Ψ **Psychology Press**
Taylor & Francis Group

Diversity during training enhances detection of novel stimuli

Cleotilde Gonzalez

Dynamic Decision Making Laboratory, Carnegie Mellon University, Pittsburgh, PA, USA

Poornima Madhavan

Department of Psychology, Old Dominion University, Norfolk, VA, USA

This research demonstrates that when individuals are expected to detect novel targets, they will be best prepared when trained with diverse categories. Participants were trained in a simulated luggage screening task, in one of three conditions of diversity: high (participants searched for dangerous objects belonging to five different categories); low (participants searched for targets belonging to one of the five categories); and no training (control condition). After training, all participants were asked to look for the same novel dangerous objects in the bags. Results show that, during training, the low diversity condition resulted in highest hit rates and fastest response times. In contrast, after training, results were reversed: participants that trained in a high diversity condition were most effective at detecting novel targets. Those with no training at all were equally poor at detecting novel targets as those that trained in a low diversity condition.

Keywords: Diversity; Robust decision making; Visual search.

Many studies in the literature of learning have demonstrated that practising a task under varied conditions (*training variability*) can enhance retention and

transfer compared to practicing under more consistent conditions. For example, Schmidt and Bjork (1992) reported the effectiveness of variation during practice versus consistent practice for transfer to some novel retention test. Their demonstrations included both motor and verbal tasks. In addition, more recent studies have also demonstrated benefits of training variability in contexts such as organisational learning (Schilling, Vidal, Ployhart, & Marangoni, 2003) and cognitive skills (Yechiam, Erev, & Gopher, 2001). But most of the benefits of training variability have been demonstrated in motor tasks rather than cognitive tasks, such as throwing objects at targets of different distances (Wulf, 1991), moving objects at different distances and speeds (Kelso & Norman, 1978), controlling the speed and accuracy of an apparatus through the motor movement of a handle (Schmidt, Young, Swinnen, & Shapiro, 1989), continuous pursuit tracking (Wulf & Schmidt, 1997), and the forehand drive in tennis (Douvis, 2005).

In contrast to the supporting evidence for positive effects of training variability, evidence also indicates that learning is highly specific to the training conditions (*specificity of training*); that is, transfer is most effective when the conditions of transfer closely match the conditions of training. Recent studies have addressed the benefits of specificity in both perceptual and motor components of a task (Healy, Wohldmann, Parker, & Bourne, 2005; Healy, Wohldmann, Sutton, & Bourne, 2006). Healy et al. (2006) found that contrary to previous findings on the advantage of training variability, individuals show durability and transfer of performance only when the mental procedures developed during training can be reinstated at testing.

Thus, the current literature is mixed regarding the benefits of training variability. This research report provides evidence of the benefits of training variability for the detection of novel targets in a visually complex cognitive task- luggage screening. Research has demonstrated support for the specificity of training in similar airport luggage screening tasks (Smith, Redford, Washburn, & Taglialatela, 2005). Specifically, researchers found that participants relied on the recognition of familiar targets and had great difficulty using category-general knowledge (Smith, Redford, Gent, & Washburn, 2005; Smith, Redford, Washburn, & Taglialatela, 2005). In those studies, performance improved when the transfer images were the same as the ones used during training. Performance dropped when unfamiliar targets from the same categories appeared. A similar pattern of effects was first found by McCarley, Kramer, Wickens, Vidoni, and Boot (2004).

We hypothesise that a possible reason for these results is the small number of training categories used in that study. To form category-general knowledge and effectively transfer this knowledge to unfamiliar targets, training must incorporate items belonging to a larger and more diverse set of categories. Thus, variability in this research refers to the categorical diversity of elements experienced during training.

In the current experiment, the number of target categories from which the targets were drawn during training was manipulated. The goal of this research was to examine the benefits of this categorical diversity on the detection of novel items at transfer in a complex visual detection task. The detection of novel or unusual targets is important in many visual detection tasks. Some examples of such tasks are a doctor trying to identify a tumour on an X-ray image, a soldier attempting to determine the presence of a combatant in unfamiliar terrain, and a security officer trying to find dangerous items in passenger luggage. We expect that training variability will lead to better detection of unknown targets in these scenarios.

EXPERIMENT

Much of threat detection at airports is still conducted by visual inspection of the bag, rather than by automated methods. This method is used partly because the rigidity of automated algorithms has difficulty adapting to the uncertainty and variability associated with threats. Humans seem more capable than automated aids at extrapolating from previous knowledge and engaging in adaptive decision making (i.e., thinking "outside the box") when faced with novel threat targets. Therefore, at the practical level, a goal of this research is to determine ways in which we can improve the accuracy of human detection of potentially dangerous items and optimise detection time through the study of skill acquisition and learning.

We tested the effects of training variability with a luggage screening task on the effectiveness of the detection of novel targets at transfer. We used targets during the training phase that varied in the number of categories from which they were drawn. We then transferred participants to a condition that involved novel targets, i.e., targets that were categorically different from those used during training and that were not shown to the participants ahead of time.

The Luggage Screening Task

We developed a luggage screening simulation to represent the visual search in luggage screening, which has also been used in other recent studies (Brunstein & Gonzalez, 2010; Lacson, Gonzalez, & Madhavan, 2008; Madhavan, Gonzalez, & Lacson, 2007). This simulation involves complex visual images constructed from individual X-ray images of dangerous and nondangerous items. The complex visual images were built by manually compiling X-ray images of individual objects provided by the Transportation Safety Administration (TSA). The TSA provided a set of individual X-ray images identified as potentially dangerous and nondangerous items. We generated a set of multiple categories of the potentially dangerous objects (knives, cutting

objects, glass objects, etc.) according to a pretest, explained later, and we manually inserted one potentially dangerous item in half of the bags to be used in the simulation. An example of a compiled image and individual images of potentially dangerous targets used in the simulation are shown in Figure 1.

During training, this task requires individuals to memorise a set of potentially dangerous objects presented on the screen. Then individuals are asked to determine the presence of any of those weapons embedded in luggage X-ray images. Other studies have used similar luggage screening images to examine the visual search aspects of a screener's performance (McCarley et al., 2004), issues related to operator trust in automated decision support systems (Madhavan & Wiegmann, 2005), and categorisation and specificity of training (Smith, Redford, Gent, & Washburn, 2005). The task provides a compilation of the actual visual stimuli that screeners are subjected to and provides similar time pressures (our study's participants have a limited amount of time to inspect the bag). We hypothesised that a larger categorical diversity of target objects used during training would result in better transfer to novel items.
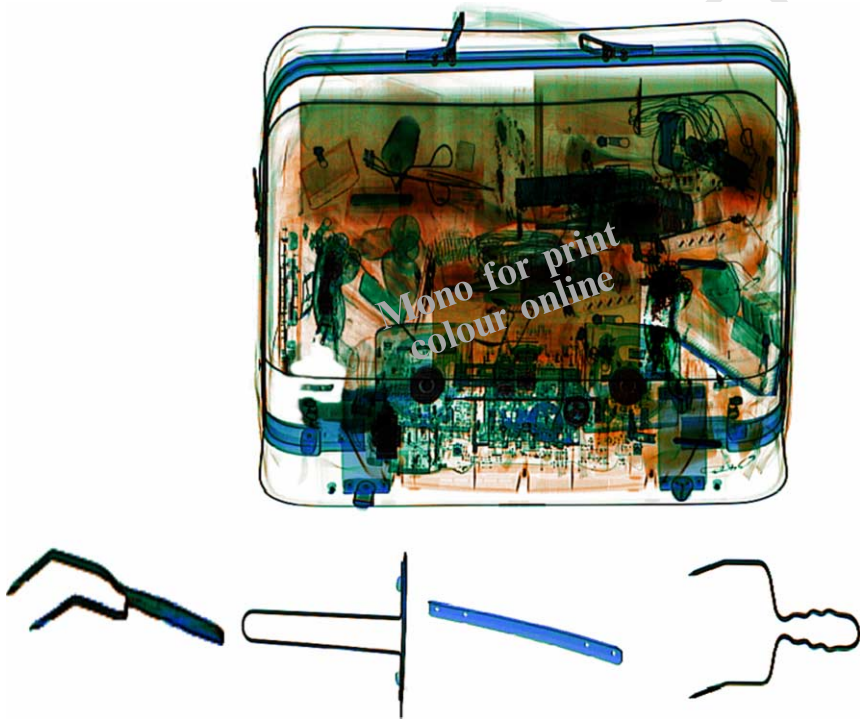

Figure 1. An example of an X-ray image of luggage and a subset of the "novel" targets selected for the transfer phase of the experiment.

Method

Pretest.   The purpose of the pretest was to select categories of potentially dangerous objects that would be used to train participants in the actual luggage-screening experiment and to identify the novel targets to be used at transfer. This pretest of images was also used for a second study, and the full methods and results of this pretest are also reported by Brunstein and Gonzalez (2010). In order to create the categories used in this experiment, a pilot group of participants ($n = 8$) was presented with 40 icons of potentially dangerous objects on a screen, one at a time. They were asked to cluster the icons on a two-dimensional, x-y coordinate screen on the basis of "shape complexity" (x-axis) and "colour complexity" (y-axis), using a computer program specifically created to record the final coordinates of the objects as participants arranged them on the screen. Objects with coordinates falling within a 500-pixel radius were classified into one category. The results clearly distinguished five categories that were used in the training phase: metal objects, knives, guns, scissors, and glass objects. These category clusters are depicted in Figure 2. Other objects that did not clearly belong to any category according to these criterion were classified as "novel" and were used in the transfer phase. Thus, these transfer targets were categorically dissimilar to the training targets, and included objects such as pointing and wire objects of different shapes, and possible cutting objects of no specific shape. Figure 1 shows some examples of the objects classified as "novel" and used in the transfer phase of the experiment.
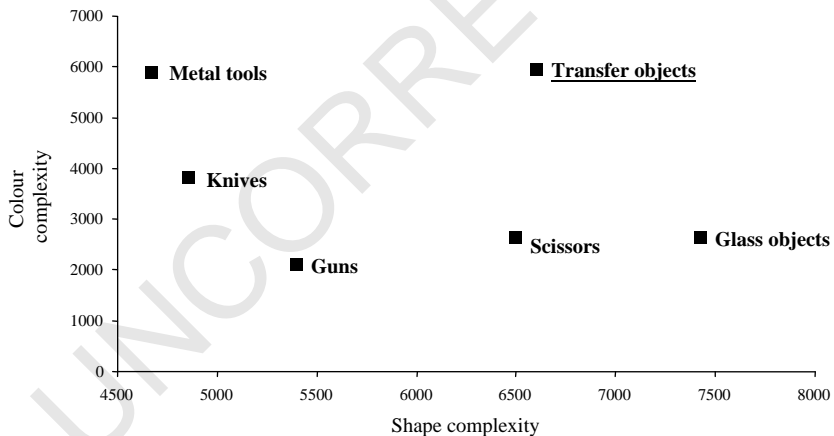

Figure 2.   Graph depicting target categories formed by the pretest classification from participants. The "transfer objects" were selected because they are not classified as part of any of the other categories.

Experimental design.   Participants were randomly assigned to one of three possible training conditions: high diversity, low diversity, and no training. After the training phase, participants were transferred to a condition that involved the detection of novel targets (as identified in the pretest). The experiment ran for two consecutive days. The training phase was completed during Day 1, and the transfer phase during Day 2. The control group received no training, and the transfer phase was completed during Day 2. During the training phase, participants were shown a "memory set" that included the targets to look for in the following block of bags. During the transfer phase, participants were not shown the targets to look for, which belonged to the "novel" category, and they were asked to use their "best judgement" to find possible targets in the bags. Based on the pretest described earlier (and in **AQ1** Brunstein & Gonzalez, 2010), categorical training diversity varied at two levels: (1) *low diversity*—five targets in the memory set were randomly drawn (with replacement) from only one category randomly selected from the five possible categories (e.g., only knives, or only guns, etc.) and this was the only category used throughout the training phase; and (2) *high diversity*—five targets in the memory set were randomly drawn (with replacement) from all five categories (one object from each of the five categories: knives, guns, scissors, glass objects, metal tools) in each block.

The transfer phase consisted of only one block of 100 trials. The bag images were the same during training and transfer, but the targets changed at transfer. However, the possibility of participants implicitly associating a particular bag with a target was extremely low given the complexity of the bags themselves and the high diversity of bags and associations of the backgrounds and the targets used. The bags were also scaled for comparability in physical similarity and difficulty, to prevent specific bag-target associations during training. This procedure of scaling the bags is explained **AQ1** in detail in Brunstein and Gonzalez (2010). The base rate of targets was 50% in both the training and transfer phases (e.g., only 50 out of 100 bags had a target).

Participants.   A total of 36 undergraduate and graduate students completed the experiment. Of these, 12 participants were randomly assigned to the high diversity condition, 12 to the low diversity condition, and 12 to the control condition. All participants were right-handed, had normal colour vision, and had normal or corrected-to-normal visual acuity. Participants were 23.35 years old on average ($SD = .61$), 35% were females and 65% were males. All participants were recruited from local universities and were paid a total of $15 for their participation. The total participation time did not exceed 1 hour on Day 1 (training phase) and 30 min on Day 2 (transfer phase) of the experiment.

Procedure.   During the training phase, participants were presented with 400 luggage images in four blocks of 100 trials each. During the transfer phase, participants were presented with 100 luggage images in one block of 100 trials. The training phase lasted about 1 hour and consisted of four blocks with 100 trials each and a base rate of 50% for targets. That is, only 50 of the 100 trials in each block had a target. During the training phase, participants were asked to memorise a set of five targets before each of four blocks. Then, they were asked to search for any member of that set during the 100 trials of the block.

On Day 1, participants were asked to memorise a set of targets at the beginning of each block. On each trial, a luggage image appeared in the centre of the screen for 4 s. Participants were required to search for any member of the memory set that appeared in the luggage image and click on the target when they detected it. If they did not click on the image, the trial timed out after 4 s, followed by a text message indicating whether they had generated a correct diagnosis or not.

On Day 2, for the transfer phase, the procedure was essentially the same; however, the targets used in the transfer phase were different from those used in the training phase. In the transfer phase, the targets were randomly selected from the "novel" category (as per the pretest explained earlier). Also, the participants were not shown these targets prior to search. Participants were not shown a memory set at transfer to keep targets novel and unknown to the participants. Participants were only told that their task was to look for possible dangerous objects in the bags using their best judgement and were not given feedback. They were not given any information about the categories and they were not told that the items to look for were "novel" or different from what they had searched for during training. They were only asked to use their best judgement to detect what the targets were during the transfer phase and they did not receive feedback. Participants in the control group performed only on the transfer phase without prior training.

The dependent variables were hit rate, false alarm rate, and detection time on correct detections (in seconds). A hit was defined as clicks on images in which a target was present and in which the participants clicked on the correct location of the target (the target was defined as a rectangular area surrounding the target). We used repeated-measures ANOVAs to investigate the effect of categorical diversity during training, one-way ANOVAs for the transfer phase, and *t*-tests for comparison between the conditions and the control condition during transfer.

Results

Training data were analysed in a 2 (categorical diversity) × 4 (block) mixed design. Transfer data were analysed for the effect of the different training

conditions and means were compared against the control condition. We provide results from the analyses of variance and *t*-tests for hit rates, false alarm rates, and detection time in both the training and transfer phases later.

Training phase. During training, there was a clear advantage for individuals in the low diversity condition over those who trained in the high diversity condition, $F(1, 22) = 38.01$, $p < .001$, $\chi^2 = .63$. Figure 3 (top panel)
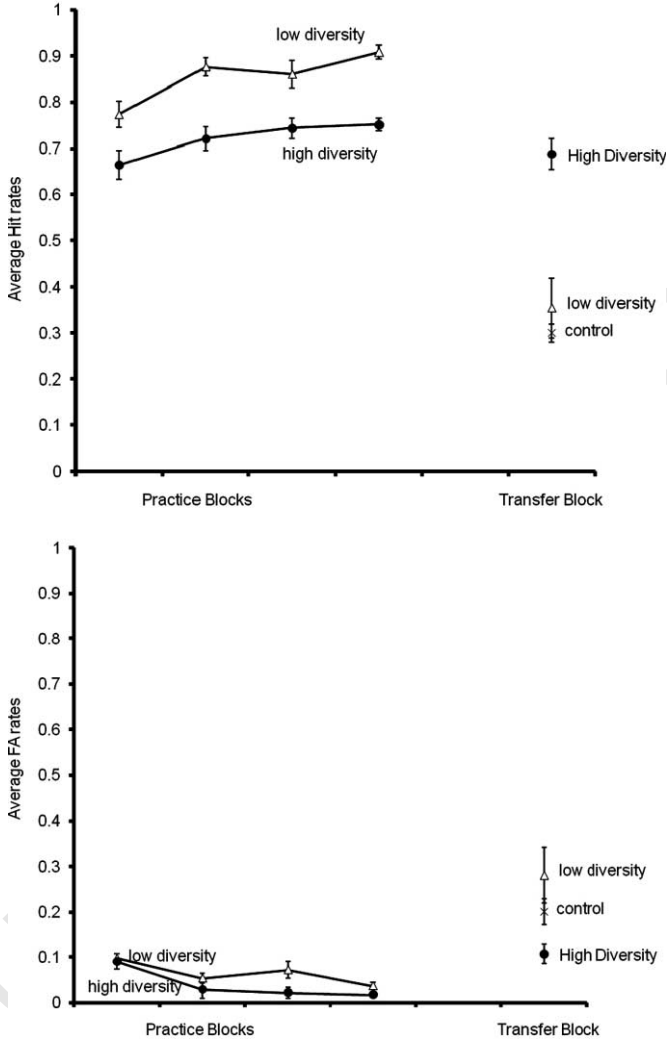

Figure 3. Average hit rates (top panel) and false alarms rates (bottom panel) per block and diversity condition during the training and transfer phases. Error bars represent standard errors.

presents the average hit rates for the training and transfer phases. On average, training with items drawn from only one category resulted in higher hit rates ($M = 0.85$, $SD = 0.06$) compared to training with items drawn from all five categories ($M = 0.72$, $SD = 0.04$); there was an improvement of hit rates over time as shown by the significant effect of the block, $F(3, 66) = 10.14$, $p < .001$, $\chi^2 = .32$, and there were no significant interactions, $F(3, 66) = 0.68$, $p > .05$, where $\chi^2 = 0$.

Categorical diversity did not have a significant effect on participants' false alarm rates, $F(1, 22) = 4.34$, $p > .05$, $\chi^2 = .17$. The average false alarm rates (shown in Figure 3, bottom panel) were low during training and no significant difference was present between the low diversity ($M = 0.06$, $SD = 0.03$) and the high diversity conditions ($M = 0.04$, $SD = 0.03$). False alarm rates significantly decreased across blocks, $F(3, 66) = 11.77$, $p < .001$, $\chi^2 = .35$, but there was no significant interaction between the blocks and the diversity condition, $F(3, 66) = 1.16$, $p > .05$, $\chi^2 = .05$.

Again, during the training phase, there was an advantage for individuals who trained in the low diversity condition, $F(1, 22) = 6.48$, $p < .05$, $\chi^2 = .22$. Figure 4 shows the average detection time per block and diversity condition
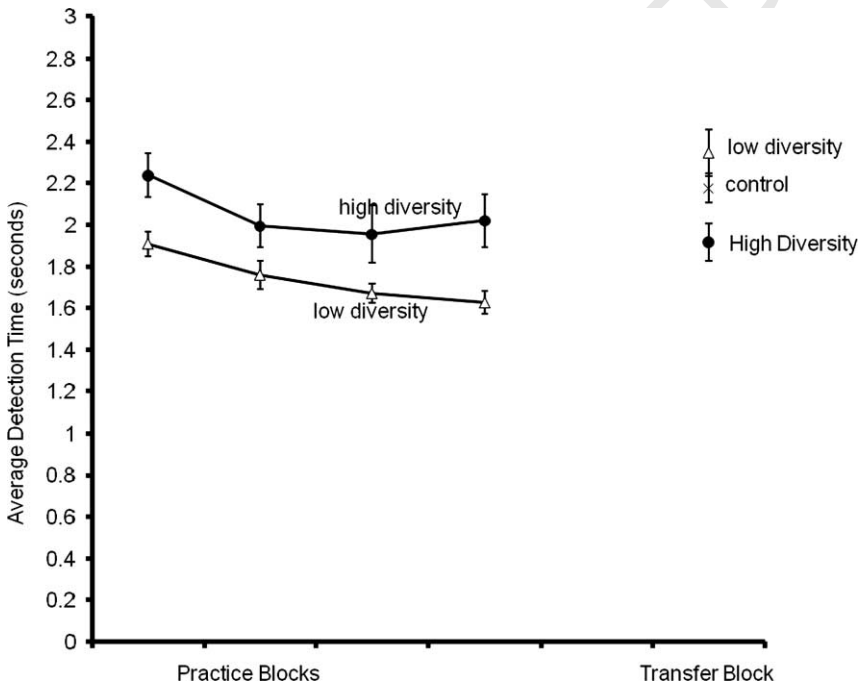

Figure 4. Average detection time per block and level of diversity during the training and transfer phases. Error bars represent the standard error.

for the training and transfer phases. On average, low diversity resulted in faster detections ($M = 1.74$ s, $SD = 0.12$) compared to high categorical diversity ($M = 2.05$ s, $SD = 0.13$). Detection time also decreased significantly over practice for all groups, $F(3, 66) = 17.31$, $p < .001$, $\chi^2 = .44$, and there was no significant interaction between training and the diversity condition, $F(3, 66) = 1.36$, $p > .05$, $\chi^2 = .06$.

Transfer phase. During the transfer phase, the diversity conditions (including the control condition) resulted in significantly different hit rates, $F(2, 57) = 21.02$, $p < .001$, $\chi^2 = .49$; false alarm rates, $F(2, 57) = 7.35$, $p < .05$, $\chi^2 = .25$; and detection time, $F(2, 57) = 8.94$, $p < .01$, $\chi^2 = .29$, compared to the training phase. The hit rates were higher for the group that trained in the high categorical diversity condition ($M = 0.69$, $SD = 0.03$) rather than the low categorical diversity condition ($M = 0.35$, $SD = 0.06$), $t(46) = 4.58$, $p < .001$. Compared to low categorical diversity, high diversity also had lower false alarm rates, $t(46) = -2.71$, $p < .05$; and shorter detection times, $t(46) = -2.99$, $p < .01$. The high diversity condition produced significantly higher hit rates ($M = 0.69$, $SD = 0.03$) than the control condition ($M = 0.30$, $SD = 0.02$), $t(22) = -9.86$, $p < .001$; significantly fewer false alarms ($M = 0.11$, $SD = 0.02$) than the control condition ($M = 0.20$, $SD = 0.03$), $t(22) = 2.67$, $p < .05$; and lower detection time ($M = 1.92$ s, $SD = 0.09$) than the control condition ($M = 2.18$, $SD = 0.07$), $t(22) = 2.27$, $p < .05$. In contrast, the hit rates in the low diversity condition ($M = 0.35$, $SD = 0.06$) were no different than the control condition ($M = 0.30$, $SD = 0.02$), $t(34) = -0.79$, $p > .05$; neither were the false alarm rates ($M = 0.28$, $SD = 0.06$ for low diversity and $M = 0.20$, $SD = 0.03$ for the control), $t(34) = -1.19$, $p > .05$; and the detection time ($M = 2.71$ s, $SD = 0.57$ for low diversity and $M = 2.18$, $SD = 0.24$ for the control), $t(34) = -1.29$, $p > .05$.

Discussion

This research tested the benefits of training variability on the detection of novel targets at transfer. Training with exemplars from diverse categories produced higher hit rates, lower false alarms rates, and faster detection times at the detection of novel targets, than training with exemplars from only one category.

There is evidence that humans are sensitive to exemplar diversity when learning new categories: Category variability makes it harder to learn than less variable categories (e.g., Fried & Holyoak, 1984;). Our results during training support these findings. High diversity consistently produced lower hit rates than low diversity during training. There is also evidence that variable categories help in generalising to novel members of that category,

especially if those members are outside the range of the trained category (e.g., Cohen, Nosofky, & Zaki, 2001; Hahn et al., 2005). Our results advance these result by showing that categorical training variability helps in generalising to novel members of a novel category. Higher hit rates, lower false alarms rates, and lower detection time were found at transfer as a result of categorical diversity training compared to one-category training.

Our explanation of the results is that training variability increased the ability to categorise novel dangerous objects as threats. Although the exact explanation of how this benefit of categorical diversity develops through diverse training will need further empirical work, a possibility is that training variability improved some cognitive process such as learning to create higher level categories from exemplars. This possibility suggests that the benefits of training variability would still hold if a set of unambiguous targets from a well-known category were used at transfer. By definition, training and transfer situations have to share some aspects to allow *transfer* of acquired skills, but they must be dissimilar to some degree to allow transfer of skills to *novel* conditions. The balance between similar and dissimilar aspects is central to transfer theories (Healy et al., 2005, 2006). The Instance-Based Learning Theory predicts that variability of experiences increases the chances of retrieving instances that are different but similar to the experiences obtained during training (Gonzalez, Lerch, & Lebiere, 2003). For our results, this means that training variability enhances the cognitive processes in a way that more encompassing categories are created. This is a possibility given the current evidence that variable categories help in generalising to members that are outside the range of the trained category but belong to the same category (e.g., Cohen, Nosofsky, & Zaki, 2001; Hahn, Bailey, & Elvin, 2005). Thus, the benefit of training variability would hold if we use novel items from a well-known threat category at transfer (e.g., Tools).

Another possibility is that training variability improved some perceptual or attentional processes that helped the differentiation of targets and nontargets in bags cluttered with irrelevant objects. So far, the visual search and categorisation literature are rarely brought together (cf. Smith et al., 2005; Wolfe et al., 2007) because categorisation often requires identifying targets presented in isolation, whereas visual search requires discriminating targets from simultaneously presented distractors. Presumably, if training variability improved the visual differentiation of targets within cluttered bags, a possibility is that the benefit of training variability would disappear in the absence of cluttered images. Clearly, the clutter in the bag would determine the effectiveness of target discrimination. However, it is unlikely that the use of bags with different clutter would change our results because of two reasons. First, the same background bag images (thus, the same clutter) used

at training were also used at transfer in this experiment. Thus, the only visual items that changed from training to transfer were the targets participants searched for. This suggests that the high-diversity benefit would exist even if the target objects were presented in isolation. Second, the objects used at transfer do not "look like" traditional weapons. In fact, a reason why these objects were selected for the transfer phase is that our pretest demonstrated that the novel objects used at transfer were not classified into a single category, while the objects used in the training phase were clearly identified in different categories such as "knives", "guns", etc., based on their visual appearance.

The exact explanations of how the benefits of categorical diversity develop through training will need further empirical work. An interesting one is the explanation of the dynamics involved in enhancing human processing through diverse categories. The computational model reported in Gonzalez et al. (2003) indicates that similarity represents one of the most important ways in which performance can improve or worsen in dynamic tasks. For example, according to this model, the generation of diverse cue values increased the diversity of instances, which in turn facilitated long-term learning (Gonzalez et al., 2003). In the current experiment, exemplars in the low diversity condition involved only one category of weapons (e.g., guns). The cues that represented this particular category of weapons likely became more readily available in memory with practice, and participants tended to become faster and more accurate during training while searching for items from only one category. However, on the flip side, they became less flexible due to the focus on only one set of cue values (Gonzalez et al., 2003). High diversity of instances provides a wider range of values from which to retrieve instances from memory, increasing the likelihood that if a "novel" stimulus were presented in the environment, something similar to the novel object might be found in memory.

The main practical contribution of this research is to suggest a way to improve human accuracy of detecting potentially dangerous items in complex visual images. Our results indicate that consistency in practice is damaging in those cases where operators are likely to be faced with novel situations and when rapid adaptation to new and unexpected situations is necessary. Our conclusions suggest new implications for training luggage screeners and for training in general: the best and fastest detection of novel items at transfer can be achieved through diverse rather than consistent training.

REFERENCES

AQ1   Brunstein, A., & Gonzalez, C. (2010). Preparing for novelty with diverse training. *Manuscript under review*.

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory and Cognition*, *29*(8), 1165–1175.

Douvis, S. J. (2005). Variable practice in learning the forehand drive in tennis. *Perceptual and Motor Skills*, *101*(2), 531–545.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework of classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 234–257.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory and Cognition*, *33*(2), 289–302.

Healy, A. F., Wohldmann, E. L., Parker, J. T., & Bourne, L. E., Jr. (2005). Skill training, retention, and transfer: The effects of a concurrent secondary task. *Memory and Cognition*, *33*(8), 1457–1471.

Healy, A. F., Wohldmann, E. L., Sutton, E. M., & Bourne, L. E., Jr. (2006). Specificity effects in training and transfer of speeded responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 534–546.

Kelso, J. A. S., & Norman, P. E. (1978). Motor schema formation in children. *Developmental Psychology*, *14*(2), 153–156.

Lacson, F. C., Gonzalez, C., & Madhavan, P. (2008). Framing and context effects in visual search training. In *Proceedings of the Human Factors and Ergonomics Society 52nd annual meeting* (pp. 348–352). New York City: Human Factors and Ergonomics Society.

Madhavan, P., Gonzalez, C., & Lacson, F. (2007). Differential base rate training influences detection of novel targets in a complex visual inspection task. In *Proceedings of the Human Factors and Ergonomics Society 51st annual meeting* (pp. 392–396). Baltimore, MD: Human Factors and Ergonomics Society.

Madhavan, P., & Wiegmann, D. A. (2005). Cognitive anchoring on self-generated decisions reduces operator reliance on automated diagnostic aids. *Human Factors*, *47*(2), 332–341.

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *American Psychological Society*, *15*(5), 302–306.

Schilling, M. A., Vidal, P., Ployhart, R. E., & Marangoni, A. (2003). Learning by doing *something else*: Variation, relatedness, and the learning curve. *Management Science*, *49*(1), 39–56.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 352–359.

Smith, J. D., Redford, J. S., Gent, L. C., & Washburn, D. A. (2005). Visual search and the collapse of categorization. *Journal of Experimental Psychology: General*, *134*(4), 443–460.

Smith, J. D., Redford, J. S., Washburn, D. A., & Taglialatela, L. A. (2005). Specific-token effects in screening tasks: Possible implications for aviation security. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1171–1185.

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*(4), 623–638.

Wulf, G. (1991). The effect of type of practice on motor learning in children. *Applied Cognitive Psychology*, *5*(2), 123–134.

Wulf, G., & Schmidt, R. A. (1997). Variability of practice and implicit motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 987–1006.

Yechiam, E., Erev, I., & Gopher, D. (2001). On the potential value and limitations of emphasis change and other exploration-enhancing training methods. *Journal of Experimental Psychology: Applied*, *7*(4), 277–285.

420

UNCORRECTED PROOF