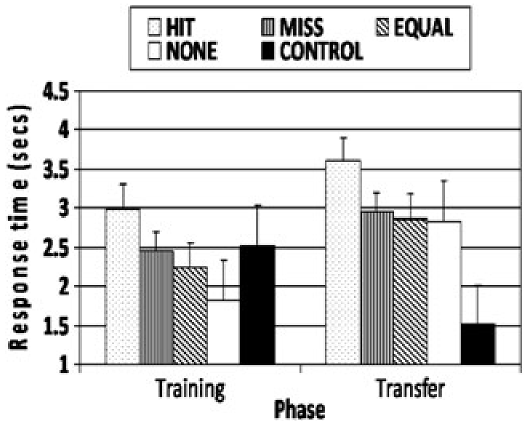

Figure 6. Response time for false alarms

Response time for hits

The results of response time for hits are illustrated in Figure 5. The 2×5 ANOVA on response times for hits revealed a significant main effect for phase alone, $F(1,69) = 9.631$, $p < .05$, partial $\eta^2 = 0.122$. The main effect for incentive frame was not significant, $F(4,69) = 0.826$, $p = .513$, partial $\eta^2 = 0.046$. The interaction between phase and incentive frame was significant, $F(4,69) = 0.826$, $p < .05$, partial $\eta^2 = 0.126$. *Post hoc* tests revealed that the HIT, MISS and control groups displayed no changes in response time for hits from training to transfer. However, the EQUAL and NONE groups demonstrated significant decreases in response time for hits from training to transfer indicating significantly faster responding during transfer (EQUAL group: $t(14) = 1.93$, $p < .05$; NONE group: $t(14) = 3.17$, $p < .01$).

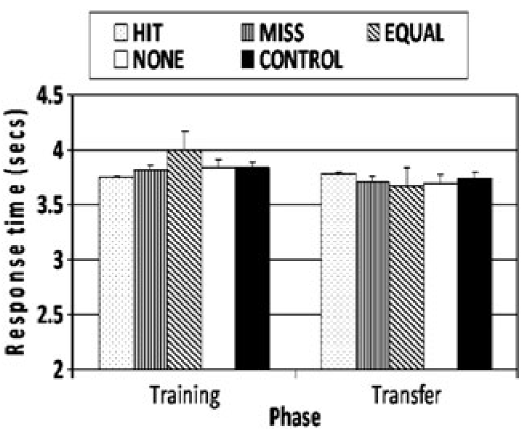

Figure 5. Response time for hits

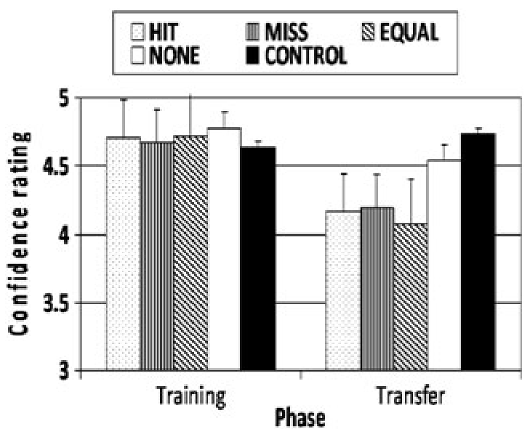

Figure 7. Confidence when generating hits

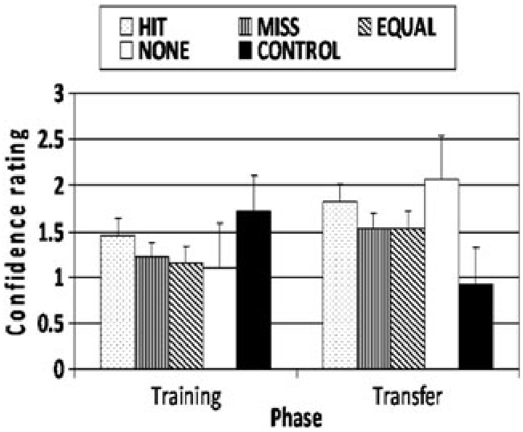

Figure 8. Confidence when generating false alarms

Response time for false alarms

The results of response time for false alarms are illustrated in Figure 6. A similar 2 × 5 ANOVA on response times for false alarms revealed a significant main effect for phase, $F(1,69) = 5.59$, $p < .05$, partial $\eta^2 = 0.075$. Contrary to the pattern for hits, response times for false alarms during training ($M = 2.41$ seconds, $SD = 0.128$) were significantly lower than response times for false alarms during transfer ($M = 2.75$ seconds, $SD = 0.138$). There was also a significant main effect for incentive frame, $F(4,69) = 4.84$, $p < .001$, partial $\eta^2 = 0.219$. These main effects can be explained further via the significant interaction between phase and incentive frame, $F(4,69) = 5.85$, $p < .001$, partial $\eta^2 = 0.253$. *Post hoc t*-tests revealed that participants in the HIT and NONE groups demonstrated either significant or marginally significant increases in response time for false alarms from training to transfer (HIT group: $t(14) = 1.77$, $p = .09$; NONE group: $t(14) = 2.95$, $p < .05$). The MISS and EQUAL groups did not demonstrate any statistically significant changes in response time for false alarms from training to transfer. However, the control group alone displayed a significant decrease in response time for false alarms from training ($M = 2.52$ seconds, $SD = 1.28$) to transfer ($M = 1.51$ seconds, $SD = 0.99$), $t(14) = 3.13$, $p < .01$.

Subjective confidence when generating hits

At the end of each trial, participants rated their confidence on a scale of 0 (not confident at all) to 5 (extremely confident). The results are illustrated in Figure 7. The 2 × 5 ANOVA on subjective confidence revealed significant main effects for phase, $F(1,69) = 51.96$, $p < .001$, partial $\eta^2 = 0.430$, and for incentive frame, $F(4,69) = 6.99$, $p < .001$, partial $\eta^2 = 0.288$.

Table 2a. Sensitivity (d')

	Sensitivity	
Incentive frame	Training	Transfer
Hit-sensitive (*HIT*)	0.75	0.05
Miss-sensitive (*MISS*)	1.14	0.76
Equal-incentives (*EQUAL*)	1.32	0.87
No-incentives (*NONE*)	1.53	0.69
Control (*self-training*)	0.98	1.67

Table 2b. Response criterion settings (c)

	Criterion setting	
Incentive Frame	Training	Transfer
Hit-sensitive (*HIT*)	0.67	0.78
Miss-sensitive (*MISS*)	0.99	1.26
Equal-incentives (*EQUAL*)	1.08	1.50
No-incentives (*NONE*)	1.16	1.48
Control (*self-training*)	1.24	1.48

These main effects are explained further by the significant interaction between phase and incentive frame, $F(4,69) = 6.88$, $p < .001$, partial $\eta^2 = 0.288$. As depicted in Figure 7, *post hoc* tests revealed that all participants who received training demonstrated significant reductions in subjective confidence when generating hits from training to transfer (HIT group: $t(14) = 4.93$, $p < .001$; MISS group: $t(14) = 4.66$, $p < .001$; EQUAL group: $t(14) = 3.99$, $p < .005$; NONE group: $t(14) = 2.35$, $p < .05$). However, control participants who self-trained demonstrated the same levels of decision confidence during transfer that they did during the training phase. This transfer confidence was significantly higher than that of any of the trained groups during the transfer phase.

Subjective confidence when generating false alarms

In general, participants were significantly less confident when generating false alarms than when generating hits. The data for false alarm confidence are illustrated in Figure 8. Because false alarm confidence was overall significantly lower than hit confidence, the scale in Figure 8 is different from that used in Figure 7. The 2 × 5 ANOVA on subjective confidence for false alarms revealed a significant main effect for phase, $F(1,69) = 166.62$, $p < .001$, partial $\eta^2 = 0.707$. There was no significant main effect for incentive frame, $F(4,69) = 0.657$, $p = .624$, partial $\eta^2 = 0.037$. However, within the training phase alone, the control group had a significantly higher level of false alarm confidence ($M = 1.72$, $SD = 0.82$) than any of the trained groups. The interaction between phase and incentive frame was significant, $F(4,69) = 8.88$, $p < .001$, partial $\eta^2 = 0.340$. As depicted in Figure 8, and contrary to the pattern for confidence on hit trials, *post hoc t*-tests revealed that most participants who received training exhibited significant or marginally significant increases in subjective confidence from training to transfer when generating false alarms (HIT group: $t(14) = 1.89$, $p = .07$; EQUAL group: $t(14) = 4.10$, $p < .05$; NONE group: $t(14) = 3.96$, $p < .005$). However, control participants displayed a significant reduction in false alarm confidence from training to transfer ($M = 0.093$, $SD = 0.68$), $t(14) = 3.25$, $p < .01$.

DISCUSSION

The purpose of this study was twofold: to examine how training with incentives compared with self-training in a simulated airline baggage screening task; and to compare the effectiveness of different incentives during training on learning and transfer of learning.

Incentive-based training versus self-training

Not surprisingly, on day 1, participants who self-trained (i.e. the control group) performed significantly worse than participants who received training. The control group generated fewer hits, was less confident when generating hits and more confident when generating false alarms during training. These results indicate the value of training procedures that include target exposures during the process of skill acquisition. However, the most interesting and perhaps startling result is that, contrary to our hypothesis, the control group demonstrated a substantial improvement in performance from training to transfer such that control participants actually outperformed participants who had received incentive-based training. This was manifested in higher hit rates, lower false alarm rates, and lower response times for false alarms at transfer. The control group also appeared to have had the highest level of performance awareness as was indicated by the significantly high confidence when generating hits and proportionately low confidence for false alarms on day 2 of the study.

It is possible that the relative improvement of the control group from training to transfer might be some function of the target exposures shown to the trained participants alone during training, which we will discuss in detail later. However, we believe that it is not entirely the case. Our earlier research that has used the training-transfer paradigm in luggage screening (Madhavan & Gonzalez, 2010) has revealed that providing a practice 'memory set' during training leads to superior performance than when such a memory set is not provided. However, the results of this study revealed the opposite, with trained participants performing worse than control participants. There could be two possible explanations for this finding, both of which are related to the incentives provided to trained participants during the first few blocks of the experiment.

First, that lack of specific incentives for the control group likely minimized the 'pressure' to perform and actually reduced cognitive demands (Tversky & Kahneman, 1981) relative to the groups that did receive incentives. Their confidence could be a clear 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. However, the control group performed even better than the NONE group, which did not receive incentives either. Therefore, a more plausible explanation for the control group's superior transfer performance could be that not priming them with specific targets during training helped participants in this group to effectively create their own visualizations of threat objects over time; this subsequently improved their ability to connect targets during training to new targets at transfer (Gonzalez, Lerch, & Lebiere, 2003). Contrary to the pattern for the control group, the introduction of novel targets hindered performance for participants in the experimental groups (who were trained with specific target exposures).

Object-based training versus category-based training

In some ways, our experimental design can be reinterpreted as 'object-based training' (wherein the trained groups searched only for the specific objects that were shown to them) versus broader 'category-based training' (wherein the control group derived their own representations of more generalized categories of targets). The performance differences between the control group and trained groups can be explained, at least in part, by the 'typicality effect' (Castelhano, Pollatsek, & Cave, 2008) that explains search processes when people are only given a category as a search cue. For instance, when a category is mentioned, prototypical exemplars (of that category) are retrieved easily from memory (Rosch, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Thus, when target search is driven by a broad category, a prototypical exemplar (from that category) might serve as the template for search. Presumably, search would be easier for targets that are prototypical than for those that are atypical. If search can be guided on the basis of only category membership, it is possible that target detection will be more efficient for people who possess a clear idea of what the 'general category (of desired objects)' represents (in our study, the control group; Smilek, Dixon, & Merikle, 2006) relative to people who are exposed to only object-specific cues (in our study, the trained groups).

Theories of similarity and decision making support the typicality effect by suggesting that the effectiveness of future decisions is directly determined by the similarity between memory traces of past decisions (instances) and the current decision situation (Gonzalez et al., 2003). In this study, presenting one set of objects consistently as targets (for the experimental groups) likely increased the memory representation of those objects as targets, thereby making it more difficult for them to generalize the knowledge to novel objects that were different from the original targets. However, when participants (in the control group) had the freedom to derive their own mental representations of targets, their definitions of 'threat objects' likely became more broadly represented in memory. This possibly increased the subjective preparedness to perceive physically different targets during transfer, which potentially led to improvements in the ability to distinguish between target and nontargets and proportionate changes in decision confidence.

Search termination

One important issue to consider here is that of search termination during unsuccessful searches (or, when no target is detected). Researchers have proposed two different strategies for deciding when a target is not present without having to exhaustively search the entire display (based on the original Guided Search model; Chun & Wolfe, 1996). The *activation threshold* hypothesis suggests that people may simply search through the distractors that have a certain likelihood of being a target and ignore those items that are less similar to the target. Another hypothesis is that as an observer performs a visual search task, he/she may develop some internal estimate of how long it takes to find a target that may enable the observer to make 'educated guesses' (because the probability of a guess being correct increases as evidence is accumulated as a search trial progresses). According to this *timing hypothesis*, observers will terminate a trial when the duration of the trial exceeds some duration threshold based on the assumption that the target should have been found by then.

Either of the aforementioned hypotheses could have impacted performance in our study. Based on the *timing hypothesis*, it is possible that self-training favored performance as participants had more time to make 'educated guesses' without the additional task of matching prototypes with the objects in the luggage. On the other hand, the *activation hypothesis* dictates that incentive-based training afforded participants better opportunities to evaluate the likelihood of an object being a target (because of their prior exposure to prototypes). It must be noted that in our paradigm, participants did not have the opportunity to 'log in' a decision on trials in which they missed a target. That is, a miss was automatically recorded when a target-present trial timed out without participants overtly making a response. In some rare cases, a miss was also recorded when participants incorrectly clicked on the wrong object within a target-present bag. Therefore, the only times when participants terminated their own search was on trials in which they believed (correctly or incorrectly) that they had successfully detected a target and clicked on it. They did not have any opportunity to terminate their own 'unsuccessful searches' making it difficult to narrow down the real reasons for search termination in this paradigm.

Incentive frames and performance

Despite the discussed evidence for the superiority of self-training over incentive-based training in this study, it is possible that self-training may not be the best option for the real world luggage screening task wherein targets are more varied and task duration is substantially longer. The results of this study also provide some support for the effectiveness of training and incentives. Specifically, results revealed that among the trained groups providing certain forms of training coupled with incentives (i.e. the four trained groups) did help transfer performance primarily by maximizing hits. This was evidenced by the fact that the HIT and MISS groups generated more hits during both training and transfer in comparison to the EQUAL and NONE groups. Evidently, giving specific directional incentives (in other words, reward or punishment) did improve transfer performance in comparison to giving equal incentives or no incentives at all.

Participants in the EQUAL group were unaware of which performance outcome was more beneficial—generating more hits and risking false alarms, or, alternatively, curtailing false alarms by allowing misses. As a result the EQUAL group generated neither the maximum hits nor the fewest false alarms relative to the other incentive groups, during either phase of the experiment. The NONE group in this study appeared to have simply performed the task without focusing on a specific outcome as is indicated by the lack of a clear pattern of performance for this group. This group demonstrated a significant drop in hit rates from training to transfer that was not accompanied by corresponding change in false alarm rates; contrary to this pattern, they demonstrated a dramatic increase in response time for false alarms from training to transfer but their response time was not impacted for hits; yet, these participants demonstrated unusually high confidence during transfer relative to training when generating false alarms, whereas the opposite was observed for hits. These dramatically differing response patterns are difficult to interpret and suggest that contrary to hypotheses, the NONE group does not appear to have been systematically influenced by context or base rate of target stimuli.

Although the incentive frames for HIT and MISS increased participants performance leading them to consciously maximize their hits; rewarding hits implicitly primed HIT participants to say 'yes' more often. This resulted in a significantly higher proportion of false alarms particularly at transfer compared with all other groups. Rewarding hits also led participants to spend a significantly longer time searching for weapons that were not present, indicated by their lengthy response times for false alarms. Therefore, although rewards were effective in motivating participants to detect targets, they certainly hindered performance by priming participants to invest more time and cognitive resources in searching for weapons in bags that did not contain weapons.

More importantly, the HIT group was not significantly higher in hit rates in comparison with the MISS group, implying that deducting points for misses (or, using punishment as incentive) was just as effective for detecting weapons as providing rewards. However, penalizing (or punishing) participants for misses (the MISS group) was also effective in curtailing their 'yes' response and consequently succeeded in reducing their false alarm rates. Additionally, the time spent searching for weapons was significantly lower when points were deducted for misses. Interestingly, punishment in the form of point deduction for misses had a more powerful effect on balancing hits and false alarms than rewards in the form of point allotment.

Gain frames versus loss frames

The better performance of the MISS group over the HIT group is evidence of the *loss frame superiority* effect where decision makers naturally lean toward being loss averse and are willing to take extra measure to avoid a loss or punishment rather than achieve a gain of the same size (Tversky & Kahneman, 1983). A recent study by Kern and Chugh (2009) revealed that the cognitive state created by a loss frame leads people to cheat, take mental 'shortcuts', and demonstrate a general propensity toward unethical behavior. This is frequently exacerbated by the presence of time pressure or a demanding deadline (Stanovich & West, 2002). Although there was little opportunity for cheating or unethical behavior in this paradigm, participants in the loss frame (i.e. the MISS group) did demonstrate a higher level of conservatism in their decision making by curtailing their 'yes' responses in situations where they were perhaps unsure of the presence of a target. This can be interpreted as a manifestation of loss aversive or punishment-aversive decision-making behavior.

Contrary to the findings of Lacson et al. (2005), providing both hit and miss sensitive information simultaneously (i.e. for the EQUAL group) did not impact performance as strongly as the loss or gains frame alone. Based on signal detection theory, equal incentives in this study were operationalized such that the actual point change was + or −1 point and not + or −350 as in the case of the HIT and MISS groups, respectively. It is possible that the relative differences in points allotted across groups diluted the strength of the framing effect for the EQUAL group relative to the HIT and MISS groups. Alternatively, this suggests that implementing both reward and punishment simultaneously weakens the effect of incentives on goal-directed performance.

Some researchers have likened the cognitive effect of framing effects to the development of automaticity (De Martino, Kumaran, Seymour, & Dolan, 2006)—under conditions of high time pressure, decision makers tend to take mental shortcuts or heuristics to make decisions. These findings are supported by the original premise of Tversky and Kahneman (1981) that framing effects are immediate and automatic and tend to disappear when decision makers are given time and opportunity to deliberate and fully process information. It is therefore logical that in the absence of complete information processing, humans tend to react more easily to the implied threat of losses versus the implied utility of gains.

The airline baggage screening paradigm is a classic example of a highly time pressured task wherein operators have little time to process information in detail. The recent study by Madhavan and Gonzalez (2010) has shown supporting evidence for such automaticity development in the luggage screening process. They found that, often, decisions to stop or pass a piece of luggage are based on heuristics (or, an extrapolation that a new object 'appears' similar to an object seen earlier) rather than on any certainty that a weapon is indeed present. This provides a possible explanation for why participants in this study demonstrated better performance when incentives were framed as punishments rather than rewards.

An alternative explanation for the loss-superiority (or, punishment-superiority) effect in this study can be based on the use of affective and cognitive processes in the framing effect (Gonzalez et al., 2005). This theory emphasizes an inherent need of decision makers to make emotionally acceptable choices. In this study, the MISS group was faced with a 'sure loss' situation if they made an incorrect decision. The associated negative affect possibly caused the MISS group to engage in more thorough consideration of their decision options and consequently served as a strong motivator to improve performance. Overall, the fear of a loss evidently had a stronger effect on performance than the anticipation of a gain when decision making was inherently heuristic-based.

CONCLUSIONS AND IMPLICATIONS

There are some limitations in the design of this study, which affect the generalizability of the results to actual luggage screening contexts. The study used a relatively contrived situation where the participants were college students. Consequences of wrong decisions (i.e. textual feedback) do not compare with those faced when a threatening object is undetected by security systems at an actual airport. The experimental session was longer than the amount of time spent by screeners searching X-ray images at real airports (typically 30 minutes), and the orientation of targets was held constant, which, again, is not characteristic of targets in a real screening task. Research has revealed that orientation information can be used effectively in guiding target search, and it is important to vary target orientation realistically to disentangle training effects from orientation-related search in future paradigms.

The base rate of targets in this study (20% for statistical purposes) was higher than the base rate in the real world. Wolfe, Horowitz, and Kenner (2005) found that at a target prevalence (or base rate) of 50%, participants failed to detect targets on 7% of trials. The errors increased as prevalence decreased; 10% prevalence produced 16% errors, and errors soared to 30% at 1% prevalence. The authors reasoned that the 'low prevalence effect' occurred for low target base rate situations because as observers repeatedly respond with correct rejections (accurately indicating the absence of a target), they begin to terminate their searches more quickly, consequently missing targets on the rare occasions when they are present. The 20% target base rate in our study could potentially have influenced participants' response patterns to an extent and led to some portion of the errors observed.

Despite these limitations, the present study is novel in that it highlights the role of decision frames and biases in airline luggage screening (as opposed to earlier studies that focus solely on visual search issues). Our results suggest that when equal incentives are provided for hits (gains frame) and misses (losses frame), decision makers have an implicit tendency to respond more strongly to loss frames than gain frames. However, despite the effectiveness of incentives, the best transfer performance was achieved when participants self-trained and possibly formed 'category general' mental representations of stimuli without the constraints imposed by a finite training set. This research draws attention to the tradeoff between training for optimal training and the potential for maximal transfer when conditions of transfer differ from the conditions of training in a multitude of ways. Therefore, training programs must focus not only on the effectiveness of initial learning but also on the durability and transferability of the knowledge acquired when the training variables are modified.

ACKNOWLEDGEMENTS

This research was partially supported by the Multidisciplinary University Research Initiative Program (MURI; N00014-01-1-0677) grant to Cleotilde Gonzalez. The experiment reported here was done at Carnegie Mellon University in the Dynamic Decision Making Laboratory. We thank Varun Dutt for programming the visual search simulation used in this study. An earlier version of this work was reported at the Human Factors and Ergonomics Society meeting by Lacson, Gonzalez, and Madhavan (2008).

REFERENCES

Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy?: The effect of task demands and motivational factors on scanning strategy in visual search. *Journal of Vision*, *9*(3), 1–16.

Boot, W. R., McCarley, J. S., Kramer, A. F., & Peterson, M. S. (2004). Automatic and intentional memory processes in visual search. *Psychonomic Bulletin & Review*, *5*, 854–861.

Bregman, N. J., & McAllister, H. A. (1983). Skin temperature biofeedback: Evaluation of positive and negative monetary incentives. *The International Journal of Neuroscience*, *22*, 119–124.

Brockmole, J. R., & Henderson, J. M. (2006). Using real world scenes as contextual cues during search. *Visual Cognition*, *13*, 99–108.

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review*, *15*, 795–801.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Christ, S. E., & Abrams, R. A. (2006). Abrupt onsets cannot be ignored. *Psychonomic Bulletin & Review*, *13*, 875–880.

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.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*, 684–687.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *The American Psychologist*, *49*, 709–724.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology. Human Perception and Performance*, *18*, 1030–1044.

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 317–329.

Fridrici, M., Lohaus, A., & Glab, C. (2009). Effects of incentives in web-based prevention for adolescents: Results of an exploratory field study. *Psychology & Health*, *24*, 663–675.

Gelder, J. L., Vries, R. E., Pligt, J. (2009). Evaluating a dual-process model of risk: Affect and cognition as determinants of risky choice. *Journal of Behavioral Decision Making*, *22*, 45–61.

Gonzalez, C., Dana, J., Koshino, H., & Just, M. A. (2005). The framing effect and risky decisions: Examining cognitive functions with fMRI. *Journal of Economic Psychology*, *26*(1), 1–20.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*, 591–635.

Itti, L. (2006). Quantitative modeling of perceptual salience at human eye position. *Visual Cognition*, *14*, 959–984.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Keppel, G., & Wickens, T. (2007). *Design and analysis*, 5th edn. Upper Saddle River, New Jersey: Prentice Hall.

Kern, M. C., & Chugh, D. (2009). Bounded ethicality: The perils of loss framing. *Psychological Science*, *20*, 378–384.

Lacson, F. C., Gonzalez, C., & Madhavan, P. (2008). Framing and context effects in incentive structure training for visual search tasks. *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society*, (pp. 333–337). Santa Monica: CA.

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 Meeting 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.

Madhavan, P., & Gonzalez, C. (2010). The relationship between stimulus-response mappings and the detection of novel stimuli in a simulated luggage screening task. *Theoretical Issues in Ergonomics Science*, *11*(5), 461–473.

Maxwell, S. E., & Delaney, H. D. (2003). *Designing experiments and analyzing data: A model comparison perspective*, 2nd edn. London: Routledge Academic.

Meyerowitz, B., Chaiken, S. (1987). Effect of message framing on breast self-examination, attitudes, intentions and behavior. *Journal of Personality and Social Psychology*, *52*(3), 500–510.

Peterson, M. S., & Kramer, A. F. (2001). Contextual cueing reduces interference from task-irrelevant onset distractors. *Visual Cognition*, *8*, 843–859.

Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology. General*, *104*, 192–233.

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Scott, D., & Goldwater, B. C. (1998). Increasing attendance at sport training sessions. *Perceptual and Motor Skills*, *87*, 515–518.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Slovic, P., Peters, E., Finucane, M., & MacGregor, D. G. (2005). Negative affect, risk, and decision making. *Health Psychology*, *24*, 35–40.

Smilek, D., Dixon, M. J., & Merikle, P. M. (2006). Revisiting the category effect: The influence of meaning and search strategy on the efficiency of visual search. *Brain Research*, *1080*, 73–90.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421–440). New York, NY: Cambridge University Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and psychology of choice. *Science*, *211*, 453–458.

Tversky, A., & Kahneman, D. (1983). Choices, values, and frames. *The American Psychologist*, *39*, 341–350.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance*. 3rd edn. Upper Saddle River, New Jersey: Prentice Hall.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*, 439–440.

