Article:

Dowsett, Andrew J. and Burton, A. Mike (2015) Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. British journal of psychology. pp. 433-445. ISSN: 0007-1269

https://doi.org/10.1111/bjop.12103

In press: British Journal of Psychology, to be published 2015

Unfamiliar Face Matching: Pairs out-perform individuals and provide a route to training

Andrew J. Dowsett & A. Mike Burton
School of Psychology, University of Aberdeen, UK

Acknowledgements
This work was supported by awards from the Economic and Social Research Council, UK, and the European Research Council to Burton (ES/J022950/1; ERC-ADG 323262).

Correspondence to:
Mike Burton, School of Psychology, University of Aberdeen, AB24 3FX, UK
m.burton@abdn.ac.uk   Tel. +44 1224 273930

Running head: Face matching in pairs and individually

Word Count (excluding abstract and figure legend): 5655

Unfamiliar Face Matching: Pairs out-perform individuals and provide a route to training

Abstract

Matching unfamiliar faces is known to be difficult. Here we ask whether performance can be improved by asking viewers to work in pairs, a manipulation known to increase accuracy for low-level visual discrimination tasks. Across four experiments we consistently find that face matching accuracy is higher for pairs of viewers than for individuals. This 'pairs advantage' is generally driven by adopting the response of the higher-scoring partner. However, when the task becomes difficult, both partners' performance is improved by working in a pair. In two experiments we find evidence that working in a pair can lead to subsequent improvements in individual performance, specifically for viewers whose accuracy is initially low. The pairs technique therefore offers the opportunity for substantial improvements in face matching performance, along with an added training benefit.

Keywords

Face recognition; unfamiliar face matching; identity verification; perception in groups.

Photographic identification is commonly used to prove who we are. Whether this is to buy age restricted items, access the workplace or cross international borders, there is a considerable reliance on this method of facial comparison for proving identity. However, it has consistently been shown just how poor we are at matching unfamiliar faces. For example, in the Glasgow Face Matching Task (GFMT; Burton, White, & McNeill, 2010) where participants simply have to decide if two simultaneously presented images depict the same person or two different people, participants typically make 20% errors. Similarly, in a more realistic setting, Kemp, Towell, and Pike (1997) asked supermarket cashiers to match photographic credit cards to live shoppers. Accuracy was as low as 50% when these were used "fraudulently" by someone presenting the wrong card.

Although such errors are now well-documented, it is only recently that attempts have been made to find ways to reduce them. Studies examining the use of feedback training for face matching have produced mixed results. White, Kemp, Jenkins, and Burton (2014) showed a benefit for participants trained with trial-by-trial feedback, while Alenezi and Bindemann (2013) demonstrated only that this technique halted an overall decline in performance across a series of trials. Attempts have been made to improve unfamiliar face matching in other ways, such as through increased motivation (Moore & Johnston, 2013), facial caricaturing (McIntyre, Hancock, Kittler, & Langton, 2013) and adding multiple images to photo-ID (White, Burton, Jenkins, & Kemp, 2014), all with limited success.

In this study, we examine the potential benefit of matching by groups: specifically, we ask whether pairs of viewers achieve higher performance than individuals. In the face matching literature, there is rather little precedent for this. Bruce, Henderson, Newman and Burton (2001; Experiment 3), showed participants 7s video clips of faces. Participants viewed these either in isolation, or in pairs, where they were encouraged to discuss the faces. On a subsequent matching test, viewers who had originally been exposed to the faces in pairs showed higher accuracy. More recently, White, Burton, Kemp and Jenkins (2013) tested a *Wisdom of Crowds* approach (Kerr & Tindale, 2004; Surowiecki, 2004). They found the standard poor performance on average for individuals, but this was much improved when groups of responses were pooled. With pools of 32, accuracy was near perfect – far higher than any individual. This is, of course, a statistical property of uncertain responses: the groups in White et al's study were notional, in that they all made their responses individually.

However, combined with the Bruce et al (2001) study, it does suggest that it may be worth exploring face matching using multiple viewers.

In the studies below, we ask pairs to discuss two images, and come to a joint decision as to whether they depict the same person or two different people. While this has not been directly explored in face matching, Bahrami, Olsen, Latham, Roepstorff, Rees, and Frith (2010) have demonstrated the benefits of this communicative, paired decision making using low level visual stimuli. In a series of experiments, participants saw two separate displays of six vertically oriented Gabor patches, one after the other. Their task was to determine which of the two displays contained a Gabor patch with elevated contrast. All participants performed the task as a pair. They first offered their own individual response, and in the event of a disagreement, they were instructed to discuss their choice, before coming to a joint decision. Joint responses were consistently better than those of individuals and, interestingly, this was true even in the absence of trial-by-trial feedback. Bahrami et al (2010) hypothesise that this benefit is due to the sharing of subjective estimates of confidence, allowing the least confident of the pair on any trial to defer. In support of this hypothesis, Koriat (2012) replicated the ‘pair advantage’ in the absence of any dyadic interaction, simply by selecting the more confident participant on each trial in a virtual dyad.

While the above research demonstrates improved decision making in groups, it is important to note classic findings which demonstrate that, due to conformity, groups may not always perform best. First, Asch (1955) demonstrated using a simple perceptual task based on line measurements that individuals as part of a group of confederates will make a response which they do not necessarily believe in order to comply with group norms. Second, Sherif (1936) has shown that groups of individuals can often make false reports because they are dependent on one another for accurate information, which is not always available. These findings have had a large impact in explaining the problems surrounding eyewitness identification. It is often found that if eyewitnesses are allowed to discuss the identity present during a crime or a scene, then they are likely to conform to one another’s reports, leading to inaccuracies which could have profound results (Wright, Memon, Skagerberg, & Gabbert, 2009). Although these results on social conformity are important, and demonstrate a negative impact on face perception, we should note that face memory and face matching are two very different tasks which social conformity may affect in different ways.

In the current study we seek to investigate whether unfamiliar face matching can be improved by having participants perform as a pair. In experiment 1 half of our participants performed the standardised GFMT alone, and half completed the same task as a pair. In experiment 2, again using items from the GFMT, we tested the effects of paired viewing within-subjects. Participants performed a baseline test alone, followed by a test in pairs, and finally a third test alone. In experiment 3, we replicated experiment 2 using new and more difficult matching items in order to generalise results. To anticipate, we consistently found a pairs advantage, by comparison to individual performance. We also report evidence that working in pairs can lead to some subsequent individual improvement in performance.

Experiment 1

In this experiment, viewers completed the GFMT either in pairs or individually. Pairs were asked to discuss the images on screen, and to come to a joint decision as to whether they depicted the same person or two different people. There were two groups of individual respondents. In one group, the participants completed the test entirely alone. In the second, participants carried out the test in the presence of the experimenter. We hypothesised that any effect of motivation, which could potentially underlie a pairs advantage, would be reduced by asking participants to make their decisions with the experimenter present. The simple presence of the experimenter may help in maintaining participants' focus. If performance in this third condition is poorer than in a pairs condition, this would provide some support for Bahrami et al (2010) in proposing that an advantage arises through a rich exchange of information, through which participants can make optimal decisions.

Methods

Stimuli

This experiment used the short version of the Glasgow Face Matching Test (Burton et al, 2010). This comprises 20 same and 20 different face pairs, which are established to be relatively difficult. The face images are all front facing, and cropped of any background. The two images in any pair were taken minutes apart, but with different cameras. Here they were presented on-screen, side-by-side in greyscale at a resolution of 1000x700 pixels each.

Participants

60 undergraduate students (40 female; mean age 22) took part in the experiment, in exchange for a small payment.

Procedure

Participants were randomly assigned to *pair, individual* or *individual-with-experimenter* conditions. Those in the pairs condition were randomly seated, side-by-side, in front of a single computer screen placed in the middle of the two participants, at a distance of approximately 50cm. Following a two second central fixation cross, a face pair would appear centrally on screen until the participants had made their response. Participants in the pairs condition were instructed to discuss the images on screen, and to come to a joint decision as to whether the face pair depicted a match or mismatch. Responses were made using one of two keyboard buttons – and pairs were told that either member could press the appropriate button. Accuracy was emphasised and the order of trials was randomly determined by the experimental software.

Participants in the "individual" condition performed the experiment as above, except they did it alone and so could not discuss the stimuli with anyone. Participants in the "individual-with-experimenter" condition, made individual decisions, but gave their response verbally to the experimenter who was seated opposite them in the room.

Results and Discussion

Mean accuracy across conditions is shown in figure 1. A one way ANOVA demonstrated a significant effect of experimental condition ($F(2,42)=7.88$, $p<0.005$). Follow-up Tukey HSD tests showed a significant advantage for the pairs condition over both individual conditions (pairs vs individual, $F(2,42) = 10.18$, $p < 0.001$; pairs vs individual-with-experimenter, $F(2,42) = 13.25$, $p < 0.001$), but no difference between the two individual conditions ($F<1$).

FIGURE 1 HERE PLEASE

These results show quite clearly that unfamiliar face matching can be improved by asking participants to do the task together as pairs. It is therefore interesting to ask how this pairs advantage comes about. One hypothesis is that the presence of another person improves the motivation of participants to improve their performance. To examine this possibility, we included an ‘individual-with-experimenter’ condition, which provided the extra person, but no extra interaction. Interestingly, this provided no benefit over simple individual performance, suggesting that the locus of the effect lies in the interaction itself. We explore this further in the next experiment.

Experiment 2

Experiment 1 demonstrated that accuracy in an unfamiliar face matching task can be improved by having participants make decisions as a pair. However, the experiment does not allow us to establish whether the advantage is restricted to the pairs setting, or whether it might generalise to subsequent performance by the individuals involved. A study by Maciejovsky, Sutter, Budescu and Bernau (2013), has demonstrated that enhanced performance for “groups” can be used as a training method for subsequent individual performance, when solving the Wason selection task. Maciejovsky et al (2013) suggest that this transfer in knowledge is due to team members challenging one another, which subsequently results in deeper and more critical levels of analysis and thinking. Of course, this reasoning task is very different from our perceptual judgement task, but nevertheless it would be interesting to establish whether a similar knowledge transfer might occur between “paired” and subsequent “individual” face matching.

The purpose of the current experiment was to establish whether unfamiliar face matching in pairs can act as a training mechanism for subsequent individual unfamiliar face matching. To do this, an extended version of the short GFMT was used, consisting of 60 trials overall. Participants first undertook 20 pre-training trials to determine their baseline performance. Next, they undertook a further 20 trials as part of a pair. Finally, participants performed 20 trials individually, to investigate any effect of “paired training”. This design provides useful extensions of Experiment 1. First, it allows us to replicate the individuals vs pairs comparison using a within subjects design. On the basis of the original experiment, we predict that performance in the pairs phase will exceed that in the initial individual phase. Second, it allows us to establish whether there might be a learning effect in the experiment – either due to practice or to working in pairs. Comparison between individual performance in phases 1 and 3 will establish whether there is any general improvement through the study.

Methods

Stimuli

As above, stimuli comprised face pairs. These were the 40 trials (20 match, 20 mismatch) used in experiment 1, plus a further 20 trials (10 same and 10 different) selected from the long version of the GFMT. These 60 trials were divided into three sets of equal difficulty, based on performance data from 300 participants in the original GFMT test construction (see Burton et al, 2010).

Participants

36 undergraduate students, with a mean age of 22, took part in this experiment in exchange for a small payment.

Procedure

All participants were invited to the lab in pairs. The general procedure for each of the three blocks remained the same. Face pairs were presented on screen until a response had been made using one of two keyboard keys, at which point the images were replaced by a two second central fixation cross, before proceeding to the next trial. The order of trials within each block was randomly determined by the experimental software and accuracy was emphasised.

Participants completed three blocks: phase 1 individual; phase 2 pairs; phase 3 individual. For phases 1 and 3 participants were seated opposite each other at separate computers. This ensured that they could not see one another's screens, nor could they communicate. For phase 2 they were brought together around one computer, such that both were looking at the same screen at the same time. During this phase participants were instructed to discuss the images on screen and come to a joint decision as to whether they depict the same person, or two different people. Stimulus sets used for each block were counterbalanced, so that across the experiment, each face-pair appeared equally often in the three blocks (see *Stimuli* above).

Results and Discussion

Mean accuracy across conditions is shown in Figure 2. One-way repeated measures ANOVA demonstrated a significant effect of experimental condition ($F(2,70)=12.73$, $p<0.001$). Follow-up Tukey HSD tests showed a significant advantage for the pairs condition over both individual conditions (phase 2 vs phase 1, $F(2,70) = 23.55$, $p < 0.001$; phase 2 vs phase 3, $F(2,70) = 13.13$, $p < 0.001$), but no difference between the two individual conditions (phase 1 vs phase 3, $F(2,70) = 1.51$, n.s.).

FIGURE 2 HERE PLEASE

These results replicate the findings of Experiment 1, within subjects, suggesting that when collaborating, pairs of participants are more accurate than individuals. However, there was no significant difference between baseline accuracy and post-test accuracy (phases 1 vs 3), suggesting that there is no general improvement over time, and no transfer of knowledge between paired performance and subsequent individual performance.

One possible explanation for this paired improvement is that pairs are only as good as their best member. Mean performance for the individual blocks may conceal individual differences such that some individuals are actually performing rather well. Once paired, it may be that discounting the weaker individual response improves performance overall. To test this possibility, we compared accuracy at baseline (phase 1) for the best performing member in each pair ("high performers") with the accuracy in the pairs block (phase 2). Mean accuracy in phase 1 for the "high performers" was 90.56% (SD=6.4%) compared with mean accuracy of 93.89% (SD=2.2%) for the pairs in phase 2. A paired samples t-test demonstrated that this difference was not significant ($t(17)=1.14$, $p=0.27$).

These results suggest large individual differences in participants' ability on this unfamiliar face matching task, a result which is already well founded within the literature (Burton et al, 2010; Megreya & Burton, 2006). Therefore, a third set of analyses sought to investigate whether working in pairs had affected the baseline performance of high or low performers. To investigate this, we examined the performance of the higher and lower scoring member of each pair at baseline (phase 1) and again at phase 3. Figure 3 shows mean accuracy across these groups. The higher performing subjects showed no significant effect of having worked in a pair (phase 1 vs phase 3: $t(17)=1.1$, $p=0.28$). However, the lower performing subjects did show an improvement ($t(17)=2.3$, $p<0.05$).

FIGURE 3 HERE PLEASE

These results demonstrate that while those participants who are good at the task do not get any better (or worse), those participants who were originally bad at the task do improve their accuracy. Our suggestion here is that these "low performers" have learned something about the task from the "high performers" during the pairs block, which has benefitted their subsequent individual performance. Alternatively, these participants may simply be improving due to practice effects. This possibility is explored below.

Experiment 3

Both Experiments 1 and 2 have shown that unfamiliar face matching can be improved by having participants perform the task in pairs. Experiment 2 also extended this finding to show that while this "pairs effect" may be driven, in part, by the high performing members of a pair, these members appear to be training lower performing individuals. However, one problem with this interpretation of the result is that accuracy in the GFMT is already quite high, and trial numbers in each block are relatively low. This could plausibly result in a ceiling effect for the high performing subjects. In short, this particular test may be obscuring an effect of working in pairs, even for the higher performing viewers. To examine this, in experiment 3a, we replicate Experiment 2 using a new stimulus set which is designed to be more difficult and contain more trials. This allows us to replicate the basic pairs-effect with new stimuli, and also to examine any putative learning effects across sessions. Experiment 3b is a control experiment in which subjects carry out three consecutive sessions of matching as individuals. Taken together, the aim is to establish whether there are general benefits to working in pairs, and whether these can be separated from simple improvements through practice.

Stimuli

The stimuli used in Experiment 3 were 180 full colour images of 135 male models, downloaded from a single portfolio website for professional models. The images were cropped to display the face only. Images were selected which were free of occlusions, and showed front facing views. They were presented in full colour at a size of 300x420 pixels. The 180 full colour images were separated into 45 match trials and 45 mismatch trials. These pairings had been predetermined by asking 6 individuals to sort 384 face images of male models into piles, based on their perceived similarity, in a procedure similar to that used by Bruce, Henderson, Greenwood, Hancock, Burton, & Miller (1999). This test was then piloted with 36 new participants, resulting in a mean accuracy of 77%. From this large set, the 90 most difficult trials were chosen, and these were divided into three separate blocks of equal difficulty, each containing 15 match and 15 mismatch trials.

Experiment 3a: Methods

Participants

36 undergraduate students, with a mean age of 22, took part in this experiment in exchange for a small payment.

Procedure

The procedure used for Experiment 3a was identical to that described in Experiment 2. Once again, stimulus sets used for each block were counterbalanced, so that across the experiment, each face-pair appeared equally often in the three blocks (see *Stimuli* above).

Results and Discussion

Figure 2 shows mean accuracy across conditions. One-way ANOVA demonstrates a significant effect of experimental condition ($F(2,70)=29.6$, $p<0.001$). Follow-up Tukey HSD tests showed a significant advantage for the pairs condition over both individual conditions (phase 2 vs phase 1, $F(2,70) = 57.02$, $p < 0.001$; phase 2 vs phase 3, $F(2,70) = 25.53$, $p < 0.001$), and also a significant advantage for later over earlier individual blocks (phase 1 vs phase 3, $F(2,70) =6.24$, $p < 0.05$).

These results replicate the previous findings that pairs are more accurate than individuals, even in a new, more difficult matching task. Furthermore, these results appear to suggest that some learning has occurred overall between paired and post-test performance. As with Experiment 2, we next compared baseline (phase 1) performance for the best performing member of each pair with the pairs block (phase 2). Mean accuracy in phase 1 for the "high performers" was 80.56% compared with mean accuracy of 86.67% for the pairs in phase 2. A paired samples t-test showed this difference to be significant ($t(17)=3.97$, $p<0.001$). This result appears to show that even the better performers benefited within the pairs phase. The absence of this effect in Experiment 2 is possibly due to a ceiling effect, such that highly accurate participants were already performing as well as possible, leaving no room for a benefit of pairing. However, in this more difficult version of the face matching test, they do show a benefit.

Finally, as with experiment 2, we ask whether the benefits of the pair phase lead to subsequent improved performance for individuals. Figure 2 suggests a general benefit, and Figure 3 shows a direct comparison between phases 1 and 3, separately for the higher and lower performers in each pair. The higher performing subjects showed no significant effect of having worked in a pair (phase 1 vs phase 3: $t(15) < 1$). However, the lower performing subjects did show an improvement ($t(15)=2.6$, $p<0.05$). (Note, in two of the pairs, both participants recorded the same accuracy at baseline, and have therefore been omitted from this analysis).

Together, these results suggest that while "high performers" take no subsequent benefit from doing the matching task as a pair, "low performers" do, replicating the findings in Experiment 2. Nevertheless, the lack of improvement in the high performers (between phases 1 and 3) is not due to a ceiling effect, because in this experiment, the pairs outperformed the higher performers. It appears then, that the lower performers are somehow being trained to do better, during the paired phase, and that this leads to a subsequent benefit, resulting in an overall difference between phases 1 and 3 (see figure 3). In comparison, the high performers are performing as well as any individual can, yet this is still not as good as the pairs. However, in order to establish whether this benefit is genuinely a result of the paired phase, rather than simple improvement over time, we conducted a final experiment in which subjects were never paired. Instead, in experiment 3b they completed three different blocks of trials, each time individually.

Experiment 3b: Methods

Participants

18 undergraduate students (mean age 24) took part in experiment 3b.

Stimuli and Procedure

Both the stimuli used and the general procedure were exactly the same as that in Experiment 3a. However, here participants were recruited individually, and completed all three sections of the experiment on their own. Once again, stimulus sets used for each block were counterbalanced, so that across the experiment, each face-pair appeared equally often in the three blocks.

Results and Discussion

Figure 2 shows mean accuracy across conditions. One-way ANOVA showed no significant difference between the blocks ($F(2,34)=1.05$, $p=0.38$). These results demonstrate that despite participants experiencing an increasing number of trials throughout the experiment, allowing them to get more practice at the task, their accuracy does not improve.

As our previous experiments have shown a difference in ability between high and low performers, we next explored this for the individual data. We separated subjects into high and low performers at phase 1, using a median split. Figure 4 shows mean accuracy across conditions. A 2 (High vs Low Performers) x 3 (Phase 1 vs Phase 2 vs Phase 3) mixed factor ANOVA showed a significant difference between "high" and "low performers" ($F(1,32)=58.92$, $p<0.001$), as expected. However, there was no main effect of phase, and no interaction ($F(2,32)=0.17$, ns, and $F(2,32)=1.01$, ns, respectively). These results show that regardless of their ability, participants' accuracy did not improve over time. This again suggests that practice does not improve performance in the task, even when participants are initially poor performers.

FIGURE 4 HERE PLEASE

Taken together, the results of experiment 3 suggest that any improvement in individual performance across phases, as seen in Experiment 3a, must be due to experience in a paired phase. Although this proposal is based on group means, we have suggested that such improvements may be driven by learning in the member of the pair whose performance starts weakest, and there is some support for this in the results of Experiment 3a.

General Discussion

Overall, our results replicate the finding that individuals are very poor at matching pairs of unfamiliar faces. In experiment 1, mean accuracy for the short version of the GFMT was 80.3%, with a range from 50-95%, results which closely replicate the original description of the test (Burton et al, 2010). This figure fell to a mean accuracy of 72.2% (Experiment 3b) when harder stimuli were used. Note, however, that these were all good quality images, presented with no time limits, and viewers simply had to make same/different responses. These results add to the growing evidence that confirming a person's identity based on matching face images is a very difficult task, on which people often fail. Despite these problems, we still rely heavily on face matching on a day-to-day basis using photo-ID.

These results show that performance on unfamiliar face matching tasks can be improved by having people make *same/different* decisions in pairs. Experiment 1 demonstrated this basic finding, between subjects, consistently showing that pairs of participants were more accurate than their individual counterparts. The same pattern of results was observed in both Experiments 2 and 3a, when participants were first tested individually and subsequently tested as a pair on novel items. These results are important as they suggest that in order to improve security protocols, such as in passport control, or for the purchasing of age restricted items, it would be better for individuals to seek advice from a second person. Analogous results have been reported by White et al (2013), who show that aggregating responses over large groups can also significantly improve accuracy on such face matching tasks. The results presented here complement the White et al (2013) study, as they confirm that groups do outperform individuals. Yet our results further suggest that one does not need very large groups to achieve this improvement. Providing they are able freely to discuss the images, groups of just two people also show significant accuracy improvements.

In order to establish the basis for these effects, we asked whether pairs were only as good as their best member. While this does appear to be the case in experiment 2, with accuracy for "high performers" at baseline being equal to subsequent paired performance, it was not true in experiment 3a. Once potential ceiling effects had been eliminated, all participants appeared to benefit from working in pairs. The fact that both group members appear to contribute to the increase in accuracy in Experiment 3a further demonstrates the need for discussion in producing accuracy improvements in unfamiliar face matching tasks.

Social facilitation would suggest that individuals in our experiments should perform better in the presence of another person (Zajonc, 1965). However, Experiment 1 demonstrated that accuracy in the GFMT was no different when individual participants were asked to give their response verbally to the experimenter and when they performed the task alone. These results suggest that the accuracy improvements observed for pairs were not a result of increased motivation, but instead that the exchange of information between pairs is what provides the benefit.

The distinction between "high performers" and "low performers" in a pair led to the suggestion that this technique could be used as training for improving individual unfamiliar face matching accuracy. Options for training have been explored before, using different techniques, with limited success (White et al, 2014; Alenezi & Bindemann, 2013; Moore & Johnston, 2014; McIntyre et al, 2014). In experiments 2 and 3a, we investigate training effects of pairs by having participants perform a final face matching block on their own. Our results consistently demonstrated that the accuracy of "low performers" in a pair improved considerably in the subsequent individual matching task, compared with the original baseline task. Furthermore, experiment 3b shows that these results are not due to practice effects, as individual performance in all three blocks remained the same throughout the task. Taken together these results demonstrate a promising avenue for training individuals at unfamiliar face matching tasks, whereby people who are good at face matching can improve the performance of those that are bad.

What is the mechanism for this pairs advantage? Bahrami (2010) suggests that communication of confidence is a key component, such that accuracy improves because pairs are able to rely on the more confident group member. Indeed, Koriat (2012) showed accuracy improvements for pairs by taking the response of the more confident member, without any need for discussion. Support for the importance of communication of confidence is provided by Fusaroli et al (2012). These authors transcribed the conversations of pairs performing the Bahrami et al (2010) task and showed that almost all conversations related to a confidence in their own individual decision. In fact, those participants who adapted to each other's expressions of confidence recorded a higher level of accuracy in the perceptual task.

While previous studies may go some way to explain our basic finding, that pairs outperform individuals, they do not necessarily provide an insight into the training effects observed in experiments 2 and 3. In order for participants to become better at the task on their own, it would not be enough simply to choose the response of the most confident partner. Instead, there seems to be evidence that viewers are learning how to become more accurate during the pairs phase of the study. Our experiments reported here are not able to establish exactly what is learned, nor how it is learned. For example, might an explicit discussion draw some viewers' attention to aspects of the stimuli they had previously neglected? On the other hand, might viewers simply reflect more deeply on their own responses, when confronted with a more confident partner? In future research, it will be necessary to study the nature of the interaction between participants. This could be achieved either by restricting the levels of interaction available (perhaps restricting exchanges to reporting of confidence). Alternatively, a full content analysis of the interactions may be revealing. Either way, the nature of the training effect observed here certainly requires future exploration.

Finally, in planning future work , it is important to note a major difference in methodology between the experiments described in the current paper and those preceding it. In the Bahrami et al (2010) task, participants discuss their joint decision without being able to see the stimuli – it disappears before discussion. Therefore, the only information they can communicate would be confidence. Conversely, in our task, participants are able to see the stimuli at all times and so discussions based on their appearance is possible. It therefore seems plausible that discussions of *both* the viewers' confidence and the visual stimuli themselves provide benefits in this task. Once again, future research should seek to differentiate between these two types of communication in order to create an effective form of unfamiliar face training.

In summary, our results show that while unfamiliar face matching is a difficult task, accuracy can be improved by having participants make these *same/different* decisions in pairs, a result which is consistent across all three experiments. We also show that this result is not due to motivational factors, the abilities of high performers or practice effects, but instead that accuracy improvements arise from pairs of participants simply being able to discuss the images on screen. Finally, our results demonstrate that the technique of pairing participants in a face matching task also works as a training mechanism for "low performers", whose accuracy improves in a subsequent individual matching task.

References

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*(6), 735–753. doi: 10.1002/acp.2968

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193*(5)*,* 31-35. doi: 10.1038/scientificamerican1155-31

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science, 329,* 1081-1085. doi: 10.1126/science.1185718

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of Face Identities From Images Captured on Video. *Journal of Experimental Psychology: Applied, 5*(4), 339-360. doi: 10.1037/1076-898X.5.4.339

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching Identities of Familiar and Unfamiliar Faces Caught on CCTV Images. *Journal of Experimental Psychology: Applied, 7*(3), 207-218. doi: 10.1037//1076-898X.7.3.207

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42*(1), 286-291. doi: 10.3758/BRM.42.1.286

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science, 23*(8), 931-939.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, *11*(3), 211-222. doi: 10.1002/(SICI)1099-0720(199706)11:3

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*(1), 623–655.

Koriat, A. (2012). When Are Two Heads Better than One and Why?. *Science, 336,* 360-362. doi: 10.1126/science.1216549

Maciejovsky, B., Sutter, M., Budescu, D. V., Bernau, P. (2013). Teams Make You Smarter: How Exposure to Teams Improves Individual Decisions in Probability and Reasoning Tasks. *Management Science, 59*(6), 1255-1270. doi: 10.1287/mnsc.1120.1668

McIntyre, A. H., Hancock, P. J. B., Kittler, J., & Langton, S. R. H. (2013). Improving Discrimination and Face Matching with Caricature. *Applied Cognitive Psychology, 27*(6), 725-734. doi: 10.1002/acp.2966

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition, 34*(4), 865-876. doi: 10.3758/BF03193433

Moore, R. M., & Johnston, R. A. (2013). Motivational Incentives Improve Unfamiliar Face Matching Accuracy. *Applied Cognitive Psychology, 27*(6), 754-760. doi: 10.1002/acp.2964

Sherif, M. (1936). *The psychology of social norms.* New York: Harper Collins.

Surowiecki, J. (2004). *The wisdom of crowds*. US: Random House.

White, D., Burton, A. M., Kemp, R. I., & Jenkins, R. (2013). Crowd Effects in Unfamiliar Face Matching. *Applied Cognitive Psychology, 27*(6), 769-777. doi: 10.1002/acp.2971

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning Photo-ID to Improve Unfamiliar Face Matching Performance. *Journal of Experimental Psychology: Applied, 20*(2), 166-173. doi: 10.1037/xap0000009

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100-106. doi: 10.3758/s13423-013-0475-3

Wright, D. B., Memon, A., Skagerberg, E. M., Gabbert, F. (2009). When Eyewitnesses Talk. *Current Directions in Psychological Science, 18*(3), 174-178. doi: 10.1111/j.1467-8721.2009.01631.x

Zajonc, R.B. (1965). Social Facilitation. *Science, 149,* 269-274.

Figure 1. Mean accuracy by condition in Experiment 1. Error bars are SE.

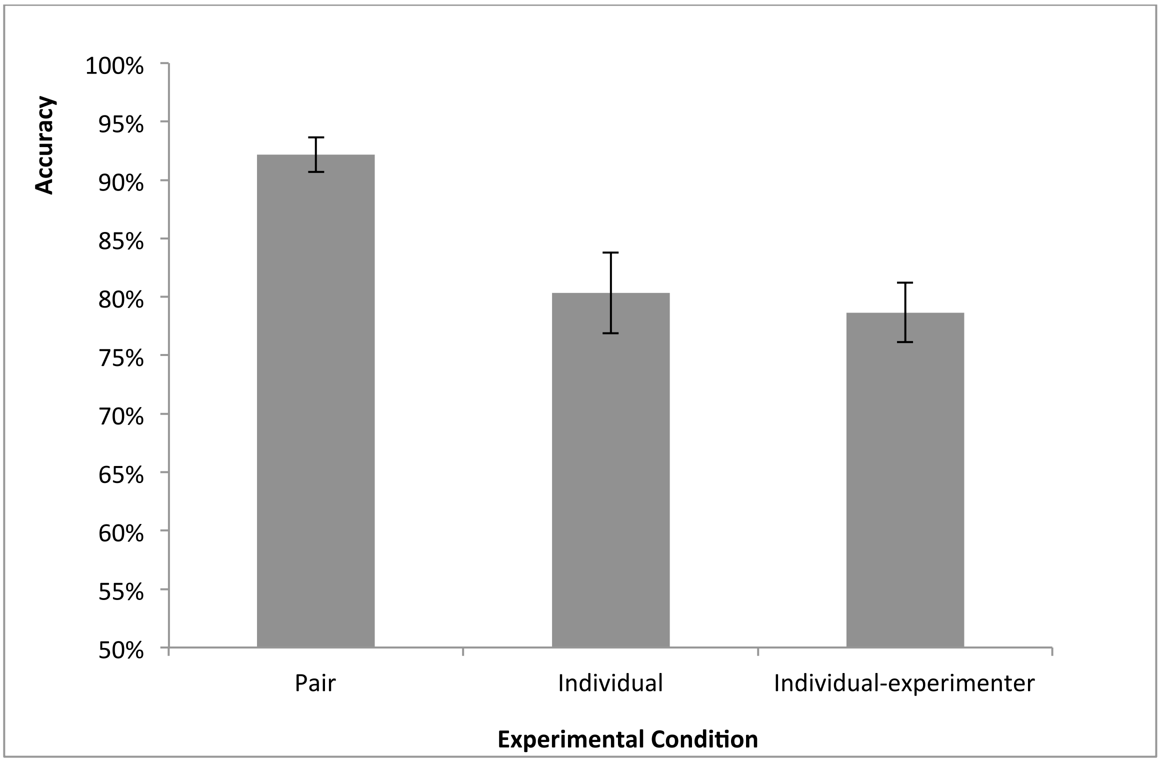


Figure 2. Mean accuracy by condition in Experiments 2 and 3.. Error bars are SE.

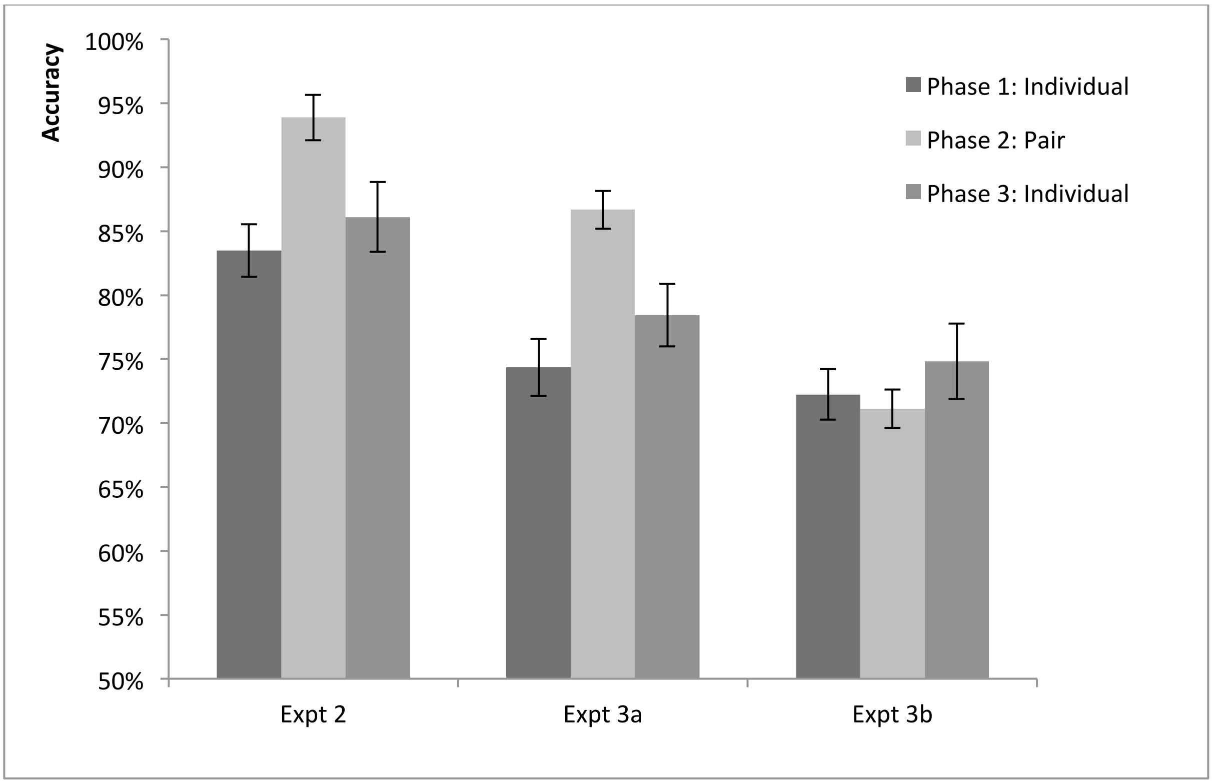
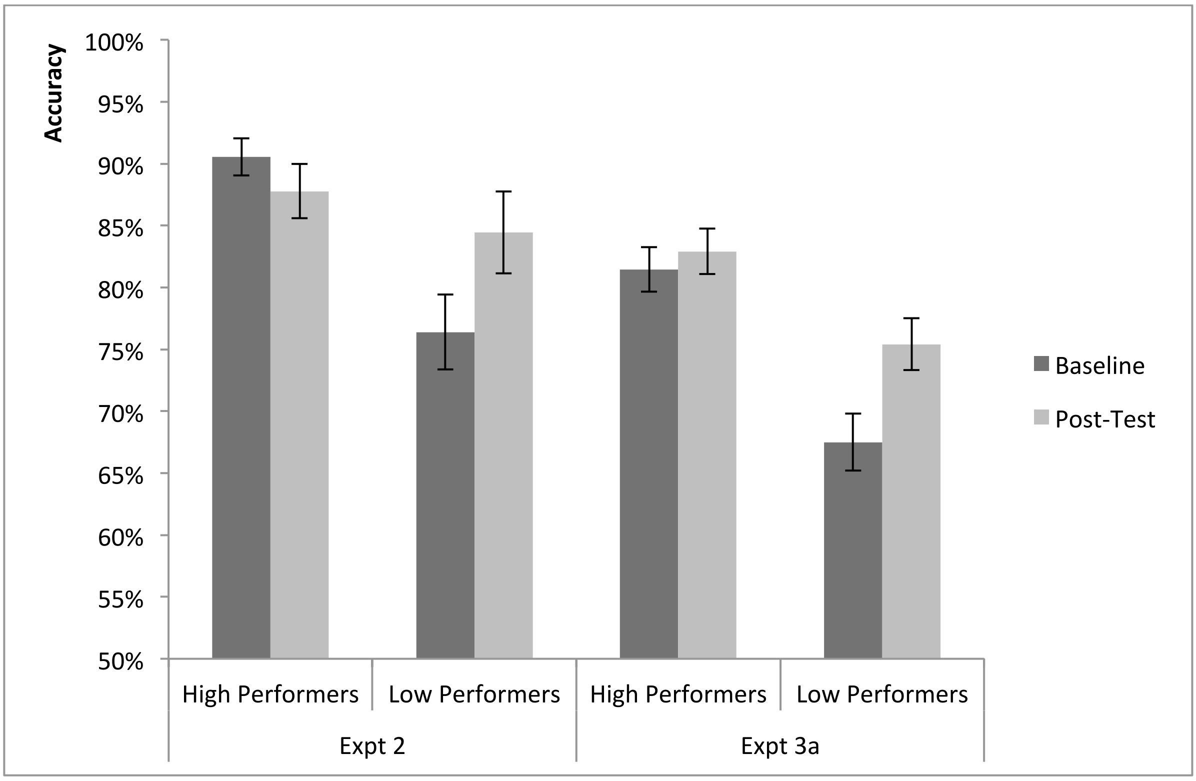


Figure 3. Mean individual accuracy for higher and lower performers in Experiments 2 and 3. Error bars are SE.



t 3.

Figure 4. Mean accuracy by condition in Experiment 3b, for high and low performers. Error bars are SE.