Article:

Mike Burton, A. orcid.org/0000-0002-2035-2084 (2013) Why has research in face recognition progressed so slowly? The importance of variability. Quarterly journal of experimental psychology (2006). pp. 1467-1485. ISSN 1747-0226

https://doi.org/10.1080/17470218.2013.800125

In Press: Quarterly Journal of Experimental Psychology, 2013

Why has research in face recognition progressed so slowly? The importance of variability

A. Mike Burton
School of Psychology
University of Aberdeen
Aberdeen AB24 3FX, UK

Tel. 01224 273930
Email. m.burton@abdn.ac.uk

Short title: Progress in face recognition

Key words: face recognition; identity; variability;

Abstract

Despite many years of research, there has been surprisingly little progress in our understanding of how faces are identified. Here I argue that there are two contributory factors: (i) our methods have obscured a critical aspect of the problem, within-person variability; and (ii) research has tended to conflate familiar and unfamiliar face processing. Examples of procedures for studying variability are given, and a case is made for studying real faces, of the type people recognize every day. I argue that face recognition (specifically identification) may only be understood by adopting new techniques which acknowledge statistical patterns in the visual environment. As a consequence, some of our current methods will need to be abandoned.

Face perception is now a popular, mainstream, focus for research in psychology. From its roots in eye-witness identification, the study of face processing has developed to cover a very wide range of topics, from social interaction to neural implementation. Much progress has been made: for example a great deal is now known about the perception of emotion, gaze and attractiveness. However, despite the volume of this research, there has been surprisingly little progress towards understanding *identification;* this key aspect of face *recognition* appears to have resisted incremental solution. Since there is both a theoretical and a practical imperative to understand identification, it can be seen as a failure of research effort within psychology.

In this article, I will suggest reasons for our lack of progress. I will make the following arguments. 1. Experimental methods have become popular which direct us away from studying perception of real faces. The use of artificial or tightly controlled stimuli not only limits the generalisablity of results, but may actually undermine a theoretical understanding of the problem. As a consequence the field has come to rely on poorly-specified notions, such as 'configural processing', which have hindered rather than helped. 2. A proper analysis of face recognition requires a theoretical understanding of the differences between familiar and unfamiliar face processing. Failure to draw this distinction properly has also restricted progress in the field. 3. A key part of familiar face recognition is an understanding of how any particular face may vary. Different pictures of the same face are highly variable, and we cannot understand how to recognize someone without an understanding of that variance.

A theme running through the paper will be the necessity properly to analyse the task of face recognition. Bruce and Young's (1986) well-known paper on this topic remains influential many years after it was published because it provides a thorough analysis of the tasks involved in different aspects of face processing (see Young & Bruce, 2011 for a modern reflection on this). However, as the field has widened, theoretical coherence has sometimes been hard to maintain. The study of particular laboratory-based phenomena can be taken as a proxy for studying face recognition itself, and researchers have sometimes been guilty of following chains of experimental reasoning, without serious reflection on how these relate to our remarkable ability to recognize one another in daily life.

I will start this analysis with a consideration of the stimuli we use in our research. However, before doing so, it is important to point out that the arguments below are not aimed at any specific researcher or research groups. Much of what follows is critical of methods used in face recognition research, and in almost every case my own research is vulnerable to these criticisms. However, it seems undeniable that progress in understanding recognition has been disappointingly slow, and the following analysis is an attempt to unpack the reasons for this.

Face recognition research using the faces people recognise

A lay person visiting a face research lab might be surprised to see the types of images used in our experiments. That person can recognize Paul McCartney or Meryl Streep across a huge range of different conditions, as an internet search will demonstrate instantly. And yet the images we find in many experiments are much more tightly constrained than this. In many research projects photographs of unfamiliar people are taken in highly controlled conditions, specifically for the experiments planned. Images are typically captured with the same camera, under constant (or systematically varied) lighting conditions, in the same (or systematically varied) pose, and so forth. We often crop face images of their hair and perhaps normalize them for whatever low-level image characteristics we think important at the time. In short, we attempt to control for all noise, and to eliminate variance in all dimensions except those we wish to manipulate. In this section I will argue that choice of experimental stimuli has sometimes misled us. Through a well-intentioned effort to eliminate spurious variability in our stimuli, we have often asked the wrong question of our experimental participants.

FIGURE 1 HERE

The first point to make is that it is almost never appropriate to conflate face recognition with image recognition. Figure 1 shows two different matching tasks, one is trivially easy, the other is hard. It is perhaps surprising that after so many years of research in this field, this conflation continues. Figure 1 presents the problem in a starker way than normal, and the issue of same-photo recognition is usually buried in experimental procedure, perhaps as an item to be remembered, or to be recognized among distractors. However, it is *always* easier to recognize a picture than to recognize a face, and treating face recognition as a special case of image recognition will deliver the wrong answer. In fact, we have argued that unfamiliar face recognition is heavily influenced by image-level characteristics (Hancock, Bruce, & Burton, 2000; Megreya & Burton, 2006). When confronted with two images of the same unfamiliar person, one does not recruit face-related processing, but instead relies on unsophisticated image-comparison techniques, which might be used for any pattern. To this extent, one's success in matching two photos of the same face relies on how similar are the images of that face – and of course they are maximally similar in the case of identical photos.

FIGURE 2 HERE

Although many researchers do try to avoid conflating image and face recognition, it is not straightforward to know what to do about this. One common response is to use different photos of the same people taken in the same photographic session, and with *the same camera,* but with a change of pose. Good examples can be seen in the Benton Test (Benton et al, 1994) and the Cambridge Face Memory Test (Duchaine & Nakayama, 2006). However, the effect of a camera on the appearance of a face is actually rather great, as are superficial differences in presentation. If one is to use *image-matching* as a strategy to compare unfamiliar faces, then images taken with the same camera, using the same shutter and lens settings, will by definition share much in common. To illustrate, figure 2 shows different images of the same person. A and B are different images with the same camera. C and D match these poses but with a different camera, and with a minor change in presentation: the subject has let her hair down. All the photos were taken within minutes of each other. It is clear that a within-camera, same-hair-state match is much simpler than a between camera match. However, these superficial differences are completely transparent to familiar viewers. So, evidence from studies using highly similar images cannot simply be extended to face recognition in general.

FIGURE 3 HERE

Differences in camera characteristics are well-known to photographers. To illustrate just one important dimension, figure 3 (after Harper & Latto, 2001) shows the effect of manipulating a single variable, within a camera: focal length. Images of the same person, taken within the same session, look rather different, simply through having been taken from different distances. These differences should remind us that trying to capture the nature of a face by taking measurements in the picture plane are doomed to failure – because such measurements change under commonplace changes in the capture situation. A simple photographic variable, distance between subject and camera, is enough to make severe changes to images, and this may contribute to the difficulty of unfamiliar face matching. Such differences are, of course, much larger over changes in camera, and entirely missed if one restricts oneself to photos taken under tightly controlled conditions. The danger is that these experimental restrictions obscure the true nature of face recognition.

In fact, the problem of constrained stimuli is often more severe than use of a database of experimental images. It has become quite common in the literature to edit photos or even use artificial stimuli, with an assumption that perception of the resulting images is the same as perception of faces. One example can be seen in research comparing the recognition of identity and sex. It is well-established, across a large number of different studies, that people are highly accurate and very fast at judging the sex of a face. A judgement of familiarity is typically longer. (Values vary according to stimuli and experimental conditions, but RTs of 600ms and 800ms respectively are typical of early work – e.g. Ellis et al, 1990). Using logic typical of this field, Bruce et al (1987) asked whether these two judgements might lie in series or in parallel. They showed participants famous male faces which had been judged androgynous or highly masculine by a separate group of viewers. Bruce et al found that androgynous faces took longer to judge as male, but that this had no knock-on effect to judgements of identity, giving strong evidence for an independent, rather than sequential, arrangement of the two processes.

This independence of gender and identity perception has been challenged in the literature, but interestingly, it has always been challenged using stimuli deliberately edited to remove some aspects of the face. For example, Goshen-Gottstein & Ganel, (2000) demonstrated priming onto a sex judgement – an effect which had not previously been reported in the literature. Furthermore, the pattern of priming did suggest an association with identity processing, in contrast to the Bruce et al study. However, these patterns existed *only* when the faces were severely cropped, to strip away all hair and face-outline; results for intact faces were consistent with the earlier 'independent route' theories. Despite this, the results were taken to imply "evidence for a common route for the processing of sex and identity" - the sub-title of Goshen-Gottstein & Ganel's paper. In similar vein, (Rossion, 2002) used morphs of faces with no hair or external features, and showed that photos of experimentally familiarized stimuli (called 'familiar' here), gave rise to faster sex judgements than novel hair-less outline-less faces. Once again, the conclusion drawn is that gender and identity processing are not independent.

What do we learn from such studies? My contention here is that one should not generalize experimental results from highly artificial stimuli. Both these papers are interesting methodologically, and both, very helpfully, provide examples of their stimuli. However, in both cases, conclusions are drawn about our normal face recognition system on the basis of images which deliberately exclude some features of faces. The most that can be claimed in such circumstances, is that it is possible to edit faces in such a way as to demonstrate a particular pattern of effect. However, such demonstrations do not speak to claims based on face recognition in general (e.g. Ellis et al, 1990) deriving from experiments using full images of faces. In fact, it has become so common to use artificial or highly edited faces, that acknowledgement of this is often omitted from published titles or abstracts. There is no doubt that it is sometimes appropriate to use graphically manipulated stimuli, depending on the question being asked; but this is such an important experimental decision that readers might reasonably expect to see it clearly flagged in any précis of the research.

FIGURE 4 HERE

Another extreme example of using manipulated images arises when researchers employ artificial stimuli, which can be manipulated in computer graphics to have exactly the characteristics required. Good examples are stimuli in important and influential papers by Loffler, Yourganov, Wilkinson, & Wilson, (2005) and Leopold, O'Toole, Vetter, & Blanz, (2001). Figure 4 shows illustrations from these papers, providing examples of the artificial stimuli used in each. In these cases, faces vary within a 'face space': dimensions are derived from a statistical analysis of examples, or in some cases from first principles, and individual people's faces are defined as having values on each of these dimensions. While such a characterization is useful for some purposes, it is severely restricted in others. Under these notions, faces have a singleton point representing their 'true' value. Any movement through space must represent a move to another identity, or to a less accurate representation of that person, for example, a neighbourhood of 'error' might be defined.

Why object to the use of such stimuli? One simplistic complaint is that they may reduce the generalisablity of experimental results. Real faces do vary, and studying artificially derived faces in these examples is, in fact, another example of confounding a single image with a face. There is no single image which truly represents any real person. A counter-example might be the Mona Lisa, but the contention here is that understanding recognition of the Mona Lisa does not make a good model for understanding recognition of Hillary Clinton. However, there is a potentially more serious problem which can arise. These stimuli can give false prominence to characteristics of faces which are important for restricted sets, but not for the world. To illustrate this, I will consider this issue of 'configural processing' in face recognition.

Configural processing is a theoretical concept which is very often recruited in the face recognition literature, and yet it is normally poorly defined. A clear statement is given by Tanaka & Gordon (2011): ' We use the term "configural processing" … to refer to encoding of metric distances between features (i.e. second-order relational properties)". So, the proposal is that we can differentiate between the people we know, because they have a characteristic set of relations between the metric distances within their faces.
One significant problem with this notion is that it is never operationalized. How do we know *which distances* to measure? What are the key distances (or relations between these) which allow us to recognise Hillary Clinton in thousands of pictures of her? No-one has come close to providing a working definition which allows this, and computer-based attempts to recognize people in this way have a thirty-year history of failure. Furthermore, attempts to operationalize this notion in forensic contexts has not worked (Kleinberg, Vanezis, & Burton, 2007). It is possible that the key set of relations between measurements uniquely characterizing a face is yet to be discovered, but the evidence that this will eventually produce reliable results is not good.

FIGURE 5 HERE

In fact, there are two good reasons to believe that configural processing, defined in this way, probably does not underlie our perception of identity. The first is that 'metric distances between features' is not stable across different photos of the same person. This is illustrated in Figures 2, 3 and 5. For example, in Figure 3, the face has different configurations according to how the photograph was taken, in this case distance from the subject. Such differences are completely unnoticed in images of familiar faces (fig 5), but undermine attempts to use metric distances computationally in any straightforward manner.

The second reason to be wary of the explanatory power of configural processing is that familiar face recognition is not damaged at all by image distortions which severely affect configuration. The important work of Hole (Hole, George, Eaves, & Rasek, 2002) shows that familiar faces may be stretched up to twice their original height with no effect on subjects' ability (or speed) to recognise them. While this is a severe distortion, affecting all distances, ratios and angles in the face except simple 1d measures, it appears that our face recognition system is blind to such severe configural changes. Further support for this finding is found in studies of the ERP component N250r, which is sensitive to repetitions of familiar faces. The component is affected maximally by repetitions of the same image of a face, allowing one to ask, what counts as 'same image' for this purpose. Bindemann, Burton, Leuthold, & Schweinberger, (2008) showed that repeating images in which one is a stretched version of the other elicits the same N250r response as repeating the identical image – i.e. whatever function is being observed by this technique appears robust to quite severe geometric distortions.

My proposal here is simple: if we had taken as our starting-point, the images of faces which people recognize every day, from newspapers, the internet or TV, then we would almost certainly not have gained the impression that what differentiates our faces is a pattern of 2d measurements in the picture plane. It would be hard to hold that position, given normal exposure to images, because it is clear that there is as much variability within individuals as there is between them. In fact, the position is only tenable if one deliberately excludes much of the variability found in everyday images of faces. As a result, current theories give us a good understanding of how to discriminate two specific pictures, but not how to discriminate between real faces, with all their complex variability. In a later section, I will develop this idea further, arguing that variability itself should be the focus of future research.

Although I have used configural processing as an example, the point is more general: a well-motivated desire to use properly controlled images, free from inconvenient noise, can nevertheless obscure some aspects of the problem one wishes to study. In general, the choice of experimental stimuli will certainly constrain one's thinking about any problem. In this particular case, I have argued that it may actually guide one's theorizing *away* from an understanding of face perception. In the next section, I turn to a different problem: the important issue of familiarity.

Differences between familiar and unfamiliar face recognition

In this section I will argue that there are qualitative differences between perception of familiar and unfamiliar faces, and that a failure to incorporate this into modern theories has further contributed to the lack of progress in face recognition. In previous work, we have suggested that unfamiliar faces are perceived, for the purposes of *identity*, simply as patterns – visual images with no privileged ('special') method of processing (Hancock et al., 2000; Burton & Jenkins, 2011). Based on evidence presented in a paper provocatively titled 'Unfamiliar faces are not faces' (Megreya & Burton, 2006), we argued that the processing of familiar and unfamiliar faces dissociate, to some extent. If this claim is true, then it is important, because almost all contemporary theories of face recognition fail to distinguish between familiar and unfamiliar stimuli, instead treating faces as a unitary class to be compared with other objects of visual recognition.

It has been known for many years that recognition memory tests show an advantage for familiar over unfamiliar faces, both in accuracy and in speed (e.g. Bruce, 1986; Ellis et al, 1979; Klatzky & Forrest, 1984). Furthermore, changes in expression, lighting and viewpoint have all been shown to damage recognition memory for unfamiliar faces more than for familiar faces (e.g. Patterson & Baddeley, 1977; Hill & Bruce, 1996; O'Toole et al, 1998; Roberts & Bruce 1989). These results consistently demonstrate a superiority for familiar faces in memory, but do not in themselves indicate a fundamental difference between familiar and unfamiliar faces. However, more recent studies using matching rather than memory, provide stronger evidence.

FIGURE 6 HERE

In a series of experiments, Bruce et al (1999, 2001) set out to establish a baseline for unfamiliar face recognition, uncontaminated by fallible memory. Participants were shown 1-in-10 line-ups (see figure 6), in which the target face (top) may or may not be present in the array of ten possible matches. All faces were young men, Caucasian, short-haired and clean-shaven. Surprisingly, at the time, viewers performed this task poorly. Despite unlimited time to study the faces, and simultaneously present target and arrays, subjects typically achieved only 70% accuracy – for both target-present and target-absent arrays. This finding has been replicated many times, using these and other stimuli (e.g. Megreya & Burton, 2006, 2008). Recent studies have tended to use simple pair-wise matching. Two faces are presented simultaneously, and subjects respond that the photos show the same or different people. Again, viewers are surprisingly poor at this task – making 20% errors on a typical test (Megreya & Burton, 2007; Burton, White, & McNeill, 2010). This difficulty matching unfamiliar faces is not restricted to photographs. There have now been a number of experimental demonstrations that viewers are similarly poor at matching a live person to their photo (Davis & Valentine, 2009; Kemp, Towell, & Pike, 1997; Megreya & Burton, 2008).

In contrast to unfamiliar faces, people are very good indeed at recognizing familiar people. For example, Burton, Wilson, Cowan, & Bruce, (1999) showed real CCTV images, of low resolution, to familiar and unfamiliar viewers. The unfamiliar viewers were very bad at identification, performing almost at chance level. However, familiar viewers performed almost perfectly with the same images. Interestingly, a group of police officers, specialising in identification, performed no better than the unfamiliar group – despite reporting significantly greater confidence in their (very poor) performance. In general, we are excellent at recognizing familiar faces, across a very wide range of settings, and even in difficult viewing conditions.

Given the large overall performance differences between familiar and unfamiliar faces, how might we established whether there are qualitative processing differences? One approach is to use correlational studies – a technique which has recently begun to be exploited in the literature (DeGutis, Wilmer, Mercado, & Cohan, 2013; Wang, Li, Fang, Tian, & Liu, 2012). While performance with unfamiliar faces is poor on average, there are actually very large individual differences in the population (Burton et al., 2010; Megreya & Burton, 2006; Russell, Duchaine, & Nakayama, 2009). It is therefore possible, in principle, to ask whether these performance differences correlate with differential abilities to recognise *familiar* faces. In practice, this is not straightforward, because viewers' high levels of performance with familiar faces restrict variability on that task. For this reason, researchers have typically used familiar face tasks which are contrived to be hard, and do not correspond well to the unfamiliar tasks with which they are compared. Results are mixed, with some researchers showing no association between familiar and unfamiliar face recognition (Megreya & Burton, 2006) and others showing significant associations (Russell et al., 2009). However, patterns of association within variants of the same task are interestingly different across levels of familiarity. For unfamiliar faces, only, there is a strong association between people's ability to match faces when they are upright and when they are inverted (Megreya & Burton, 2006; Russell et al., 2009). However, this association disappears when using familiar faces.

To interpret this pattern of results, it is necessary to consider current theories of the well-known face inversion effect. It is often suggested that there is some face-specific processing which is engaged by upright, but not inverted faces (Murray, Yong, & Rhodes, 2000; Rossion, 2008). Furthermore, inversion is sometimes held to damage a viewer's ability to process configural, or holistic information. However, evidence for such positions is typically based on judgements made to unfamiliar faces. In contrast, the basic phenomenon, demonstrated in undergraduate classes throughout the world, is that we often fail to recognise *familiar* faces when presented upside down. There are very significant arguments in the literature about whether inversion has quantitative or qualitative effects on face perception (e.g. Richler, Mack, Palmeri, & Gauthier, 2011; Sekuler, Gaspar, Gold, & Bennett, 2004), and almost all the evidence recruited in this debate comes from unfamiliar face research. However, recent evidence based on individual differences suggests an association between inverted and upright face processing for unfamiliar but not familiar faces (Megreya & Burton, 2006; Russell et al., 2009). It seems that, in this case, a tendency to conflate all faces together may actually be obscuring the true nature of inversion – it is quite possible that it affects familiar and unfamiliar faces differently.

Another example of the dissociable behaviour of familiar and unfamiliar faces is shown by the absence of a mirror effect when matching unfamiliar faces (Megreya & Burton, 2007). Recognition memory tests often show a mirror effect, such that the same items are easily recognized (when present) and easily rejected (when absent). This effect, rather counter-intuitively, is never observed for unfamiliar faces – it seems that the faces we find easy to recognize as having been present, are unrelated to those we find easy to reject as having been absent. This has been a puzzle in the literature for some time (Deffenbacher, Johanson, Vetter, & O'Toole, 2000; Vokey & Read, 1992). When originally demonstrated, this finding was influential in eliminating the unidimensional view of facial 'distinctiveness' which had been held to that point. Vokey & Read replaced this with a two-component model of distinctiveness, based on (i) memorability of a face; and (ii) context-free familiarity (i.e. the way certain faces appear familiar, whether or not they are). In our work, (Megreya & Burton, 2007) we demonstrated two significant additions. First, a mirror effect is absent in unfamiliar face *matching*, just as it is in *memory* - a finding which seems to undermine an account based on 'memorablity'. Second, the mirror effect is very strongly present for familiar faces – even those with rather small levels of experimentally-induced familiarity. This large discrepancy supports the idea of a systemic dissociation between familiar and unfamiliar faces.

If there are qualitative differences between processing familiar and unfamiliar faces, then there are serious theoretical and practical implications. A failure to differentiate between these, risks confounding processes which are particular to only one of the classes of faces, hence contributing to the lack of progress in understanding identification. I will highlight two areas of research (somewhat related) which are relevant here. The first is a long-standing debate about whether or not faces are ‘special’ in the sense of requiring dedicated processing resources, distinct from those used in other object recognition tasks. This is a debate which has exercised a great many people but which, after many years of extensive research effort, has produced no consensus (Gauthier & Bukach, 2007; McKone, Kanwisher, & Duchaine, 2007). One side of this debate holds that faces are not special, but can appear to be so because people have acquired great experience in processing faces over the course of their lives. In other words, we are ‘face experts’. However, researchers concerned with differences between familiar and unfamiliar faces find it odd to claim that people are experts in face recognition. The extremely poor performance of viewers on apparently straightforward matching tasks suggests that we are actually rather poor at face recognition – in particular circumstances. So, for example, the Glasgow Face Matching Test requires subjects to view two photos of the same or different people, taken minutes apart in very similar lighting and pose, and with two high quality cameras. Face-pairs remain visible until a response is made, and there is no time-restriction. Nevertheless, people make 20% errors in one version of the task, and 10% in another. This hardly seems like expertise. On the other hand, when these images are familiar, the task is trivial – viewers consistently score at 100% accuracy.

The key to understanding this problem is that we *are* experts at recognizing *some* faces – i.e. those we know. However, and most importantly, our ability to match or recognize familiar faces *does not* allow us to generalize this performance to unfamiliar faces. The central point is that our long-experience of seeing our friends and family allows us to generalize recognition of those people over a very diverse range of conditions. However, that experience does not allow us to generalize across pictures of unfamiliar faces – as can easily be seen from poor performance in unfamiliar face matching. We might speculate that this discrepancy lies behind the failure to discriminate between familiar and unfamiliar faces in psychological theory. We all have the impression from daily life that face recognition is an easy perceptual task, when in fact it is only easy in some circumstances.

These observations do not speak to either side of the 'faces are special' debate, and are certainly not intended to support one position or the other. Instead, the point to note is that any arguments about faces *in general*, are susceptible to the problem that faces are not a single perceptual category. The distinction between familiar and unfamiliar faces is ignored in this debate, and so an attempt to cohere the disparate evidence often involves a blending of different types of effect.

I now turn, briefly, to the very large literature on the neuroscience of face processing. There are some well-established phenomena associated with viewing faces as compared to other objects. (This statement is slightly contentious in the light of the debate described above, but for the purposes of this part of the argument it is enough to note that there are some highly dependable effects which can be observed in response to presentation of a face, by comparison to most other objects.) The N170 ERP component is modulated in a characteristic way by faces, and this is highly reliable (e.g. Bentin, et al, 1996; see Eimer, 2011 for an overview). Furthermore, evidence from fMRI and MEG shows three face-sensitive cortical areas, which are commonly reported: the FFA, OFA and STS. These observations, and careful experimental procedure, have led researchers to models of face processing such as that proposed by Haxby et al, (Haxby, Hoffman, & Gobbini, 2000). However, it is worth noting here that there is rather little differentiation between familiar and unfamiliar face processing in the neuroscientific literature.

In ERP research, the large majority of studies show no modulation of N170 by familiarity. In fact, it is rather hard to find any discrimination between these classes of faces in component-based ERP analysis. The most robust candidate to date is the N250r – a component which is sensitive to *immediate repetitions* of familiar, as compared to unfamiliar faces (Schweinberger, Huddy, & Burton, 2004; Wiese & Schweinberger, 2008). In fMRI, most research is conducted with unfamiliar faces – as a way of avoiding activation of person-specific knowledge which may implicate non-face cortical areas. However, in the relatively few studies which have directly compared familiar and unfamiliar faces, there is rather little evidence to suggest any modulation at all, in any of the three candidate brain regions. Specifically, the FFA, which is usually taken to process identity (on the basis of indirect evidence) does not appear to show any great sensitivity to this dimension. (See the review by Natu and O'Toole, 2011, for detailed evidence on this point.) For example, Davies-Thompson, Gouws, & Andrews, 2009) used an fMR-adaptation technique to demonstrate adaptation in FFA. As expected, this region was sensitive to face images, but the adaptation was tied to the specific image used. There was no adaptation to different images of the same face: these exhibited the same response as images of different people. Such demonstrations fail to show image-invariant adaptation in FFA – and hence show no effect of familiarity.

The lack of a clear marker for familiarity in neuroscientific face research is interesting. Given the large behavioural effects, it is somewhat surprising that no such marker has been found. This might reflect the absence of focus on the issue of familiarity – a conflation between classes of faces which we have already noted in behavioural work. On the other hand, it may reflect a more complex mechanism underlying familiarity by comparison to other information conveyed by faces. For example, Davies-Thompson et al propose a distributed network for representing familiarity. Whatever the solution, there is a significant discrepancy between behavioural research on faces, in which the effects of familiarity are very large, and neuroimaging research, in which these effects are hard to detect at all. This discrepancy needs to be addressed.

Studying variability

In an earlier section, I argued that face recognition research has ignored the variability inherent in pictures of the same person. This omission has led to a focus on what differentiates us, which is only part of the problem in face recognition. In fact, just as with any statistical inference, two pieces of information are required to discriminate between samples: information about the differences between, and also within, the samples. By analogy, current methodological approaches to face recognition are like comparing two groups of data simply on the basis of their means, with no regard for their variances. We take pains to explain to our students that a t-test (say) compares differences between group means in the context of within-group variability, but our methodological approach to face recognition has concentrated on just one of these components.

FIGURE 7 HERE

If this proposal is true, then we need an agenda for studying face recognition which includes the missing component. In recent work, we have begun this, in the first instance simply by documenting examples of very large within-person variability. For example, Jenkins et al (2011) showed viewers multiple photos of foreign celebrities whom they did not know. Asked to make an attractiveness judgement to each photo, it transpired that variability within each person considerably exceeded variability between people. Figure 7 shows that some celebrities were judged more attractive *on average* than others, but these differences are much less than the differences within photos of the same person. This suggests that attractiveness is just as much a property of a photo as it is a property of a person. Similar demonstrations are possible with other rated dimensions. For example, Jenkins et al also show big within person differences on ratings of 'likeness' for familiar faces (how good a photo is this of Bill Clinton?). It seems reasonable to propose that such variability exists across a wide range of perceived facial dimensions.

If we are to take seriously this injunction to study variability, then it will be necessary to go beyond simple demonstrations, and provide operational methods for quantifying this. I will finish this paper by describing briefly how this might be done for (i) perceptual aspects of faces; and (ii) physical aspects of facial images. Jenkins et al (2011) describe a technique for measuring one aspect of variability based on a sorting task. We presented viewers with 40 passport-sized photos, comprising 20 photos of two different Dutch celebrities. Participants were asked to sort these into piles, one per identity, but were not told how many identities to expect. The results were very clear; UK viewers who did not know these celebrities, sorted the photos into many piles – 9 on average. On the other hand, almost all Dutch viewers performed perfectly, correctly sorting the images into two piles.

The effect of familiarity here is very marked, but it is important to note that the difficulty confronted by unfamiliar viewers was not discriminating between individuals. The number of piles containing two identities was very small: less than one per participant on average across all unfamiliar viewers. So, these participants do not have problems telling faces apart, they have problems 'telling faces together.' Note that the stimuli for these experiments were gathered from an internet search. We did not attempt to constrain the range of photos used, except by very loose criteria (no occlusion of the face, sufficient resolution for printing). Within these criteria, we used the first twenty photos returned from each internet search. While we have no way to estimate how this variability relates to the actual variability of all existing photos of these celebrities, it is certainly the case that a wide range emerged, and that these were nevertheless well-recognised by familiar viewers. This is an easily-replicable result, which will allow other groups to witness the level of variability with their own searches.

Although this is research at a relatively early stage, the technique appears to give an index of variability which is properly sensitive to familiarity. Previous research by Clutterbuck & Johnston (2002, 2004, 2005) has shown that one's ability to match two different images of the same person is highly sensitive to familiarity, and here a sorting task seems to offer a generalization of that result. In future work, we intend to track performance on this sorting task through familiarization.

FIGURE 8 HERE

It is also possible to examine the variability of photos themselves. Some time ago, we proposed a scheme for understanding face familiarity based on the computation of face averages (Burton, Jenkins, Hancock, & White, 2005). Figure 8 shows the average of two people's faces, constructed from a set of individual photos. Our proposal was that averaging individual instances of faces forms a good model of acquiring familiarity with a new person. As the person is encountered more often, the sample on which the average is based becomes a better estimate of the 'population mean' for that person. In practice, this has the effect that superficial image characteristics become less important in the representation as familiarity increases. Figure 8 shows, for example, how differences in lighting direction are 'washed away' in the average images.

We have argued (Burton et al., 2005; Burton, Jenkins, & Schweinberger, 2011) that viewers' difficulties in matching unfamiliar faces arise because, without a stable representation, they are forced into an unsophisticated image-matching strategy, which is inherently error-prone. For example, viewers unfamiliar with RJ, top half of Fig 8, often perform poorly when asked to match two images taken at random from the set of individual images. This is because the images themselves are rather poorly matched – there are very large differences between them, and these cause difficulties for the unfamiliar viewer. However, these differences are eliminated by the average, which we refer to as a 'stable' face representation. This preserves aspects common to the sample (eye-brows, feature shapes etc), while eliminating variability in incidental dimensions (e.g. lighting, expression, etc). Familiar viewers, we claim, do not find matching difficult, because they are able to avoid unsophisticated image-matching, and instead compare individual photos to a stored representation.

It is interesting to note that this proposal, based on averaging, is completely consistent with a common idea in the literature on familiar vs unfamiliar face processing. Originally proposed by Ellis, Shepherd, & Davies, (1979), it is held that recognition of unfamiliar faces is based on external features, such as hair, to a greater extent than familiar faces, where internal features are used comparatively more often. The 'averaging' proposal is consistent with this if we assume that external features occupy a large part of any uncropped image, and so dominate a strategy based on image-matching. At the same time these aspects of the face are more variable than internal features; one's hair changes more often than one's nose, say. This is an interesting comparison. The feature-based account of familiar vs unfamiliar face processing assumes some kind of intelligent strategy for making identifications, based on knowledge of the structure of a face: a general face expertise, perhaps. In contrast, our own proposal reaches the same prediction based entirely on the statistical nature of face images as one encounters them – it relies on no expertise in the structure of faces.

These proposals can also be applied to automatic face recognition. Although there has been steady progress in these systems, none is currently available with useable levels of accuracy. O'Toole, An, Dunlop, & Natu, (2012) demonstrate that the best available systems at present can achieve accuracy levels comparable to *unfamiliar* viewers, but as we have seen, these are far from perfect. We have proposed that in order to improve these systems further, they should incorporate aspects of *familiar* face recognition in humans – currently none do. To demonstrate this, we have shown that systems in which novel images are compared to a person's average face, typically out-perform systems in which two separate instances are compared (Jenkins & Burton, 2008; Jenkins, Burton, & White, 2006). Importantly, this effect appears to be robust over different strategies for matching images. The technical aspects of the match algorithm itself tends to dominate the engineering literature on face recognition. However, our proposal is that a focus on what is to be matched may improve performance.

So far, I have described the statistical approach to face recognition, using only the barest of statistical descriptions, the mean of a sample of images. However, the behavioural research described above forces a consideration of variability in the face images themselves: it appears that people's faces have both a characteristic centroid and *characteristic variability*. How might it be possible to quantify this variability in photos of a particular person? In fact, there is a long tradition studying variability in face research, using techniques such as principal components analysis (PCA), Independent Components Analysis (ICA) and other related techniques (Bartlett, Movellan, & Sejnowski, 2002; Kirby & Sirovich, 1990; Turk & Pentland, 1991). However, the use of these techniques is almost always aimed at discriminating *between* different faces. Large corpora of face images are subject to these techniques, one image per person, and analyses such as PCA extract dimensions of variability between people. More recently, we have used these same techniques to examine the nature of variability within a person. By subjecting a sample of images of the same person to techniques like PCA, it is possible to examine quantitatively the dimensions on which that person varies.

FIGURE 9 HERE

Details of the PCA approach to understanding within-person variability can be found elsewhere (Burton et al., 2011; Jenkins, White, Van Montfort, & Burton, 2011). The approach, while in its infancy, appears promising. In particular, our explorations appear to demonstrate that dimensions of variation are to some extent shared between people: for example transformations such as pose, which are common to all faces, always emerge early in the analysis. On the other hand, these analyses appear to show idiosyncratic variability too – particularly as manifested in non-rigid transformational dimensions. As an example figure 9 illustrates the first five dimensions which emerge from a PCA on the shape of different pictures of Harrison Ford. Early components seem to capture pose; for example, dimension 1 is a relatively straightforward coding of left-right rotation – and this seems to be the case whenever such analyses are performed on sets of pictures of the same individual. However, later dimensions incorporate non-rigid transformations, seen for example in the expressive change visible in dimension 4.

This example (figure 9) is illustrative, and gives an indication of future work. There are a number of non-trivial issues which will need to be addressed; for example, how to quantify the extent of idiosyncratic variability in individuals' faces, and how properly to sample images for statistical analysis. However, the main point here is that it is in principle possible to operationalize the concept of within-person variability. The eventual aim of this line of research is to establish a representation of familiar faces analogous to a *confidence interval* for that person. That is, we aim to establish what range of photos can be recognised as a particular individual. Importantly, derivation of this 'confidence interval' is entirely statistical. Familiarity is not a binary dimension: we know some people better than others, and we have a range of different types of exposure (for example to celebrities, colleagues or family). These predict different representations, depending on the particular sample of images to which one has been exposed. This clearly has testable empirical implications, as well as being consistent with our daily experience in face recognition.

Concluding Remarks

I have argued that there are two serious problems with our current approach to face recognition, as it applies to perception of *identity*. First, our experimental methods do not acknowledge the inherent variability of different images of the same face. Second, our theoretical approaches fail to acknowledge the major differences between familiar and unfamiliar face processing. These two problems are not independent. With colleagues, I have argued that unfamiliar face recognition is, to a large extent, a problem of picture matching – because we do not have access to information about how individual unfamiliar faces vary. On the other hand, we do have access to variability in familiar faces (through repeated exposure), and we have argued that this information is a key part of familiar face recognition. To be clear, we are not arguing that variability is a novel and neglected part of face recognition, whose study will add to our knowledge in a cumulative way. Instead, we suggest that variability is a key part of familiar face recognition – and without it, one may not be able to understand the process at all. In fact, Bruce (1994) anticipated this argument many years ago in this journal. She suggested that the non-rigid motion of faces may make them *easier* to recognise by comparison to objects without such variation – a radical alternative to the view that familiar face recognition is so good because of specialized processing, innate or acquired. Unfortunately, this suggestion has not been taken up in the literature.

As we have seen, an acknowledgement of the importance of variability is not a counsel of despair. There are well-defined procedures for studying this directly. Of course, no-one would claim that specific techniques (PCA, ICA or related procedures) are truly representative of the way in which human perceivers code the range of images which can be recognized as a single person. These are simple statistical tools to analyse images stored in particular ways on computers. As the study of variability develops, it will be necessary to bring more sophisticated approaches to modeling the statistics of images – almost certainly incorporating more realistic constraints from low level vision. Furthermore, the statistical approaches themselves may need to be more sophisticated. For example, it is surprising how much progress is possible using simple linear techniques, but longer-term these may need to be replaced with more complex statistical analyses. However, despite all these caveats, the study of variability is possible, as demonstrated above – and we have argued that it is also necessary.

If variability is to be taken seriously in face recognition, then some current methodological approaches will also need to be abandoned. Where possible, it will be necessary to use *ambient images*, i.e. the pictures of faces which people normally recognize. It is tempting not to do this when testing the effect of a particular independent variable on recognition, because it seems natural to construct stimuli which vary *only* on that IV. However, we have argued that eliminating natural variability may be misleading – at worst leading one to believe that some dimension is important, when it is in fact only important within an artificially-constrained set of laboratory stimuli – perhaps only *detectable* within such a set. Wherever possible, we therefore suggest that face recognition researchers use naturally-occurring stimuli. While this is difficult, it can sometimes be more tractable than it first seems. For example, when comparing familiar with unfamiliar faces, one might wish to control the two groups of stimuli for certain visual properties. However, it is sometimes possible to avoid this by using a two-site design in which only local faces are familiar –

hence eliminating the need for systematic matching of familiar and unfamiliar faces, since all faces can be used in familiar and unfamiliar conditions.

Finally, I should reiterate that the criticisms leveled at face research in this paper can all be applied to our own previous work. There is no intention here to be critical of specific research programmes or (still less) research groups. Face recognition is an exciting international research topic, and it has developed its own vocabulary and practices, just as any other research area. However, I have argued that it has developed some unhelpful shared assumptions. One of these is the idea that the problem of face recognition is exclusively the problem of how we tell each other apart. After all, many of us have pointed out to students that faces are essentially all the same, and so recognition must somehow capture the ways we are able to perceive rather subtle differences between people. I have argued here that this way of conceiving the problem is not consistent with our everyday experience of face recognition, which not only copes with variability, but exploits it. 'Telling people together' is a key component of telling them apart. This paper calls for a new approach to face recognition; it remains to be seen whether the research community finds the arguments it contains compelling.

Acknowledgements

This paper was written while in receipt of an ESRC Fellowship (ES/J022950/1). All the arguments it contains have been developed over several years in collaboration with Rob Jenkins. I am grateful to him and to Stefan Schweinberger for reading an earlier draft. I would also like to express my thanks to Vicki Bruce and Andy Young for career-long support in thinking about these issues.

References

Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, *13*(6), 1450–1464. doi:10.1109/TNN.2002.804287

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. doi:10.1126/science.1598577

Benton A.L., Sivan A.B., Hamsher K.deS., Varney N.R., Spreen O. (1994). Contributions to neuropsychological assessment. 2nd edition. New York: Oxford University Press.

Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. R. (2008). Brain potential correlates of face recognition: Geometric distortions and the N250r brain response to stimulus repetitions. *Psychophysiology*, *45*(4), 535–544.

Bruce, V. (1994). Stability from variation: The case of face recognition. *Quarterly Journal of Experimental Psychology Section A*, *47*(1), 5–28.

Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception*, *15*(4), 387–397.

Bruce, V., Ellis, H., & Gibling, F. (1987). Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology*, *41*(4), 510–520.

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., 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.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British journal of psychology*, *77*(3), 305–327.

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. In A. J. Calder, G. Rhodes, M. Johnson, & J. Haxby, (Eds.) *Oxford Handbook of Face Perception*, pp 287–306. Oxford: Oxford University Press.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943–958.

doi:10.1111/j.2044-8295.2011.02039.x

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*(3), 256–284. doi:10.1016/j.cogpsych.2005.06.003

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

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.

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*(8), 985–994. doi:10.1068/p3335

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition*, *11*(7), 857–869.

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.

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*(6), 1627–1635.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology*, *23*(4), 482–505. doi:10.1002/acp.1490

Deffenbacher, K. A., Johanson, J., Vetter, T., & O'Toole, A. J. (2000). The face typicality-recognizability relationship: encoding or retrieval locus? *Memory & Cognition*, *28*(7), 1173–1182.

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. doi:10.1016/j.cognition.2012.09.004

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. doi:10.1016/j.neuropsychologia.2005.07.001

Eimer, M. (2011). The face-sensitive N170 component of the event-related brain potential. In A. J. Calder, G. Rhodes, M. Johnson, & J. Haxby, (Eds.) *Oxford Handbook of Face Perception*, pp 287–306. Oxford: Oxford University Press.

Ellis, A. W., Young, A. W., & Flude, B. M. (1990). Repetition priming and face processing: Priming occurs within the system that responds to the identity of a face. *The Quarterly Journal of Experimental Psychology Section A*, *42*(3), 495–512. doi:10.1080/14640749008401234

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*(4), 431–439.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition*, *103*(2), 322–330. doi:10.1016/j.cognition.2006.05.003

Goshen-Gottstein, Y., & Ganel, T. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgment task: Evidence for a common route for the processing of sex and identity. *Journal Of Experimental Psychology-Learning Memory And Cognition*, *26*(5), 1198.

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

Harper, B. & Latto, R. (2001). Cyclopean vision, size estimation, and presence in orthostereoscopic images. *Presence*, *10*(3), 312–330.

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–232.

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

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, *31*(10), 1221–1240. doi:10.1068/p3252

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science*, *319*(5862), 435.

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1671–1683. doi:10.1098/rstb.2010.0379

Jenkins, R., Burton, A. M., & White, D. (2006). Face Recognition from Unconstrained Images: Progress with Prototypes. *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 25–30.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313–323. doi:10.1016/j.cognition.2011.08.001

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.

Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *12*(1), 103–108.

Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory & Cognition*, *12*(1), 60–70.

Kleinberg, K., Vanezis, P., & Burton, A. M. (2007). Failure of Anthropometry as a Facial Identification Technique Using High-Quality Photographs. *Journal of forensic sciences*, *52*(4), 779–783.

Leopold, D., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, *8*(10), 1386–1391. doi:10.1038/nn1538

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*(1), 8–15. doi:10.1016/j.tics.2006.11.002

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

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception & psychophysics*, *69*(7), 1175.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364–372.

Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science*, *11*(6), 492–496.

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British journal of psychology*, *102*(4), 726–747. doi:10.1111/j.2044-8295.2011.02053.x

O'Toole, A. J., An, X., Dunlop, J., & Natu, V. (2012). Comparing face recognition algorithms to humans on challenging tasks. *ACM Transactions on Applied Perception*, *9*(4:16), 1–13. doi:10.1145/2355598.2355599

O'Toole, A.J., Edelman, S., and Bülthoff, H.H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research, 38***,** 2351–263.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(4), 406–417.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, *51*(3), 333–342. doi:10.1016/j.visres.2010.11.014

Roberts T. and Bruce, V. (1989). Repetition priming of face recognition in a serial choice reaction-time task. *British Journal of Psychology, 8,* 201–211.

Rossion, B. (2002). Is sex categorization from faces really parallel to face recognition? *Visual Cognition*, *9*(8), 1003–1020. doi:10.1080/13506280143000485

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *128*(2), 274–289. doi:10.1016/j.actpsy.2008.02.003

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, *16*(2), 252–257. doi:10.3758/PBR.16.2.252

Schweinberger, S. R., Huddy, V., & Burton, A. M. (2004). N250r: a face-selective brain response to stimulus repetitions. *NeuroReport*, *15*(9), 1501–1505. doi:10.1097/01.wnr.0000131675.00319.42

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion Leads to Quantitative, Not Qualitative, Changes in Face Processing. *Current Biology*, *14*(5), 391–396. doi:10.1016/j.cub.2004.02.028

Tanaka, J. W., & Gordon, I. (2011). Features, configuration, and holistic face processing. In A. J. Calder, G. Rhodes, M. Johnson, & J. Haxby, (Eds.) *Oxford Handbook of Face Perception*, pp 177–194. Oxford: Oxford University Press.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal Of Cognitive Neuroscience*, *3*(1), 71–86.

Vokey, J. R., & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, *20*(3), 291–302.

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual Differences in Holistic Processing Predict Face Recognition Ability. *Psychological Science*, *23*(2), 169–177. doi:10.1177/0956797611420575

Wiese, H., & Schweinberger, S. R. (2008). Event-related potentials indicate different processes to mediate categorical and associative priming in person recognition. *Journal Of Experimental Psychology-Learning Memory And Cognition*, *34*(5), 1246–1263. doi:10.1037/a0012937

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British journal of psychology*, *102*(4), 959–974. doi:10.1111/j.2044-8295.2011.02045.x

Figure 1. A same-image match (top row) and a same-person match (bottom row)

Figure 2. Four pictures of the same person taken within minutes of each other. Photos A and B were taken with one camera. Photos C and D were taken with a second camera, against a different background.

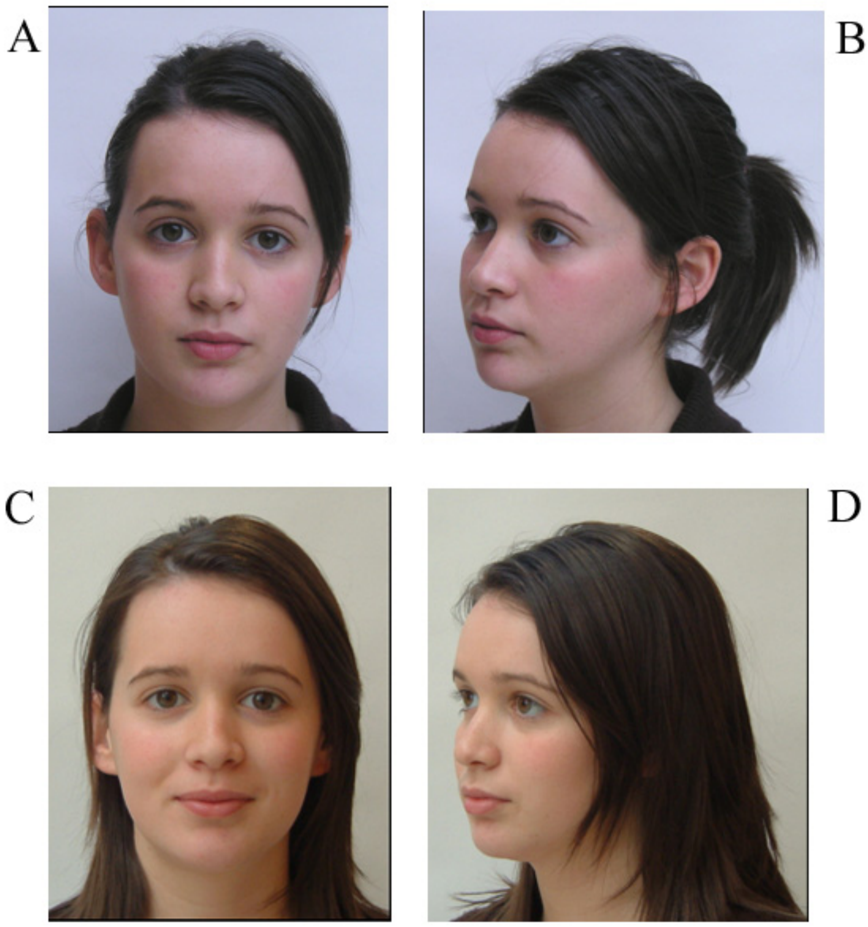


Figure 3: After Harper & Latto (2001). Images of the same person taken at different distances (c. 0.5m to 3m).

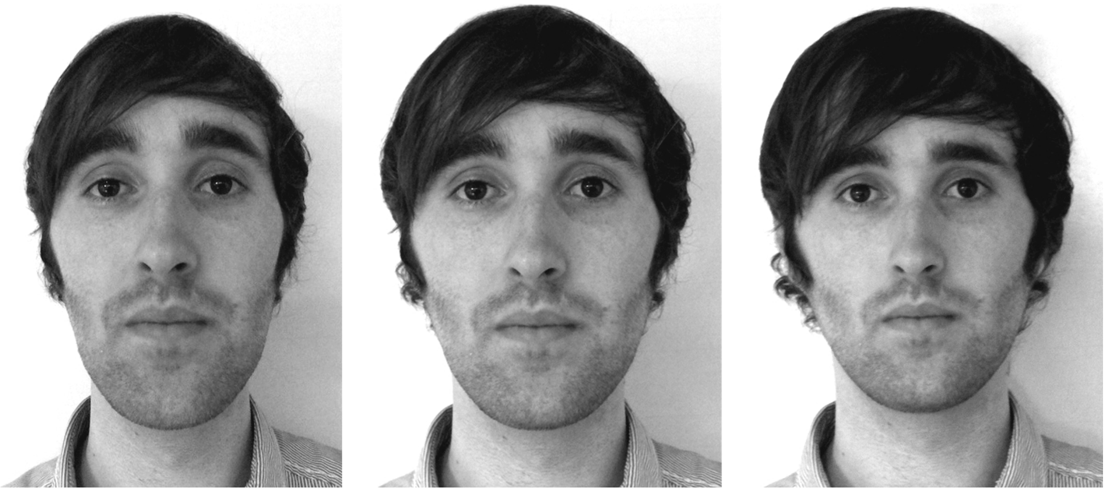
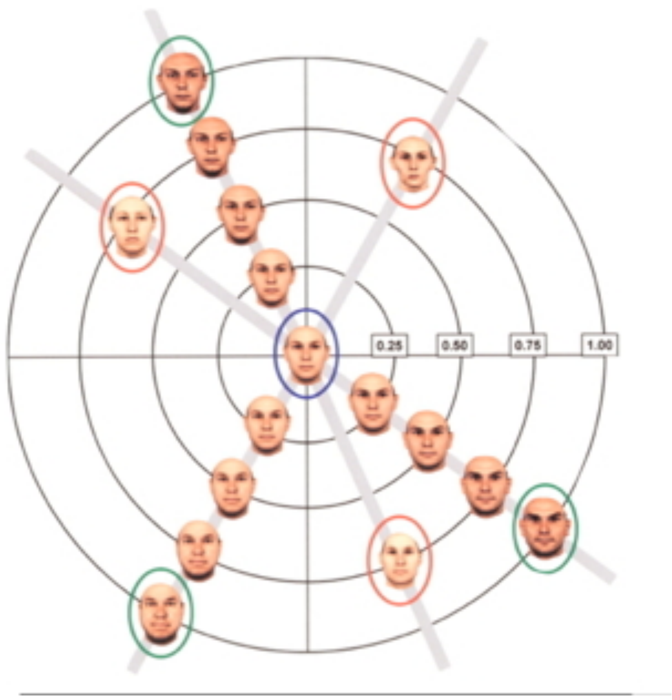


Figure 4: Artificial face images from (a) Leopold et al., (2001) and (b) Loffler et al., (2005)



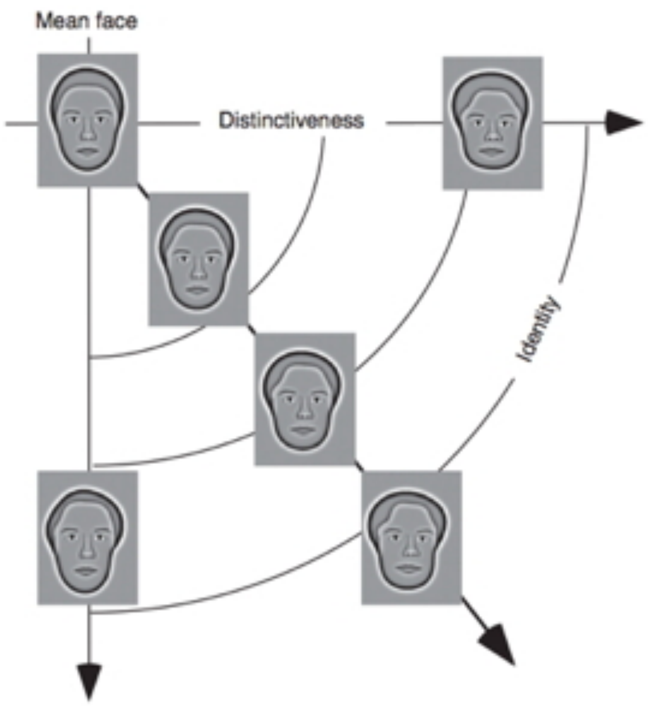


Figure 5: Ambient photos of the same face. (From Jenkins et al., 2011)

Figure 6. Examples of the 1-in-10 face matching arrays from Bruce et al (1999). The person shown at the top may or may not be one of the ten below.

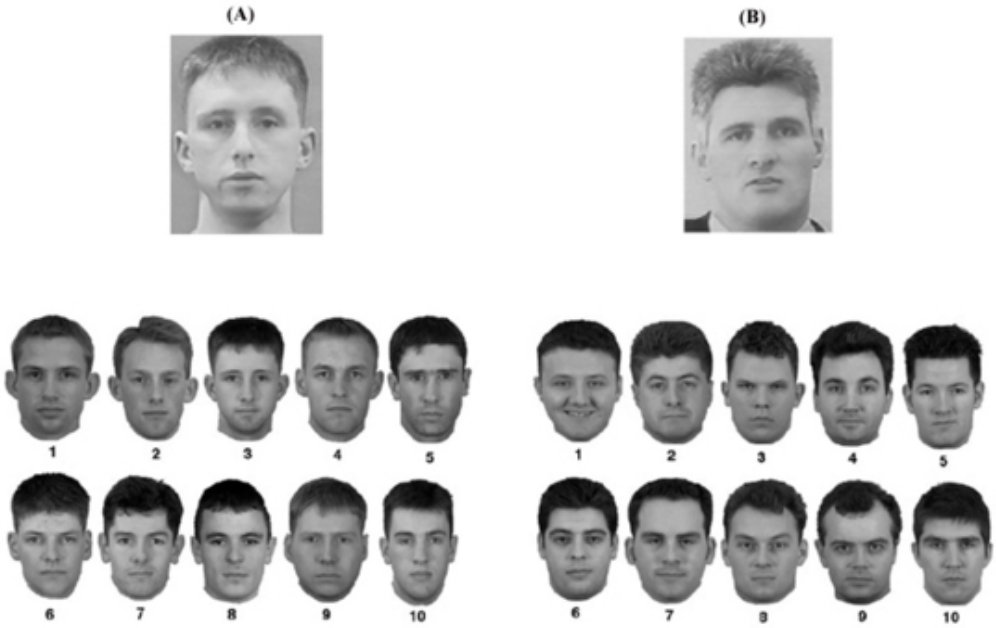


Figure 7: Rated attractiveness for unfamiliar female faces. Each column represents a single person, and dots represent ratings to individual photos of that person. From Jenkins et al., (2011)

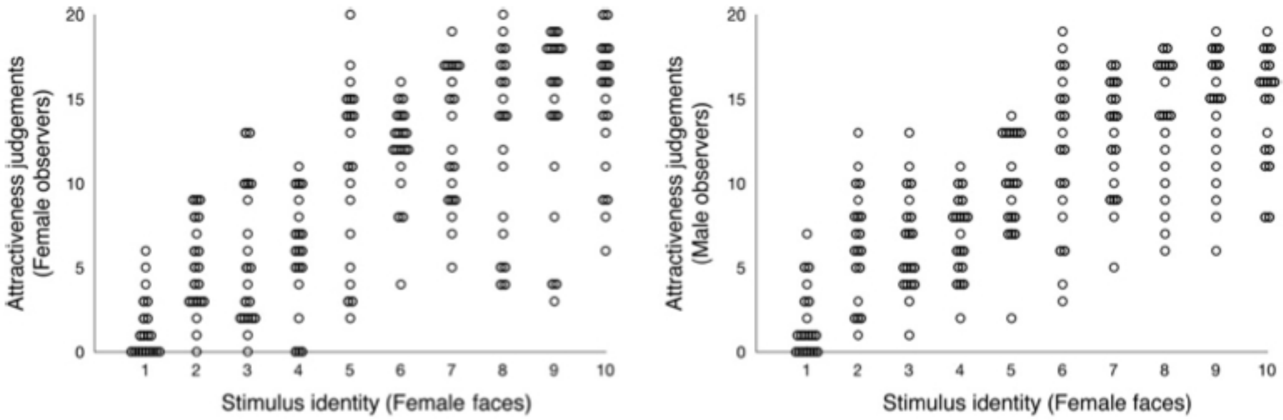


Figure 8: Average images and their constituent photographs for two people, RJ & AMB. From Jenkins & Burton, (2011)

Figure 9: Shape components derived from a PCA on 48 photos of Harrison Ford. Columns show the first five components (left to right), with values $z = +1$ above, and $z = -1$ below. The average texture from the original photos has been mapped to these shapes in each case. From Burton et al., (2011)