Telling faces together: Learning new faces through exposure to multiple instances

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID:	QJE-STD 14-265.R2
Manuscript Type:	Standard Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Andrews, Sally; University of Aberdeen, Jenkins, Rob; University of York, Cursiter, Heather; University of Glasgow, Burton, Mike; University of Aberdeen,
Keywords:	Face recognition, face perception, face learning, identity, stable representations

Telling faces together: Learning new faces through exposure to multiple instances

Sally Andrews[1], Rob Jenkins[2], Heather Cursiter[3] & A. Mike Burton[1,2]

1. School of Psychology, University of Aberdeen, UK
2. Department of Psychology, University of York, UK
3. School of Psychology, University of Glasgow, UK

Address for correspondence

Mike Burton,
Department of Psychology, University of York, York YO10 5DD, UK
mike.burton@york.ac.uk

Keywords: Face recognition, face perception, identity, face learning, stable representations

Abstract

We are usually able to recognise novel instances of familiar faces with little difficulty, yet recognition of unfamiliar faces can be dramatically impaired by natural within-person variability in appearance. In a card-sorting task for facial identity, different photos of the same unfamiliar face are often seen as different people (Jenkins, White, van Montfort & Burton, 2011). Here we report two card sorting experiments in which we manipulate whether participants know the number of identities present. Without constraints, participants sort faces into many identities. However, when told the number of identities present, they are highly accurate. This minimal contextual information appears to support viewers in 'telling faces together'. In Experiment 2 we show that exposure to within-person variability in the sorting task improves performance in a subsequent face-matching task. This appears to offer a fast route to learning generalizable representations of new faces.

Telling faces together: Learning new faces through exposure to multiple instances

Introduction

Face recognition has generally been approached from the perspective of telling faces apart. Within the homogenous category of faces, subtle differences in the appearance of different people allow us to identify thousands of individuals successfully. Yet, our understanding of this ability often ignores the fact that two instances of the same person might not look the same; that is, taken under different conditions, two images of the same person can look very different. For example, Adini, Moses and Ullman (1997) demonstrated that it is possible to find two instances of the same person that appear more visually distinct than two images of different people. In order to identify a face successfully, one must not only discriminate between *different* people, but also code information that is stable across different instances of the *same* person (Jenkins, White, van Montfort & Burton, 2011).

The face of an individual may vary for a number of reasons. Within a single interaction, expressions and speech alter the relative appearance of features, as do rotation of the head along horizontal or vertical axes (e.g. nodding or turning). Between encounters, there may be differences in the source and nature of lighting, both of which affect the apparent texture and general appearance of the face. Over longer periods, effects of ageing, health and adiposity also influence appearance dramatically. In photographs, an additional subset of variability is encountered due to characteristics of the capture device (focal length, perspective, etc). Different images typically contain a combination of variability along these dimensions.

While within-person variability is unavoidable in normal interactions, it is rarely a problem for recognition of familiar faces, which is highly robust to within-person variability. However, introducing even one dimension of variability to newly learned faces can dramatically reduce recognition accuracy. For example, a few degrees of head rotation can reduce accuracy close to baseline in simple identity tasks (Longmore, Liu & Young, 2008; Krouse; 1981; O'Toole, Edelman & Bültoff, 1998; Favelle, Palmisano & Avery, 2011). Changes in lighting conditions have a similar disruptive effect (Hill & Bruce, 1996; Longmore et al., 2008). In recognition memory tasks, simply changing expression from smiling to neutral between learning and test can introduce errors (Bruce, 1982). The difficulty is not restricted to memory; matching simultaneously-present high quality images of unfamiliar people turns out to be a very difficult task (Bruce et al., 1999; Burton, White & McNeill, 2010).

In the light of these findings, it has been proposed that familiar and unfamiliar face processing rely, to some extent, on separate processes (e.g. Hancock, Bruce & Burton, 2000; Megreya & Burton, 2006). It is well documented that familiar face recognition occurs with remarkable ease and accuracy (Bruce, 1982; Burton, Wilson, Cowan & Bruce, 1999), whereas unfamiliar face identification is much harder – even in seemingly ideal situations (e.g. Bruce et al., 1999; Clutterbuck & Johnston, 2002; 2005). In a demonstration of how difficult unfamiliar face recognition is, Jenkins and colleagues (Jenkins et al., 2011) asked participants to sort a set of 40 unfamiliar faces into piles, one pile per identity. Although there were only two identities present, unfamiliar viewers sorted them into many piles (mean 7.5, mode 9), whereas viewers familiar with the faces performed almost perfectly. Interestingly, the mistakes made by unfamiliar viewers were not mistakes of confusing two people: there were very few piles containing both identities. Instead, people seemed to have great difficulty in cohering different photos of the same person into a single identity (i.e. 'telling people together' rather than 'telling them apart').

The ability to reconcile instances that look visually distinct as the same person may require additional cues. In an unconstrained situation, where many different faces may be present, there is no reason to suspect that two instances of a face belong to the same person. Yet for unfamiliar and low-familiarity faces, context may provide the mechanism to overcome variability and to perceive different instances as the same person. An early diary study, documenting failures of identification in ordinary situations, provides an indication of the importance of context (Young, Hay and Ellis, 1985). Approximately 16% of the occasions when *familiar* faces went unrecognized were due to meeting in an unexpected context. When faces that initially seemed familiar were correctly rejected as actually being *unfamiliar*, approximately 87% of the time this was because it would be implausible or impossible for the candidate person to appear in that context.

Learning context has a clear effect on subsequent recognition, such that learned faces are less likely to be recognized if the context is altered (Dalton, 1993). This holds for contexts such as geographical location, semantic contexts (e.g. job title), and even the presence of an additional face presented at the same time as the target face (Dalton, 1993; Watkins, Ho &

Tulving, 1976). Some context effects may be related to the expectancy of seeing a particular face. For example, manipulating expectancy can affect performance in a change detection task for facial identity (Austen & Enns, 2003; Simons & Levin, 1998). In fact, the role of expectancy in face processing is large enough to disrupt processing of even highly familiar faces. In an entertaining demonstration of this, Sinha and Poggio (1996) duplicate the internal features from Bill Clinton's face onto the face of Al Gore. Perhaps because of the expectancy of seeing both men at the podium, and external features serving as a visual cue, most people do not detect the alteration. This demonstration provides useful clues to the utility of context in face processing.

In the experiments reported here, we examine the acquisition of stable face representations from varied images. In Experiment 1, we show that participants in a sorting task perform almost perfectly when they are told that only two faces are present, but generate many piles when this instruction is withheld. In Experiment 2, we use the same card-sorting to fast-track face learning. We show that the sorting task enhances incidental learning, such that participants develop robust representations of new identities that can be recruited in subsequent recognition. This is a novel approach to face learning. Previous studies have tended to emphasise number or duration of training encounters. Here we show that assimilation of inherent within-person variability is a key part of face learning, and that a simple technique can support efficient acquisition of a new face identity.

Experiment 1: Sorting unfamiliar faces

Experiment 1 builds on the finding that observers find it difficult to integrate different photos of an unfamiliar face into a single identity (Jenkins et al., 2011). To test whether this difficulty can be overcome by manipulating observers' expectations, we compared sorting performance when viewers are given no information about the number of identities to when they are told (correctly) that there are only two.

Method

Participants

Participants were 40 students (26 female) from the University of Aberdeen, who participated for course credit or a small fee. The mean age was 22.05 (SD = 9.02, range = 18 - 60). All had normal or corrected to normal vision.

Materials

We are particularly concerned here to use images whose appearance varies naturally, as a result of environmental differences such as lighting, general health and focal distance. These images are referred to as ‘ambient images’, (Jenkins et al., 2011; Sutherland et al., 2013). Images for sorting were gathered from an internet search, reflecting a normal range of variability over which target people are typically recognized. For this reason, we chose two celebrities from The Netherlands, for whom many pictures exist, but who are unknown by our UK participants (Bridget Maasland, BM, and Chantal Janzen, CJ). From an internet search using the names of these Dutch celebrities, images were selected within the pose range ¾ to full face. Images were included as long as the faces were not occluded in any way, and had sufficient resolution (defined as a minimum of 285 x 190 pixels for these purposes). Using these criteria, the first 65 images for each search target were retained. Photographs were converted to greyscale, and cropped to a size of 285 x 190 pixels, such that the whole face was visible and filled the majority of the slide. Forty of these images (20 BM and 20 CJ), chosen using a random number generator were then scaled to passport size (35mm x 45mm), printed in high quality greyscale and laminated.

Procedure

All participants were handed a pile of shuffled cards of the two identities (BM and CJ) and were asked to sort the images into separate piles so that all the instances of the same person were together. 20 participants were given no indication about how many identities were present (free sort), and the remaining 20 were informed that the images were of two different people, taken at different times (two-sort). Emphasis was placed on accuracy, with no time restriction on the task. Participants were encouraged to place instances of the same person alongside one another so that all images were visible at the same time, rather than on top of one another, and were free to move photos between piles as many times as they liked.

Results and Discussion

Free Sort: Participants sorted photos into a mean of 6.8 different identity piles (median = 6, range = 2 - 16). A one-sample t-test revealed that this difference was significantly greater than the two that were actually present ($t(19) = 5.10$, $p < .001$, $d = 2.34$). The two different identities were rarely confused for each other. We define an intrusion error as an instance of a face appearing in a pile containing the majority of the second face. So, a single instance of face A in a pile containing a majority of face B is counted as one intrusion. Two face A pictures in a pile with majority of face Bs, is counted as two intrusions. Intrusions were summed across all a participant's piles, and these ranged from 0 to 3 across participants (median 1). One participant made one pile containing one instance of face A and one of face B. This was the only situation where there was a pile consisting of 50% of each identity, and was counted as ~~one identity with one~~a single misidentification error.

Two-sort: Ten participants sorted the two identities perfectly; median and mode intrusion errors were zero, with range 0-11. Only two participants made more than two intrusion errors per identity, and t~~. T~~here were no piles consisting of 50% of each identity.

An independent samples t-test revealed there was no significant difference in the number of misidentification errors made between two-sorters and free sorters ($t(38) = 1.73$, $p > .05$, $d = .56$).

These results confirm that free sorting is a very difficult task for unfamiliar viewers. However, simply instructing participants that there are only two identities radically improves their performance. Given minimal context information in the form of expectancy (Austen & Enns, 2003; Simons & Levin, 1998), unfamiliar observers are largely able to overcome the natural variability in faces, and discern identity-specific information, with~~ the majority of observers making~~ no errors of misidentification the most common solution.

This is a particularly interesting finding, because it illuminates viewers' difficulty with unfamiliar faces. The problem of face recognition is typically posed as one of distinguishing between faces. For example, the notion of 'configural processing' is often recruited to propose that the spatial layout of facial features is used to distinguish between faces (e.g. Maurer, Le Grand & Mondloch, 2002). Within this context, one might expect that the well-known difficulty of unfamiliar face recognition might lie in failure to discriminate individuals. However, these findings show that, for this particular task, participants have little difficulty in distinguishing between faces. Instead, they have considerable difficulty cohering different instances of the same face into a single representation, when given no expectation of how many individuals they will see. The fact that this task suddenly becomes very easy with the addition of this information demonstrates that faces *can* be cohered on the basis of visual information – but only when viewers are specifically encouraged to do so.

The results are important in understanding the difference between familiar and unfamiliar face recognition. However, they also suggest a novel approach to face learning: the transition from unfamiliar to familiar. During the two-sort task participants appear to learn to cohere superficially different stimuli into a single representation—and to do so relatively quickly (i.e. over a single experimental session). In the next experiment we test the possibility that this task can accelerate face learning, and we briefly review this topic below.

Experiment 2: Face Learning

It is well established that recognition of familiar faces is considerably more robust than recognition of unfamiliar faces. While unfamiliar face processing is relatively image-bound (e.g. Hancock et al., 2000; Jenkins & Burton, 2011), familiar face recognition is much more generalizable across changes in image. There has been a long-standing interest in the nature of visual representations underlying familiar face recognition, denoted Face Recognition Units (FRUs) by some researchers (Bruce & Young, 1986; Burton, Bruce & Johnston, 1990). Recently, there has been renewed interest in FRUs, and how they code stable, generalizable representations (Davies-Thompson, Gouws & Andrews, 2009; Davies-Thompson, Newling & Andrews, 2013; Jenkins & Burton, 2011; Carbon & Leder, 2005). However, surprisingly little is known about how such representations are formed. It has been shown that seeing a face for longer (Reynolds & Pezdek, 1992) or more often (Xue, Dong, Chen, Lu, Mumford & Poldrack, 2010) can improve subsequent recognition. Furthermore, learning faces alongside semantic or personal information leads to more accurate recognition (Klatzky, Martin & Kane, 1982), and accompanying differences in neural activation (Kaufmann, Schweinberger & Burton, 2009). Faces that are successfully associated with semantic information are recognized more confidently, spanning multiple sessions (Bonner, Burton, Jenkins & McNeill, 2003). However, while these findings provide some useful foundations, recognition in these experiments typically involves testing *the same image* that was initially learned, or a still from a seen video. But faces encountered outside of experimental settings vary greatly, as discussed above. We can see from sorting and matching tasks that seeing one face image is not adequate for successful face learning (Jenkins et al., 2011; Bruce et al., 1999; Burton et al., 2010). If it were possible for observers to successfully learn a face from only one instance, observers would be able to learn one face from a pair (or set), which would then allow them to successfully identify different instances as the same person. Thus, the challenge for any account of face learning is to accommodate this within-person variability.

In fact, it has been suggested that exposure to within-person variability may be necessary for face learning. Bruce (1994) proposed that by experiencing variability in an individual face, we become better equipped to determine which information is common between encounters and may be diagnostic of identity. One approach to understanding face variability is to attempt a parameterization of the stimuli used: for example it is highly likely that changes in lighting direction, saturation, pose, focal length, age, and other factors will each impact learning. However, we have argued (Burton, Jenkins, Hancock & White, 2005; Burton, 2013) that such attempts may obscure important information, which is often 'controlled out' of experimental stimuli. In the following experiments we examine learning using ambient images of unfamiliar faces, of the kind recognized by viewers in daily life.

In Experiment 2, we asked participants to sort photos by identity, under free or constrained (two-sort) conditions, as in Experiment 1. However, we then followed up this sorting task with an unexpected matching task. Clutterbuck and Johnston (2002, 2005) have shown that pairwise matching of faces is a good measure of familiarity, and is a sensitive index such that accuracy steadily improves with increasing familiarity. Here, we tested matching performance using unseen photos of the faces from the sorting task (seen IDs), and photos of previously unseen faces (unseen IDs). If sorting provides a good way of learning a new face, then we expect this to result in greater matching accuracy for seen than unseen IDs at test.

Method

Participants

Participants were 54 students (20 male) from the University of Aberdeen, who all reported normal or corrected to normal vision (mean age = 21.7, range = 17-41). Participants who took part were given course credit, or reimbursed a small fee for their time.

Design & Stimuli

The experiment comprised two phases: a sorting task followed by a matching task. The stimuli and procedure for the sorting task were identical to Experiment 1. Two groups of participants (18 in each) completed the free-sort and two-sort versions of the task respectively, as described in Experiment 1. A third group of participants (n=18), did not take part in the sorting phase, and proceeded directly to phase 2. In the second phase, all participants carried out a face matching task in which they judged whether pairs of photos showed the same or different people. Faces in the matching test comprised new pictures of the people used in sorting phase, as well as completely novel identities.

For the matching task, 240 additional images were collected using the method described in Experiment 1. None of these 240 photos had been presented in the card-sorting task. 90 were novel photos of the sorting-task identities (45 BM, 45 CJ). A further 90 were photos of two previously unseen identities—Wendy van Dyke (45 WD) and Tatjana Simic (45 TS). The remaining 60 were photos of foil identities, selected based on their similar age and hair color to the four target identities (BM, CJ, WD, and TS). Same-person pairs comprised 2 images of the same person, which were randomly selected independently for each participant. Different-person pairs comprised 2 images; one of the target ID, and one of a foil ID. Once again, these were randomly selected independently for each participant.

One hundred and twenty matching trials were completed in total. These were 15 same-ID and 15 different-ID trials for each of BM and CJ (sorting-task identities), and 15 same-ID and 15 different-ID trials for each of WD and TS (novel identities). Order of presentation was mixed (unblocked) and randomly ordered, independently for each participant.

Results

Sorting Phase

In the free sort condition, observers generated a median number of 6.5 identities (mode = 3; range = 3-16). A one-sample t-test revealed that participants in this condition generated significantly more identities than the two that were actually present ($t(17) = 23.33$, $p < .001$, $d = 11.32$). As in Experiment 1, misidentification errors were low (mean =1; median=0; mode=0; range = 0-7).

The median number of errors in the two-sort condition was 0.5 (mode = 0; range = 0 - 11). An independent samples t-test revealed that there was no reliable difference in the number of misidentification errors made between the two-sort and the free sort conditions ($t(34) = 1.85$, $p = .07$, $d = .63$). In neither the free sort nor the two-sort condition did any participant make any piles consisting of 50% of each identity.

While response time was not a primary dependent variable in this study, an independent-samples t-test was conducted to establish whether participants spent longer looking at the images in one condition than the other. This revealed no significant difference in the time taken to sort stimuli in two-sort and free sort conditions. (7.44 vs 7.5 minutes respectively, $t(34) = .03$, $p = .98$, $d = .01$).

Face-matching Phase

	Novel IDs	Sorting-task IDs
Free sort	.73 (.01)	.81 (.01)
Two-sort	.76 (.01)	.86 (.01)
No sort	.73 (.02)	.75 (.01)

Table 1. Mean face matching accuracy following different types of sort (standard error in parentheses)

Accuracy for the face matching test is shown in Table 1. A 3 (sort type: free sort, two-sort, no sort) x 2 (ID type: novel vs sorting-task identities) mixed design ANOVA showed main effects of sort type ($F(2, 51) = 5.15$, $p < .01$, $\eta_\rho^2 = .17$) and ID type ($F(1, 51) = 45.36$, $p < .001$, $\eta_\rho^2 = .47$). There was also a significant interaction between these factors ($F(2, 51) = 5.22$, $p < .01$, $\eta_\rho^2 = .17$).

Simple main effects analyses showed significant effects of ID-type for free sort ($F(1, 51) = 20.52$, $p < .001$, $\eta_\rho^2 =.29$) and two-sort conditions ($F(1, 51) = 33.46$, $p < .001$, $\eta_\rho^2 = .40$), but not for the no sort condition ($F(1, 51) = 1.83$, $p > .05$, $\eta_\rho^2 = .03$). Further, there was a significant effect of sort type for sorting-task IDs ($F(2, 102) = 9.22$, $p < .001$, $\eta_\rho^2 = .15$) which was not present for novel IDs ($F(2, 102) = 1.11$, $p > .05$, $\eta_\rho^2 = .02$). For sorting-task IDs, Tukey's HSD revealed a significant difference between no sort and two-sort conditions ($p = .009$), which was not evident between no sort and free sort conditions ($p = .373$). There was also no significant difference between two-sort and free sort conditions ($p = .198$).

Discussion

The clearest finding from this experiment is that the faces of people seen during the sorting phase were matched more accurately during the later test phase. There are three important points to note in order to understand these data: (i) the matching test employed entirely new images, and so this advantage is not brought about by memory for particular photos; (ii) the matching test was unexpected, participants were not trying to commit faces to memory during the initial phase, but were focused on the sorting task; (iii) the matching test was ostensibly independent of the sorting phase – participants were not asked to remember anything, merely to say whether pairs of simultaneously-presented faces were the same or different.

These data give quite a clear indication that the sorting task leads to incidental learning of the identities involved. Consistent with Clutterbuck and Johnston (2005), we find that the matching task is sensitive to levels of familiarity, and that familiarity is enhanced for the identities that participants have encountered previously. It seems then, that exposure to a range of very different images of these people helps to form a representation of them, which can be used to recognize new, unseen images when they are encountered later. Furthermore, this occurs without any deliberate attempt to learn the new faces, but as a side-effect of encountering them in this context. This provides evidence to suggest that stable representations have begun to form for these faces, and that these representations are sufficiently flexible to be useful in recognition of novel instances.

The comparison of the different types of sort is also interesting, though less conclusive. There is some evidence that the two-sort might lead to better learning of the identities than the free sort. This is suggested by the trend for the two-sort participants to show better performance than the free sort participants - a trend which is present only for the (already seen) sorting-task identities. However, support for this idea is weakened by the nonsignificant pairwise comparisons, and can therefore only be regarded as tentative from these data.

Finally, we note the design used here means that participants learn only a small number of identities. Because the experimental constraints require a large number of instances for four people, the same identities were used for learning throughout the experiment. To address the possibility that there is an inherent difference in the difficulty of face matching between those identities we used for the sorting task, and those used for comparison (the unseen identities), note that the 'no sort' group showed no difference in performance between these groups. Instead, differences in performance at test seems to rely entirely on previous exposure to these identities.

General Discussion

Using sorting and matching tasks, the present experiments explore the influence of within-person variability on face identification. In Experiment 1, we observe that the debilitating effects of within-person variability when identifying unfamiliar faces (Jenkins et al., 2011) can be rather simply overcome. With additional information about the number of targets to expect, unfamiliar observers are able to categorize different face instances as the same person, with quite a high degree of accuracy. In Experiment 2 we used the sorting task as an incidental learning procedure, to explore the development of stable representations. By using a sensitive measure of familiarity (simultaneous matching task; Clutterbuck & Johnston, 2002; 2005) we observe that stable face representations can form through experience of within-person variability.

If stable representations form through experience with natural variability, one might expect that reconciling more instances as the same person results in stronger individual face representations (Jenkins & Burton, 2011). That is, reconciling different instances as the same person requires that transient within-person variability that occurs between different instances is disregarded, and that the stable information is retained into a representation. Experience of more instances should therefore promote the extraction of stable, identity-specific information. More flexible face representations would then be more likely to be activated by a completely novel instance of that person's face. We find some evidence to suggest this from our data. Free sorters generate far more identities than two-sorters, which necessarily means that each 'identity' generated comprises fewer instances in the free sort than the two-sort condition (even though all observers see the same number of images of each identity). We observe a trend for free sorters to perform less well at matching learnt faces than do two-sorters, while performing no differently in accuracy to novel faces. However, these data are somewhat nuanced, and it will be an important topic of future research to establish the degree to which learning can be encouraged with top-down constraints, as opposed to simple exposure to instances.

We rarely notice the influence of within-person variability in normal social situations, yet under experimental conditions variability becomes an obvious problem (Bruce et al., 1999; Burton et al., 2010; Jenkins et al., 2011; Longmore et al., 2008). In Experiment 1 we show evidence to suggest that this is because of the addition of context information. Context information is known to influence the likelihood that the same person will be recognized on a subsequent occasion (Dalton, 1993; Watkins et al., 1976). Similarly to Young et al.'s (1985) early diary study, we observe that even context in the form of expectancy is adequate to identify different instances as the same person – and in doing so rarely mistake one person for another.

Familiar faces, on the other hand, rarely require context information for recognition (Burton et al., 1999; Ramon, Caharel, Rossion, 2011), as stable representations enable recognition of previously unseen instances (Davies-Thompson et al., 2009; 2013; Jenkins & Burton, 2011). Here we show that such representations can form through incidental learning, from experience of natural variability. This is consistent with Bruce's (1994) earlier proposal that experience of within-person variability is necessary, as it allows us to determine which information is consistent – and therefore identity-specific – and which information is transient. Notably, stable representations were formed in the absence of explicit instructions to remember faces, while familiarity was measured using entirely new instances of the faces, thus we can be confident that these effects are not merely the result of conscious image learning.

Matching accuracy for previously sorted faces was greater for two-sorters relative to no-sorters. However, the corresponding difference between free-sorters and no-sorters was not found. This observation~~The difference between two-sorters and free sorters in face-matching performance that is observed for previously sorted faces~~ is consistent with the idea that experiencing a wide range of different images results in a more stable representation. Yet it is notable that accuracy is high for learnt faces, regardless of whether they are learnt from a few, or many instances. This suggests that perhaps the first few encounters of a face actually form a reasonable representation, while additional instances that are subsequently included then produce small increases in stability. If few instances are required in order to determine which information is identity-specific (Bruce, 1994), this would result in representations that are highly tolerant to variability even at an early stage. When new instances are used to update existing representations (Carbon et al., 2007; Leopold, Rhodes, Müller & Jeffery, 2005), this seemingly results in finer-tuned representations, which show incremental tolerance to variability, until faces are highly familiar and identification accuracy is at ceiling (Burton et al., 1999).

These data offer a novel perspective on understanding face learning; past research has focused largely on how we are able to differentiate *between* identities, and how we are able explicitly to remember faces over different encounters (Bonner et al., 2003; Longmore et al., 2008; Reynolds & Pezdek, 1992; Xue et al., 2010). Here we address the problem of how we identify difference instances as the same person, and show that experience of variability is necessary in forming stable face representations that are useful for recognising previously unseen instances of faces. It therefore appears evident that *between*- and *within*-person variability make different, but equally important, contributions to the problem of face recognition. In order comprehensively to understand face learning processes, we need to consider experience of natural variability alongside other factors.

References

Adini, Y., Moses, Y. & Ullman, S. (1997). Face recognition: the problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 1-12.

Austen, E. L., & Enns, J. T. (2003). Change detection in an attended face depends on the expectation of the observer. *Journal of Vision*, *3*, 64-74.

Bonner, L., Burton, A. M. & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, *10*(5), 527-536.

Bonner, L., Burton, A. M., Jenkins, R. & McNeill, A. (2003). Meet The Simpsons: Top-down effects in face learning. *Perception*, *32*, 1159-1168.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105-116.

Bruce, V. (1994). Stability from variation: The case of face recognition - The M.D. Vernon memorial lecture. *The Quarterly Journal of Experimental Psychology Section A*, *47*(1), 5-28.

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.

Bruce, V. & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305-327.

Burton, AM. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*(8), 1467-1485.

Burton, A. M., Bruce, V. & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*(3), 361-380.

Burton, A. M., Jenkins, R., Hancock, P. J. B. & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*, 256-284.

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

Burton, A. M., Wilson, S., Cowan, M. & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243-248.

Carbon, C.-C. & Leder, H. (2005). Face adaptation: Changing stable representations of familiar faces within minutes? *Advances in Experimental Psychology*, *1*(1), 1-7.

Clutterbuck, R. & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*, 985-994.

Clutterbuck, R. & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17*(1), 97-116.

Dalton, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory and Cognition*, *21*(2), 223-234.

Davies-Thompson, J., Gouws, A. & Andrews, T. J. (2009). An image-dependent representation of familiar and unfamiliar faces in the human ventral stream. *Neuropsychologia*, *47*, 1627-1635.

Davies-Thompson, J., Newling, K. & Andrews, T. J. (2013). Image-invariant responses in face-selective regions do not explain the perceptual advantage for familiar face recognition. *Cerebral Cortex*, *23*, 370-377.

Favelle, S. K., Palmisano, S. & Avery, G. (2011). Face viewpoint effects about three axes: The role of configural and featural processing. *Perception*, *40*(7), 761-784.

Hancock, P. J. B., Bruce, V. & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*(9), 330-337.

Hill, H. & Bruce, V. (1996). The effects of lighting on the perceptual of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 986-1004.

Jenkins, R. & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society: B*, *366*, 1671-1683.

Jenkins, R., White, D., van Montford, X. & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323.

Kaufmann, J. M., Schweinberger S., R. & Burton, A. M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of Cognitive Neuroscience, 21*(4), 625-641.

Klatzky, R. L., Martin, G. L. & Kane, R. A. (1982). Semantic interpretation effects on memory for faces. *Memory and Cognition, 10*(3), 195-206.

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, *66*(5), 651-654.

Lander, K. & Bruce, V. (2003). The role of motion in learning new faces. *Visual Cognition*, *10*(8), 817-912.

Leopold, D. A., Rhodes, G., Müller, K.-M. & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society: B*, *272*(1566), 897-904.

Longmore, C. A., Liu, C. H. & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 77-100.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

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

O'Toole, A. J., Edelman, S. & Bültoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*, 2351-2363.

Ramon, M., Caharel, S. & Rossion, B. (2011) The speed of recognition of personally familiar faces. *Perception*, *40*(4), 437-449.

Reynolds, J. K. & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology*, *6*, 279-292.

Simons, D. J. & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review*, *5*(4), 644-649.

Sinha, P. and Poggio, T. (1996). I think I know that face… *Nature*, *384*(6608), 404.

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118.

Watkins, M. J., Ho, E. & Tulving, E. (1976). Context effects in recognition memory for faces*. Journal of Verbal Learning and Verbal Behavior, 15*, 505-517.

Xue, G., Dong, Q., Chen, C., Lu, Z. L., Mumford, J. A. & Poldrack, R. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *327*, 97-99.

Young, A. W., Hay, D. C. & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology*, *76*, 495-523.