Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale
2017, Vol. 71, No. 1, 32–39

1196-1961/17/$12.00 http://dx.doi.org/10.1037/cep0000124

Training Perceptual Experts: Feedback, Labels, and Contrasts

Rachel A. Searston and Jason M. Tangen
The University of Queensland

Are strategies for learning in education effective for learning in applied visual domains, such as fingerprint identification? We compare the effect of practice with immediate corrective feedback (feedback training), generating labels for features of matching and mismatching fingerprints (labels training), and contrasting matching and mismatching fingerprints (contrast training). We benchmark these strategies against a baseline of regular practice discriminating fingerprints. We found that all 3 training protocols—feedback, labels, and contrasts—resulted in a significantly greater ability to discriminate new pairs of prints (independent of response bias) than the baseline training protocol. We also found that feedback and labels training produced significantly lower rates of bias (i.e., learners in these groups were less likely to overcall matches) compared with baseline training. Our results demonstrate 3 different ways to boost expertise with matching prints, and have direct application to training perceptual expertise.

Keywords: perceptual expertise, categorisation, feedback, transfer, forensic science

There is a growing literature aimed at understanding the study behaviours and metacognitive abilities (and illusions) of learners (Bjork, Dunlosky, & Kornell, 2013; Dunlosky & Lipko, 2007)—as well as a surge of research aimed at pinpointing the best and most generalisable ways to practice (Bjork et al., 2013; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Hattie, 2009). Several robust learning strategies—typically referred to as *desirable difficulties*—have been identified, which result in superior performance during transfer and after a delay (e.g., interleaving, retrieval practice, elaborative interrogation; Bjork & Bjork, 2011; Schmidt & Bjork, 1992). Other work has focused on the role of feedback (Butler, Karpicke, & Roediger, 2008; Hattie & Timperley, 2007; Pashler, Cepeda, Wixted, & Rohrer, 2005). A large portion of these studies, however, have relied on stimuli geared toward education in schools or material intended to be memorised (e.g., English-Swahili word pairs). Here, we extend these learning strategies to the applied visual domain of fingerprint identification.

Fingerprint examiners typically refer to their expertise as being based on "training and experience" with matching and mismatching fingerprints (Busey & Parada, 2010); they spend their days determining whether a fingerprint collected at a crime scene belongs to the same finger or different fingers as a candidate print. Their years of training and experience has been the benchmark for courts to accept fingerprint evidence. Remarkably, this benchmark has existed with very little pressure to empirically demonstrate the quality and effectiveness of current training programs to produce genuine expertise (Tangen, Thompson, & McCarthy, 2011). Several prominent scientific bodies have now encouraged the development of research programs on human observer performance in forensic examinations (Campbell, 2011; National Research Council, Committee on Identifying the Needs of the Forensic Science Community, 2009; National Institute of Standards and Technology (NIST), 2012). In particular, these reports highlight a need to establish improved, empirically validated, and standardised training programs for forensic examiners. We test the effect of three learning strategies for developing fingerprint expertise, comparing them to "individuation training" or practice with matching prints without feedback, labels, or any additional categorical information.

Rachel A. Searston and Jason M. Tangen, School of Psychology, The University of Queensland.

Correspondence concerning this article should be addressed to Rachel A. Searston, School of Psychology, The University of Queensland, St Lucia QLD, Australia 4072. E-mail: rachel.searston@gmail.com

Feedback

Immediate feedback is often prescribed, alongside deliberate practice, as an effective training tool for developing expertise (Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson & Lehmann, 1996). In the education and learning literatures, however, the specific role of immediate feedback is less clear (Hattie & Timperley, 2007). Some researchers have found learning interventions to be more helpful when feedback is gradually reduced during practice (Wulf & Schmidt, 1989), when learners have a choice to skip feedback in favour of more retrieval practice (Hays, Kornell, & Bjork, 2010), and when immediate feedback is provided only after responding incorrectly (Karpicke & Roediger, 2007, 2008; Pashler et al., 2005), responding correctly with little confidence (Butler et al., 2008), or responding correctly with a delay (Kulhavy & Anderson, 1972; Smith & Kimball, 2010; but see Metcalfe, Kornell, & Finn, 2009). Immediate feedback has also been shown to have an undesirable effect on learners' metacognitive judgments (Kornell & Rhodes, 2013).

These studies show that the effect of feedback varies substantially across conditions, but few studies have examined its use for developing visual expertise (see White, Kemp, Kemp, Jenkins, & Burton, 2014 for one exception), and we were unable to find any previous studies that have directly compared the effect of feedback training with other learning strategies in an applied visual domain. We compared the effect of feedback training to two other training protocols (in addition to our baseline, no feedback protocol) as a tool for learning to distinguish fingerprints. White et al. (2014) found immediate corrective feedback to be more effective for learning to distinguish unfamiliar faces by identity than practice with no feedback, and we predicted a similar result with fingerprints.

Labels

Learners in many applied domains typically begin by consulting guides, textbooks, and standards that provide annotated prototypical examples of to-be-learned categories. Medical students study lists of clinical features that are typical of various diseases (Kulatunga-Moruzi, Brooks, & Norman, 2004), the latest guidelines for forensic face recognition promote a careful and deliberative process of describing and comparing the features of a face (eyes, ears, etc.; Facial Identification Scientific Working Group (FISWG), 2012), and fingerprint examiners are encouraged to identify and articulate the features of a print early on in training (e.g., ridge events, creases, and scars; Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2012).

Generating informational features or verbal descriptors like "polygonal shaped papule" in dermatology, "detached earlobes" in face recognition, or "ridge ending" in fingerprints, likely directs learning of the particular instantiations of a category (Brooks & Hannah, 2006). Indeed, feature lists and verbal descriptors have been shown to improve the diagnostic acumen of novices in medicine (Brooks, LeBlanc, & Norman, 2000; Norman, Brooks, Regehr, Marriott, & Shali, 1996). Providing the correct diagnosis provokes novices to identify more features (Brooks et al., 2000), and their performance improves when they list features after they have made a diagnosis (Norman, Brooks, Colle, & Hatala, 1999). Similarly, in education, prompting learners to generate an explanation for an answer or stated fact—a process referred to as *elaborative interrogation*—results in better cued-recall performance than reading a pregenerated explanation (Pressley, McDaniel, Turnure, Wood, & Ahmad, 1987; see Dunlosky et al., 2013 for a review). We devised a *labels* training procedure to extend these manipulations to fingerprints: learners view pairs of fingerprints labelled as a "Match" or a "No Match" (equivalent to a correct diagnosis in medicine or a stated fact in education), and ask participants to generate their own labels or descriptors that best characterise the similarities or differences between the prints. We predicted that this process of generating descriptors for labelled fingerprint pairs would result in a similar boost to discrimination ability on a test of transfer as feedback training, by directing learners toward more diagnostic visual dimensions.

Contrasts

Another strategy found to be effective for learning visual categories in particular (e.g., learning to recognise the style of an artist), is interleaved practice. This strategy involves presenting exemplars (e.g., paintings by Cézanne) in a sequence that is intermixed with exemplars of other categories (e.g., paintings by Matisse or Monet; Hatala, Brooks, & Norman, 2003; Kornell & Bjork, 2008; Zulkiply & Burt, 2013). Interleaving is effective for learning even when the different categories are presented simultaneously (Kang & Pashler, 2012; Wahleim, Dunlosky, & Jacoby, 2011), suggesting that such learning effects are not explained by spacing exemplars in time. Interleaving is thought to aid generalisation by promoting attention to the differences between categories (Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010), and by decreasing the fluency of processing (Bjork et al., 2013). Conversely, presenting two exemplars of the same category, either back to back or simultaneously (i.e., massing or blocking), is thought to draw attention to the commonalities among members of a category. Consistent with this view, other work has demonstrated the benefits of interleaved practice to be most pronounced for homogenous categories (Carvalho & Goldstone, 2014).

In fingerprints, novices tend to overcall matches when the mismatching pairs are highly similar (Tangen et al., 2011), and such cases are increasingly common in practice because of the use of computer algorithms to help narrow down the list of candidate prints (Dror & Mnookin, 2010). Increasing the saliency of *between*-finger differences through a process similar to interleaving should, therefore, be particularly beneficial for learning to distinguish highly similar distractor prints. We devised a contrast training protocol that allows learners to compare a fingerprint alongside a different matching print and a mismatching print at the same time. In other words, learners can contrast *within*-finger variance with *between*-finger variance simultaneously. Our goal here is to help learners distinguish between the visual information that is because of differences within the same print (i.e., distortion, slippage, pressure, etc.), and the visual information that is because of differences between different fingers.

Method

Participants

The participants were 100 undergraduate psychology students from The University of Queensland with no prior experience viewing fingerprints. Course credit was provided in exchange for participation. Each participant was randomly allocated to one of four training conditions: individuation training (17 women; *Mean Age* = 20.1), individuation training with feedback (13 women; *Mean Age* = 20.8), labels training (12 women; *Mean Age* = 19.8), or contrast training (17 women; *Mean Age* = 19.3). There were 25 participants in each condition.

Stimuli

The fingerprint set used in this experiment were 100 fingerprint trios. Each trio consisted of an exemplar fingerprint, a fingerprint left by the same finger on a separate occasion (for match trials), and a highly similar distractor print returned from a search of the Australian National Automated Fingerprint Identification System (for mismatch trials). The highly similar distractors were the most highly ranked mismatching candidate print returned by the search algorithm available in the Queensland Police hard-copy archives, which contains approximately 3.3 million prints (see Tangen et al., 2011 for full details). The fingerprints were photographed and cropped to 512 × 512 pixels with the impressions isolated in the centre on the image. Some of the fingerprints were fully rolled and others were slaps, where the finger is pressed down and not rolled from one side to the other, introducing greater variation between instances of the same finger.

For each participant, 50 fingerprint trios were randomly allocated to the training phase of the experiment and the remaining 50 to the test phase. From the pool of 50 training trios, a random 25 were paired with the impression from the same finger (generating the matching pairs), and the remaining 25 with the highly similar impression from another individual (generating mismatching pairs). This same method was used to generate matching and mismatching pairs for the test. Each exemplar print had a chance to be paired with its match or its distractor, and to appear during training or test, which also varied for each participant so that any differences observed between training groups are unlikely to be because of low level image characteristics.

Procedure

The experiment was displayed on 23 in. iMac computers with headphones. Learners in all four training groups first read an information sheet and listened to an instructional video about the nature of the training before completing one of the four training sessions. In each of the instructional videos, learners were shown an example of a matching and mismatching fingerprint pair, and were shown what their respective training environments looked like. All learners were instructed that they would be tested on their ability to discriminate fingerprints later in the experiment before completing the training phase. After completing the training phase, learners in each group watched another instructional video showing what their test environment looked like (including an example matching and mismatching trial) before completing the test. The test involved judging 50 novel pairs of fingerprints as belonging to the *same* finger or *different* fingers and responses were indicated on the same 12-point confidence rating scale used during training. None of the prints were presented previously and no feedback was provided during the test—resembling real-world casework conditions, where ground truth information (e.g., knowledge of whether the prints were left by the same finger or two different fingers) is not available. Progression through the test was self-paced, such that participants only advanced to the next case after providing a rating and clicking "OK."

Individuation training (baseline). *Baseline* learners viewed 50 pairs of fingerprints (half matching and half mismatching), one after the other, and were instructed to judge whether the two impressions in each case belonged to the *same* finger or whether they belonged to two *different* fingers (see Figure 1 for an example of matching and mismatching fingerprints). The fingerprints were the only source of information and they remained on the screen until learners provided a rating and moved on to the next case. As with each of the training conditions, the sequence of matching and mismatching trials was random for each participant. Learners used a 12-point, forced choice confidence rating scale which appeared below the fingerprints in each case. This scale is the same used in previous experiments (e.g., Searston, Tangen, & Eva, 2015; Tangen et al., 2011) and ranges from 1 (*sure different*) to 12 (*sure same*). Ratings of 1 to 6 indicated a "no match" decision (i.e., the learner thought the two prints belong to two different fingers) and ratings 7 to 12 indicated a "match" decision (i.e., the learner thought the two prints belong to the same finger). Forcing one of two decisions allowed us to examine changes in learners discrimination ability or accuracy as well as changes in their decision strategy or response bias (Green & Swets, 1966).

Feedback training. *Feedback* learners judged 50 pairs of fingerprints (half matching and half mismatching) as either belonging to the *same* finger or two *different* fingers using the same 12-point rating scale as the baseline training group. In addition, they received trial by trial corrective feedback on their decisions in line with previous manipulations of feedback training (e.g., White et al., 2014). Immediately after providing a rating in each case, the fingerprints were removed and they were presented with the word "Correct" in bold green font in the middle of the screen for correct decisions (i.e., correctly calling or rejecting a match) or the word "Incorrect" in bold red font for incorrect decisions (i.e., falsely calling or missing a match). The feedback remained on the screen for a further 3 s before participants advanced to the next case.

Labels training. Learners in the *labels* training group viewed 50 pairs of fingerprints, one after the other (half matching half mismatching), but they did not explicitly engage in individuation training. Instead, these learners were provided with the correct responses as labels above each pair of fingerprints. On matching trials, the word "Match" appeared in bold font above the prints, and on mismatching trials, "No Match" appeared in the same font and location. In each case, learners were instructed to list the similarities and dissimilarities between the two prints in a text box below them, resembling manipulations of elaborative interrogation in education research (e.g., Norman et al., 1999; Pressley et al., 1987). Learners could not advance to the next trial if the text box was empty.

Contrast training. Learners in the *contrast* training group viewed 50 sets of four fingerprint on the screen, one after the other. In each case, the same exemplar fingerprint appeared

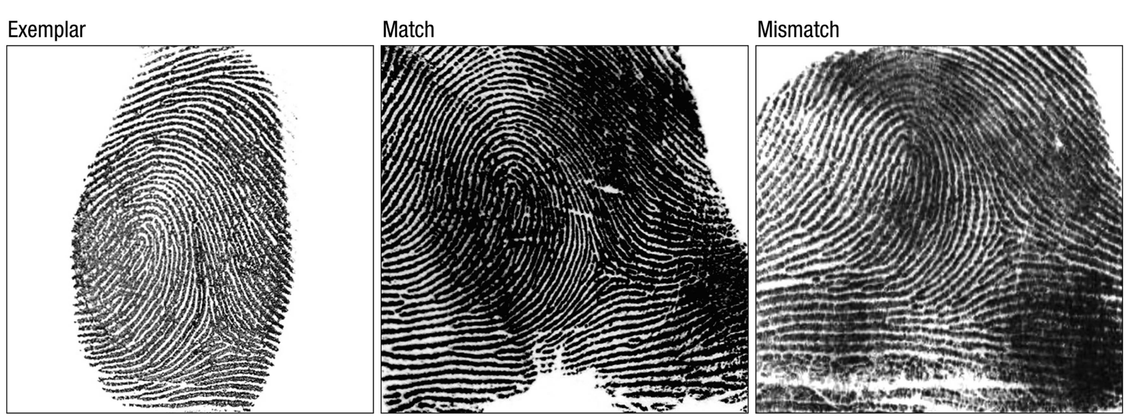

Figure 1. A sample trio consisting of an exemplar, a matching, and a mismatching impression.

twice on the screen: once alongside a different fingerprint left by the same finger (i.e., matching pair), and again alongside a highly similar fingerprint from another individual (i.e., mismatching pair), which resulted in a simultaneous interleaving protocol (Kang & Pashler, 2012; Wahlheim et al., 2011). Learners were instructed to compare and contrast the two pairs before judging each pair individually as a match or a mismatch without feedback and were not able to advance to the next trial until providing a judgment about both pairs (a 12-point rating scale was located on the right of each pair in each case). Whether the matching pair appeared at the top or the bottom of the screen was randomised on each trial and for each participant. Participants were not instructed that each trial consisted of one matching and one mismatching pair.

Results

Learners who engaged in individuation training with feedback made the most correct decisions on the transfer test (83.23%), followed by learners in the labels training group (79.46%), learners in the contrast training group (76.79%), and learners in the baseline group (i.e., individuation training without feedback; 67.00%). To examine differences in discriminability independent of response strategy (i.e., tending to say "match" or "no match" more regardless of the correct response), we computed the average discrimination ability (A') and response bias (B''_D) separately for each individual learner in all four training groups (see Donaldson, 1992 on the use of A' and B''_D as nonparametric measures of performance). Again, learners in the individuation with feedback training group were the most accurate on the test (*Mean* $A' = .89$) followed by learners in the labels training group (*Mean* $A' = .86$), learners in the contrast training group (*Mean* $A' = .84$), and learners in the baseline training group (*Mean* $A' = .75$). An interesting find was that learners in the individuation with feedback group (*Mean* $B''_D = 0.00$), labels group (*Mean* $B''_D = .13$), and contrast training group (*Mean* $B''_D = -.18$), also displayed less bias on the test than learners in the baseline group (*Mean* $B''_D = -.53$). Learners in the baseline training group tended to say "match" more than any other group regardless of whether the fingerprints actually matched or not (see Figure 2).

Signal Detection Analyses

We first subjected learners A' scores and B''_D scores on the transfer test to two separate one-way, between-subjects analyses of variance. These analyses showed a significant main effect of Training Type (baseline, feedback, labels, or contrast) on discriminability, $F(3, 96) = 7.10$, $MSE = 0.01$, $p < .001$, $\eta^2_G = .18$, and response bias, $F(3, 96) = 8.85$, $MSE = 0.23$, $p < .001$, $\eta^2_G = .22$. Tukey's pairwise comparisons using learners A' scores further revealed that, compared with the baseline group, the feedback, $t(48) = 4.34$, $p < .001$, $d = 1.05$, labels, $t(48) = 3.50$, $p = .004$, $d = .84$ and contrast training, $t(48) = 2.90$, $p = .023$, $d = .70$, groups were all significantly more accurate on the transfer test. Tukey's pairwise comparisons of learners' B''_D scores also revealed that learners in the feedback, $t(48) = 3.90$, $p = .001$, $d = 1.12$, and labels group, $t(48) = 4.85$, $p < .001$, $d = 1.54$, were significantly less (liberally) biased than those in the baseline training group. The difference between the B''_D scores in the baseline group and contrast training group was not significant, $t(48) = 2.60$, $p = .052$, $d = .68$. None of the comparisons between the feedback, labels, and contrast training groups were significant.

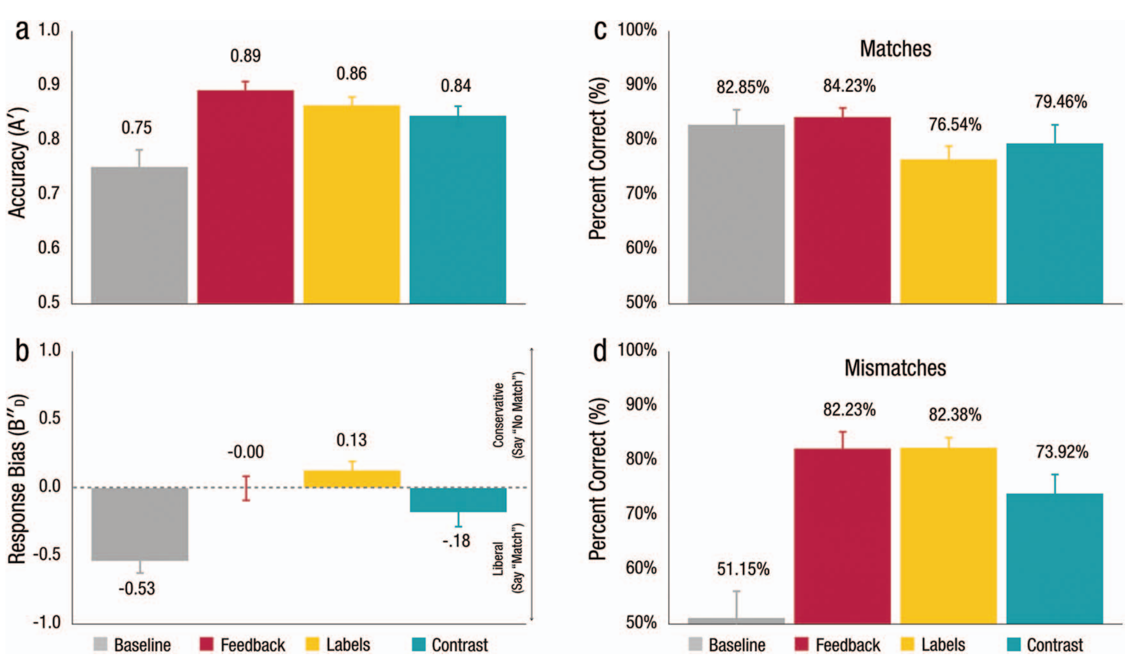

Figure 2. Mean test discriminability (A'; a) and response bias (B''_D; b) for each training group, as well as their mean percent correct for matching (c) and mismatching trials (d). The error bars indicate the *SEM*. See the online article for the color version of this figure.

Percent Correct Analyses

We also examined differences in learners' percent correct scores for matching and mismatching cases by conducting a Training Type (baseline, feedback, labels, or contrast) × Trial Type (match or mismatch) mixed analysis of variance. We found a significant main effect of Training Type, $F(3, 96) = 9.07$, $MSE = 0.03$, $p < .001$, $\eta^2_G = .12$, Trial Type, $F(3, 96) = 13.03$, $MSE = 0.03$, $p < .001$, $\eta^2_G = .06$, as well as a significant interaction between the two, $F(3, 96) = 12.39$, $MSE = 0.03$, $p < .001$, $\eta^2_G = .16$. To investigate this interaction further, we conducted Tukey's pairwise comparisons between each of the training groups for matching and mismatching trials. For matching trials, there was no significant difference in percent correct on the test between the baseline learners and feedback learners, $t(48) = .35$, $p = .985$, between the baseline learners and labels learners, $t(48) = 1.59$, $p = .392$, or between the baseline learners and contrast learners, $t(48) = .85$, $p = .830$. For mismatching trials on the other hand, there was a significant different between the baseline training group and the feedback training group, $t(48) = 6.01$, $p < .001$, $d = 1.50$, between the baseline training group and the labels training group, $t(48) = 6.04$, $p < .001$, $d = 1.65$, and between the baseline and contrast training groups, $t(48) = 4.40$, $p < .001$, $d = 1.04$. None of the comparison between the feedback, labels, and contrast training groups were significant.

Time Taken to Complete the Experiment and Response Times at Test

Baseline learners completed the experiment in the shortest period of time (*Mean Time to Complete Experiment* = 23 min and 46 s), followed by feedback learners (*Mean Time to Complete Experiment* = 30 min and 34 s), contrast learners (*Mean Time to Complete Experiment* = 31 min and 24 s), and then the labels learners (*Mean Time to Complete Experiment* = 39 min and 43 s). A look at learners' mean response times at test can provide further insight into where each of these groups are spending their additional time. Baseline learners responded the fastest at test, on average (*Mean Response Time* = 10.23 s), but the feedback (*Mean Response Time* = 10.77 s), labels (*Mean Response Time* = 11.27 s), and contrast learners (*Mean Response Time* = 11.49 s) responded with similar speed. We found no significant differences in response times at test between the four groups of learners, $F(3, 96) = 1.53$, $p = .213$, $\eta^2_G = .05$.

Finally, we also examined whether the 25 fastest learners (*Mean Time to Complete Experiment* = 14 min and 51 s) and the 25 slowest learners (*Mean Time to Complete Experiment* = 53 min and 00 s) differed in their ability to discriminate pairs of prints on the test, irrespective of their training group. Despite spending substantially more time on training trials, the 25 slowest learners displayed only slightly better discrimination ability on the test (*Mean* $A' = .84$), compared with the 25 fastest learners (*Mean* $A' = .83$), and this difference was not significant, $t(48) = .47$, $p = .642$, $d = .09$.

Discussion

Despite a reliance on training and experience as an index of expertise in courts, there is a paucity of research examining whether certain training protocols are more effective than others at producing genuine perceptual expertise in forensic domains. Here, we produce such data, drawing on well-established learning strategies used in education. We have shown that corrective feedback is a powerful tool for learning to discriminate between highly similar fingerprints, resulting in significantly greater discrimination ability and lower rates of bias on a test of transfer than individuation training alone. We have also shown two other equally effective ways to boost fingerprint discrimination performance: generating labels for similar and dissimilar features between prints, and contrasting matching and mismatching image-pairs. Learners in the labels and contrast training groups also produced lower rates of bias on the test compared with a baseline group. However, this reduction in bias was only significant for those who received the labels training, who produced a slightly conservative response bias (i.e., tending to overcall mismatches) on the test compared with the liberal bias (i.e., tending to overcall matches) displayed by baseline and contrast learners. The benefits of *feedback*, *labels*, and *contrast* training were isolated to mismatching trials, which were selected for their high similarity.

The result of learners showing above baseline discrimination ability across the three training groups, suggests the effects we observed for mismatching trials were partly because of an increase in sensitivity to information that is diagnostic of highly similar, but mismatching prints. This finding is consistent with previous work demonstrating that fingerprint experts with years of experience perform particularly well compared with novices at identifying highly similar distractors (Tangen et al., 2011; Thompson, Tangen, & McCarthy, 2014). Furthermore, while baseline learners spent considerably less time completing the experiment compared with the other groups, time taken to complete the experiment did not have a significant bearing on learners discrimination ability on the test. Learners' response times on the test also did not differ significantly between each of the groups. These results indicate that feedback, labels, and contrast learners spent substantially more time examining fingerprint pairs during training compared with baseline learners, but this additional exposure is not likely to account for the effects we have observed.

The reduced bias observed for the feedback and labels training groups suggests that these two protocols further helped to calibrate learners' decision strategy. It is possible that the added information of "Correct"/"Incorrect" in the feedback training protocol and "Match"/"No Match" in the labels training protocol enabled learners to tune their response bias so as to reflect the underlying base rates of matching and mismatching pairs. This interpretation is in line with previous work showing that recognition memory false alarm rates are dependent on base rate information when feedback is provided (Estes & Maddox, 1995), or when base rate information is provided in advance (Ratcliff, Sheu, & Gronlund, 1992). In practice, the base rates of matching and mismatching pairs are likely to differ from the equal base rate scenario we set up in our experiment, and it is worth considering how injecting truthful information into an examiners' workflow (e.g., by providing feedback on cases where the ground truth is known) might affect their response strategies across the board: boosting discrimination ability through feedback or labels might push around an examiners' response bias in unintended ways. If this is the case, contrast training may be a viable alternative.

Our present design does not allow us to determine whether the greater sensitivity displayed by learners in the labels training group was because of the "Match"/"No Match" information, generating descriptors for similar and dissimilar features, or a combination of the two. One way to test this would be to include a condition where learners rate the prints as matching or mismatching without feedback before listing similarities and dissimilarities. Prior work in education certainly suggests the process of describing why something is true—such as describing why two fingerprints belong to the same finger—is more helpful for learning than passively gathering truthful information (Pressley et al., 1987; see Dunlosky et al., 2013). Our labels training protocol was based on this elaborative interrogation technique and it is likely that generating similar and dissimilar features encouraged learners to attend closely to particular instantiations of generic features that are diagnostic of a correct response (Brooks & Hannah, 2006).

One other caveat concerns the contrast training protocol. We chose to equate the number of training trials rather than the number of images, meaning that the increased sensitivity observed for contrast learners may be because of their exposure to a greater number of prints during training. It is difficult to draw any firm conclusions about the source of the contrast learning effect we observed with fingerprints without further research that equates the number of exemplars. However, previous work has shown that viewing images of different bird species in pairs versus singles aids classification of new birds, even with the same number of exemplars across conditions (Wahlheim et al., 2011). It is possible that the simultaneous presentation of matching and mismatching prints contributed to the greater sensitivity we observed by highlighting differences between matching and a highly similar distractor prints. This explanation is consistent with discrimination accounts of spacing effects in category learning (Kornell & Bjork, 2008; Kornell et al., 2010).

Practical Implications

We rely on forensic examiners to detect passport fraud or to identify individuals who were at the scene of a crime from images of faces, fingerprints, shoe prints, tool marks, firearm impressions, and bite marks. The purpose of the present study was to broadly compare different strategies for training novices in these applied visual domains to accurately classify members of a category, using fingerprint identification as a testbed. As a starting point, we examined variants of three strategies that have been shown to be effective in education: corrective feedback, generating lists of features (similar to previous elaborative interrogation strategies; Norman et al., 1999; Pressley et al., 1987), and contrasting category exemplars (similar to simultaneous interleaving interventions in previous work; Kang & Pashler, 2012; Wahlheim et al., 2011). We also set a high bar for these learning strategies by comparing them to individuation training or regular practice with matching fingerprints. Identifying and discriminating exemplars of a category or object (e.g., a finger) without feedback or labels encourages retrieval of similar prior instances (Searston et al., 2015), and the more salient dimensions that were diagnostic in those cases are likely to come to mind more readily on retrieval than less diagnostic details. Retrieval practice certainly has a well-documented positive effect on learning in education (Bjork et al., 2013; Karpicke & Roediger, 2007, 2008), and our *feedback*, *labels*, and *contrast* training protocols all surpassed this formidable benchmark, showing promise as training tools in applied visual domains.

Further work is needed to clarify the specific mechanisms underlying the learning effects we observed, as well as their generality to other domains of perceptual expertise (e.g., unfamiliar face recognition or radiology). Future investigations may wish to examine our hypothesis that feedback serves to direct learners attention toward underlying base rate information. Combining the learning strategies we have adapted here might also produce a cumulative learning benefit, above and beyond the benefit provided by each strategy separately, which may help to shed light on whether each strategy is contributing to developing different aspects of perceptual expertise. Practically, it would be useful to explore the effects of feedback, labels and contrast training on learning in other forensic domains that rely on visual comparison of images. Little is known about whether humans are able to develop genuine expertise in discriminating shoe prints, tool marks, firearm impressions or bite marks, and further research aimed at identifying general learning strategies that are effective across these different disciplines will help to address concerns regarding the validity of these domains (Campbell, 2011; National Research Council, Committee on Identifying the Needs of the Forensic Science Community, 2009; NIST, 2012). Our results demonstrate three different ways to boost expertise with matching fingerprints, serving as a first step toward addressing the need to establish improved, empirically validated, and standardised training programs for forensic examiners.

Résumé

Les stratégies d'apprentissage utilisées en éducation sont-elles aussi efficaces pour l'apprentissage de domaines visuels appliqués, comme l'identification d'empreintes digitales? Nous évaluons l'effet de la pratique par rapport à la rétroaction corrective immédiate (formation sur la rétroaction), à la génération d'étiquettes relatives à des caractéristiques d'empreintes digitales appariées et non appariées (formation sur l'étiquetage) et à la comparaison entre des empreintes digitales appariées et non appariées (formation sur la comparaison). Nous comparons ces stratégies à des pratiques régulières de discrimination d'empreintes digitales. Nous avons constaté que les trois protocoles de formation – rétroaction, étiquettes et comparaison – ont entraîné une plus grande habileté à discriminer les nouvelles paires d'empreintes (indépendamment du biais méthodologique) par rapport au protocole de formation de base. Nous avons également constaté que les formations basées sur la rétroaction et sur les étiquetage produisaient des taux de biais considérablement inférieurs (par ex., les apprenants de ces groupes étaient moins susceptibles de manquer les paires appariées) par rapport aux formations de base. Nos résultats présentent trois différentes façons de renforcer l'expertise en matière d'appariement d'empreintes digitales et ont une application directe sur la formation de l'expertise perceptuelle.

Mots-clés : expertise perceptuelle, catégorisation, rétroaction, transfert, sciences judiciaires.

References

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth Publishers.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64,* 417–444. http://dx.doi.org/10.1146/annurev-psych-113011-143823

Brooks, L. R., & Hannah, S. D. (2006). Instantiated features and the use of "rules". *Journal of Experimental Psychology: General, 135,* 133–151. http://dx.doi.org/10.1037/0278-7393.32.6.1416

Brooks, L. R., LeBlanc, V. R., & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychological Science, 11,* 112–117. http://dx.doi.org/10.1111/1467-9280.00225

Busey, T. A., & Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin & Review, 17,* 155–160. http://dx.doi.org/10.3758/PBR.17.2.155

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918–928. http://dx.doi.org/10.1037/0278-7393.34.4.918

Campbell, A. (2011). *The fingerprint inquiry report*. Retrieved from http://www.thefingerprintinquiryscotland.org.uk/inquiry/3127-2.html

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition, 42,* 481–495. http://dx.doi.org/10.3758/s13421-013-0371-0

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121,* 275–277. http://dx.doi.org/10.1037/0096-3445.121.3.275

Dror, I. E., & Mnookin, J. L. (2010). The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk, 9,* 47–67. http://dx.doi.org/10.1093/lpr/mgp031

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16,* 228–232. http://dx.doi.org/10.1111/j.1467-8721.2007.00509.x

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions From cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406. http://dx.doi.org/10.1037/0033-295X.100.3.363

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273–305. http://dx.doi.org/10.1146/annurev.psych.47.1.273

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1075–1095. http://dx.doi.org/10.1037/0278-7393.21.5.1075

Facial Identification Scientific Working Group (FISWG). (2012). *Guidelines for facial comparison methods version 1.0*. Retrieved from https://www.fiswg.org/document/viewDocuments

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education: Theory and Practice, 8,* 17–26. http://dx.doi.org/10.1023/A:1022687404380

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81–112. http://dx.doi.org/10.3102/003465430298487

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin & Review, 17,* 797–801. http://dx.doi.org/10.3758/PBR.17.6.797

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97–103. http://dx.doi.org/10.1002/acp.1801

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151–162. http://dx.doi.org/10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968. http://dx.doi.org/10.1126/science.1152408

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19,* 585–592. http://dx.doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging, 25,* 498–503. http://dx.doi.org/10.1037/a0017807

Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied, 19,* 1–13. http://dx.doi.org/10.1037/a0032147

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2004). Using comprehensive feature lists to bias medical diagnosis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 563–572. http://dx.doi.org/10.1037/0278-7393.30.3.563

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 63,* 505–512. http://dx.doi.org/10.1037/h0033243

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition, 37,* 1077–1087. http://dx.doi.org/10.3758/MC.37.8.1077

National Research Council, Committee on Identifying the Needs of the Forensic Science Community. (2009). *Strengthening forensic science in the United States: A path forward*. Washington, DC: The National Academies Press.

National Institute of Standards and Technology (NIST). (2012). *Latent print examination and human factors: Improving the practice through a systems approach: The report of the expert Working Group on human factors in latent print analysis*. Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology.

Norman, G. R., Brooks, L. R., Colle, C. L., & Hatala, R. M. (1999). The benefit of diagnostic hypotheses in clinical reasoning: Experimental study of an instructional intervention for forward and backward reasoning. *Cognition and Instruction, 17,* 433–448. http://dx.doi.org/10.1207/S1532690XCI1704_3

Norman, G. R., Brooks, L. R., Regehr, G., Marriott, M., & Shali, V. (1996). Impact of feature interpretation on medical student diagnostic performance. *Academic Medicine, 71*(Suppl.), S108–S109. http://dx.doi.org/10.1097/00001888-199601000-00059

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychol-*

ogy: Learning, Memory, and Cognition, 31, 3–8. http://dx.doi.org/10.1037/0278-7393.31.1.3

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 291–300. http://dx.doi.org/10.1037/0278-7393.13.2.291

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535. http://dx.doi.org/10.1037/0033-295X.99.3.518

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. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00029.x

Searston, R. A., Tangen, J. M., & Eva, K. W. (2015). Putting bias into context: The role of familiarity in identification. *Law and Human Behavior, 40,* 50–64. http://dx.doi.org/10.1037/lhb0000154

Smith, T. A., & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 80–95. http://dx.doi.org/10.1037/a0017407

Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST). (2012). *Standard for the documentation of analysis, comparison, evaluation, and verification version 2.0*. Retrieved from http://www.swgfast.org/Documents.html

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science, 22,* 995–997. http://dx.doi.org/10.1177/0956797611414729

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior, 38,* 84–93. http://dx.doi.org/10.1037/lhb0000051

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39,* 750–763. http://dx.doi.org/10.3758/s13421-010-0063-y

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review, 21,* 100–106. http://dx.doi.org/10.3758/s13423-013-0475-3

Wulf, G., & Schmidt, R. A. (1989). The learning of generalized motor programs: Reducing the relative frequency of knowledge of results enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 748–757. http://dx.doi.org/10.1037/0278-7393.15.4.748

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41,* 16–27. http://dx.doi.org/10.3758/s13421-012-0238-9

Received June 26, 2016
Accepted December 14, 2016 ■