

## The Impact of Target Base Rate on Training and Transfer of Learning in Airline Luggage Screening: An Examination of Three Base Rate Scenarios

POORNIMA MADHAVAN<sup>1</sup>\*, CLEOTILDE GONZALEZ<sup>2</sup> and JEREMY R. BROWN<sup>1</sup>

<sup>1</sup>346C Mills Godwin Building, Old Dominion University, Norfolk, VA, USA

<sup>2</sup>Porter Hall 223C, Carnegie Mellon University, Pittsburgh, PA, USA

**Summary:** We examined the extent to which transfer of training is impacted by target probability and physical similarity of training and transfer stimuli. Participants performed a luggage screening task where they trained on three base rates of weapon presence as follows: 100%, 50% and 20%; at transfer, participants detected the same weapons (Study 1) or novel weapons (Studies 2 and 3) at a base rate of 20%. In Study 3, we modified training such that all participants observed the same absolute number of weapons ( $n = 100$ ) across base rate conditions. In Study 1, different base rates did not significantly impact transfer. In Study 2, 100% base rates helped hit rates at transfer but increased false alarms via liberal shifts in decision criteria and slowed detection times. Study 3 revealed no effect of base rates on transfer hit rates; however, training on 100% base rates still led to liberal responding and inflated false alarms. Copyright © 2013 John Wiley & Sons, Ltd.

### INTRODUCTION

Target detection tasks are pervasive. They are as common as looking for fruit defects at the grocery store or finding an open teammate on the basketball court. Also, many of these target detection tasks are important and relevant for our health and society such as a radiologist searching for a tumour on an X-ray image, a soldier determining the presence of a combatant in unfamiliar terrain or more common to our current activities and an airport security officer looking for dangerous items in passenger luggage. Typically, low base rate is one of the major factors underlying the high error rates in visual inspection tasks such as luggage screening (Parasuraman, Warm, & Dember, 1987; Wolfe, Horowitz, & Kenner, 2005; Wolfe et al., 2007). In search tasks, low prevalence creates challenges for observers by providing fewer instances of target presence and consequently fewer opportunities to evaluate decisions regarding target presence or absence; these challenges are ultimately manifested in shifts in signal detection performance (Van Wert, Horowitz, & Wolfe, 2009). Therefore, it is important to understand what base rates to incorporate during training to optimise the detection of low base rate targets while balancing false alarms and detection time.

Existing research on base rates in visual search paradigms has revealed that higher base rates facilitate the probability of detection (Rhodes & Jacoby, 2007). Ideally, increasing the number of signals among distractors should increase the ‘attention attracting tendency’ of these signals (Shiffrin & Schneider, 1984), thereby leading to improved detection performance. Logically, the resulting enhanced detection performance should transfer better to low base rates compared with training under low target probabilities. This was supported by a recent study by Wolfe and colleagues (Wolfe et al., 2005), who examined this issue of low base rates (which they refer to as the ‘prevalence effect’) in a baggage screening paradigm. The researchers found that at 50% target prevalence,

participants failed to detect targets on 7% of trials. The errors increased consistently as prevalence decreased; 10% prevalence produced 16% errors and errors soared to 30% at 1% prevalence. These errors were primarily misses; false alarms were negligible, although they were found in later more complex screening tasks (Wolfe et al., 2007). Overall, reducing base rates led to a fourfold increase in error rate. The authors reasoned that the ‘low prevalence effect’ occurred for low base rate situations because as observers repeatedly respond with correct rejections, they begin to terminate their searches more quickly, consequently missing targets on the rare occasions when they are present.

If the discrimination of target and distractor is difficult, observers would be more likely to conclude that an ambiguous item was a distractor, if target prevalence is low. This would bias observers toward absent responses and miss errors in difficult tasks and would be manifested as a criterion shift (Rich et al., 2008). Therefore, in signal detection terms, the prevalence effect has been explained as a criterion shift and not a change in sensitivity (Wolfe et al., 2007). Although several efforts to induce observers to adopt a better criterion have failed, it has been proposed that a regime of brief retraining periods with high prevalence and full feedback will allow observers to hold an optimal criterion during periods of low prevalence (Wolfe et al., 2007).

Fleck and Mitroff (2007) extended the research by Wolfe et al., to examine the motor underpinnings of the ‘low prevalence effect’. They modified the paradigm Wolfe et al. by giving observers the option of correcting a previous response in order to differentiate between errors of perception and errors of action. They found that a large percentage of errors under low prevalence were errors of action or execution rather than errors of perception, that is, observers were cognizant of the fact that a target existed. However, they responded too quickly to the non-presence of targets because they were habituated to respond negatively because of the low prevalence of targets. We refer to this as the ‘premature search termination theory’.

There are two potential problems with the aforementioned research. First, the study by Wolfe and colleagues used the same targets at training and transfer. This is significantly less

\*Correspondence to: Poornima Madhavan, 346C Mills Godwin Building, Department of Psychology, Old Dominion University, Norfolk, VA 23529, USA.

E-mail: pmadhava@odu.edu

complex than an actual luggage screening task where threat stimuli can appear in various physical forms that might differ substantially from the objects used to train screeners. The only study that addressed the issue of unfamiliar targets specific to luggage screening (McCarley, Kramer, Wickens, Vidoni, & Boot, 2004) found that recognition skills developed by practice largely generalised to unfamiliar targets; hit rate declined by only about 4% following the introduction of unfamiliar targets. However, all targets in the Mc Carley study, both familiar and unfamiliar, were drawn from the same category of objects (knives) and thus did not differ dramatically in appearance. Research has shown that simultaneous search for targets (in luggage) that differ in shape and colour as well as orientation may result in costs in terms of speed, accuracy or both (Menner, Barrett, Phillips, Donnelly, & Cave, 2007). Therefore, it follows that stimulus unfamiliarity could be far more detrimental in real-world circumstances where targets vary greatly in category as well as in physical appearance.

Second, from Fleck and Mitroff's explanations of premature search termination, it is unclear whether training on high base rates will transfer appropriately to situations where a response is only required to indicate target presence but not absence. In other words, errors of perception and action could become indistinguishable during actual luggage screening when a physical response is required only when a threat is detected but not otherwise. Thus, there are two situations that are characteristics of an actual luggage screening task that have not been addressed in studies of target prevalence: first, situations in which target absence does not require an explicit response and second, situations in which new and unfamiliar targets (from different object categories) are encountered during the course of search.

The luggage screening paradigm in this research addresses both issues discussed earlier. First, our task required an overt response only when a target was detected (leading to a hit or false alarm) but not when a target was not detected (thereby not requiring a physical response for a miss or correct rejection). Second, our paradigm included transfer targets that were novel (in category and physical appearance) and differed from those used during training. The purpose of the series of studies presented in this article is to examine the 'optimal' training base rate that will transfer effectively to a low base rate situation.

## STUDY 1—EXAMINING TRANSFER TO SAME STIMULI

This experiment was the first step toward establishing whether the probability (or, base rate) of the detection state during training influences the detection of targets at transfer. In this experiment, the transfer stimuli were identical to training stimuli. The only notable difference between training and transfer phases was the target base rate. We hypothesised that a low number of detection searches during training would be detrimental during transfer, because a lower 'attention attracting tendency' would develop during practice (Shiffrin & Schneider, 1984). A higher base rate could be more beneficial during transfer because it would

increase the attention attracting tendency of the individual targets. On the other hand, it is also possible that because the training and transfer targets are identical, training on a base rate identical to the transfer base rate might be more helpful to performance thereby suggesting the need for 'realistically low' base rates during training.

### Pre-test

The purpose of the pre-test was to ensure that the stimuli that would be used during training and transfer were of comparable difficulty. We created a laboratory version of the airline luggage screening task. 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. Colours in these images code material type, with blue indicating metal, orange indicating organic material (plastics, clothing and food) and green indicating materials of intermediate density. Using Adobe Photoshop, we generated artificial packed bags. For each trial, we began with an image of an empty bag and added objects into it that could overlap in a transparent manner (our algorithm for superimposing objects attempted to mimic an actual packed bag as closely as possible). Objects were 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 in the succeeding text.

### *Scaling luggage images for clutter*

A total of 500 images of bags were generated. First, these images were scaled to ensure that all images were cluttered to a comparable degree before a target was superimposed. For this, 30 participants observed all 500 images and rated the perceived clutter of each image on a scale of 1 (not cluttered at all) to 5 (extremely cluttered). There was no time limit for this task, but most participants finished the task in less than 2 hours. '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 (or 'uncluttered areas') in the entire bag. Bags with 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 previously mentioned) 40 times in succession on a computer screen. A neutral grey screen appeared between trials to minimise carryover effects. On each of the 40 trials, a potentially dangerous object (described in the following text) was embedded in the luggage image. The task was to click on the target (in the luggage image) as soon as it was detected. In addition, participants rated the perceived difficulty of detecting each target on a scale of 1 (cannot detect at all) to 5 (can detect 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 completed the task in less than an hour and a half. In a separate test, participants ( $n=10$ ) categorised each of the 40 targets into five possible categories based on the dimensions of colour 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 pre-test. An example of an X-ray image of luggage and the selected targets are presented in Figure 1. Within these two categories, we selected a subset of 10 targets with 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 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 final timed luggage screening task.

## Method

### Participants

Thirty-six undergraduate students completed all sections of the experiment. The experiment lasted approximately 1.5 hours, for which the participants received 1.5 credits towards research participation requirements.

### Tasks and procedures

Participants were randomly assigned to one of three training conditions, with an equal number of participants in each training condition. Each training condition consisted of four blocks with 100 X-ray images per block; the fifth block was the transfer block. The three training conditions differed in the base rate of target presence, with the base rates being 100%, 50% and 20%, respectively. The transfer block for all participants had a target base rate of 20%. Participants were not informed of the base rate during training and transfer.

At the beginning of each training block, five ‘training’ targets were shown on the screen (drawn from the set

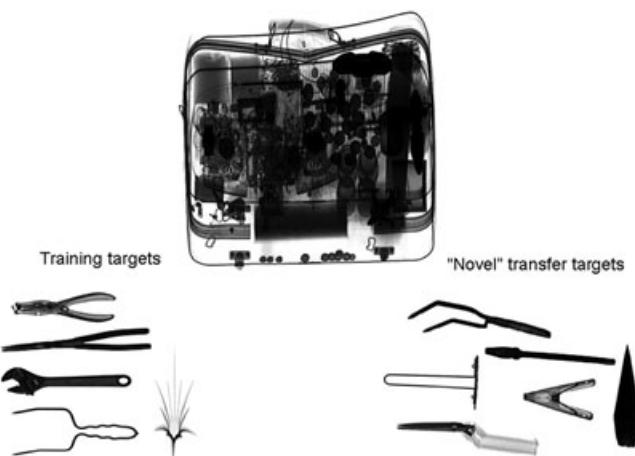


Figure 1. Sample luggage image and targets

presented in Figure 1). After memorising the targets, the participant pressed the spacebar to begin the training block. At the beginning of each trial, an X-ray image of luggage appeared in the centre of the screen for 4 seconds. If participants detected a target, they clicked on the target. The image then disappeared and feedback was provided in the form of a text message on the screen. If the participants did not see a target, they let the trial time out, followed by feedback. Only one target was present in any given bag at a time; all five targets appeared an equal number of times per block in random order.

During training, each participant was randomly assigned to one of three base rate conditions. The 100% base rate group saw a target on every trial, a total of 400 targets across four blocks. The 50% base rate group saw a target on 50 trials in each block of 100, a total of 200 targets across four blocks. The 20% base rate group saw a target on 20 trials in each block of 100, a total of 80 targets across four blocks.

The transfer block was set up similarly to the training blocks. The targets were exactly the same as in the training block. The only difference was that the target base rate for all participants was now 20%. Also, the transfer block comprised only one block of 100 trials because we were interested in studying the immediate effects of transfer alone. The measures of interest were (i) hit rate; (ii) false alarm rate; (iii) response time for correct detections (hits); and (iv) response time for incorrect detections (false alarms). We used signal detection indices of performance (sensitivity and response criterion setting) only in the transfer block because one of the experimental groups (the 100% base rate group) was not exposed to any noise (or nonsignal) trials during the training blocks in all three studies.

## Results

The data were analysed using analyses of variance. Results with alpha values of .05 and below are treated as statistically significant and alpha values between .06 and .09 are discussed as marginally significant because of their theoretical relevance. Independent sample *t*-tests were used to conduct post-hoc comparisons of means. Effect sizes for all statistically significant differences between groups are reported as Cohen’s *d*. Two sets of analyses are reported below—the first set includes two-way mixed ANOVAs (and corresponding post-hoc tests) on performance variables (hit rate, false alarm rate and response time) across all five blocks; the second set includes one-way ANOVAs on signal detection variables (sensitivities and criterion settings) only for the transfer block.

### Performance analyses.

#### Hit rates

The hit rate for each group was calculated as a proportion based on the base rate of target presence for each group. The results for hit rates are illustrated in Figure 2. A 5 (trial block)  $\times$  3 (base rate: 100%, 50% and 20%) mixed ANOVA on hit rates revealed significant main effects for base rate,  $F(2, 33)=12.99$ ,  $p=0.0001$  and for block,  $F(4, 132)=28.99$ ,  $p=.0001$ . The interaction between block and base rate was also significant,  $F(8, 132)=7.45$ ,  $p=.0001$ . As can be seen in the figure, across the four training

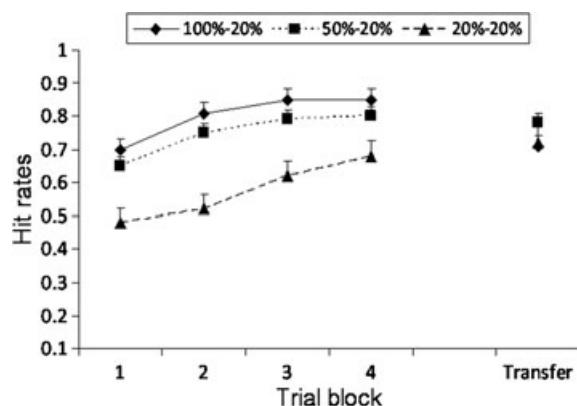


Figure 2. Progression of hit rates by block (Study 1)

blocks, the 100% (average  $M=0.80$ ,  $SD=0.03$ ) and 50% (average  $M=0.75$ ,  $SD=0.04$ ) base rate groups generated significantly more hit rates than the 20% group (average  $M=0.57$ ,  $SD=0.72$ );  $d$  (100% vs 50%) = 1.41;  $d$  (50% vs 20%) = .35,  $d$  (100% vs. 20%) = .45. However, during transfer, the 100% group demonstrated a significant decrease in hit rate ( $M=0.71$ ,  $SD=0.04$ ) relative to the last training block ( $t(11)=3.88$ ,  $p=.003$ ),  $d=2.34$ , whereas 20% group demonstrated a marginally significant increase in hit rate ( $M=0.72$ ,  $SD=0.04$ ) relative to the last training block, ( $t(11)=1.88$ ,  $p=.06$ ),  $d=1.13$ ; this equated the transfer hit rate for these two base rate conditions. The 50% base rate group generated significantly more hits ( $M=0.78$ ,  $SD=0.04$ ) at transfer compared with the 100% ( $t(22)=2.51$ ,  $p=.048$ ),  $d=1.07$  and 20% base rate groups ( $t(22)=2.30$ ,  $p=.049$ ),  $d=.98$ .

#### False alarm rates

Again, the false alarm rate for each group was calculated as a proportion based on the target base rate for each group. Because the 100% group did not perform any ‘target absent’ trials during training (and consequently had no opportunity for false alarms), we first analysed the false alarm data for the 50% and 20% groups in a mixed ANOVA during

training and transfer. This was followed by a one-way ANOVA for the transfer block alone, which included the 100% base rate group.

A 5 (trial block)  $\times$  2 (base rate: 50% vs 20%) repeated measures ANOVA on false alarm rates for the 50% and 20% groups alone revealed a significant main effect for block,  $F(4, 88)=10.90$ ,  $p<.001$  but not for base rate,  $F(1, 22)=1.042$ ,  $p=.32$ , nor an interaction between the two,  $F(4, 88)=1.98$ ,  $p=.10$ . False alarm rates generally decreased from training to transfer for both the 50% and 20% base rate groups. A one-way ANOVA on false alarm rates in the transfer block, which included the 100% base rate group revealed no significant differences among the three base rate conditions,  $F(2, 33)=1.503$ ,  $p=.237$ . The false alarm rates for all conditions are presented in Table 1(a).

#### Response time

(a) Correct detections: The results for response time for correct detections (i.e. hits) are illustrated in Figure 3. The 5 (trial block)  $\times$  3 (base rate) mixed ANOVA on response times for correct detections revealed a significant main effect for base rate,  $F(2, 33)=11.8$ ,  $p=0.0001$  and for block,  $F(4, 132)=10.11$ ,  $p=.0001$ . The interaction between base rate and block was also significant,  $F(8, 132)=4.67$ ,  $p=.0001$ . As can be seen in Figure 3, across training blocks, the 100% (average  $M=1.78$  seconds,  $SD=0.04$ ) and 50% (average  $M=1.76$  seconds,  $SD=0.04$ ) base rate groups detected targets significantly faster than the 20% group (average  $M=2.11$  seconds,  $SD=0.04$ );  $d$  (100% vs 50%) = .50;  $d$  (50% vs 20%) = 2.75,  $d$  (100% vs 20%) = 2.75. However, during transfer, the 100% ( $M=1.94$  seconds,  $SD=0.07$ ) and 50% groups ( $M=1.89$  seconds,  $SD=0.73$ ) demonstrated a significant increase in response times relative to the last training block (100%: ( $t(11)=1.88$ ,  $p=.03$ ),  $d=1.13$ ; 50%: ( $t(11)=1.06$ ,  $p=.018$ ),  $d=.64$ ); on the other hand, the 20% group continued to demonstrate an improvement in response times in the transfer block ( $M=1.86$  seconds,  $SD=0.04$ ),

Table 1. a. False alarm rates—Studies 1 and 2 (standard deviations in parentheses)

STUDY 1	Block 1	Block 2	Block 3	Block 4	Transfer
100% group					0.08 (0.12)
50% group	0.19 (0.21)	0.10 (0.18)	0.08 (0.06)	0.06 (0.13)	0.04 (0.07)
20% group	0.09 (0.11)	0.06 (0.07)	0.05 (0.06)	0.03 (0.03)	0.02 (0.02)
STUDY 2					
100% group					0.42 (0.28)
50% group	0.07 (0.04)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.19 (0.20)
20% group	0.09 (0.06)	0.03 (0.03)	0.02 (0.02)	0.02 (0.10)	0.02 (0.01)

b. Response times (in seconds) for incorrect detections—Studies 1 and 2 (standard deviations in parentheses)

STUDY 1	Block 1	Block 2	Block 3	Block 4	Transfer
100% group					2.59 (0.91)
50% group	2.36 (1.15)	1.87 (1.33)	1.92 (1.24)	1.80 (1.40)	1.95 (1.28)
20% group	2.02 (1.28)	2.43 (0.89)	2.22 (0.82)	2.36 (0.57)	1.89 (1.01)
STUDY 2					
100% group					2.59 (2.83)
50% group	2.05 (1.57)	1.94 (1.69)	1.86 (1.44)	1.82 (1.70)	2.28 (1.82)
20% group	2.55 (1.16)	2.26 (3.26)	2.16 (3.17)	2.11 (1.92)	2.12 (2.65)

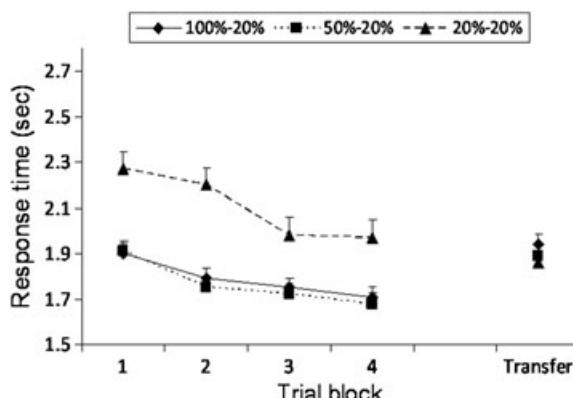


Figure 3. Progression of response time for correct detections by block (Study 1)

( $t(11)=3.88, p=.003$ ),  $d=2.34$ . This equated the transfer response times of all three groups during transfer.

(b) Incorrect detections: Analyses revealed no significant effects during training and transfer. Response times for incorrect detections are presented in Table 1(b).

Signal detection analyses (transfer block).

#### *Sensitivities ( $d'$ prime)*

A one-way ANOVA on sensitivities in the transfer block revealed no significant differences among base rate conditions,  $F(2, 33)=1.27, p=.29$ ; 100% group ( $M=1.95, SD=0.09$ ), 50% group ( $M=2.52, SD=0.11$ ) and 20% group ( $M=2.63, SD=0.50$ ).

#### *Response criterion settings (c)*

Likewise, the one-way ANOVA on response criterion settings in the transfer block revealed no statistically significant difference among base rate conditions,  $F(2, 33)=1.97, p=.18$ ; 100% group ( $M=1.53, SD=0.09$ ), 50% group ( $M=1.63, SD=0.14$ ) and 20% group ( $M=2.08, SD=0.06$ ).

#### **Summary and discussion**

The results of Study 1 revealed that during training, high base rates resulted in higher hit rates and faster response times for correct detections compared with low base rates; high target prevalence likely provided more learning opportunities during training leading to higher probabilities of target detection accompanied by low detection times on instances when participants generated a ‘target present’ response. However, in a signal detection task such as luggage screening, high hit rates and low detection times by themselves do not necessarily constitute evidence of good performance as they could potentially reflect a decision criterion biased toward ‘yes’ responses. Furthermore, the learning advantage (in terms of more hits) for participants trained with high base rates did not transfer effectively to a low 20% base rate. Participants trained on low base rates gradually improved their performance over training blocks such that they transferred effectively to the same base rate that they were trained on; conversely, participants trained on higher base rates demonstrated a reduction in hit rates at transfer when target prevalence

decreased. As a result, all base rates groups performed comparably at transfer.

The results suggest that when the stimuli during training and transfer are the same, high base rates facilitate the generation of hits but do not necessarily translate into more effective transfer when the base rate during transfer decreases relative to training. The training condition that was most similar to the transfer condition (20% base rate) led to as effective transfer as the higher training base rates (as evidenced by comparable sensitivities and response criterion settings at transfer among the three base rate conditions), although there was no significant advantage for the 20% base rate condition over and above the high base rate conditions. It appears that, when training stimuli were the same as transfer stimuli, just a few target exposures were sufficient to create a stable mental representation of what constitutes as ‘target’. It is important to note that the targets used in this study were not common weapons (such as guns and knives; see Figure 1). Therefore, these target categories were not pre-acquired and did require learning during the training phase.

In the case of luggage screening, however, targets during transfer are rarely the same as those encountered during training. Decision making is extremely challenging owing to the number and diversity of threat objects that might potentially be embedded in a piece of luggage (Rich et al., 2008). Therefore, in Study 2, we examined what base rates would lead to optimal transfer when the stimuli during transfer are physically dissimilar and unique from those used during training.

#### **STUDY 2—EXAMINING TRANSFER TO NOVEL STIMULI**

Study 1 revealed that when training and transfer stimuli are identical, low base rates were sufficient to lead to create a good mental representation of targets, possibly without the redundancy that is a characteristic of higher base rates. However, it remains to be seen if transfer performance improves, degrades or remains the same when transfer stimuli are not the same as training stimuli and target detection requires generalisation of knowledge from the limited number of target exposures during training. In this study, we used a paradigm similar to Study 1 wherein participants were trained on one of three base rates—100%, 50% and 20%—for the training portion of the study and then transferred to a low base rate of 20%. Unlike in Study 1, the transfer targets in Study 2 were novel and not identical to those used in training (details are presented in the Method section).

We predicted one of two possible outcomes. First, in keeping with the pattern for Study 1, low base rates during training would lead to optimal transfer to low base rate conditions due to the similarity of training and transfer conditions. This could lead to comparable transfer performance for all base rate groups. Conversely, given the dissimilarity of training and transfer stimuli, training with high base rates could be more beneficial than low base rates by providing a better mental representation of what constitutes a target, eventually facilitating generalisation of knowledge to novel targets. This is based on the results of a study by Gonzalez

and Madhavan (2011), which found that training under diverse conditions (or with a variety of target exposures) leads to better generalisation of knowledge to novel situations than training under less diverse conditions. However, higher base rates could also lead to high false alarm rates during transfer reducing detection sensitivity.

## Method

### Participants

Thirty-three undergraduate students completed all phases of the experiment. The duration of the experiment was approximately 1.5 hours and participants were paid a total of \$15.

### Tasks and procedures

Participants were randomly assigned to one of three base rate conditions ( $n=11$  per group). The only difference from Study 1 was that the transfer targets were novel and different from those seen in the training blocks. These objects were also drawn from the original set of 10 targets from the pre-test (described in Study 1). 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 pre-tested 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. Instead, they were instructed to use their knowledge gained during the training phase to detect the novel transfer targets. Similar to Study 1, each participant completed four training blocks followed by one transfer block, for a total of five blocks with 500 X-ray images.

## Results

### Hit rates

The results for hit rates are illustrated in Figure 4. A 5 (trial block)  $\times$  3 (base rate: 100%, 50% and 20%) mixed ANOVA on hit rates revealed significant main effects for base rate,  $F(2, 30)=20.65, p=0.0001$  and block,  $F(4, 120)=22.07,$

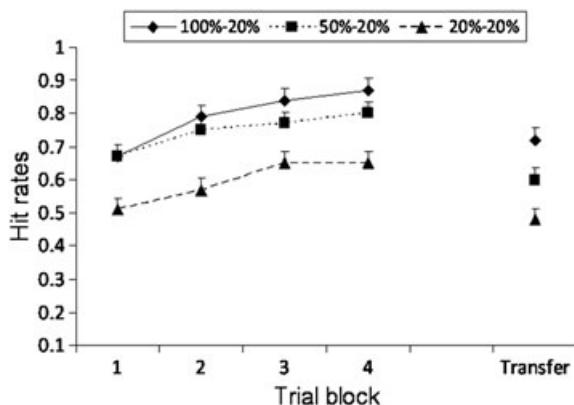


Figure 4. Progression of hit rates by block (Study 2)

$p=.0001$ . The interaction between block and base rate did not reach statistical significance,  $F(8, 120)=.99, p=.45$ . As can be seen in the figure, hit rates generally decreased from training to transfer. The 100% base rate group consistently generated significantly more hits during training (average training:  $M=0.79, SD=0.03$ ) as well as transfer ( $M=0.72, SD=0.05$ ) relative to the 50% base rate (average training:  $M=0.75, SD=0.03, d=1.33$ ; transfer:  $M=0.56, SD=0.05, d=3.20$ ) and 20% base rate conditions (average training:  $M=0.60, SD=0.03, d=4.00$ ; transfer:  $M=0.48, SD=0.05, d=3.88$ ). Likewise, the 50% base rate group generated significantly more hits than the 20% group at training ( $d=3.00$ ) as well as transfer ( $d=1.60$ ).

### False alarm rates

Similar to Study 1, we first analysed the false alarm data for the 50% and 20% groups in a mixed ANOVA during training and transfer. This was followed by a one-way ANOVA for the transfer block alone, which included the 100% base rate group. The repeated measures mixed ANOVA on false alarm rates for the 20% and 50% groups revealed a significant main effect for block alone,  $F(4, 88)=9.32, p=.006$  but not for base rate,  $F(1, 22)=.90, p=.35$ , nor an interaction between the two,  $F(4, 88)=.82, p=.45$ . In general, false alarm rates increased from training to transfer [see Table 1(a)].

The one-way ANOVA on false alarm rates for the transfer block, which included the 100% group revealed significant differences between base rate groups,  $F(2, 30)=4.62, p=.018$ . The 100% group generated significantly more false alarms during transfer ( $M=0.419, SD=0.28$ ) than the 50% group ( $M=0.190, SD=0.20$ ),  $t(20)=2.09, p=.05, d=.93$  and the 20% group ( $M=0.167, SD=0.11$ ),  $t(20)=2.73, p=.013, d=1.22$ . The false alarm rate for the 50% group was not significantly different from that of the 20% group.

### Response time

(a) Correct detections: The results for response time for correct detections (i.e. hits) are illustrated in Figure 5. A 5 (trial block)  $\times$  3 (base rate) mixed ANOVA on response time for correct detections revealed significant main effects for base rate,  $F(2, 30)=5.95, p=0.07$  and block,  $F(4, 120)=30.95, p=.0001$ , as well as a significant interaction between base rate and block,  $F(4, 88)=10.22, p=.0001$ . As can be seen in Figure 5, during training

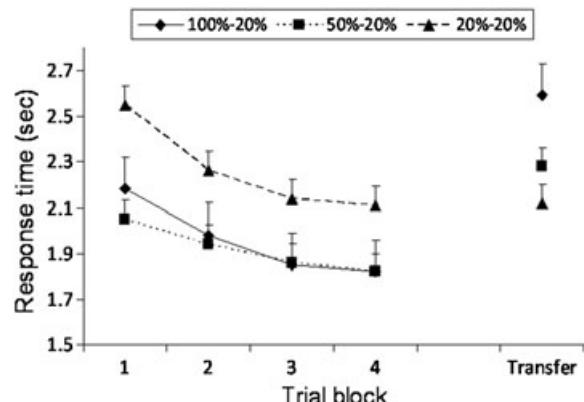


Figure 5. Progression of response time for correct detections by block (Study 2)

both the 100% (average training:  $M = 1.96$  seconds,  $SD = 0.06$ ) and 50% base rate groups (average training:  $M = 0.1.92$  seconds,  $SD = 0.06$ ), although not significantly different from each other, detected targets significantly faster than the 20% group (average training:  $M = 2.27$  seconds,  $SD = 0.06$ );  $d$  (50% vs 20%) = 3.83,  $d$  (100% vs 20%) = 4.16. However, during transfer, the 100% ( $M = 2.59$  seconds,  $SD = 0.07$ ) and 50% base rate groups ( $M = 2.28$  seconds,  $SD = 0.08$ ) demonstrated a significant slowing down in response times relative to the last training block (100% group:  $t(10) = 3.37$ ,  $p = .002$ ,  $d = 2.13$ ; 50% group:  $t(10) = 3.28$ ,  $p = .002$ ,  $d = 2.07$ ), whereas the 20% group maintained response times that were comparable with and not significantly different from the last training block ( $M = 2.12$  seconds,  $SD = 0.07$ ).

(b) Incorrect detections: Analyses revealed no significant effects for response time of incorrect detections (false alarms). The data are presented in Table 1(b).

#### Signal detection analyses (transfer block).

##### *Sensitivities (d prime)*

A one-way ANOVA on sensitivities in the transfer block revealed no significant differences among base rate conditions,  $F(2, 30) = .271$ ,  $p = .76$ ; 100% group ( $M = 0.78$ ,  $SD = 0.60$ ), 50% group ( $M = 1.02$ ,  $SD = 0.20$ ) and 20% group ( $M = 1.01$ ,  $SD = 0.10$ ).

##### *Response criterion settings (c)*

A one-way ANOVA on response criterion settings in the transfer block revealed a statistically significant difference among base rate conditions,  $F(2, 30) = 4.86$ ,  $p = .031$ . In keeping with the pattern for hits and false alarms, the 100% group demonstrated a significantly more liberal bias in responding ( $M = 0.83$ ,  $SD = 0.41$ ) relative to the 50% ( $M = 1.41$ ,  $SD = 0.60$ ),  $d = 1.13$  and 20% groups ( $M = 1.65$ ,  $SD = 0.24$ ),  $d = 2.44$ . The difference between the 50% and 20% base rate conditions did not reach statistical significance.

### Summary and discussion

Similar to Study 1, training on high base rates led to higher hit rates during training. However, unlike Study 1 where the advantage of high base rate did not transfer when the base rate decreased to 20%, the findings of this study continued to show an advantage of the highest base rate condition (100%) with respect to the high probability of hits generated during transfer to novel targets. However, training on a very high base rate also led to a higher incidence of false alarms during transfer, which was reflected in the significantly more liberal decision criterion settings for the 100% group. High base rates possibly created higher expectations of signals being present, which in turn led to a liberal response bias; ultimately, this bias potentially led participants in the 100% group to generate more 'target present' responses during transfer. This suggests that high target prevalence did not necessarily *help* transfer because there were no relative improvements in sensitivities for high base rate groups relative to low base rate groups at transfer; instead, high base rates resulted in biassing effects on response patterns that were not accompanied by improvements in actual signal detection abilities.

Similar to the negative pattern for false alarms, the 100% group demonstrated the slowest response times for correct detections during transfer. It appears that the higher expectation of a target being present led the participants to spend more time scanning each image, thereby slowing their response time. This pattern was more salient in this study than in Study 1 possibly because participants were unfamiliar with transfer targets; consequently, transfer performance (correct detections as well as detection time) required more cognitive resources, which, in turn, were influenced by the training base rates.

Relative to the 100% and 20% base rate conditions, training with a base rate of 50% resulted in a balance between hits and false alarms as well as quicker response times (comparable with the 20% group). This is possibly because, on the one hand, a 50% base rate sufficiently increased the 'attention attracting tendency' (Shiffrin & Schneider, 1984) of targets during training, leading to a higher hit rate relative to the 20% group. On the other hand, a moderate base rate of 50% during training likely led to lower expectations of target presence during transfer, thereby circumventing the problems associated with high false alarm rates and liberal criterion settings for the 100% group. Overall, a low base rate of 20% was the least effective in accurately detecting novel targets despite the fact that the base rate for this condition was identical during training and transfer.

### STUDY 3—EXAMINING TRAINING AND TRANSFER TO NOVEL STIMULI WHEN CONTROLLING FOR NUMBER OF TARGET EXPOSURES

In Studies 1 and 2, the base rate of targets was manipulated by keeping the number of trials constant. As a result, groups with higher base rates were exposed to a larger absolute number of targets during training than those with lower base rates. Research has suggested that in complex visual search tasks, target prevalence influences performance more critically than just the number of target exposures per se (Wolfe et al., 2005). However, there exists little evidence to indicate that the effect of base rates is the same when the number of targets is actually increased over time such that the number of target exposures is the same despite different base rates.

The purpose of this study was to examine whether the target prevalence effects demonstrated in Studies 1 (during training) and 2 (during training and transfer) remains robust when the total number of target exposures during training is equalised across base rate conditions. Specifically, we examined the hypothesis that showing all participants the same number of targets during training (by manipulating the number of trials in each block and maintaining the original base rates) would lead to the same level of transfer performance regardless of different base rates.

### Method

#### *Participants*

Thirty-four undergraduate students completed all phases of the experiment. The duration of the experiment was approximately 1.5 hours and participants received 1.5 credits fulfilling research participation requirement.

### Tasks and procedures

Study 3 was structured similarly to Study 2. However, the key difference was the number of training trials completed by each base rate group. The objective was to equalise the number of target exposures across all three base rate groups to 100 by adjusting the number of training trials for each group without altering the base rates. The 100% base rate group saw a target on every trial in the training block; therefore, they completed 100 trials during training to achieve a total of 100 target exposures. The 50% base rate group saw a target on 50% of the trials during training; therefore, they completed 200 trials during training to achieve the same total of 100 target exposures. The 20% base rate group saw a target on 20% of training trials; therefore, in this study, they completed 500 trials to achieve the 100 target exposure at the end of training. In short, depending on the base rate group to which each participant was assigned, they completed 100 training trials (100% base rate), 200 training trials (50% base rate) or 500 training trials (20% base rate), until all participants had observed a total of 100 targets. This was followed by the transfer block wherein they completed another 100 trials with novel targets at 20% base rate, similar to the procedure in Studies 1 and 2. It is important to note that, similar to Study 2, the transfer targets in Study 3 were novel and had not been encountered by participants during training and were drawn from the same categories as the training targets (illustrated in Figure 1).

## Results

### Hit rates

Because the number of training blocks performed by the three base rate groups was unequal, we averaged the hit rates across training blocks for the 50% and 20% groups and compared them to the hit rates on the single training block performed by the 100% group. A 2 (trial block: training vs transfer)  $\times$  3 (base rate: 100%, 50% and 20%) ANOVA on hit rates revealed significant main effects for both block,  $F(1, 31)=23.24, p < .001$  and base rate,  $F(2, 31)=11.001, p < .001$ , as well as a significant interaction between the two,  $F(2, 31)=14.14, p < .001$ . As illustrated in Figure 6, during training, the 20% group (average training:  $M=0.53, SD=0.02$ ) had a significantly lower hit rate than the 100% ( $M=0.64, SD=0.16$ ),  $d=.96$  and the 50% base rate groups ( $M=0.60, SD=0.17$ ),  $d=.58$ . However, the

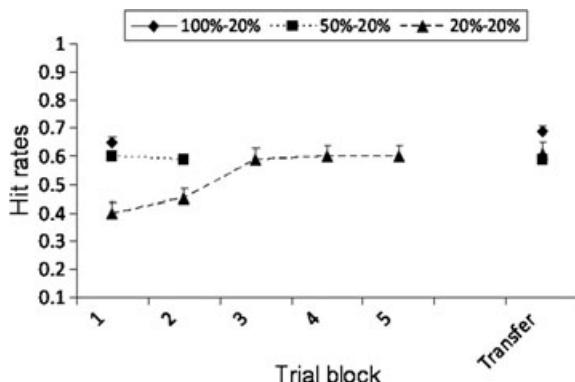


Figure 6. Progression of hit rates by block (Study 3)

20% group demonstrated an almost monotonic increase in hit rates across the five training blocks culminating in a hit rate of 0.6 in the final training block. In the transfer block, all three base rate groups demonstrated hit rates not significantly different from their last training block: 100% group:  $M=0.69, SD=0.18$ , 50% group:  $M=0.59, SD=0.25$  and 20% group:  $M=0.61, SD=0.25$ . Consequently, the transfer hit rates of the three base rate groups did not differ significantly during transfer.

### False alarm rates

Similar to the analyses for hit rates, we first averaged the false alarm rates across training blocks for the 50% and 20% base rates groups. Then, we compared the 50% and 20% groups with each other during training and transfer and next with the 100% group during transfer. The results of the 2 (trial block: training vs transfer)  $\times$  2 (base rate: 50% vs 20%) mixed ANOVA on false alarms for the 50% and 20% groups revealed no significant main effects or interactions. False alarm rates are presented in Table 2(a).

The one-way ANOVA on false alarm rates that included all three base rate groups in the transfer block revealed statistically significant effects for base rate,  $F(2, 33)=3.213, p=.05$ . As illustrated in Figure 7, during transfer the 100% group generated more false alarms ( $M=0.33, SD=0.23$ ) than the 50% ( $M=0.20, SD=0.14$ ) and 20% base rate groups ( $M=0.16, SD=0.11$ ). However, only the difference in false alarm rates between the 100% and 20% groups was statistically significant,  $t(20)=2.25, p=.036, d=.94$ . There were no significant differences between the 100% and 50% and 50% and 20% base rate conditions.

Table 2. a. False alarm rates—Study 3 (standard deviations in parentheses)

	Training	Transfer
100% group		0.33 (0.23)
50% group	0.11 (0.09)	0.20 (0.14)
20% group	0.01 (0.01)	0.16 (0.11)

b. Response times (in seconds) for correct detections—Study 3 (standard deviations in parentheses)

	Training	Transfer
100% group	1.93 (0.22)	1.78 (0.43)
50% group	1.90 (0.21)	1.97 (0.57)
20% group	2.06 (0.12)	1.79 (0.29)

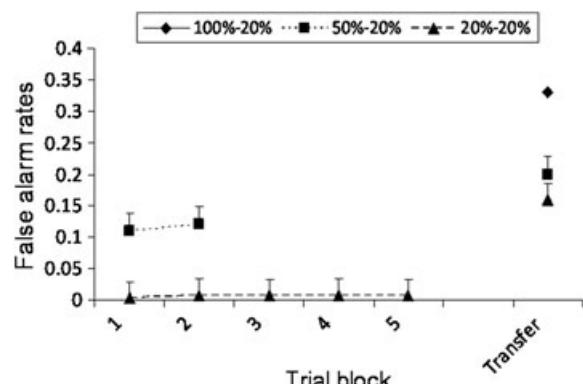


Figure 7. Progression of false alarm rates by block (Study 3)

### *Detection time*

(a) Correct detections: Analyses revealed no significant effects for base rates on speed of correct detections (i.e. hits). Data are presented in Table 2(b).

(b) Incorrect detections: Again, similar to the analyses for false alarm rates, we first averaged the response times for incorrect detections (i.e. false alarms) across training blocks for the 50% and 20% base rates groups. Then, similar to the procedure for Studies 1 and 2, we compared the 50% and 20% groups with each other during training and transfer and then with the 100% group during transfer. The results of the 2 (trial block: training vs. transfer)  $\times$  2 (base rate: 50% vs 20%) mixed ANOVA on response times for incorrect decisions for the 50% and 20% groups revealed a significant main effect for base rate,  $F(1, 21)=14.59, p=.001$  but not for block,  $F(1, 21)=.40, p=.54$ . Analysis also revealed a significant interaction between base rate and block,  $F(1, 21)=8.29, p=.009$ . Results revealed that during training the 20% base rate group ( $M=1.73$  seconds,  $SD=0.76$ ) responded significantly faster when generating incorrect decisions than the 50% group ( $M=3.00$  seconds,  $SD=0.46$ ),  $d=1.97$ . However, during transfer, the 50% group speeded up their responses ( $M=2.57$  seconds,  $SD=0.046$ ), whereas the 20% group significantly slowed their response time ( $M=2.4$  seconds,  $SD=0.084$ ), such that the transfer response times of the two groups were not significantly different from each other.

The one-way ANOVA on response times for incorrect detections including all three base rate groups in the transfer block did not reveal any statistically significant effects,  $F(2, 33)=.50, p=.61$ .

Signal detection analyses (transfer block).

### *Sensitivities ( $d'$ prime)*

Similar to the results of Studies 1 and 2, the one-way ANOVA on sensitivities in the transfer block revealed no significant differences among base rate conditions,  $F(2, 33)=.492, p=.82$ ; 100% group ( $M=0.93, SD=0.09$ ), 50% group ( $M=1.06, SD=0.14$ ) and 20% group ( $M=1.27, SD=0.06$ ).

### *Response criterion settings (c)*

The one-way ANOVA on response criterion settings in the transfer block revealed a statistically significant difference among base rate conditions,  $F(2, 33)=3.14, p=.026$ . Similar to the results of Study 2, the 100% group demonstrated a significantly more liberal bias in responding ( $M=0.97, SD=0.23$ ) relative to the 20% group ( $M=1.43, SD=0.21$ ),  $d=2.08$  but not the 50% group ( $M=1.35, SD=0.41$ ). The difference between the 50% and 20% base rate conditions did not reach statistical significance.

## **Summary and discussion**

The results of Study 3 to an extent mirrored Study 1 in that higher training base rates led to higher hit rates in the first training block. Although higher target base rate provided several learning opportunities with relatively little training (i.e. exposure to 100 targets in just one block), low base rates facilitated gradual learning as the total number of targets

added up to 100 across five blocks. As a result, low base rates resulted in transfer performance that was comparable with higher base rates by equating the hit rates of the two base rate conditions. These results suggest that the absolute number of target exposures during training played a more critical role in learning and transfer than base rate per se. Remarkably, the results of Study 3 also replicated some findings from Study 2; false alarm rates at transfer were significantly higher (as reflected in significantly more liberal criterion settings at transfer) for the 100% base rate group relative to the 20% group. It is possible that this effect occurred because of the inclusion of novel targets during transfer similar to Study 2. Interestingly, this study was the only one where significant effects were found for response time on incorrect detections, supporting findings by Wolfe and Van Wert (2010) that target prevalence impacts target-absent responses more than target-present responses due to a shift in decision criteria.

It is apparent that increasing the number of target exposures for the 20% group across blocks was successful on two counts: first, it helped participants detect targets when present (comparably with the higher base rate groups) by the end of the last training block; second, it also likely provided the participants with sufficient opportunities to assess the true base rate of target presence, ultimately helping them shift their response biases in a manner similar to Study 2. It appears that in a cognitively challenging task such as luggage screening, the primary source of challenge is the high level of variability in what constitutes a target. Under such conditions, it is most beneficial to maximise the number of training targets such that the definition of a 'target' is as broadly represented in memory as possible. However, maximising targets is most effective when it is carried out over time (i.e. across several blocks) so as to provide participants with an intuitive 'feel' for the true base rate of target presence.

## **GENERAL DISCUSSION**

The purpose of this research was to answer three important questions pertaining to how humans use base rate information in visual search tasks: (i) What training base rate leads to best transfer to low base rate conditions when training and transfer conditions are otherwise identical?; (ii) What training base rate is most effective when transferring to low base rate situations when transfer stimuli differ physically from training stimuli?; and (iii) Does base rate continue to influence transfer performance when the number of target exposures during training is equalised across different base rate conditions?

The three studies revealed mixed results. The results of Study 1 revealed that high base rates during training led to rapid and efficient generation of hits during training. However, because training and transfer targets were identical, there was no particular advantage of high base rates at transfer. In Study 2, the introduction of novel targets at transfer provided a significant advantage to participants trained on high base rates, again in terms of hit rates alone; however, high base rates also led to more false alarms and longer response times for correct detections. In Study 3, equating

the number of target exposures across base rate conditions resulted in transfer performance that mirrored a combination of Studies 1 and 2; similar to Study 1, exposure to the same number of targets across base rate conditions eliminated the training advantage of high base rates, despite the transfer targets being novel. However, similar to Study 2, the disadvantage of high base rates was replicated with significantly higher false alarm rates reflected by liberal criterion settings for the 100% group during transfer. Interestingly, a low base rate at transfer decreased false alarms when transfer stimuli were identical to training stimuli (Study 1) but increased false alarms when transferring to novel stimuli (Studies 2 and 3); these effects for false alarms were particularly strong for the groups that were at opposite ends of the base rate spectrum (100% vs 20%).

Contrary to the existing research, our three studies revealed that the behavioural effects of base rates fail to remain robust and change considerably if the total number of target exposures is manipulated over time regardless of base rate. We found only partial support for the 'low prevalence effect' Wolfe *et al.* (2005, 2007) 'low prevalence effect' wherein a reduction in base rate led to a proportionate reduction in performance efficacy. Specifically, we found evidence for the low prevalence effect during training wherein the 20% group performed worse than the higher base rate groups in terms of generating hits alone; however, this pattern changed during transfer as a function of the similarities between the training and transfer task across the three studies. Wolfe *et al.* reasoned that the negative effect of low base rates found in their research occurred because as observers repeatedly responded with correct rejections, they began to terminate their searches more quickly because of conservative shifts in decision criteria, consequently missing targets on the rare occasions when they are present.

Fleck and Mitroff (2007) contended that a large percentage of errors under low base rate conditions were motor errors or errors of action/execution rather than errors of perception, that is, observers were aware of the fact that a target existed but responded too quickly to the non-presence of targets because low base rates habituated participants to respond negatively thereby terminating their target search prematurely. We argued that such situations described by Fleck and Mitroff that presumably lend themselves to motor errors are rarely characteristic of a real luggage screening task where target absence in most cases does not warrant an action (i.e., screeners generally do not perform any action when they do not detect a threat object). Consequently, in our paradigm, correct rejections did not require the execution of a physical action; that is, when participants did not detect a target they simply allowed the trial to time out. An action (clicking on the target) was only required on trials in which a target was detected, either correctly or incorrectly. Therefore, the likelihood that theories of premature search termination under low base rates impacting performance in this paradigm is low. However, the results of Study 3 did reveal that the 20% base rate group had the shortest response times when generating false alarms during training thereby providing some evidence in support of the premature search termination theory. Combined with the significant effects in our studies for response criterion settings but not

sensitivities, our findings support the claim by Van Wert *et al.* (2009) that misses many results from motor or response errors in relatively simple search tasks; however, in more complex tasks, shifts in decision criteria appear to be an important contributor, indicating the role of complex cognition in the success of visual search (Van Wert *et al.*, 2009; Wolfe & Van Wert, 2010).

The results of our research can be explained partially by the 'attention attracting tendency' that is developed during repeated practice with one set of targets (Shiffrin & Schneider, 1984). In all three studies, it is possible that high base rates increased the attention attracting tendency of targets and provided participants with a more accurate mental representation of what constituted a 'target'. This mental representation was particularly critical in Study 2 wherein the transfer targets were novel and unfamiliar and participants had to perform complex mental rotations based on their knowledge and memory of the targets observed during training. The 'attention attracting tendency' of the training targets failed to make as significant an impact on transfer when participants encountered the same targets as during training (Study 1), as also observed in the earlier study by McCarley *et al.* (2004). This pattern of results for Study 1 suggests that just a few exposures of target stimuli (i.e. 20% base rate) during training were sufficient to create a stable mental representation of the target. From this, as hypothesised, it is possible to infer that when the transfer task required detection of familiar targets, a larger base rate during training only led to higher levels of redundancy in knowledge acquisition.

Our research presents an interesting possibility—in a complex task such as luggage screening, the most effective training methodology is not necessarily one that adapts the human to target base rate but rather provides the screener with as many exposures to the target stimuli as possible. The ability to adapt from a high base rate situation (during training) to a low base rate situation (during transfer) is undoubtedly an important one; nevertheless, the results of this research indicate that in tasks such as luggage screening, it is more critical to simply optimise the strength of skill acquisition such as to maximise the overall probability of correct detections and minimise the wastage of resources due to false alarms.

## Conclusions and implications

The results of this three-study research are important in that security personnel at a real airport are likely to encounter very rare and unexpected targets during the process of screening. The effect of low target base rates is exacerbated by the novelty and variety of potential threat objects. Therefore, studying the effect of base rates and the conditions that lead to optimal detection are of paramount importance. On the basis of the results of this research, it is difficult to make the blanket conclusion that training with one base rate provides as significant advantage over another in absolute terms. It is possible that our results were influenced by the target probability, which was much higher for all base rate conditions than that in the real world as well as in previous studies that have examined the issue of target prevalence (e.g. Fleck & Mitroff, 2007; Wolfe *et al.*, 2005). Regardless

of the relative differences in base rate, our results suggest that the most important goal in such target detection tasks is to build the appropriate memory set for potential threat objects, which, it appears, is facilitated more by the number of target exposures as opposed to base rate per se. The base rate incorporated during training must therefore be varied to ensure skill acquisition and transferability of skills to a variety of transfer situations as is the characteristic of the real world.

### ACKNOWLEDGEMENTS

This research was partially supported by the Multidisciplinary University Research Initiative Program (MURI; N00014-01-1-0677) and by the Advanced Decision Architectures Collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory (DAAD19-01-2-0009) grants to Cleotilde Gonzalez. We are also grateful for the programming assistance for the Luggage Screening task provided by Jack Lim and Varun Dutt at the DDMLab.

### REFERENCES

- Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science*, 18, 943–947.
- Gonzalez, C., & Madhavan P. (2011). Diversity during training enhances detection of unknown visual stimuli. *Journal of Cognitive Psychology*, 23, 342–350.
- McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science*, 15, 302–306.
- Menneer, T., Barrett, D. K., Phillips, L., Donnelly, N., & Cave, K. R. (2007). Costs in searching for two targets: dividing search across target types could improve airport security screening. *Applied Cognitive Psychology*, 21, 915–932.
- Parasuraman, R., Warm, J. S., & Dember, W. N. (1987). Vigilance: Taxonomy and utility. In L. S. Mark, J. S. Warm, & R. L. Huston (Eds.), *Ergonomics and human factors*. New York: Springer-Verlag, 11–32.
- Rhodes, M., & Jacoby, L. (2007). On the dynamic nature of response criterion in recognition memory: effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 305–320.
- Rich, A. N., Kunar, M. A., Van Wert, M., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring boundaries of the low prevalence effect. *Journal of Vision*, 8, 1–17.
- Shiffrin, R. M., & Schneider, W. (1984). Automatic and controlled processing revisited. *Psychological Review*, 91, 269–276.
- Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics*, 71, 541–553.
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, 20, 121–124.
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, 435, 439–440.
- Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, 136, 623–638.