Article:
Kramer, Robin Stewart Samuel, Young, Andrew William orcid.org/0000-0002-1202-6297, Day, Matthew et al. (1 more author) (2017) Robust social categorization emerges from learning the identities of very few faces. Psychological Review. pp. 115-129. ISSN: 0033-295X

https://doi.org/10.1037/rev0000048

Robust social categorization emerges from learning the identities of very few faces

Robin S. S. Kramer, Andrew W. Young, Matthew G. Day and A. Mike Burton

University of York, UK

Correspondence concerning this article should be addressed to

A. Mike Burton, Department of Psychology, University of York, York YO10 5DD, UK.

Email: mike.burton@york.ac.uk

Abstract

Viewers are highly accurate at recognizing sex and race from faces – though it remains unclear how this is achieved. Recognition of familiar faces is also highly accurate across a very large range of viewing conditions, despite the difficulty of the problem. Here we show that computation of sex and race can emerge incidentally from a system designed to compute identity. We emphasise the role of multiple encounters with a small number of people, which we take to underlie human face learning. We use highly variable everyday 'ambient' images of a few people to train a Linear Discriminant Analysis (LDA) model on identity. The resulting model has human-like properties, including a facility to cohere previously unseen ambient images of familiar (trained) people – an ability which breaks down for the faces of unknown (untrained) people. The first dimension created by the identity-trained LDA classifies both familiar and unfamiliar faces by sex, and the second dimension classifies faces by race – even though neither of these categories was explicitly coded at learning. By varying the numbers and types of face identities on which a further series of LDA models were trained, we show that this incidental learning of sex and race reflects covariation between these social categories and face identity, and that a remarkably small number of identities need be learnt before such incidental dimensions emerge. The task of learning to recognise familiar faces is sufficient to create certain salient social categories.

Keywords: Face Recognition, Social Categorization, Face Learning

Disclosures and Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.323262, and from the Economic and Social Research Council, UK [ES/J022950/1] to A. Mike Burton.

All authors contributed to work described and to the written report. All authors have read and approved the manuscript.

No authors have any conflicts of interest regarding the work in this paper.

No additional people were involved in the work described here.

Introduction

In one of the earliest studies of face recognition and eye witness testimony, Feingold (1914) noted that witness descriptions nearly always began with the sex and race of the perpetrator. Although race has since been largely discredited as a biological concept and the relation between biological sex and gender has also been recognised as complex, a century of subsequent research has confirmed the salience of sex and race as social categories for human perceivers, and the ease with which we can classify faces along these lines (Bruce & Young, 2012).

It is well-established that adults are remarkably good at telling whether faces are those of men or women, achieving very high accuracy in classifying a face they have never seen before as male or female from a single photograph (e.g., Bruce et al, 1993; Ellis Young & Flude, 1990; Martin & Macrae, 2007). Yet almost nothing is understood about where this ability comes from. Identifying how male and female faces differ has not proved straightforward in research studies. The physical features that can be used to determine sex are known to include differences in 3D shape, relative brightness and colouration of critical face regions, but the differences are relatively subtle, with overlapping ranges of variability for any specific cue (Brown & Perrett, 1993; Burton, Bruce & Dench, 1993; Campanella et al, 2001; Hoss, Ramsey, Griffin & Langlois, 2005; O'Toole et al, 1998; Russell, 2003). The cues that underlie perception of race and ethnicity include clear differences in skin colouration and facial feature shapes, so might at first seem easier to understand. Yet there is still debate about exactly how feature shapes and skin pigmentation contribute to the perception of race and ethnic background (Blair, Judd, Sadler & Jenkins, 2002; Brook & Gwinn, 2010).

A natural intuition is that learning the difference between male and female faces must involve a great deal of practice, perhaps driven by evolved mechanisms for sexual selection and assisted by more salient cues from body shape, voice, and in many cultures clothing. Indeed, the claim that we are face 'experts' is often made in the research literature, and for many people expertise will carry connotations of studying and learning. Similarly, one might think that learning about race is driven by social mechanisms concerning group membership, and such factors feature prominently in some theories of other-race effects in face recognition (Hugenberg, Miller & Claypool. 2007; Sporer, Trinkl & Guberova, 2007).

Such intuitions are consistent with a general approach to face processing based on the statistical properties of facial images (e.g. Caldara & Abdi, 2006; Cheng, O'Toole & Abdi, 20001; Haberman & Whitney, 2007; Hancock, Bruce & Burton, 1998; O'Toole, Abdi, Deffenbecher & Valentin, 1993). Although there are many variants in the specific statistical mechanisms underlying these approaches, they share the hypothesis that core information about social categories (here sex and race) can be derived from multiple face images. For example, it seems natural to hypothesise that we become expert at deriving sex from faces because we have seen very many men and women over the course of our lives. Similarly, sensitivity to race can be understood in terms of one's relative exposure to faces of different ethnicities. Consistent with this approach, these studies demonstrate how these dimensions can be derived from a large sample of face images, normally comprising a single image of each of a number of people photographed under standard conditions to minimise irrelevant image differences in lighting, pose and expression.

Here we take a different approach. We start from the observation that our everyday exposure to faces does not typically conform to the 'one image of many faces' pattern that is commonly used in statistical learning. Instead, we normally see very many 'instances' of rather few people. Moreover, we encounter familiar people in different circumstances that create huge image variability from changes in lighting, pose and expression; this variability makes natural face learning a significant perceptual challenge (Burton, Kramer, Ritchie & Jenkins, 2016). There are caveats, of course – it is certainly true that for the past couple of decades modern media have allowed us to view hundreds of different faces per day, each for rather a short time. However, over much of our evolutionary history, and for most infants, the pattern of exposure favours multiple viewings of few people – mainly limited to immediate family. Despite this, infants become sensitive to the sex and race of faces very early in life (e.g. Kelly et al, 2007; Quinn et al, 2002). Early sensitivity to these social categories seems rather puzzling from a conventional statistical learning perspective as young infants are unlikely to have encountered a large range of faces and will have little understanding of social roles.

In recent work, we have focused on the learning of *identity* from ambient images, such as those shown in Figures 1 and 2. The concept of ambient images was introduced by Burton, Jenkins, & Schweinberger (2011) to refer to naturalistic photographs of the type we see in our everyday lives; these vary in a wide range of ways that make them unlike the standardised images used in most psychological studies. Studies using ambient images offer an important perspective through allowing a particular focus on how face familiarity emerges as people become able to classify together very different examples of the same face (Burton, Kramer, Ritchie & Jenkins, 2016; Jenkins & Burton, 2011; Jenkins, White, van Montfort & Burton, 2011). This problem of 'telling faces together' (Andrews, Jenkins, Cursiter & Burton, 2015) is a significant component of familiar face recognition, which needs to be solved alongside the more commonly studied problem of 'telling faces apart' (Burton, 2013). In this paper, we ask how a perceptual system designed to *identify* individual faces from highly variable images might lead to the emergence of social categorisation without explicit training. In particular, could multiple exposures to rather few individuals give rise to robust categorisation of sex and race, which will generalise beyond these few individuals and into the world at large?

This question is very different from the traditional approach to understanding the relation between identity, sex and race. We are not concerned here with trying to establish, for example, whether perception of sex and race are independent from the recognition of identity in the human visual system as proposed in some models (e.g. Bruce and Young, 1986; though see the Discussion section, below, for further discussion of this issue). Instead, we are focussing on how sex and race perception, which are known to be highly accurate, can arise from a system built for other purposes.

FIGURE 1 HERE PLEASE

Our starting point was an aim to achieve a model of key properties of familiar face recognition, based on realistic patterns of experience. Recognising the identities of familiar faces requires being able to extract some form of invariance across the highly variable conditions in which faces are encountered in everyday life; differences in lighting, viewpoint, expression, hairstyle and so on. This variability is, of course, one of the tricky problems that is sidestepped by the traditional approach of using single, standardised images of multiple identities. One way to simulate this variability is to make use of ambient images of faces (Jenkins et al, 2011; Kramer, Ritchie & Burton, 2015; Sutherland et al, 2013), such as those shown in Figure 1; although these are all photos of the same person, the substantial differences between the images mean that this common identity is only obvious to someone familiar with that individual.

Understanding how we are able to recognise familiar face identities across such huge variability has been one of the key challenges to understanding face recognition (Bruce & Young 1986; Burton, Jenkins, Hancock & White, 2005; Jenkins et al, 2011). Importantly, although ambient images of familiar faces are usually recognised with ease, perception and recognition of the same images by viewers for whom the faces are unfamiliar is remarkably error-prone (Bruce et al, 1999; Megreya & Burton, 2006, 2008). For example, Jenkins et al (2011) created a deceptively simple but very informative sorting task involving 20 ambient images of each of two faces, such as those shown in Figure 2. Forty images were presented to participants who were invited to sort them into piles corresponding to different identities. As would be expected, participants who knew these faces created two piles, but participants to whom the faces were unfamiliar created an average of 9 piles (median 7.5, mode 9). The experiment provides a strong demonstration of the power of familiarity, and emphasises that face recognition depends on cohering superficially different representations (telling faces together) as well as discrimination between people.

FIGURE 2 HERE PLEASE

The striking differences between familiar and unfamiliar faces imply there is something different about the way in which familiar faces are represented. The simplest way to think of this is that we must become adept at recognising the defining characteristics of the face of each person we know. To simulate this, we therefore used Linear Discriminant Analysis (LDA) to classify multiple images of a set of people into their different underlying identities. LDA is a technique which is popular in engineering approaches to face classification (Etemad & Chellappa, 1997; Martinez & Zhu, 2005), and is used to establish the features which best describe a class of stimuli, while simultaneously discriminating it from members of another class. When used specifically to classify faces, the technique is sometimes called the 'Fisherface' approach (Belhumeur, Hespanha & Kriegman, 1997) because the discriminant function used is due to R.A. Fisher (1936).

In typical use, an LDA approach to face classification will involve different images of the same person which vary only rather slightly, for example multiple images taken from the same event, and with the same camera. Here we instead use an ambient image approach, in which we exercise no control over superficial aspects of the face pictures submitted to analysis. Having derived optimal discrimination by identity (i.e. finding the physical image-space in which different pictures of the same person cluster together) we then submit the model to tests of face processing with novel images of familiar (i.e. already learnt from other images) and unfamiliar faces. Using this strong test based on entirely untrained ambient images we observe high levels of classification performance with familiar, but not unfamiliar faces – corresponding to patterns that characterise human face perception. However, more strikingly, training with a very small number of different identities results in LDA dimensions that code sex and race (respectively) on the first two discriminating factors. These two dimensions arise as *emergent* properties of a system trained to classify identity for multiple images of just a few people.

Before describing the procedure in detail, it is important to be clear about what we are claiming on the basis of the specific technical implementation of our approach. We have (of course) no commitment to LDA as a model of human perception – it clearly lacks any reasonable way of capturing the multiple complexities required of a biological vision system – as do many other mathematical approaches to exploring patterns in stimulus sets. We have chosen LDA, specifically, because it is a very direct way of capturing covariance structures *in the stimuli*. The technique is an example of a large class of clustering methods used in engineering approaches to image classification (for a review see Jain, Duin & Mao, 2000). We have chosen it here because it is able to demonstrate, in a simple and replicable way, that clustering highly variable ambient images based on one property of human faces (identity) delivers a covariance structure that is potentially extremely useful for human viewers. The details of how our brains actually do this form a separate question; what we show here is that covariation centred on face identity provides critical information for the brain to exploit.

Method

Image sets

In order to model real-world exposure to faces, we collected ambient images for our analyses. These were similar in nature to the 'Labeled Faces in the Wild' database (Huang, Ramesh, Berg, & Learned-Miller, 2007), which attempts to incorporate natural variability across numerous dimensions, including pose, lighting, expression, age, and camera conditions. For practical reasons, though, we restricted the sample to reasonably high-resolution images where no part of the face was obscured (by clothing, glasses, hands, etc.). To facilitate the placement of landmark fiducial points on each image, we also limited our image poses to within approximately ±30° from full face. Figure 1 illustrates the range of images used.

Based on these criteria, we collected a main set of 30 colour images for each of 20 identities using Google Images and entering names as search terms. These identities comprised five White women, five White men, five Black women, and five Black men. We used Hollywood actors, so that 30 images would be readily available for each identity. Images were cropped to include only the head, scaled to 190 pixels wide x 285 pixels high, and represented in RGB colour space using a lossless image format (bitmap). An additional 30 images of each of four more identities (two White men, two Black men) were used for the specific purpose of modelling familiar and unfamiliar identities, as described below.

To investigate the generalizability of our models, we also needed new images of faces that never appeared during training. We therefore collected an additional set of 200 images; 50 White women, 50 White men, 50 Black women, and 50 Black men. These were also found using Google Images and employing search terms such as "White women", "White men", etc. There was one image for each identity, and there was no overlap between these new identities and our original Hollywood actors. These 200 images were also ambient images meeting the selection criteria outlined above.

General procedure

With image classification, it is common to have fewer sample vectors (images) than features (pixels). In such cases, LDA cannot be carried out without first reducing the number of feature dimensions. This can be done in a number of ways, including morphological analysis of faces leading to a reduced-dimensional description (e.g. Chen et al, 2000). A more popular approach is first to subject the faces to Principal Components Analysis (PCA) resulting in a low-dimensional description of 'eigenfaces' (e.g., Bekios-Calfa, Buenaposada & Bamela, 2011). In our studies we adopted this approach, as follows.

All images were shape-standardised by morphing them to a template derived from the average shape of the set (Burton, Miller, Bruce, Hancock & Henderson, 2001; Craw, 1995). Shape-morphing relied on the alignment of 82 fiducial points for each image (e.g. corners of eyes, corners of mouth etc – for technical details see Burton et al, 2015). Assignment of these points was carried out using a standard semi-automatic process in which only five landmarks were identified manually (top of the head, pupils, and sides of nostrils Finding the remaining points involved first applying an iterative tree-based regression model (Kazemi & Sullivan, 2014) to locate 68 points according to the schema used by the CMU Multi-PIE Face Database (Gross, Matthews, Cohn, Kanade, & Baker, 2010). This model had previously been trained over a large independent set of manually landmarked images. The final 82 points were then determined from the set of 68+5 points via Procrustes analysis.

PCA was computed on these shape-normalised images. In order to reduce the number of dimensions describing the resulting subspace without significant loss of the variability, we retained those highest components that explained 95% of the variance. This resulted in our use of only the first 88 or 89 principal components in all the models that follow. These principal components were then entered into an LDA, where each class represented an identity. The result is a subspace comprising a maximum of c-1 dimensions, where c is the number of identities.

Our basic face identity training model included all 20 Hollywood actors (half men, half white) described above. The first 20 images of each identity were used for our 'training set'. That is to say, only 400 images were used in the PCA+LDA process in order to produce a face subspace that could distinguish the 20 trained identities. The remaining 200 images (10 of each identity) could then be used as novel instances of 'known' identities, allowing us to test the generalizability of the model. This basic training method was used to establish key properties of the LDA model, while minor variations were used to answer some specific questions, as noted below.

Simulations

As described above, we are particularly concerned here with capturing face learning when the number of identities is relatively small, but the number of 'encounters' with each person is relatively large, and these encounters represent a large range of within-person variability. In the following simulations we therefore start by demonstrating that an LDA model trained on identity does indeed capture key aspects of human face recognition: novel pictures of known people are *both* easier to tell apart, and easier to 'tell together' than pictures of unknown people. We then go on to show that this training has interesting emergent properties, delivering highly accurate and generalizable discrimination of sex and race, despite no explicit training on these dimensions.

1. Identity recognition

We first investigated how well the LDA subspace created by the basic model classified the identity of images. After training our model on 400 images (20 identities x 20 images), we calculated the centroid for each identity. Each training image was then classified using a 'nearest centroid' approach (Euclidean distance). Our results showed that 100% of the training images were correctly classified, demonstrating that the technique had delivered optimal separation between training identities. For the purposes of the following simulations, these trained identities represent *familiar* people – each identity has been learned over a set of very variable images. The simulation can now be used to test either novel (i.e. untrained) pictures of these people, or to test pictures of *unfamiliar* people (i.e. the faces or previously unseen individuals).

To test generalisation to untrained images, we projected the 200 novel instances of our 20 identities (10 images of each) into the subspace. Using 'nearest centroid', 98% of these novel instances were correctly classified, despite the fact that none of these images had been used to derive the subspace. This represents very good performance, particularly for ambient images.

We also investigated model behaviour with images of new identities. We projected the set of 200 novel instances of the original training identities (as above) into the subspace, and also 200 images of new identities (one of each, described earlier). In order to allow the model to reject images, it is necessary to set a threshold neighbourhood round the identity centroids. This threshold can be used to generate signal detection measures for 'probe' faces as follows. *Hit:* known person falls within the threshold of the correct centroid; *False alarm:* unknown person falls within the threshold of a known person. Figure 3 illustrates the performance of the model (d') as threshold is varied. Peak performance is 2.89 here, showing that the LDA model has the potential, if tuned to the optimal threshold, successfully to reject images of new identities while also recognising novel instances of the original 20 identities.

FIGURE 3 HERE PLEASE

2. Clustering of familiar and unfamiliar identities

We have shown that an LDA model can correctly classify trained face identities and generalise to new photographs of the same individuals. This satisfactorily demonstrates that the model can 'tell people apart'. We also showed that the LDA model can correctly reject examples of unfamiliar faces - it does not misclassify them as one of the trained individuals. However, any candidate model of human recognition must also show effects of familiarity on 'telling people together' (as described in the Introduction). In short, human viewers are not only able to tell familiar from unfamiliar faces, but they are much better at clustering together images of familiar than unfamiliar people (Figure 2).

To investigate the effect of familiarity, ambient images of two new identities (both White men; 30 images of each) were collected using Google Images and our 'unconstrained' procedures, as detailed above. For our PCA+LDA training, we included 20 images each for our 20 identities as before. However, now with seven White men to choose from, we selected only five for inclusion in the training set, the two remaining serving as 'unknown' faces.

First, we consider how our model distinguishes between familiar men. With 20 images of each familiar White man appearing in the training set, we predict that all 30 images of each identity should cluster well even though 10 of these are novel instances that were not included during training. To test this hypothesis, we applied cluster analysis to the images of one pair of these five men. We used *k*-means clustering to best group the 60 images into 2, 3, 4, 5, and 6 clusters. The fit of these different solutions was quantified using both silhouette coefficients (Al-Zoubi & Rawi, 2008) and the Dunn index (Dunn, 1974). Both measures attempt to identify dense and well-separated clusters, reflected in higher values. Table 1 summarises the results averaged over three iterations of this procedure, where each iteration incorporated different pairs of familiar and unfamiliar identities from the set of seven White men.

TABLE 1 HERE PLEASE

Table 1 shows that when the 60 images are of two familiar men they best fit a 2-cluster model, with a mean silhouette coefficient approaching the maximum value of 1. In addition, models utilising more clusters than this demonstrate a large drop in suitability/fit. Importantly, when the 2-cluster model was applied, all 60 images were labelled with the correct identities.

Next, we consider the two unfamiliar men. We predict that these 60 images will show worse clustering into their two identities in comparison with familiar faces (20 of 30 images included during training). The results, summarised in Table 1, show that the two unfamiliar men also best fit a 2-cluster model. However, the measures of fit are far lower (the images do not cluster well) and the 2-cluster solution is not much better than solutions involving higher numbers of clusters. Therefore, the LDA-space is less able to discriminate between these two unfamiliar identities' images.

3. Emergence of sex and race

We now turn to the ability of the LDA model to classify the untrained characteristics of sex and race. To place these findings in context we will first consider the extent to which sex and race are already represented in the eigenvectors of the underlying PCA on our original training set of 400 images (20 images for each of 20 identities). There is precedent for this, for example O'Toole et al (1991) demonstrated the emergence of sex and race information from a simple autoassociative network trained on identity. However, this information was insufficient for perfect classification, and was spread across several dimensions of variability. None the less the work of O'Toole et al does point to the possibility that there is some diagnostic information available in these dimensions of variability, even prior to discriminant analysis.

Following O'Toole et al (1991), z-scores were computed for each training image along each PCA-derived dimension (eigenvector). The mean values for women vs. men and White vs. Black images were compared for each eigenvector using *t*-tests. Seven of the first ten eigenvectors provided a statistically reliable difference between the mean for men and women, though none of these produced very strong univariate separations, with 70% classification by sex being the maximum for a single dimension. Similar results were observed for classification by race, with five of the first ten eigenvectors providing significant differences between the means of Black and White faces and the best-performing dimensions providing 80% classification. These results are similar to those of previous research (O'Toole et al., 1991), whereby multiple eigenvectors contain information regarding both sex and race. To classify the sex and race of faces more accurately with PCA would clearly require weighting the contributions of multiple eigenvectors.

Analysis of the LDA space for the same images showed a much clearer and somewhat surprising pattern. Despite the fact that the analysis was trained only to cluster identity, it is clear that the first dimension derived by LDA (which explained 17.01% of the face identity discriminability) categorises people by sex (see Figure 4). Using the mean value of all cases on this first dimension as a threshold, 99.8% of the training images were correctly distinguished by sex. In addition, using the same criterion value, 99.0% of the novel instances were also correctly classified by sex.

FIGURE 4 HERE PLEASE

We also investigated how well this dimension might generalise to entirely new identities. We therefore projected the set of 200 new images of unfamiliar faces (one of each of 200 untrained identities, described above) on to this dimension, and Figure 5 illustrates where these new images fell. Again, using the criterion value calculated from the original training set, this dimension accurately classified 95.5% of the images of the new identities by sex.

FIGURE 5 HERE PLEASE

The second dimension of the LDA's subspace (which explained 13.55% of the face identity discriminability) appeared to categorise the identities by race (see Figure 6). As above, we quantified the accuracy of this race categorisation. We found that this second dimension classified 96.5% of the 400 training images correctly in terms of race. In addition, 95.5% of the 200 novel instances of these faces were also correctly classified by race.

FIGURE 6 HERE PLEASE

We also investigated how well this dimension might generalise to the set of new images of untrained faces. We therefore projected the 200 images (one of each identity, described above) onto this dimension, and Figure 7 illustrates where these new images fell. Again, using the same criterion value calculated from the original training set, this dimension accurately classified 91.0% of the images of the new identities by race.

FIGURE 7 HERE PLEASE

Our results suggest, therefore, that these first two dimensions, derived from our 20 training identities, represent sex and race. Impressively, both dimensions show high accuracy in classifying novel instances of our original identities *and* new images of unfamiliar identities. Interestingly, sex emerges first (i.e., with the highest level of discriminability) as a way to discriminate between identities, with race representing the second dimension. Finally, these LDA dimensions are clearly better suited to the task of categorisation in comparison with those dimensions produced by the original PCA.

4. Learning over racially homogenous sets

Our results so far demonstrate that sex and race are emergent properties of a model trained to discriminate only identity. When the training set contains clearly discriminable dimensions of sex and race (half men, half Black etc) the discrimination function extracts these. However, part of our motivation for this study is to ask what information can be derived from the small number of people one might encounter in early life. In many cultures, those faces would be racially homogenous.

In the next study we investigated the possible emergence of a sex dimension from images without racial variation. By comparison to the analyses above, this severely reduces (halves) the training set instances, and we are interested to establish whether sex nevertheless emerges as an early discriminator. We are also interested to establish just how generalizable any such dimension might be. For example, if sex is an emergent dimension from a categorisation of White faces, might this same dimension also discriminate sex across Black faces?

Figure 8 shows the first LDA dimension from a study in which we used only five White women and five White men (20 images of each) as our training set, resulting in 200 images for our PCA+LDA. As Figure 8 shows, both the training images and the novel instances of these identities are accurately classified by sex. Quantifying this accuracy, by calculating the criterion value as a threshold to separate the categories (as above), showed 100% of training and novel images were correctly classified.

FIGURE 8 HERE PLEASE

To further test the generalizability of this dimension, we projected the 200 new images of untrained identities (one of each identity) into this subspace (Figure 9). These included both untrained White and Black faces. We found that 91% of the new White identities were correctly classified by sex (using the criterion value calculated above), and 93% of new Black identities were correctly classified. Impressively, then, not only did this ‘sex dimension’ generalise to new images of the same race as the training set, but also to new images of a different race.

FIGURE 9 HERE PLEASE

This level of discrimination is quite impressive. Given that the LDA is not trained to make sex discriminations, the first emergent dimension nevertheless does so, having been exposed to only ten identities, five of each sex. Furthermore, there is good generalisation to other-race faces, with no previous exposure to these at all.

We repeated this analysis, training only on Black faces, and found very similar results. Full details are available in the supplementary material, but to avoid repetition here we simply report the summary data. When the LDA is trained to distinguish the 20 Black faces, the first dimension once again codes sex. Once again, 100% of training faces are correctly distinguished for sex, and 100% of the novel pictures of these people are also categorised correctly. For novel identities 72% of new White identities were correctly classified by sex, and 94% of new Black identities were correctly classified. This is an interesting result, perhaps suggesting differences between the sex dimensions produced in order to categorise White versus Black identities. While the former dimension appears to generalise well to Black faces, the latter dimension seems to show some specialisation for the race from which it was derived.

The main finding in this section is that training the model on the *identity* of a small racially homogenous set of faces delivers very highly accurate discrimination of sex – which is an almost perfect classifier for novel identities, particularly those of the same race as the learning set. Why should this occur? One way to think about the actions of the discriminant function is that it seeks dimensions which covary highly with the trained classification. For most faces, sex and identity are completely confounded – one's gender is constant, and presumably signalled very clearly, even across the wide range captured by ambient images. If this is a consistent cue which separates some identities from some others, then we would expect a mechanical discriminator to use it – and our data are consistent with this notion. This shows how a fundamental perceptual dimension can emerge from the statistical structure of the input to a system designed to compute identity.

5. Learning to classify sex: how many people is enough?

Having demonstrated that covariance with identity is a strong candidate for explaining the incidental learning of sex and race when classifying familiar face identities (Section 3 above), we now consider how many familiar identities are needed for categorical sex and race dimensions to emerge. We have already shown that five identities in each group are sufficient to produce highly accurate dimensions for categorisation. We next investigated whether such dimensions might emerge when fewer identities are learned.

With a training set of five White men and five White women, with 20 images of each, the first dimension in the LDA's subspace was able perfectly to categorise sex (see Figure 8). Simply, there is a threshold value that will lead to the correct classification of all images as belonging to two separate categories (men and women). The effect of reducing the number of identities in the training set can be seen in Table 2, for trained and novel instances, and for completely new faces. Table 2 summarises the results averaged over three iterations of this procedure, where each iteration incorporated randomly selected identities.

TABLE 2 HERE PLEASE

As Table 2 illustrates, even when fewer identities are included in the training set, the first dimension of the LDA is able to correctly classify both training images and novel instances of the same faces according to sex. However, our results suggest that the generalizability of this categorisation to new faces suffers with fewer training identities. Even so, after being trained on only a single man and woman, the performance is still above-chance for new faces of the same race (67%). Interestingly, if we purposely train the model with a very unequal number of men and women, the emerging sex classifier remains highly accurate – exposure to a very small number of people (overt multiple instances) is enough to elicit a sex classifier, even when the sexes are profoundly unbalanced in the input stimuli (i.e. 5 of one sex and one of another).

For race categorisation, we noted that using a training set of five White women and five Black women, with 20 images of each, the first dimension in the LDA's subspace was able to categorise race (see Figures S1 and S2, supplementary material). The results of reducing the number of identities in the training set can be seen in Table 3. In addition, we can determine how well this 'race classifier' dimension is able to perform with novel instances of these familiar identities, and also new identities. Table 3 summarises the results averaged over three iterations of this procedure, where each iteration incorporated randomly selected identities.

TABLE 3 HERE PLEASE

As Table 3 illustrates, even when fewer identities are included in the training set, the first dimension of the LDA is able to correctly classify both training images and novel instances according to race. Again, our results suggest that the generalizability of this categorisation suffers with fewer training identities and that if we purposely train the model with a very unequal number of White and Black women, the emerging race classifier is less accurate (see particularly, the final row of Table 3). However, the key finding is that there is evidence of the emergence of race as an important classifier even with extremely small numbers of training identities. Once again, this seems to reflect the covariance with identity: if there are physical differences within the learning set which can be used to support identity discrimination, then these will emerge during training.

6. Learning Race

To conclude our simulations, we report some potentially interesting effects which arise when considering the emergence of race as a classifier. First, for completion, we note that it is possible to repeat the analyses described in Section 4 using sexually homogenous sets, i.e. if trained on the identity of 5 Black and 5 White women, will ‘race’ emerge as the first dimension, and will such a dimension extend to classification of male faces? Although it is not clear that this exercise models many people’s learning environment, we have provided the analysis in the supplementary material. In summary, race emerges as a strong and reliable first factor when the model is trained on a multi-racial set of women, but produces less discrimination when trained on men.

Perhaps more interestingly, we asked whether learning the identities of a racially homogenous set might give rise to representations underlying the other race effect. One manifestation of this, recently reported by Laurence, Zhou & Mondloch (in press), is that viewers are poorer at ‘telling together’ different images of other-race than own-race faces. Using a sorting task of the type described by Jenkins et al (2011; see Figure 2 above), they demonstrated that other-race viewers tended to sort images of the same people into more identities than own-race viewers. We modelled this task by first training the LDA on 10 faces (5 men, 5 women) of one race only, with 20 images per identity. We then presented two novel faces, 30 images each, and measured their clustering solution in the same manner as in Section 2 (above). These novel identities were either the same or different race as the training set. Table 4 shows the indices of clustering (Sihouette Coefficient and Dunn Index) for different combinations of training and testing.

TABLE 4 HERE PLEASE

The results show evidence of the pattern reported by Laurence et al. In each case, there is a better fit for a two-cluster solution for ‘own’ rather than ‘other’ race faces. This holds for both indices (Silhouette and Dunn). Furthermore, the two-cluster solution is consistently better than multiple-cluster solutions for own-race faces, but this is not the case for other race faces where the Dunn Index shows better fits for larger numbers of clusters.

Given the very minimal training set, and the use of multiple images of completely novel test items, it is perhaps surprising that this effect of other-race clustering emerges so clearly. Of course, we are not claiming that this demonstration provides an explanation for the extensively-studied, other-race effect, with all its nuances. However, it does provide an interesting existence proof for the observation that human-like discriminability patterns can emerge from exposure to very small numbers of people, given multiple encounters.

Discussion

By applying a simple linear discrimination to highly diverse ambient images of a set of faces we were able to simulate key properties of human familiar face recognition. These include our ability to recognise familiar faces from views we have never seen before, and the fact that this excellent identification ability does not extend to unfamiliar faces. It is very interesting that the space derived by a simple mechanical discrimination on pictures offers an efficient means of classifying the identities of the trained faces (including generalisation to new exemplars) that does not extend to untrained, unfamiliar face identities. This corresponds well to human viewers' behaviour, and is consistent with the notion that variables needed to code individual face identity are not universally applicable; they are in some respects idiosyncratic to the trained faces (Burton et al, 2016; Dowsett, Sandford & Burton, 2016).

From this standpoint, it is remarkable that the first two dimensions created by an identity-based LDA do code differences that apply to untrained faces. As we have shown, these dimensions correspond to the pervasive social categories of sex and race. By applying LDA identity training to sets of images that varied only on one of these dimensions we showed that their substantial covariance with identity is responsible for this incidental learning.

We suggest that the approach based on ambient images was important to arriving at this compelling pattern of findings. Most studies of face recognition use only a single image of each face, and when more than one image is used the views are often carefully matched to vary in only a single characteristic such as head orientation or facial expression. Such studies offer useful insights into some of the variables involved in face learning (Bruce, 1982; Longmore, Liu, & Young, 2008), but these insights are dwarfed by the variability in the views of faces we encounter in our daily lives. The closest previous study to ours is that of Dahl, Malte, Bülthoff & Chen (2016), who trained an LDA to classify 120 face identities from a set of standardized images containing 5 viewpoints of each face varying in orientation. They found that their system then performed well at classifying these identity-trained images by sex or by race. Our study uses a much wider range of images that include unsystematic natural variations in lighting, pose and expression. Moreover, we tested the system's performance with novel, untrained images – so demonstrating true generalization. Finding a consistent set of cues to characterize an individual face's identity represents a substantial challenge. We have demonstrated that, paradoxically, this may be simplified by beginning with a wide range of images in the first place, as these allow a stronger separation of the intrinsic characteristics of a face from the image differences that constitute 'noise' for the recognition of identity (Bruce, 1994; Burton, 2013; Jenkins et al, 2011).

Although the face identity-trained LDA model proved (like humans) to be deficient at classifying the individual identities of images of unfamiliar faces, it turned out to be able to categorise both familiar and unfamiliar faces on sex and race, despite receiving no explicit training for these properties. By varying the types of images in the identity training set, we showed that this incidental learning of sex and race is due to the covariation of these characteristics with face identity, and that they can emerge from exposure to a remarkably small number of different faces.

The nature of this covariation between face identity and social categories of sex and race needs to be carefully considered. One of the reasons why LDA is useful for classifying ambient images of faces by identity is that it can cope with the fact that the type of variability encountered across the images of a face will differ between one face and another (Burton et al., 2016). In effect, LDA can find what is consistent across different images of face X independently from what is consistent across different images of face Y. Yet although LDA does not rely on the same combinations of cues to signal identities X or Y, we have shown that it 'finds' dimensions that correspond to face sex or race. Because classification by identity requires a much more complex use of facial information, however, we would not expect much useful information for classifying face identity to emerge from an LDA model trained only to classify face sex or race. In other words, the relation between training and performance is asymmetric; identity training will create dimensions corresponding to sex or race, but sex or race training will not create very useful dimensions of face identity. Analyses presented here in the Supplementary Materials show that this is the case.

This incidental learning of pervasive social categories from training the identities of a small number of faces has important theoretical implications. First, it shows that understanding social properties involving gender-related roles or membership of one's own social group may be largely irrelevant to acquiring the ability to classify faces along these lines. Second, the remarkably small number of familiar identities needed to create these categories suggests an obvious ontogenetic hypothesis, since initial face learning usually takes place within a family group in which the infant has to learn how to interact with a small number of different people based on their identities. So some kind of primacy of familiar face recognition makes sense. Moreover, it is also thought that any phylogenetic contribution leading to an evolved neural substrate for face recognition ability will have arisen within the context of relatively small social groups (Layton, O'Hara & Bilsborough, 2012) Whether or not there is such an evolved substrate, our findings show that learning to recognise familiar faces will create the representations that distinguish some particularly salient social categories.

Our findings are also relevant to a longstanding debate concerning the relation between the perception of a face's sex and its identity. Bruce & Young (1986) maintained that there must be some separation between the coding of sex and identity because we can easily classify the sex of unfamiliar faces, and Bruce, Ellis, Gibling and Young (1987) showed that having a gender-stereotypical appearance affected decisions about a familiar face's sex but had no effect on recognising its identity. They therefore drew a distinction between identity-specific semantic codes based on recognising a familiar face and the visually-derived semantic codes (such as sex or race) created by its appearance. However, other widely discussed models such as Haxby, Hoffman and Gobbini (2000) elide this distinction by shifting the focus onto the fact that characteristics such as sex, race, and identity represent relatively invariant facial attributes. Consistent with this type of account, some data support the idea of an integral representation of sex and identity (Goshen-Gottstien & Ganel, 2000; Rossion 2002; Zhao & Hayward, 2013).

What the present results show is that this kind of unresolved debate is largely a matter of perspective. From an overarching standpoint, the LDA creates a set of dimensions necessary both to sex and to identity (cf. an integral representation), but at a more detailed level the dimensions needed for sex and identity differ (cf. separability). Like many polarised debates in psychology, more detailed understanding reveals it as an example of Newell's (1973) aphorism that "you can't play 20 questions with nature and win". Newell (1973) pointed out that many debates in psychology begin by assuming some kind of binary opposition between competing views and then become sterile as evidence accumulates on both sides of the divide. His suggested panacea was detailed computer simulations based on an underlying task analysis, which is very much the general approach we have taken here.

Here we have used LDA to show that some of the things that at present seem mysterious about human face perception can be seen as inherently arising from solving the task in a certain way. In fact, it is somewhat surprising that a simple linear technique provides discriminability among widely varying images of the type shown in Figures 1 and 2. The extent of our commitment to LDA is that it is able to extract covariances which best describe facial images sampled within and between identities. The fact that the most prominent physical discriminators exactly correspond to human perceptions of sex and race is revealing in that it shows the power of the information available to the perceiver. How this is actually computed by the viewer is, of course, not a problem solved by this analysis.

We should also note that the pre-processing of facial images for LDA is an important part of the analysis. We have used PCA on ‘shape-free’ images here, i.e. faces which have all been morphed to the same shape. This follows many engineering approaches (e.g. Beymer, 1995; Vetter & Troje, 1995), where shape-normalisation is necessary prior to any statistical analysis of images, in order that mouths align with mouths, eyes with eyes etc. Psychological approaches have tended to treat the ‘shape’ and ‘texture’ of face images as independent sources of variation, and to analyse both (Hancock, Burton & Bruce, 1996; Itz, Schweinberger, Schulz & Kaufmann, 2014; Schutz, Kaufmann, Walther & Schweinberger, 2012). Of course, shape-normalised images retain some information about the shape of the original; a big chin and little chin morphed to an average shape give rise to different texture signals - due, for example, to lighting differences in the originals. Here we simply note that the normalised images ('texture' in some accounts) are sufficient to separate ambient images by identity. It is quite possible that this is too simplistic an approach for a more general account of face perception. For example, one can imagine taking a similar approach to study perceptions of facial expression. It may turn out that a separate analysis of shape would be necessary for an integrated model of identity and expression (Calder, Burton, Miller, Young & Akamatsu, 2001). In this paper, without loss of generality, we have simply been asking what information is *sufficient* to make the discriminations of identity, and what dimensions are necessary to do so.

In fact, we do believe that this approach may be interestingly applied to analysis of facial expression. This contrasts well with race and sex, because emotional expression is not confounded with identity in our exposure to faces – we need to recognise familiar and unfamiliar people over varying expressions. This suggests that a mechanism trained on identity will not deliver emotional expression, except to the extent that there are reliable associations (one person usually looks grumpy, another cheerful, etc). Would it be possible to produce a generalizable emotion recognition scheme from multiple instances of a few people, given appropriate learning categories? This is certainly worth exploring, and will form the basis of future work.

Finally, we return to where we started. In his discussion of why witnesses usually began by describing the sex and race of the suspect, Feingold (1914, p.50) thought this was because sex and race "constitute separate genera, and so are incapable of confusion under any circumstances". In other words, he thought they reflect particularly salient underlying physical differences. In the intervening 100 years, this simple insight has been complicated by the difficulty of finding reliable facial cues to sex (Campanella et al, 2001; Hoss, Ramsey, Griffin & Langlois, 2005; O'Toole et al, 1998; Russell, 2003) and by demonstrations that the perception of race can be influenced by social psychological factors (Hugenberg, Young, Bernstein & Sacco, 2010; Levin & Banaji, 2006; MacLin & Malpass, 2001). Now we have the evidence that Feingold was correct.

Appendix

Solution to Figure 2

ABAAABABAB

AAAAABBBAB

BBBAAABBAA

BABAABBBBB

References

Al-Zoubi, M. D. B. A., & Rawi, M. A. (2008). An efficient approach for computing silhouette coefficients. *Journal of computer science*, *4*(3), 252.

Andrews, S., Jenkins, R., Cursiter, H., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology*, *68*(10), 2041–2050.

Bekios-Calfa, J., Buenaposada, J. M., & Baumela, L. (2011). Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(4), 858–864.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *19*(7), 711–720.

Beymer, D. (1995). *Vectorizing face images by interleaving shape and texture computations*. MIT AI Lab memo 1537. Cambridge, MA: Massachusetts Institute of Technology.

Blair, I. V, Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: judging by features and categories. *Journal of Personality and Social Psychology*, *83*(1), 5–25.

Brooks, K. R., & Gwinn, O. S. (2010). No role for lightness in the perception of black and white? Simultaneous contrast affects perceived skin tone, but not perceived race. *Perception*, *39*(8), 1142–1145.

Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22*(7), 829–840.

Bruce, V. (1982). Changing faces: Visual and non‐visual coding processes in face recognition. *British Journal of Psychology*, *73*(1), 105–116.

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

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R. & Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception*, *22*(2), 131–152.

Bruce, V., Ellis, H. D., Gibling, F., & Young, A. W. (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., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327.

Bruce, V. & Young, A.W. (2012). *Face Perception.* Hove: Psychology Press.

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., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*, 153.

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*, 943–958.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: representations of faces derived from multiple instances. *Cognitive Science, 40(1)*, 202-223.

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.

Caldara, R., & Abdi, H. (2006). Simulating the "other-race" effect with autoassociative neural networks: Further evidence in favor of the face-space model. *Perception*, *35*(5), 659–670.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*(9), 1179–1208.

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition*, *8*(2), 237–262.

Chen, L. F., Liao, H. Y. M., Ko, M. T., Lin, J. C., & Yu, G. J. (2000). New LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, *33*(10), 1713–1726.

Cheng, Y. D., O'Toole, A. J., & Abdi, H. (2001). Classifying adults' and children's faces by sex: Computational investigations of subcategorical feature encoding. *Cognitive Science*, *25*(5), 819–838.

Craw, I. (1995). A manifold model of face and object recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition*. London: Routledge.

Dahl, C. D., Rasch, M. J., Bülthoff, I., & Chen, C.-C. (2016). Integration or separation in the processing of facial properties - a computational view. *Scientific Reports*, *6*, 20247.

Dowsett, A.J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Quarterly Journal of Experimental Psychology, 69(1)*, 1-10.

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, *4*(1), 95–104.

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, 42A*(3), 495–512.

Etemad, K., & Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, *14*(8), 1724–1733.

Feingold, G. A. (1914). The influence of environment on identification of persons and things. *Journal of the American Institute of Criminal Law and Criminology*, *5(1),* 39–51.

Fisher, R.A. (1936). The use of multiple measures in taxonomic problems. *Annals of Eugenics, 7(2)*, 179-188.

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.

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2010). Multi-PIE. *Image and Vision Computing, 28*(5), 807-813.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), 751–753.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (1998). A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research*, *38*(15-16), 2277–2288.

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: human perception and principal components analysis. *Memory & Cognition*, *24*(1), 21–40.

Haxby, J. V, Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Hoss, R.A, Ramsey, J. L., Griffin, A. M., & Langlois, J. H. (2005). The role of facial attractiveness and facial masculinity/ femininity in sex classification of faces. *Perception*, *34*(12), 1459–1474.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. Technical Report 07-49, University of Massachusetts, Amherst.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology*, *43*(2), 334–340.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: an integrative account of the other-race recognition deficit. *Psychological Review*, *117*(4), 1168–1187.

Itz, M. L., Schweinberger, S. R., Schulz, C., & Kaufmann, J. M. (2014). Neural correlates of facilitations in face learning by selective caricaturing of facial shape or reflectance. *NeuroImage*, *102*, 736–747.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4–37.

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

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

Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1867-1874).

Laurence, S., Zhou, X., & Mondloch, C. J. (in press). The flip side of the other-race coin: They all look *different* to me. *British Journal of Psychology*.

Layton, R., O'Hara, S., & Bilsborough, A. (2012). Antiquity and social functions of multilevel social organization among human hunter-gatherers. *International Journal of Primatology, 33*, 1215-1245.

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: the role of race categories. *Journal of Experimental Psychology. General*, *135*(4), 501–512.

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

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089.

Kramer, R. S. S., Ritchie, K. L., & Burton, A. M. (2015). Viewers extract the mean from images of the same person : A route to face learning. *Journal of Vision*, *15*(4), 1–9.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous race face effect. *Psychology, Public Policy, and Law*, *7*(1), 98.

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37*, 806–816.

Martinez, A. M., & Zhu, M. (2005). Where are linear feature extraction methods applicable ? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(12), 1934–1944.

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.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Bartlett, J. C. (1991). Classifying faces by race and sex using an autoassociative memory trained for recognition. In K.J. Hammomd & D. Gentner (Ed.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society.* (pp. 847–851). Hillsdale, NJ: Lawrence Erlbaum.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America A*, *10*(3), 405–410.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: the role of stimulus structure in recognition and classification. *Memory & Cognition*, *26*(1), 146–160.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*(9), 1109–1121.

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

Schulz, C., Kaufmann, J. M., Walther, L., & Schweinberger, S. R. (2012). Effects of anticaricaturing vs. caricaturing and their neural correlates elucidate a role of shape for face learning. *Neuropsychologia*, *50*(10), 2426–2434.

Sporer, S. L., Trinkl, B., & Guberova, E. (2007). Matching faces: Differences in processing speed of out-group faces by different ethnic groups. *Journal of Cross-Cultural Psychology*, *38*(4), 398–412.

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

Vetter, T., & Troje, N. (1995). Separation of texture and two-dimensional shape in images of human faces. In G. Sagerer, S. Posch, & F. Kummert (Eds.), *Mustererkennung 1995, Reihe Informatik aktuell* (pp. 118–125). Berlin: Springer Verlag.

Zhao, M., & Hayward, W. G. (2013). Integrative processing of invariant aspects of faces: effect of gender and race processing on identity analysis. *Journal of Vision*, *13*(1), 1–18.

Figure Legends

Fig. 1. Ambient images of the same person. What do these images share in common, which is distinct from images of every other human being?

Fig. 2. Sorting ambient face photos by identity is a difficult task unless the faces are familiar. From Jenkins et al, (2011). How many different people are there in this matrix? Solution in the Appendix.

Fig. 3. d’ as function of threshold for a model trained on 20 images of each of 20 people. The test items comprise 200 novel instances of the learned people, and 200 images of novel people. Note that the distance used to set thresholds is in arbitrary units that are determined by the dimensions of the LDA.

Fig. 4. The first dimension of the subspace resulting from LDA training using 20 images of 20 identities. Crosses represent the 400 training images (black) and Circles represent 10 novel instances of each of the 20 identities (red).

Fig. 5. Images of 200 untrained faces projected on to the first dimension of the training set subspace represented in Figure 4.

Fig. 6. The second dimension of the subspace resulting from LDA using 20 identities (400 images). Crosses represent the 400 training images (black) and Circles represent 10 novel instances of each of the 20 identities (red).

Fig. 7. Images of 200 untrained faces projected on to the second dimension of the training set subspace represented in Figure 6.

Fig. 8. The first dimension of the subspace resulting from LDA using 10 White identities (200 images). Crosses represent training images (black) and novel instances of the same identities (red).

Fig. 9. Images of untrained identities projected onto the first dimension of the training set subspace represented in Figure 8.