

This is a repository copy of *Identity From Variation: Representations of Faces Derived From Multiple Instances*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/85534/>

Version: Accepted Version

Article:

Burton, A Mike orcid.org/0000-0002-2035-2084, Kramer, Robin S S, Ritchie, Kay L et al. (1 more author) (2016) *Identity From Variation: Representations of Faces Derived From Multiple Instances*. COGNITIVE SCIENCE. pp. 202-223. ISSN 0364-0213

<https://doi.org/10.1111/cogs.12231>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Identity from variation: representations of faces derived from multiple instances

A. Mike Burton¹, Robin S.S. Kramer,¹ Kay L. Ritchie¹ & Rob Jenkins²

1. School of Psychology, University of Aberdeen, AB24 3FX, UK
2. Department of Psychology, University of York, Y010 5DD, UK

Address for correspondence:

Mike Burton,
School of Psychology, University of Aberdeen, Aberdeen AB24 3FX, UK
m.burton@abdn.ac.uk

Abstract

Research in face recognition has tended to focus on discriminating between individuals, or ‘telling people apart’. It has recently become clear that it is also necessary to understand how images of the same person can vary, or ‘telling people together’. Learning a new face, and tracking its representation as it changes from unfamiliar to familiar, involves an abstraction of the variability in different images of that person’s face. Here we present an application of Principal Components Analysis computed across different photos of the same person. We demonstrate that people vary in systematic ways, and that this variability is idiosyncratic—the dimensions of variability in one face do not generalise well to another. Learning a new face therefore entails learning how that face varies. We present evidence for this proposal, and suggest that it provides an explanation for various effects in face recognition. We conclude by making a number of testable predictions derived from this framework.

Identity from variation: representations of faces derived from multiple instances*Background*

It is a strong intuition that recognition of people from their faces must be straightforward, because we are clearly able to recognise those we know across a wide range of viewing conditions. However, psychological research over the past fifteen years has established an important qualification: people's very high accuracy in recognising faces is limited to familiar faces. When asked to recognise previously unfamiliar faces, viewers are surprisingly bad. This is a finding that has been well-established in the memory literature for many years (e.g. Bruce, 1986; Ellis, Shepherd & Davies, 1979; Klatzky & Forrest, 1984). However, more recent research has shown that viewers find it difficult to *match* different images of the same unfamiliar face, even when high quality images are presented simultaneously, and no time limit applies (Bruce et al, 1999, 2001; Megreya & Burton, 2006, 2008). This result was surprising at first, but has been replicated many times, and has been extended into real-world settings in which people have to match a photo to a video, or to a live person (Davis & Valentine, 2009; Kemp, Towell & Pike, 1997; Megreya & Burton, 2008). In all of these settings, people make very large numbers of errors in matching faces – typically in the range of 10%-30%, depending on the task.

We have previously argued that familiar and unfamiliar face recognition involve qualitatively different representations. Familiar face recognition is robust across changes in image, and seems to rely on higher-level representations, whereas unfamiliar face recognition is bound much more closely to the visual properties of the particular image one is viewing (Hancock, Bruce & Burton, 2000). In other words, the robust recognition performance seen for familiar faces does not generalise to unfamiliar faces. This observation is important, as it carries an implication that is easy to overlook: The expertise that comes with learning faces is not expertise for faces as a class of objects. It is expertise for the individual faces that have been learned. The discrepancy between familiar and unfamiliar face processing has been highlighted by Sinha et al (2006), in their important paper bridging the fields of automatic and human face recognition. Progress has certainly been made in automatic face recognition, and there are now systems that can out-perform unfamiliar human

viewers in some circumstances (O'Toole et al, 2012). However, unfamiliar face recognition is not very accurate. To make further progress in automatic face recognition, it would be helpful to understand how familiarity confers such an advantage in human face recognition. In this paper we aim to provide a model for understanding the representation of familiar faces, a representation that can be used to support both human and automatic recognition.

The importance of variability

FIGURE 1 HERE PLEASE

Recent work on face familiarity has highlighted the importance of understanding within-person variability in appearance (Burton et al, 2011; Jenkins et al, 2011). Figure 1 shows several different images of the same person. These differ for a number of reasons, including changes in the person (e.g. pose, expression, age) and changes in the capture conditions (e.g. lighting, camera settings, focal length). Despite the fact that these vary in many different ways, they are all easily recognisable to a viewer who is familiar with this person. However, it turns out that variability is a key discriminator between familiar and unfamiliar face recognition. In a recent card-sorting task (Jenkins et al, 2011) participants were given 40 face photographs and were asked to sort them by identity, so that different photos of the same person were grouped together. All of the photos depicted Dutch TV celebrities, who were unknown to the British participants. In fact, the cards comprised just two faces – 20 photos of Person A, and 20 photos of Person B. Yet participants sorted them into 9 identities on average. Dutch viewers, who were familiar with the faces, showed a completely different pattern, with almost all participants correctly sorting the cards into two piles. Again, it is familiarity with the faces concerned—not faces as a class of objects—that determines performance on this task. Jenkins et al. (2011) report that the unfamiliar participants rarely conflated the two identities – very few piles contained both people. The difficulty for these viewers is therefore not ‘telling people apart’ but ‘telling people together’.

Extension of previous work

We have proposed that an explanation of human familiar face recognition must rely on an understanding of *both* between-person variability and within-person variability. The process of familiarisation appears to support both these components of the problem, but there is almost no research available on the latter. Many face perception experiments have eliminated within-person variability entirely, by representing each face with a single image. We have recently argued that ignoring variability in this way can be misleading. If face representations are to support reliable identification, they will have to incorporate variability somehow (Jenkins & Burton, 2011). Our initial approach to this involved stabilising the variability by averaging together multiple photos of each face (Burton et al., 2005; Jenkins & Burton, 2008). This averaging process has the effect of washing away aspects of the image that change from one photo to the next, while preserving aspects of the image that are consistent across the set. The resulting images have some interesting properties. First, they stabilize quickly. Once around 20 photographs have been averaged together, adding further photos does not greatly affect the appearance of the average image. Second, they converge well. Whether the average is composed of one random set of photographs or another random set of photographs, the results of the process are similar. Third, they are robust to errors. Incorporating a few photographs of the wrong person does not make much difference to the average image (Jenkins, Burton, & White, 2006). This process tends to improve recognition accuracy because it stabilizes the representation of a person's face, meaning that the match is not destroyed by atypical images. Although this strikes us as promising, an average remains a very limited statistical summary. It provides a measure of the central tendency of a set of images, but tells us nothing about their distribution. In this paper, we develop a method for incorporating distribution into representations of a face.

The proposal we wish to develop is that individual faces have their own idiosyncratic variability. All faces vary in appearance, but they vary in different ways. At some levels, this is plainly true. A man typically varies to some extent in beard length, whereas a woman does not. However, the position we advance here is more radical. We propose that idiosyncratic variability is fundamental to face learning and familiar face recognition. An enduring idea in face recognition research is that learning a person's face involves learning key invariants – such as metric distances between features – that distinguish that face from all others (e.g., Richler et al, 2009; Tanaka &

Gordon, 2011). This idea has intuitive appeal. What limits its viability is that no such invariants have been found (Sandford & Burton, 2014). Every measure that one might refer to – from skin tone to interocular distance – is subject to within-person variability between different images. This observation suggests that extraction of invariants may be a rather poor candidate for a face learning mechanism. In this paper we invert the problem by focusing on extraction of *variants*.

A technique for investigating within-person variability

Computational work on face recognition has used a number of approaches based on the statistical properties of images. The most common of these is principal components analysis (PCA), and we will concentrate on this technique here. The core method is to derive a space of facial images based on the eigenvectors of a PCA-decomposition of a set of faces. These faces can be satisfactorily represented in a rather small number of eigenvectors (called ‘eigenfaces’ in this literature), making the PCA approach a good technique for efficient coding of face images for engineering or telecommunications applications (Kirby & Sirovich, 1990; Turk & Pentland, 1991). Once a set of faces has been used to derive a low-dimensional space, new images can be projected into this space, and matched against the stored images. This approach has been used in many automatic face recognition systems. However, the typical approach has been to analyse images of many different people, thus extracting dimensions along which different faces vary.

Our approach here is to perform PCA on images of a single person, with the goal of spanning the space of that person’s variability. If this goal were achieved, it would allow one to understand the entire visual range of a particular person’s face. For example, we should be able to characterise all possible images of Tom Cruise. This use of multiple images of an individual relates to the *dictionary learning* techniques utilised in computer vision (e.g. Patel, Wu, Biswas, Phillips, & Chellappa, 2012), which are themselves a development of class-based approaches to vision (e.g. Edelman & O’Toole, 2001). These techniques emphasise visual processes that are specific to a particular class of objects, which are learned by exposure to examples of the class, and exploit their statistical structure. Early work in this field took ‘faces’ as a class of visual object, using variability in that class to generalise to novel examples – for example generalising knowledge of changes in viewpoint or illumination to a

novel face (e.g. O'Toole & Edelman, 1996; O'Toole, Edelman, & Bühlhoff, 1998). More recently, some researchers have shown that within-person PCA, termed 'face-specific subspace' PCA, can lead to improvements in automatic recognition accuracy (e.g. Aishwarya & Marcus, 2010; Shan, Gao, & Zhao, 2003). However, the major behavioural differences between familiar and unfamiliar faces often remain unacknowledged in computational approaches to face perception, and so our purpose is to examine how and why familiarity produces familiarity benefits. The hypothesis explored here is that learning of the idiosyncratic variability of a specific face is key to becoming familiar with that person.

In common with many PCA approaches, we employ a shape-normalisation of the face images (Beymer, 1995; Vetter & Troje, 1995, Burton et al, 2001). Prior to analysis, a standard grid is placed on the face, and altered by hand to align with key points. The image is then morphed to a standard shape, which is the same for all examples. This procedure corresponds to separating two components of the face—shape and texture. (We use the term 'texture' to describe the information in the shape-free face, though it actually includes more information, including colour, reflectance, lighting etc.) Having performed this separation on many face images, we then subject the shapes and textures separately to PCA. The projection of contributing, or novel, faces onto the resulting eigenvectors is known as the 'reconstruction' of the face, and we express this reconstruction in a low dimensional space, using the early eigenvectors of shape and of texture. Reconstruction error compares an original image with its reconstructed (low dimensional) version, and this error represents an inverse measure of the accuracy with which the new space can capture any particular face.

Research questions

In the work described below, we use this approach to ask three key questions.

1. What are the dimensions of variability for a particular individual? Many previous analyses have extracted dimensions along which different faces vary (e.g. eye-shape, nose length). However, few have focused on variability within a single face. The current approach is novel in two important respects. First, it holds individual anatomy constant, in the sense that all the images contributing to a given analysis depict the same face. Second, we use images that are sampled from the real world (*ambient images*; Jenkins et al., 2011; Sutherland et al., 2013) rather than being taken under

controlled conditions. One interesting question is whether psychologically relevant dimensions emerge from this purely statistical image analysis. If no such dimensions emerge, this would push demands on structuring the image data cognitively upstream. However, if psychologically relevant dimensions do emerge, this would imply a high degree of accessible structure in the data.

2. To what extent is within-person variability idiosyncratic? This is a key question if within-person variability is to be recruited for identification. If different faces vary in different ways, then the dimensional structure of variability will be person-specific (by definition). This allows a representational scheme that is fundamentally different from some conceptions of face space (e.g. Valentine, 1991), while preserving dimensional coding. In some versions of face space, different faces populate different regions of a single space, which is defined by common axes. The alternative we advance here is that each face is represented by its own person-specific coding space, which is defined by bespoke axes. Such a scheme would entail several basic phenomena that are otherwise difficult to explain. For example, it would explain why learning a person's face requires exposure to variation (only exposure to variation reveals the dimensions of variability). It would also explain why familiarity with one face does not generalise to another face (the dimensional structures are different). We expect that if different faces vary in different ways, then a set of dimensions that codes one face well should code other faces less well. Alternatively, if dimensions of variation are common across faces, then a single set of dimensions should code different faces equally well.

3. To what extent is it possible to span the space of an individual with a small number of contributing images? Previous studies of face learning have found graded improvements in identification performance as exposure to variability increases. This is true for human face recognition (Clutterbuck & Johnston, 2002; Bonner et al, 2003; White et al, 2014), and also automatic face recognition using nearest neighbour match (Burton et al., 2005; Jenkins & Burton 2008). These converging findings, based on very different methods, imply that representations of facial identity are more effective when they incorporate more images. However, the quantity and quality of exposure required to achieve robust recognition is not known. Here we ask how many images are required in order to capture a person's variability in appearance. If the extracted

dimensions cover the full range of possible appearances for a particular person, then they should code new images of that person just as efficiently as they code old images (i.e. the images that were used to derive the dimensions). Conversely, if the extracted dimensions do not cover the person's range of possible appearances, then old images should always have a coding advantage.

Images

In order to address these questions we need to sample a range of photos of target individuals. In contrast to many previous research projects, we specifically wish to avoid controlling our stimuli for known dimensions of variability, which are sometimes regarded as 'noise'. One approach to our problem would be to sample target faces from conditions in which environmental variables (e.g. lighting, camera) and personal variables (e.g. expression, age) are controlled, or systematically varied. Of course, it is not possible to control for every variable contributing to face photos. However, in recent work we have proposed that this is not desirable scientifically. We have argued (Jenkins & Burton, 2011; Burton, 2013) that controlling stimulus variability *removes* information that is relevant to identification. For this reason, we study the range of face images over which human face recognition/identification normally takes place. Our stimuli comprise naturally occurring face images for which we had no control over capture conditions, but for which it is easy to establish recognisability.

Our technique for gathering face images is to use Internet search. The current study uses celebrity photos, ensuring that there are very many images of each person available. A celebrity's name is entered into Google Images as a search term, along with criteria specifying full colour, large, face images only. We then choose the first 35 images delivered which meet the following criteria: (i) no part of the face should be obscured (for example by clothing, glasses, or a hand); (ii) pose should be very broadly full-face in order to allow the placement of landmarks; and (iii) pose should be standing or sitting, but not lying down, in order to limit the angle of the head to relatively upright. Figure 1 provides an illustration of the range of variability allowed by these criteria.

Method: For the purpose of this paper, we performed PCA on 30 different images (the ‘training set’, randomly selected from the set of 35) of 10 Caucasian Hollywood actors (5 females and 5 males). Each image was scaled to 190 pixels wide x 285 pixels high, and represented in RGB colour space using a lossless image format (bitmap). Face shape was derived manually for each image by aligning the points of a standard grid with anatomical landmarks. The standard grid comprised 82 xy-coordinates, resulting in a vector of 164 numbers (82 points x 2 coordinates). Shape PCA was based on these shape vectors. We next generated the average shape for each actor (i.e. the identity average), by computing the mean coordinates for each landmark, across all 30 images of that person. The texture for each image was then morphed to the average shape of the corresponding person, resulting in a vector of 162450 numbers (190 width x 285 height x 3 RGB layers). Texture PCA was based on these texture vectors.

Results 1: What are the dimensions of variability for a particular individual?

Before providing quantitative comparison of images within and between people, we offer some observations derived from a visualisation tool shown in Figure 2. The tool shows an original image (left window), and its reconstruction in 30 texture and 30 shape components (right window). Sliders on either side of the tool allow values of individual eigenvectors (shape or texture) to be manipulated independently, resulting in changes to the reconstructed image. Animation tools provide a visualisation of single eigenvectors, by reconstructing the image with incremental variation of a single dimension, while leaving all others unchanged. This allows one to gain a qualitative impression of the influence of a single dimension. One can apply this visualisation technique to any training image. In the illustrations that follow, we demonstrate this technique using reconstructions of individual celebrity photos, and also using average images of these celebrities. Applying this tool to the set of images described above leads to the following observations.

FIGURE 2 HERE PLEASE

Observation 1: The first three dimensions of shape always appear to describe rigid head rotations in three-dimensional space. The visualisation of shape for all within-person analyses seems to show that the largest variance in ambient images lies in their pose, or their angle to camera. While the first dimension usually corresponds to head-rotation ('yaw'), components 2 and 3 do not always fall so neatly into one further commonplace dimension ('pitch' or 'roll') but they always introduce novel angle of projection dimensions, so that the first three dimensions span the 3D world.

Researchers in computer science have previously suggested that lighting, and to a lesser extent pose, are incorporated within these three dimensions, although they often utilise standardised photo sets where lighting and pose are systematically varied (Belhumeur, Hespanha, & Kriegman, 1997; Geng & Li, 2007).

Observation 2: Expressed in the early components of both shape and texture is a coding of left-right rotation. Within-person PCA on all the identities studied here (and in many more previously) shows an early dimension, capturing considerable variability across ambient images, coding a rotation corresponding to a movement from one three-quarter profile to the opposite profile. When manipulated as a shape component, this is visualised as an apparent head-turn. When manipulated in texture, the visualisation shows an apparent movement of directional lighting from one side of the image to another. Figure 3 illustrates this for one of the celebrities, Tom Hanks. Low and high values of the first shape and second texture dimensions are added to his average face, illustrating the contributions of these eigenvectors.

FIGURE 3 HERE PLEASE

Observation 3: Coding of non-rigid deformations typically begins at component 4. In general across identities, the first three shape components have no non-rigid component. Non-rigid deformations code changes in the face due to expression or facial speech (see Figure 4). Although there is no logical necessity for this division – for example, there is no reason why early components should not code non-rigid as well as rigid deformations – it is quite consistent across the examples we have tried that non-rigid deformations are not seen until component 4. This clear separation in the data fits the common sense view that rigid and non-rigid deformations are not

correlated. For example, viewpoint does not determine facial expression and vice versa.

FIGURE 4 HERE PLEASE

Observation 4: From component 4 onwards (in both shape and texture), variability is idiosyncratic. We observe some common non-rigid variability across the identities we have tried, for example mouth opening as in Figure 4, a left–right eye movement, and a facial expression such as a smile. These can be coded on a single component, or combined with a tilting of the head or a change in lighting. Where we do see a ‘smile component’ in common between individuals, each person’s smile varies to differing degrees, reflecting their own idiosyncratic range of expression. We do not observe the components expressing variability in the same order, or to the same extent for each ID. For example, Figure 5 shows the sixth component for two different celebrities, both with interpretable coding, but different in each case.

FIGURE 5 HERE PLEASE

There are some differences between the components that emerge for male and female identities. For example, some of the women show a texture component that appears to code presence or absence of makeup, typified by skin becoming more orange and lips more red, but also a darkening of the eyelids. Figure 6 gives an example. Similarly some men exhibit a ‘facial hair’ texture component with a beard or moustache darkening and lightening. However, these components are still idiosyncratic in the following senses. First, some women clearly vary on a makeup dimension (and some men clearly vary on a facial hair dimension) but others do not. Second, where such a dimension is coded, it is coded by ordinally different components for different individuals (e.g. the 7th component versus the 9th component). Third, the particular information coded by the relevant component is different for different individuals. For example, a makeup component might code mainly changes around the eyes for one person, but mainly changes around the lips for another person.

FIGURE 6 HERE PLEASE

These observations provide good evidence for the idiosyncrasy of people's variability, but it is also possible that they arise, to some extent, through image sampling. By choosing 30 images from a web search, albeit in a consistent manner, it is possible that variation in that particular sample of any individual's face will be specific to that set of photos, rather than to that person. To address this possibility, we collected a larger set of images for one of the celebrities, in order to derive separate analyses from non-overlapping sets. We gathered 90 images of Tom Cruise, using the method described above, and divided these at random into three sets of 30. We then performed PCA on each of these three sets separately. The dimensional structure for the three independent analyses of Tom Cruise photos was strikingly similar. Figure 7 shows for each of these analyses the average of the set modulated by ± 1.5 SDs on dimension 6. We had identified this as a component coding mouth opening for this person in Figure 4, and this seems to be coded too in each of the new analyses. As mentioned above, each of the sets has converged to a very similar average face, but it is interesting that more complex statistical structure also survives different sampling.

FIGURE 7 HERE PLEASE

Results 2: Characterising idiosyncratic variability

In this section we take observations described above, and use a quantitative analysis to address two questions: (i) To what extent is within-person variability idiosyncratic?; and (ii) How well can we capture someone's variability within a small number of dimensions? To answer these questions, we derived separate, person-specific PCA spaces from 30 images of each actor – we call these the 'training set' for that actor. We then reconstructed face images within these spaces, using mean square error (MSE) between original and reconstructed face as a measure of the goodness of encoding of the image within these derived dimensions. For each identity we computed the following reconstructions: (i) each of the 30 training set images, in terms of the training set components; (ii) 5 novel pictures of the same actor, in terms of the training set components; (iii) the same 5 novel images of that actor, in terms of the components derived from each of the other four same-gender actors.

We expect high quality reconstructions of the original images, illustrating that these faces can be represented efficiently within a low dimensional space. Of more interest are the reconstructions of new images. If the PCA is genuinely capturing information about the particular person, rather than faces in general, then new images of that person should be reconstructed better in components derived from images of him or herself, rather than components derived from another actor. Furthermore, if the original PCA is spanning the space of that actor's face well, then reconstructions of new instances of that person will be comparatively good.

To avoid bias due to any particular image set, we carried out 5 iterations of this process, using different sets of 30 images to derive PCA space, with the remaining 5 used as novel same-identity images. Mean MSE for these runs is shown in Figure 8, separately for reconstruction of texture and reconstruction of shape. The results show that in every case, reconstruction error for novel faces is smaller using their personal training set components than components derived from other actors. This is true of both texture and shape components. Of course, reconstruction error in texture and shape are different magnitudes, representing the vastly different size of the data contributing to each (164 points in the shape vector vs 162,450 points in the texture vector). Nevertheless, it appears that the within-person PCA is genuinely capturing some variance that is specific to that person – both characteristic shape and characteristic texture. This lends support to the proposal that people not only differ, but differ idiosyncratically: the ways in which one face varies are not the same as the ways in which another face varies.

FIGURE 8 HERE PLEASE

Figure 8 also shows that we have not captured (spanned) the entire space of each individual. Novel images of a person always give rise to larger reconstruction errors than the images used to build the space. The goal here is to capture as large a range as possible of a particular person's variability – in other words, to describe the space of 'all possible images of Paul McCartney' (for example). In the perfect case, this would be indicated by reconstruction errors for novel images of the same person being no larger than those for an original face. We do not see that pattern here, which is perhaps unsurprising – it seems unlikely that 30 images of an actor from an internet

search would entirely describe the range of variation in that actor. However, neither face recognition nor face learning requires reconstruction of new images to be as good as reconstruction of old images. All that is required is for reconstruction error to be lower using the same person's PCs than using another person's PCs, and that is the pattern seen here.

In the earlier section describing observations, we noted that visibly idiosyncratic components tend to emerge only at rank 4 and beyond. We therefore asked whether the first three components of any individual's PCA code universal physical dimensions, or whether there is evidence of idiosyncrasy even in these early components. To do this, we repeated the analysis above, but this time reconstructing all images in a smaller number of components. Figure 9 gives an example of this analysis for one actor, though results for all actors were qualitatively the same.

FIGURE 9 HERE PLEASE

The data show two interesting patterns, and these are common across reconstructions of shape and texture. First, reconstruction error asymptotes very quickly to a low value. In other words, the variance in these pictures is captured by rather few components. As we might expect, error falls almost to zero for the training set images, but reconstruction of images in non-training set components stabilises very quickly too. More interestingly, there is an advantage for reconstructing a person in his or her own components immediately. There is some degree of idiosyncratic variability present even in the very first component derived from within-person PCA. At first sight, it is not obvious why this should be so. Although it is easy to accept that, say, a person's smile might be idiosyncratic (Harrison Ford's smile and Jack Nicholson's smile transform a face in different ways), the intuition is not so clear for viewpoint. Surely a 10° turn of the head transforms any face in the same way? The key point is to note that PCA here is not applied to 3D objects; it is applied to 2D projections of 3D objects, and identical turns of different heads cause different changes in the 2D projection. Suppose two people vary in nose length. Each person turns his head 10° to the left. This 3D movement in the world translates the tip of the long nose much further than the tip of the short nose, with consequences for both 2D shape and texture information. A 10° change in lighting direction would likewise have idiosyncratic

effects, as noses of different length cast different shadows (see Beveridge, Draper, Chang, Kirby, Kley, & Peterson, 2009).

These effects demonstrate the benefit of treating different instances of a single person as a perceptual ‘class’. While our approach is derived from earlier proposals (for example the class-based approaches described above), these results demonstrate the benefit that can be gained from an appropriate choice of the conceptual level over which one generalizes. If we regard ‘faces’ as a class, and extract statistical regularities from these, then this will not provide all the information we need in order to understand variations in novel faces. The fact that idiosyncratic information is evident even in the very early components, demonstrates the added benefit of this approach over those class-based positions which exploit only variability between people.

Discussion

In this paper we have illustrated how it is possible to explore within-person facial variability. In previous work (Jenkins et al, 2011; Burton et al, 2011) we have argued that face recognition relies *both* on between-person and within-person differences. The broad history of research on face recognition has almost entirely focussed on discriminating between individuals – and this is true both for theories of human perception and for computational approaches. In most cases, this reduces the problem to discriminating between specific images of individual people. This emphasis, we have argued, has impeded progress in the field (Burton, 2011). Differences between people must be interpreted in the context of differences within people.

By applying PCA, a standard computational tool in face recognition, we have provided a way of operationalizing study of within-person differences, and this appears to be a promising start in understanding a number of difficult problems in the field. The key to this understanding lies in the proposal that learning a new face (becoming familiar) involves not just repeated exposure to the same stimulus, but incorporating many superficially different stimuli into a common representation. In so doing, one is able to move from a simple image-dependent recognition strategy to a

more sophisticated, abstractive recognition strategy that generalises to novel instances of the person.

An important difference between this and previous approaches is that faces are not represented as points in some face-space (Leopold et al, 2001; O'Toole, 2011; Rhodes, 1996). There is no idealised, Platonic Form of Barak Obama's face that is the 'true' value of that face on some set of dimensions. More sophisticated proposals which keep separate the different aspects of a face image (face shape, texture, illumination, view etc) go some way to conceptualising faces as 'regions' rather than points (e.g. Lando & Edelman, 1995). However, even these tend to assume independence between variations in the world (lighting etc), and variations in the photographic subject (expressions etc). We propose that it is not merely that different faces load differently onto a set of common dimensions; nor even that a smile dimension codes different smiles for different people. Rather, the very dimensionality of representational space is different for different people – within-person variability is idiosyncratic.

This observation provides an explanation for some difficult problems in face recognition, specifically those surrounding familiar/unfamiliar differences. For example, when matching two images of a known person, the task can be reduced to whether each image lies within the region occupied by the person. When matching two images of an unknown person, one has no knowledge of how that face varies – and so one cannot make appeal to within-region mapping. Instead, one must rely on a strategy that is more image-bound, making direct comparison between specific aspects of each image (Hancock et al, 2000).

The work presented here is just a start in trying to understand within-person variability. We have no particular commitment to PCA, and there are many alternatives available which may provide a better account (see Wija, Uchimura & Zhencheung, 2009). We have chosen the technique simply because it is very common in the face recognition literature, and has been used in cognitive as well as perceptual models (Burton, Bruce & Hancock, 1999; Nestor, Plaut & Behrmann, 2011). Since it has been popular in trying to understand the problem of telling people apart, it seems a promising place to start in understanding telling people together. However, PCA

has well-known limitations when used for recognition (see Zhao et al, 2003, for a review including a comparison of this technique with others). For example, in order to recognise an individual, PCA-based systems need to compute similarity of a target face to all those in the database – possibly too inefficient an approach to be useful. Furthermore, in order for the visual system to have the possibility of using a within-person approach, it must necessarily have already succeeded in recognising an initial set of instances of an individual. While context provides a large degree of top-down constraint over seen variations of an individual (consider the different views one has during a conversation), there is some circularity in the notion of using already-recognised images to build representations for recognition. For these reasons, we have limited ourselves above to a consideration of within-person variability only, with no implied commitment to a PCA-based *recognition* system computing within and between person variability in the same way. Despite the early stages of research in this field, there are already some clear predictions emerging from the work presented here.

First, the nature of one's exposure to a face should have a clear and predictable effect on subsequent recognition. For example, cinemagoers have very wide exposure to Tom Cruise, having seen many different images of him. But this exposure is still limited. A member of his family will have an even wider range of visual exposures, having seen him in different states of health, at different ages, and so on. The generalisation of one's representation is clearly based on the statistical properties of one's exposure. This should be a testable prediction: exposure to a face over a range of ages should improve recognition of that face across changes in age, but not across changes in health (and vice versa). Second, the efficiency with which one learns a face should be directly related to the variability of the exposure, and not, for example, the number of encounters, or time spent encoding a new face. Once again, this is straightforwardly testable: twenty diverse images of a face should result in more generalizable learning than twenty similar images of that face. Third, the account allows one to incorporate different levels of familiarity. It is clear that in daily life we have differing levels of familiarity to faces. However, perceptual research typically makes only a binary familiar/unfamiliar distinction. This has made research in face learning rather difficult, and in many cases, the measure of familiarity lacks sensitivity (Clutterbuck & Johnston, 2002). Here, the proposal linking the statistics of

one's exposure to the robustness of one's representation reflects more directly our everyday experience, and leads once again to testable predictions: the quantity and quality of image variability should have graded effects on recognition performance.

Finally, we consider how this work might inform the larger field of face perception, incorporating representations of familiar and unfamiliar faces, and the many different types of information available to the viewer (Bruce & Young, 1986). We have shown that performing PCA on individual faces can produce useful representations of each individual's variability. In turn, by assessing how well that representation, or region of space, incorporates new instances of a face, we have gone some way towards establishing how we might conceptualise the classification of novel images. To develop this account further, we need to consider how these representations might be related to other conceptual issues in face recognition. For example, we have claimed that representations of unfamiliar faces are more image-bound than those for familiar faces, but this cannot be the whole story. While we have emphasised the idiosyncratic nature of facial variation, it is clear that we can interpret an expression or read the facial speech even from an unfamiliar face.

Our general hypothesis is that the relation between different facial variables is an empirical one, properly studied by statistical analysis of the range of faces we typically encounter. Attempts to relate human perception of different types of facial information (for example identity and expression, see Young & Bruce, 2013) typically rely on systematic manipulation of stimuli to vary these dimensions only, eliminating apparently spurious noise. However, we have argued that such a systematic approach may 'control away' important aspects of the problem. Modern computer-based techniques for face recognition exploit covariation, derived from more naturally occurring images of faces than traditionally studied in psychology (e.g. see Beveridge, Givens, Phillips & Draper, 2009; Phillips & O'Toole, 2014). While these approaches have not been incorporated into psychological models of face perception, it seems clear that there would be benefit in doing so.

In sum, we have presented a technique for studying within-person variability, an aspect of face recognition that is both little-studied, and, we argue, very important. We have presented a number of observations derived from this technique, using

famous actors as examples. We have also drawn out some implications from this study, and listed some testable predictions. We hope to have convinced readers that a critical part of our understanding of face recognition has been largely ignored, and to have made a start in addressing this problem.

Acknowledgements

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

Attributions for Figure 1 (top row left to right, bottom row left to right): The Department for Culture, Media and Sport [CC-BY-2.0]; Zasitu (Own work) [CC-BY-SA-4.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; Richard J. Cole [CC-BY-SA-3.0]; Umakanth Jaffna (Own work) [CC-BY-SA-4.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; Willwal.Willwal at en.wikipedia [GFDL]; Russell Watkins/Department for International Development [CC-BY-SA-2.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; The Department for Culture, Media and Sport [CC-BY-2.0]; English Foreign and Commonwealth Office [Open Government Licence v1.0]; Xtrememachineuk at en.wikipedia [Public Domain].

References

- Aishwarya, P., & Marcus, K. (2010). Face recognition using multiple eigenface subspaces. *Journal of Engineering and Technology Research*, 2, 139-143.
- Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 711-720.
- Beveridge, J. R., Draper, B. A., Chang, J. M., Kirby, M., Kley, H., & Peterson, C. (2009). Principal angles separate subject illumination spaces in YDB and CMU-PIE. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31, 351-363.
- Beveridge, J. R., Givens, G. H., Phillips, P. J., & Draper, B. A. (2009). Factors that influence algorithm performance in the Face Recognition Grand Challenge. *Computer Vision and Image Understanding*, 113(6), 750-762.
- Beymer, D. (1995). Vectorizing face images by interleaving shape and texture computations. MIT AI Lab memo 1537. Massachusetts Institute of Technology, Cambridge, MA.
- Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, 10(5), 527-536.
- Bruce, V. (1986). Influences of familiarity on the processing of faces. *Perception*, 15(4), 387-397.
- 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. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305-327.
- Burton, A. M. (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., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, 23(1), 1-31.

- Burton, A. M., Miller, P., Bruce, V., Hancock, P. J. B., & Henderson, Z. (2001). Human and automatic face recognition: a comparison across image formats. *Vision Research*, *41*(24), 3185–3195.
- 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.
- Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943–958.
- Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception*, *31*, 985–994.
- 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
- Edelman, S., & O'Toole, A. J. (2001). Viewpoint generalization in face recognition: The role of category-specific processes. In (M. Wenger & J. Townsend, Eds.) *Computational, geometric, and process perspectives on facial cognition*. Lawrence Erlbaum, 397-428.
- 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.
- Geng, X., & Li, M. (2007, December). Individual discriminative subspace for face recognition under uncontrolled conditions. In M. J. Cree (Ed.), *Proceedings of Image and Vision Computing*. Paper presented at Image and Vision Computing, University of Waikato, Hamilton, 5-7 December (pp. 13-18). Hamilton, New Zealand.
- Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337.
- 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.
- Jenkins, R., Burton, A. M., & White, D. (2006). Face Recognition from Unconstrained Images: Progress with Prototypes. *Proc. 7th IEEE International Conference on Automatic Face and Gesture Recognition*. pp. 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.
- 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.
- Lando, M. & Edelman, S. (1995). Receptive field spaces and class-based generalization from a single view in face recognition. *Network: Computation in Neural Systems*, *6*, 551-576.
- 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.
- 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. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364–372.
- Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences*, *108*(24), 9998–10003.
- O'Toole, A.J. (2011). Cognitive and computational approaches to face recognition. In A.J. Calder, G. Rhodes, M.H. Johnson & J.V. Haxby (Eds.) *The Oxford Handbook of Face Perception*, pp:15-30. Oxford: OUP.
- 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.
- O'Toole, A. J., & Edelman, S. (1996). Face distinctiveness in recognition across viewpoint: An analysis of the statistical structure of face spaces. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996.*, (pp. 10-15). IEEE.

- O'Toole, A. J., Edelman, S., & Bühlhoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*, 2351-2363.
- Patel, V.M., Wu, T., Biswas, S., Phillips, P.J., & Chellappa, R. (2012). Dictionary-based face recognition under variable lighting and pose. *IEEE Transactions on Information Forensics and Security*, *7*, 954-965.
- Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, *32*(1), 74–85.
- Rhodes, G. (1996). *Superportraits: Caricatures and Recognition*. Hove: Psychology Press.
- Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research*, *49*(23), 2856–2861.
- Sandford, A. & Burton, A.M. (in press). Tolerance for distorted faces: Challenges to a configural processing account of familiar face recognition. *Cognition*.
- Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE*, *94*(11), 1948–1962.
- Sandford, A. & Burton, A.M. (2014). Tolerance for distorted faces: Challenges to a configural processing account of familiar face recognition. *Cognition*, *132*, 262-268.
- Shan, S., Gao, W., & Zhao, D. (2003). Face recognition based on face-specific subspace. *International Journal of Imaging Systems and Technology*, *13*, 23-32.
- Sutherland, C.A., Oldmeadow, J.A., Santos, I.M., Towler, J., Burt, M.D., & Young, A.W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105-118.
- Tanaka, J.W. & Gordon, I. (2011). Features, configuration and holistic face processing. In A.J. Calder, G. Rhodes, M.H. Johnson & J.V. Haxby (Eds.) *The Oxford Handbook of Face Perception*, pp:15-30. Oxford: OUP.
- Turk, M. A., & Pentland, A. (1991). Eigenfaces for recognition. *Journal Of Cognitive Neuroscience*, *3*(1), 71–86.

- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology Section A*, 43(2), 161–204.
- Vetter, T., & Troje, N. (1995). Separation of texture and two-dimensional shape in images of human faces. In *Mustererkennung 1995* (pp. 118-125). Springer Berlin Heidelberg.
- White, D., Burton, A.M., Jenkins, R. & Kemp, R.I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, 20, 166-173.
- Wija, I.G.P.S., Uchimura, K. & Zhencheng, H. (2009). Why the alternative to PCA provides better performance for face recognition. IEE: *WIAMIS 2009*. 149-152.
- Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, 102(4), 959–974.
- Zhao, W. Chellappa, R., Phillips, P.J., & Rosenfield, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), 399-458.

Figure 1 Different images of the same face, all identifiable to a familiar viewer (see Acknowledgements for attributions)



Figure 2. Visualisation tool. This figure shows an image of Tom Cruise (left window) and its reconstruction (right window). The sliders on the left and right show the reconstruction values for this image for each of the 30 texture components (left) and shape components (right).

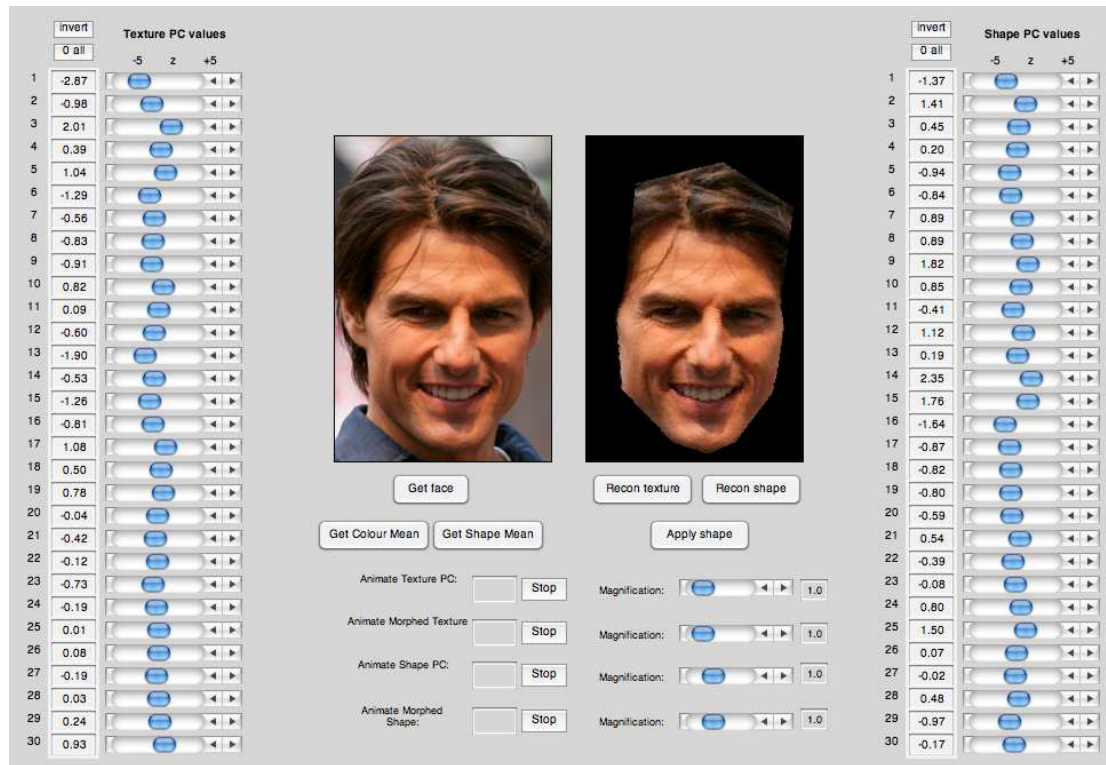


Figure 3. Variance captured by the first component of shape and the second component of texture for one of the celebrities. The contribution of these components is illustrated by adding a low and high value (± 2 SDs) to the person's average.

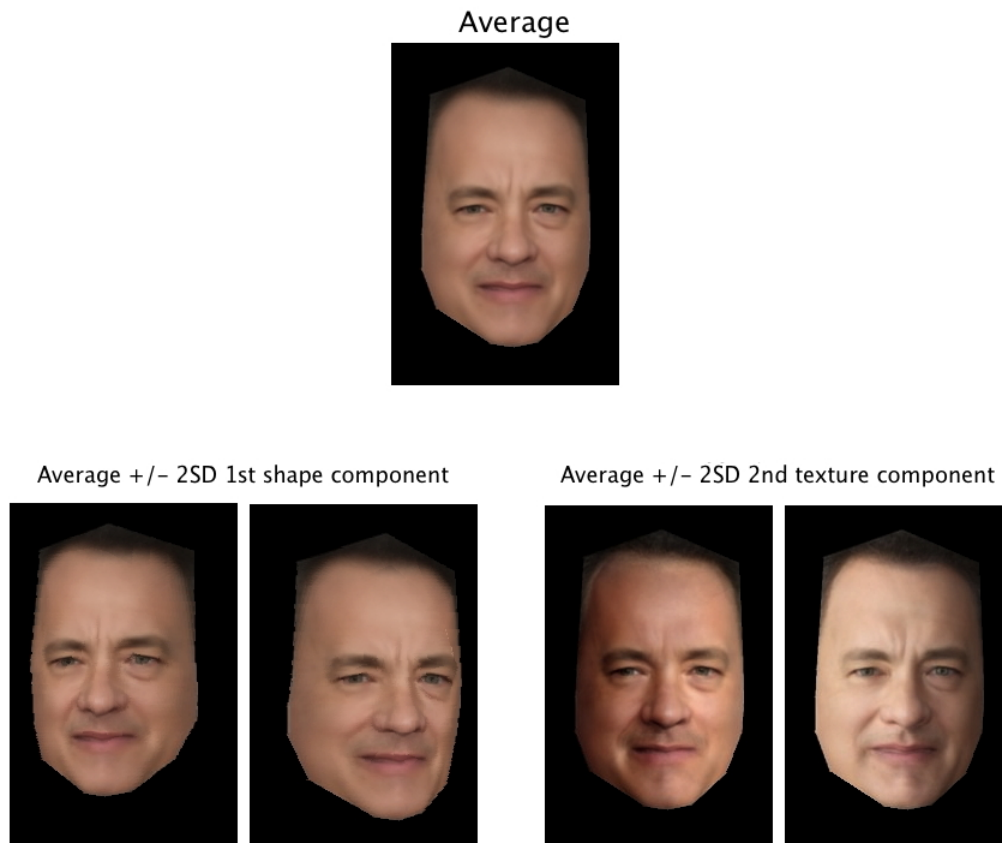


Figure 4. A reconstructed image of Tom Cruise (left) and the effects of manipulating values on 1st, 2nd and 6th shape eigenvectors, while holding all other values constant. Earlier components show rigid motion while later dimensions introduce idiosyncratic non-rigid motion.

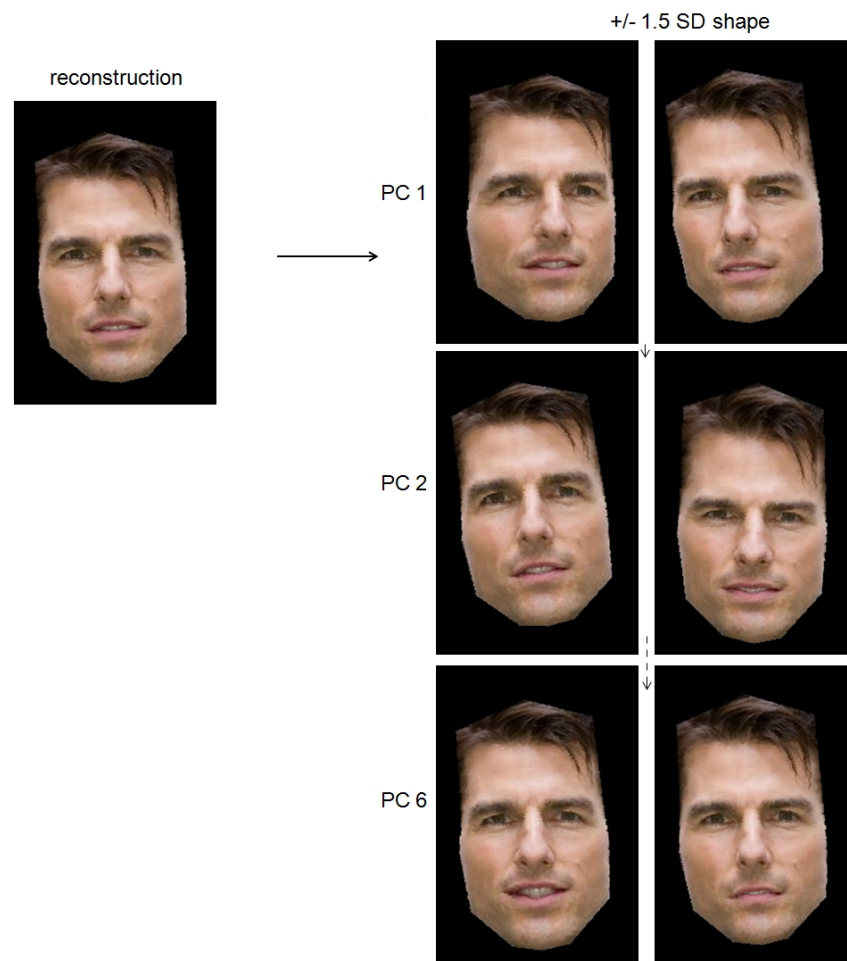


Figure 5. Reconstructed images using ± 1.5 SDs on the sixth shape component. **Top:** For Kiera Knightley, this component shows a clear left-right eye movement, combined with a slight opening of the mouth. **Bottom:** For Gwyneth Paltrow, this same component shows a mouth opening, combined with a slight increase in distance to camera.



Figure 6. Reconstructed images using ± 2 SDs on the seventh texture component for Kiera Knightly. The component appears to correspond to application of make-up, one source of variation in images of this celebrity.



Figure 7: PCA on three non-overlapping sets of Tom Cruise images (one set per row, 30 images per set). Images show the result of adding the set average to ± 1.5 SDs of eigenvector 6 in each analysis.

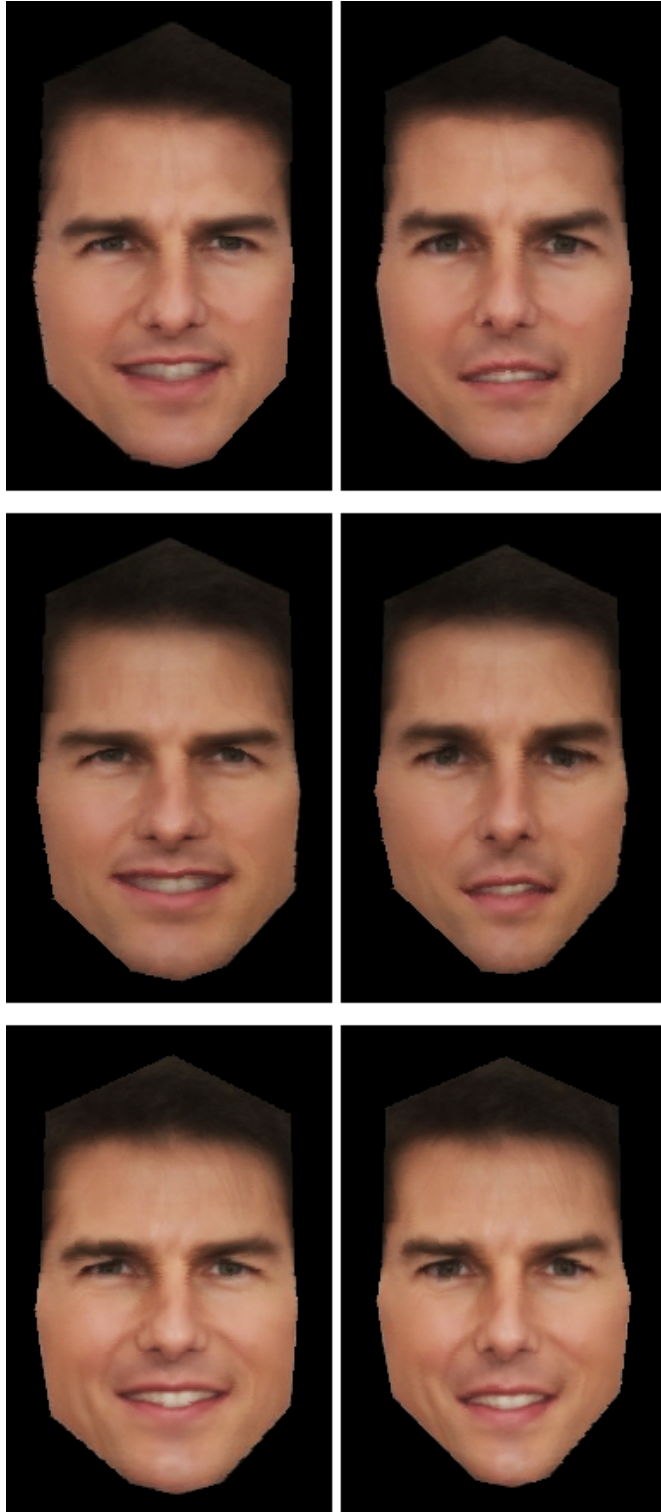
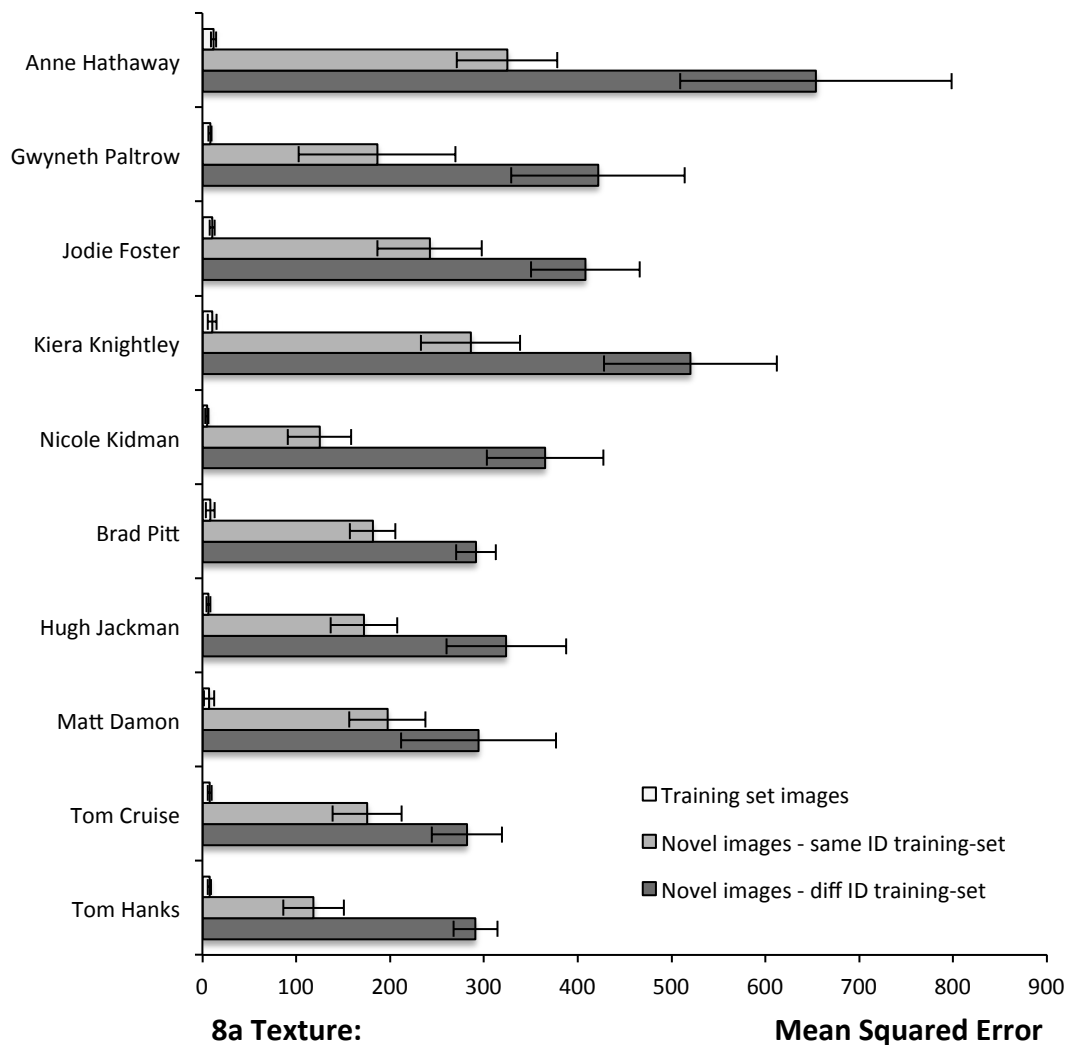


Figure 8 Reconstruction errors (MSE) for images of each of the ten actors. 8a reconstruction of texture; 8b reconstruction of shape. Figures show mean MSE over five runs with different images contributing to training and test sets. ‘Different-ID’ training sets are the average of reconstructions using each of other same-gender actors.



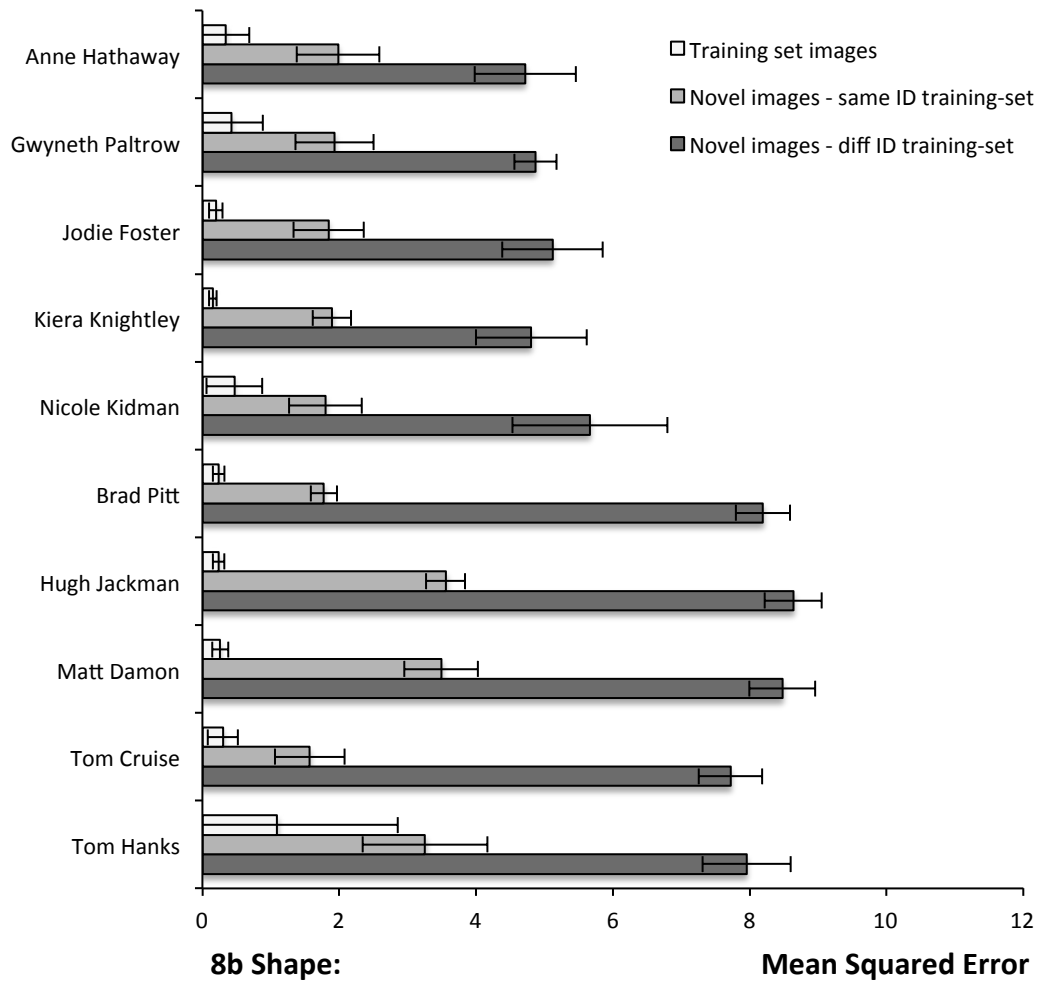


Figure 9: Reconstruction error for one actor using varying numbers of components.

9a reconstruction of texture; 9b reconstruction of shape

