Article:

BURTON, MIKE (2026) Representations for face recognition: The 53rd Bartlett Lecture. The Quarterly Journal of Experimental Psychology. pp. 531-545. ISSN: 1747-0226

https://doi.org/10.1177/17470218251396729

Representations for face recognition: The 53rd Bartlett Lecture

A. Mike Burton
Department of Psychology, University of York, UK
Faculty of Society and Design, Bond University, Australia

Abstract

Models of human face recognition rely on the notion of representation, but rarely describe this in detail. Here, I will argue that our conception of face representations is often 'essentialist' – assuming that there is some fixed set of values that captures a particular person's face. However, this conception is inadequate for the purpose of familiar face recognition, and I will suggest that representations instead need to incorporate the statistical properties of our exposure to all the faces we know, including variability and sampling. I will review findings from empirical and simulation research suggesting that the idiosyncratic properties of each perceiver results in a unique set of representations, which can be difficult to understand using traditional experimental approaches. Methodological diversity seems to offer the best route for understanding face recognition – a problem that remains stubbornly unsolved.

Acknowledgements

I would like to thank Rob Jenkins and Andy Young, both of whom read drafts of this paper and have been responsible for many of the ideas it contains over years of collaboration. I am equally grateful to many advisers, collaborators and colleagues, too numerous to mention here, but I would particularly like to thank Vicki Bruce, Stefan Schweinberger, David White, Markus Bindemann and Holger Wiese.

Introduction

Perhaps the most puzzling aspect of recognising a face is how easy it is. In common with many other cognitive and perceptual tasks (conversation, driving) we only really notice we are doing it when it goes wrong. On the occasions when we fail to recognise a family member or friend, we are often amused or embarrassed – while no congratulations accrue for our multiple daily successes. And yet, the typical accuracy of familiar face recognition hides its complexity and leads us to make other mistakes. For example, we have come to rely on photo-ID as a means of confirming someone's identity – presumably because we find it so easy to recognise the people we know. However, as we shall rehearse below, face recognition is surprisingly difficult for viewers *unfamiliar* with a face. The fact that we can match a friend's face to her passport does not generalise well to other viewers who do not know her – even those trained for that specific purpose are prone to error in this task.

The history of psychological research in face recognition does not start with the question of how we recognise one another in daily life. Instead, two key findings, apparently in contradiction to one another, dominated early research. Both concerned face *memory* rather than face *perception*. First, studies of visual memory suggested that faces were particularly easy to remember, by comparison to other classes of stimulus. For example, recognition memory tests showed higher performance when recognising pictures of faces compared to other 'homogenous' stimuli, such as snowflakes or inkblots (Goldstein & Chance, 1970). This is a secure finding, often repeated over the intervening years - for example Kapsetaki & Zeki (2022) show better memory for pictures of human faces compared to pictures of buildings, animal faces and more general non-face images.

The second finding, of more concern in wider society, was that eye-witness memory for people can be unreliable, even when accompanied by high confidence. In the 1970s, across several jurisdictions, it became clear that miscarriages of justice were occurring due to incorrect eyewitness identification, and various changes were suggested to legal practice to take this into account. These real cases were consistent with experimental findings, showing that experimental participants witnessing a staged crime were often poor at subsequently picking the people involved from a lineout (Leippe et al., 1978; Wells et al., 1979).

This apparent contradiction – good performance in memory tests, but poor lineup results – drew the attention of cognitive psychologists, and was resolved in a way that is interesting for our current purposes for at least three reasons: (i) because the resolution relies on a distinction between picture recognition and face recognition; (ii) because this distinction is fundamental to understanding face familiarity; and (iii) because it continues to be ignored in large portions of the academic literature on face recognition, several decades later.

The key to understanding this problem is to recognise the fact that *picture* memory is generally very accurate. We are good at recognising a previously-viewed picture of a pastoral scene just as we are good at recognising a previously-viewed picture of a face. This was established in recognition memory studies in the 1960s and 70s (Galper & Hochberg, 1971; Shepard, 1967; Standing et al., 1970) and our large capacity for visual image memory has continued to be explored and extended (Brady et al., 2008). However, our ability to remember having seen specific individuals, independently of the particular image used, is critically dependent on our *familiarity.* If shown a picture of Barack Obama (say) in the first phase of an experiment, and he also appears in a later test phase, then viewers are typically highly accurate at test, and unaffected by whether the same picture is used twice, or a new photo of the former US President is used on the second time around. However, for somebody unfamiliar to viewers, a change of photo between learning-phase and test-phase can be catastrophically damaging for accuracy. Bruce (1982) showed that a change in expression or pose between first exposure and test substantially reduced recognition accuracy, even for photos from the same studio session (i.e. with the same camera and lighting), and this has been replicated and confirmed across many different settings (e.g., Longmore et al., 2008; O'Toole et al., 1998; Patterson & Baddeley, 1977).

It is now widely agreed that familiar and unfamiliar faces are perceived in different ways. To present the contrast starkly, we have argued that unfamiliar faces are processed in fundamentally image-bound fashion, whereas familiar faces are processed in a more abstract manner (Hancock et al., 2000; Megreya & Burton, 2006). Some authors have suggested that these differences are qualitative (Ramon & Gobbini, 2018), though one could argue that certain types of quantitative change could also accommodate the effects of familiarity. However, the fundamental point is that there are large behavioural differences between familiar and unfamiliar face perception, with familiar face recognition being more generalisable across viewpoints, settings and changes in the individual being recognised. In the remainder of this paper, I will concentrate on the representations of familiar faces themselves – what characteristics would these representations need, to support recognition in the lab and in daily life?

Familiar face representations: early days

The concept of a 'face recognition unit', or FRU is common to a number of theoretical accounts dating from the 1980s. Bruce and Young (1986) write that an FRU "will respond when any view of the appropriate person's *face* is seen, but will not respond at all to his or her voice, the name …[etc])". In a subsequent implementation of the Bruce and Young model (Burton et al., 1990), we wrote that's FRUs "store the visual structural descriptions which allow views of one known face to be discriminated from views of other faces, whether known or unknown". It will be clear then, that FRUs are doing a highly sophisticated job – indeed, for practical applications like automatic face recognition, they capture the whole problem of individuation. But how might they achieve this?

While proponents of early functional models of face recognition remained agnostic about the nature of an FRU representation, many authors implicitly adopted an "essentialist" interpretation. Under this view, an FRU for Hilary Clinton (say) codes the core essence of her visual characteristics. For someone highly familiar to the viewer, this essence will have been extensively refined over many exposures, leading to a representation that can be triggered by "any view of the appropriate person's face" (ibid.). The intellectual environment in the 1980s encouraged this interpretation: FRUs were held to be analogous to "logogens", specialised functional units for recognising individual words in the highly influential model of lexical recognition (Morton, 1969).

A very clear essentialist view of familiar face recognition can be found in the proposal that we recognise the people we know by representing the "configuration" of their features. This "configural processing" account holds that each person has a unique layout of features, which can be used for recognition by the perceiver. For example, "recognition of an individual person … must depend on recognition of the subtle differences in the features and their configuration that form the impression of a unique face" (Tanaka & Gordon, 2011). Similarly, "Because faces are made from common features (eyes, nose, mouth etc.) arranged in the same general configuration, subtle differences in spatial relations between face features being encoded [are] particularly useful for successful recognition of a given face." (Richler et al., 2009). We have argued elsewhere that this appeal to configural processing as a way of understanding familiar face recognition is unwarranted, based on both the available evidence and the theoretical underpinnings of the approach (Burton et al., 2015). We will not rehearse those arguments here, but will only point out that the approach emphasises the critical importance of "telling people apart" (what makes a face different to all others), ignoring the equally pressing component of the problem "telling people together" (how are many different images of the same face cohered), to which we shall return below.

While a configural processing account is clearly an essentialist view of familiar face recognition, it is not the only explicit example in the literature. In our own work, we have proposed two other essentialist accounts – both rather different from one another. In one proposal (Burton, Bruce, et al., 1999), we derived a statistical description using principal components analysis (PCA) on single images of many faces. This allows one to imagine a "face space" in which particular individuals are located at particular points, and was popular in early computer-based recognition systems (Kirby & Sirovich, 1990; Turk & Pentland, 1991). Under this procedure, it is important to provide a representative image of each face, which can be used as a basis for subsequent recognition of new images of the same person.

Our second essentialist approach to deriving familiar face representations appealed to the idea of "face averages" (Burton et al., 2005). We observed that multiple images of the same face could differ very significantly, but that averaging these together (in the sense of image morphing) quickly produced a stable representation. So, for example, given roughly fifteen images of a single person, their average would be largely unaffected by subsequent additions. Furthermore, the average for a given person was essentially the same, regardless of which particular images were selected to construct it.

In subsequent work, we have come to disavow the essentialist approach to familiar face recognition. I will argue, below, that a more sophisticated, more plastic, and less structured representational apparatus is necessary to explain the large range over which we can achieve face recognition. However, before spelling this out, we need briefly to review some of the large differences between familiar and unfamiliar faces which underlie this view.

More on familiar and unfamiliar faces

I described, above, how memory for familiar faces is largely unaffected by superficial image changes between first and second encounter in an experiment. In later work, it turned out that robust recognition over multiple images is a core characteristic of familiar face perception, and one that holds much more widely than simple recognition memory experiments.

In a series of studies led by Vicki Bruce around the turn of the 21st Century, we showed participants line-ups of faces, inspired by eyewitness studies. However, in these studies, the participants were not required to remember anything. Instead, a target face was presented simultaneously with the line-up candidates, and viewers were given unlimited time to decide whether the target was present, and if so to pick that person out (Bruce et al., 1999, 2001; Henderson et al., 2001). Performance for unfamiliar faces was typically poor - in these experiments around 70% accuracy for images taken in good visual conditions, showing similar pose and taken on the same day. A 30% error rate seems surprising in these circumstances, which in many ways are optimal for matching two images of the same person. While these experiments were not the first to demonstrate poor unfamiliar face *matching* (rather than memory), they were influential in undermining an assumption that humans are good at recognising faces generally. Here was an example where we were surprisingly bad at face recognition. Furthermore, it had important practical implications – we prove our identity with Face-ID all the time, not only at passport-checks, but in everyday settings such as buying age-restricted goods. If face matching is unreliable, why do we do this?

Over the subsequent years, it has become clear that unfamiliar face matching is generally very poor, and this extends to many settings beyond the line-ups first used in these experiments. Viewers are similarly poor at matching pairs of unfamiliar faces (Burton et al., 2010; Fysh & Bindemann, 2018) and remain poor when matching a live person to a photo (Davis & Valentine, 2009; Kemp et al., 1997; Megreya & Burton, 2008). Furthermore, people on whom society relies to make accurate matches, and who have often received extensive training, are typically no better than inexperienced control participants. Such failure of expertise has been demonstrated for passport officers (White et al., 2014), police officers (Burton, Wilson, et al., 1999; Towler et al., 2019), bank tellers and notaries (Papesh, 2018) bar staff and bouncers (Robertson & Burton, 2021). While there are often large individual differences in face matching ability (White & Burton, 2022), this ability appears to resist training, and performance by professionals is generally much lower than might be expected.

Fig 1: Pairs of items from the Glasgow Face Matching Test (Burton et al., 2010). The top pair show two different people, while the bottom pair who the same person. People often find these tests difficult.

In contrast to unfamiliar face matching, we are remarkably good at familiar face matching, even in highly degraded or distorted images (Burton, Wilson, et al., 1999; Hole et al., 2002). In fact, our ability to match different images of the same person is a good index of the level of familiarity we have with that person (Clutterbuck & Johnston, 2004). This disparity appears to explain why we have come to rely on photo-ID, when the evidence suggests it is unreliable. We appear to over-generalise our abilities with familiar faces – assuming we will be just as good at judging the identity of unfamiliar faces. In fact, it is very difficult to put oneself in the position of an unfamiliar viewer when we know a face ourselves; participants asked to judge the difficulty of particular face matching items, so that they could be used for a test with unfamiliar viewers, nonetheless rated the faces they happened to know as being generally easier (Ritchie et al., 2015) – a failure to de-centre which seems to have affected our whole social world.

These large discrepancies between perception of familiar and unfamiliar faces are now widely accepted in behavioural psychology. However, they are not so prominent in neuroscientific investigations of face perception. The influential model by Haxby et al (Gobbini & Haxby, 2007; Haxby et al., 2000), proposes a 'core network' of face regions, including the Fusiform Face Area (FFA), a visual region held to code a view-invariant representations of faces. However, fMRI studies comparing familiar and unfamiliar face perception have not consistently found effects of familiarity in the FFA (Gobbini & Haxby, 2006; Gobbini et al., 2004; Leveroni et al., 2000). Instead, such studies have typically reported familiarity effects in areas beyond the core network, including non-visual areas involved in affective responses, semantic and episodic memory (Gobbini et al., 2004; Leveroni et al., 2000; Noad et al., 2024). It is clear that recognition of a familiar person will generally evoke retrieval of memories about that person, and possibly an emotional response (Zhou & Jenkins, 2022), so it is not surprising to see familiarity effects in these regions. Nevertheless, it is perhaps more surprising that the large behavioural effects, always observed, do not translate into a simple region-based effect in FFA, the area held to code identity.

One possibility to explain this pattern of responses is that familiarity itself does not manifest visually. Perhaps we can only identify a known person by recruiting semantic and emotional information - an explanation that is gaining some traction in the current literature (Kovács, 2020; Noad & Andrews, 2024; Noad et al., 2024). Having said that, it seems odd to suggest that there is no visual learning of a specific person's face, which develops over familiarisation. While not denying the importance of the extended system for person recognition, rejecting the acquisition of a purely visual representation for familiar faces seems to be an unnecessarily radical response. An alternative is to propose that MRI is not well suited to exploring such a representation. In fact, ERP evidence is much stronger for *both* visual and extended-system familiarity effects, broadly in the 250-400ms and 400-600ms range, respectively, post stimulus onset (Schweinberger et al., 2002; Wiese et al., 2024; Wiese, Tüttenberg, et al., 2019).

Familiar face representations: modern times

In this section I will outline an operationalisation for a familiar face representation. To be clear, I have in mind the same goals as those who first postulated the FRU – a representation that will capture any recognisable view of a known person, and will allow access to higher-level representations (memories, emotional responses) but will be visual in nature, i.e. not accessible by voices, names etc. In doing so, I will depart from the "essentialist" position outlined above, and will start by enumerating the requirements for such a representation.

First, a representation should be able to handle *variability* in input. It is trivial to point out that a single face may give rise to very many different images. Photos of Taylor Swift can be very different, one from another. While this much is obvious to the non-face-researcher, it has sometimes escaped our theorising (including my own). Face recognition has all the same challenges as any object recognition problem, in needing to account for viewing angle, lighting etc. However, face recognition has a further issue, in that faces deform non-rigidly, through speech, expression and so forth. Furthermore, they change, again non-rigidly, with health, ageing, visits to the barber. While initially considered extra problems for our perceptual systems to solve, it was Vicki Bruce who first pointed out that these variations could, in fact, *support* recognition (Bruce, 1994). If a viewer experiences variability, for example within a conversation, this variability delivers a great deal of information about what changes within a face leave its identity unchanged. Establishing *which* variation is non-diagnostic of identity for this person helps to encapsulate a more sophisticated, and generalisable representation.

At this point, we need to return to the idea of "telling people apart", which is so often emphasised in the literature. To this, we add the problem of "telling people together", cohering together multiple images, often very different, of the same person into a single representation. Studies are often designed to avoid this "telling-together" problem. In the worst scenario, the same image might be used as a learning and a test item. In studies with more convincing designs, images may be changed between learning and test or non-identical images used in matching tasks. But it is very rare that the full range of variance is acknowledged – the range of images that could be Taylor Swift, say. In an attempt to remedy this, we designed a sorting task, in which participants were presented with multiple images of two people, and asked to sort them into piles for each identity, without being told how many piles to expect (Jenkins et al., 2011). In one version of this task, we used two Dutch TV presenters, twenty photos of each, and gave the task to UK viewers who did not know the presenters, and Dutch viewers who did. The results were very striking: the UK viewers sorted the two people into multiple piles, an average of 7.7 in this case, while the Dutch viewers were almost perfect in sorting them into two piles. Furthermore, the UK viewers very rarely confused the two presenters, sorting a photo of both identities into the same pile less than once per participant, on average. The problem these unfamiliar viewers had was not telling the people apart, but telling them together. The Dutch viewers, however, found this a trivially easy task, their familiarity ensured that they could cohere these superficially disparate images easily. We need our representation of familiar faces to be able to do that.

Second, a working representation should be able to incorporate *degree* of familiarity straightforwardly. In face recognition studies (including most of my own), a contrast is usually made between 'familiar' and 'unfamiliar' faces, as though this were a simple categorical distinction. Of course, it is self-evident that the range of familiarity in daily life is very large, from close family members to people seen occasionally on a daily commute. But models of face recognition typically do not incorporate the apparently continuous dimension of familiarity. For example the Bruce and Young (1986) model and its later variants (Burton, Bruce, et al., 1999) acknowledge the need for graded familiarity, but do not operationalise it, instead concentrating on the familiar/unfamiliar distinction through the presence or absence of an FRU.

Modern analyses acknowledge that the degree of familiarity is a critical aspect of face recognition (Kovács, 2020; Ramon & Gobbini, 2018), and ignoring it will lead to a partial account. Nevertheless, it is difficult to incorporate this into experimental procedure. One of the problems is that familiar face recognition is so good – it is challenging to design a recognition task in which accuracy is not at ceiling for familiar faces, even those with lower familiarity. Clutterbuck and Johnston (2004) offered a solution to this problem, demonstrating that face matching ability serves as a good index of familiarity - though this has not been widely adopted in the literature. Despite these practical issues, our daily experience tells us that we know some people better than others, and we know from diary studies that familiar face recognition is not perfect – we do sometimes make mistakes, and these can be more frequent for the people we know least well (Young, Hay, & Ellis, 1985). It is therefore important that our theoretical models incorporate representations explicitly capable of incorporating large ranges of familiarity.

Third, representations of familiar faces should show *plasticity*. This requirement is necessary both for initial face learning and for subsequent forms of updating. It is clear that we learn new faces throughout life, and like most forms of learning, it seems reasonable to propose that this is gradual and modulated by exposure, and indeed this has been demonstrated many times (MacLin et al., 2001; Read et al., 1990). All familiar faces were once unfamiliar, and the transition to familiarity remains a problem for accounts arguing for qualitative processing differences. Furthermore, representations should be updatable, even for people we know well. If we meet an old friend, not seen for several years, it will be necessary to update their appearance to accommodate changes with age. Even within the same time period, we may need to recognise someone we have not previously seen wearing makeup, with a new hairstyle, exercising or singing. The range of variability over which we have encountered someone has somehow to be incorporated into our representation of that person – a flexibility that will allow further generalisation in subsequent recognition.

Common to these requirements is the idea that a perceiver's representation of a familiar face reflects their exposure to that particular person. The range of this exposure will differ between viewers, even for the same very famous person. For example, we may have had extensive exposure to a particular actor or singer, through following them on media or at live events. But our exposure will not be the same as the actor's family members, who experience a different range. The goal of the perceivers in this example is not the essentialist programme of trying to establish the true essence of that person's face. Instead, it is to gain an understanding of the range of visual events that "count" as that actor, within the world of the perceiver. Psychologically speaking, there is no core essence to establish, only data to accrue.

To operationalise these ideas, we have offered a *statistical* account of familiar face recognition. In the next section, I will describe some of the ways this can be used to understand familiar face recognition. In this account, we have used simple statistical methods, mostly Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA). The intention here is not to claim that these particular techniques, or even this class of linear methods, capture the operations of the brain. However, they do deliver apparently high-level descriptions of picture-level information, useful for examining the processes underlying face perception and classification, and sensitive to their pattern of exposure. We therefore use these techniques in a functional manner, for explanation purposes and without commitment to specific human implementation.

Modelling face recognition

Idiosyncratic variability.

In the first of a series of papers exploring individual variability, we performed PCA on multiple images of the same face (Burton et al., 2016). We used Hollywood actors, because it is easy to source multiple different images of these people, and performed PCA on pictures of Tom Cruise, another PCA on pictures of Tom Hanks, and so forth. This is an unconventional use of PCA, which has previously been used as a way of understanding the dimensions on which people differ, and so is based on images of multiple people, using one or few images of each. In contrast, we were interested to establish the dimensions on which images of a single person vary, and to ask whether the same dimensions emerge for different individuals.

To summarise, we found that image-spaces for all individuals shared early components of variability in common. The first three dimensions of all people's spaces coded viewpoint, capturing head rotation, head nodding and distance from the camera. These early dimensions all capture rigid transformations, with apparently no non-rigid deformations of the face contributing to their variance. Later dimensions had two interesting properties: (i) they coded non-rigid transformations of the face; and (ii) they were idiosyncratic. For example, while dimension 4 of Tom Hanks might code a smile, dimension 4 of Tom Cruise might code a different mouth movement. Dimensions of variability ceased to be common after the early components.

While the divergence of derived components between people suggests that faces *vary in different ways*, we might not expect exact dimension-by-dimension correspondence, as this may reflect the specific images chosen for any individual. Instead, a more interesting question to ask is whether the same space is spanned by the full set of components derived from multiple images of each individual. To do this, we projected a novel image of an individual into a space derived from photos of that same person, or a space derived from images of someone else, see Fig 2. In all cases, a face was substantially better coded by their own dimensions than somebody else's, even when both faces had similar descriptions (e.g. two young white men with short dark hair and clean-shaven). In short, it is possible to build an accurate novel image of a person from dimensions derived from a range of pictures of that person – but dimensions derived from someone else do not provide the building blocks necessary.

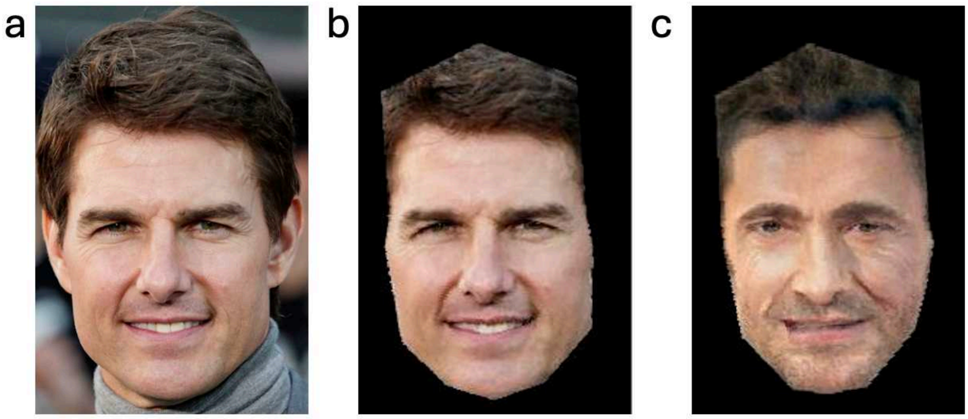

Figure 2: A photo of the actor Tom Cruise (a), reconstructed using dimensions derived from multiple images of (b) Tom Cruise, and (c) Hugh Jackman. (*Image Acknowledgement CC BY-NC 4.0)*

These results show that individual variation is substantially *idiosyncratic.* People differ not only photo-to-photo, or average-to-average, but the *ways* in which Hugh Jackman varies are different from the *ways* in which Tom Cruise varies. In a more formal sense, people's individual variability does not span a common space. This apparently simple observation has quite profound implications. It means that a viewer's knowledge of one person, and the range over which they can appear in photos, does not necessarily help in recognising variation in another person.

Recall that unfamiliar face matching is highly error-prone. This is because the viewer does not know the range of variation relevant *to that person* and knowing someone else's range is not helpful. Note further, that the early components of people's individual spaces, which by definition capture the most variance, are held in common. This means that most of the variance in a pair of images for matching is coding information not relevant to identity – but in fact superficial pictorial aspects such as pose. Fig 2c retains the general pose, distance from camera and other superficial characters of the original photo, because both actors (and all others) share the early dimensions of space which capture these aspects of the image. Most of the information in the image does not code identity, which inevitably makes unfamiliar face matching difficult.

This observation also implies that becoming familiar with a person requires a viewer to experience a wide range of variation across which that person can appear. Rather than learning the core essence of someone (for example an accurate geometrical coding of their feature spacing), learning requires one to sample across a range of images which will span the space of possible images for that individual. This is a prediction which has subsequently been confirmed in experiments showing that face learning is enhanced by exposure to wide-ranging images, taken under varying conditions (Murphy et al., 2015; Ritchie & Burton, 2017) – an advantage that persists on top of the usual predictors of learning such as duration and frequency of exposure. The very information which is usually discarded in psychological experiments on face recognition and learning – noisy, uncontrolled variance between images – turns out to be critical to the process.

Graded familiarity.

While it is useful to derive "face-spaces" for individuals, this approach cannot capture a perceiver's face knowledge overall. In subsequent work, we developed a model based on statistical descriptions of large image sets, containing several thousand photos and comprising multiple images of multiple people (Kramer et al., 2018). Some faces were represented by many photos, some by fewer, and some by just one. Instead of imposing highly controlled, systematic differences between the frequency of representation, we allowed a more natural variation in the degree to which faces were familiar, captured here by the number of images encountered for each person. We were thus able to move away from categorical comparisons between familiar and unfamiliar faces, and instead adopt a more graded approach, observing any correlates with the degree of familiarity.

This model performed in a manner consistent with our knowledge of familiarity in behavioural tasks. For example, higher familiarity led to better recognition of new photos of known people, better matching of previously-unseen image pairs and better tolerance to image degradation. These are perhaps unsurprising results but confirm that the derivation of spaces containing multiple images of multiple people nevertheless confer an advantage for those faces sampled over the greatest variance. This seems like a more natural analogue of human perception – we know some people well, some people less well, and do not divide our visual world into simple familiar/unfamiliar categories.

While standard effects of familiarity develop naturally within this model, some results were perhaps more surprising. One example is that the more familiar a face, the more recognition relies on the *internal features*. This is a result that has been observed for many years in the behavioural literature (Ellis et al., 1979; Young, Hay, McWeeny, et al., 1985) but, to our knowledge, had not been captured in face recognition simulations. The explanation of this effect is quite natural in this model: if one's representation of a person includes multiple images, then the internal features (eyes, nose, mouth) are typically more consistent across images than the external features (face shape, pose, hair). On the other hand, coding only a few images of a person leaves one with no knowledge of whether the hairstyle (say) is highly characteristic of that person or not. Furthermore, the external features simply occupy more of the image than the internal features, and so an image-matching recognition strategy, which we hold to be the only option for unfamiliar face matching, is likely to favour external features.

Finally, this simulation highlighted the importance of top-down influences on face recognition. When we see multiple images of a single person, by what mechanism are they cohered together? We argue that there is often considerable top-down support for person recognition. The task is generally supported by multiple redundant processes – in daily life, context, voices and clothing all support recognition. In fact, a psychology experiment is one of the few settings in which faces are presented for identification in the absence of other information. Outside the lab, we usually have the information to decide that a new image should be classed with previously-seen images of that person. We model this top-down cohesion with simple clustering processes based on LDA, the Fisherface approach which has proved useful in earlier models of face processing (Bekios-Calfa et al., 2011; Belhumeur et al., 1997). As one might expect, support from top-down processes significantly enhances model performance – an observation that is usually excluded from behavioural experiments, but is completely consistent with the involvement of higher-level processes observed in neuroscientific studies of face perception (Gobbini & Haxby, 2007; Kovács, 2020).

Extreme variation – faces across the lifespan.

Some of the people whose faces are familiar to us are associated with specific periods of life – perhaps college friends, or actors in a once popular show. However, for other people, such as family members, recognition can be lifelong. Our daily task is to recognise someone as they currently appear, but we can nonetheless recognise past photos, even those we have not previously seen. In the media, certain faces appear over their whole adult lives – for example Paul McCartney has been photographed extensively from his early twenties to his eighties, and many people around the world can recognise those photos, seemingly independently of his age. How can this be achieved? From an essentialist point of view, this is a very challenging problem – does the core essence of a Paul McCartney representation favour a particular point in life, and what is common between 20-year-old Paul and 80-year-old Paul? Is it perhaps necessary to develop more than one representation – i.e., FRUs for young and old Paul?

To study this problem, we examined the faces of 1960s popstars for whom we could source multiple images over a fifty year period. (Mileva et al., 2020). Figure 3 shows examples from the young and old time periods for two pop stars. Using the standard PCA and LDA approach, we derived models based on images sampled from across the whole lifespan, and others based on images from a single period. We found that models based on image similarity (i.e. PCA, but no top-down 'clustering' support from LDA) could achieve moderate levels of age-independent recognition. In other words, there was enough image similarity between photos of a popstar at young and old ages to achieve some levels of recognition – a result that supports the idea that we do not need multiple representations of each face. However, recognition was considerably enhanced in models that incorporated very simple clustering, based on LDA. Recognition was possible over a huge range of images, with generalisation across decades, if the faces had initially been encoded to support "telling faces together". For example, to ensure recognition of a novel photo of Paul McCartney in his 60s, an excellent strategy was to cohere his images together, even when all these images were taken in his 20s. Telling together seems to generalise well, even to these extreme changes over age.

Figure 3, example images from the highly diverse image set showing young and old photos of (a) Brian Wilson and (b) Frankie Avalon. (*Image Acknowledgements, left to right: Capitol Records [Public domain], Takahiro Kyono [CCBY2.0], Film Studio [Public domain], John Mathew Smith [CCBY-SA2.0]*)

A second observation from this study was the idiosyncrasy of the ways in which people age. As with other dimensions of variability, described at the start of this section, ageing affected different people differently. Of course, this is well-known outside face research – we describe people as “aging well”, or not, presumably meaning that they change a little or a lot over the years. However, psychological research has tended to focus on generic aspects of ageing (Burt & Perrett, 1995; Geng et al., 2007). These capture the effects of changes in face-shape and skin texture in the general population on average, and can be used to make a face look older graphically. However, they cannot capture the very substantial idiosyncrasy in aging including, for example, health, lifestyle and self-presentational choices such as facial hair or make-up. Human recognition achieves this for familiar faces, through use of image variability as a source of information.

Emergent properties of recognition

I have described an approach to face representation that is statistical, and contrasted this with an “essentialist”, all-or-nothing conception popular in early models. However, a statistical approach requires us to consider sampling. What exactly is our exposure to faces? In fact, rather little is known about this. Based on data from recall and recognition of personally familiar and media-familiar faces, we have estimated that the modern person knows about 5000 faces (Jenkins et al., 2018). By contrast, current DCNN-based face recognition systems, which operate with extremely high accuracy, base their learning on orders of magnitude greater number of faces – for example, *Clearview AI*, a system currently in operation around the world, is based on 50 billion face images (https://www.clearview.ai).

These numbers are puzzling. For most of evolutionary history, humans lived in small groups of 40-100 (Dunbar, 1993; Schiffels & Durbin, 2014) and while we do not currently have a good estimate of human *capacity* for face recognition, the observations of how may faces we actually know make it clear that our capacity vastly outweighs the demands of small-group living. We also observe that early in life, infants see a small number of faces multiple times, typically those of their caregivers and immediate community. While a few researchers have made an admirable start to examine the true statistics of face exposure (Fausey et al., 2016) this is a very difficult problem to address in practice, and is more usually ignored. Furthermore, while there are attempts to compare human and automatic face recognition systems (Phillips & White, 2025) these do not take into account the vastly dissimilar patterns of sampling by the two types of recogniser; for example, to see the 50 bn face images used to train *Clearview AI*, at a rate of one per second, a human would have to watch a continuous stream of faces for over 1500 years.

To examine the statistical generalisations available within a small number of faces, we built models based on just twenty people (Kramer et al., 2017). Using the standard PCA and LDA approach, we trained a model to categorise new images of the known twenty people highly accurately. In itself, this is a simple task, but we next examined the nature of the dimensions extracted by the model, in order to cohere together multiple images of the same person, while also distinguishing between people. It turned out that these dimensions were socially interpretable in an interesting way.

In one model, we constrained the set of twenty known individuals such that half were men and half women. Further, orthogonally, half the faces were Black and half White. In the model itself, these dimensions were not labelled – the task was simply to recognise identity over multiple images. However, it turned out that the first derived dimension coded sex almost perfectly. That is to say, over 99% of all learned and novel images fell on one end or the other of a single dimension, according to their sex. Figure 4 shows values for training and test items on the first dimension of this model. When novel faces were projected into this space, unrecognisable because the model had seen no previous instances of these people, they were nevertheless coded by sex on this dimension with over 95% accuracy. It seems then, that a system designed entirely to compute identity, had "discovered" sex as a major dimension on which to classify them. The second dimension behaved similarly, this time coding the dimension Black/White highly accurately for both known and previously unknown faces.

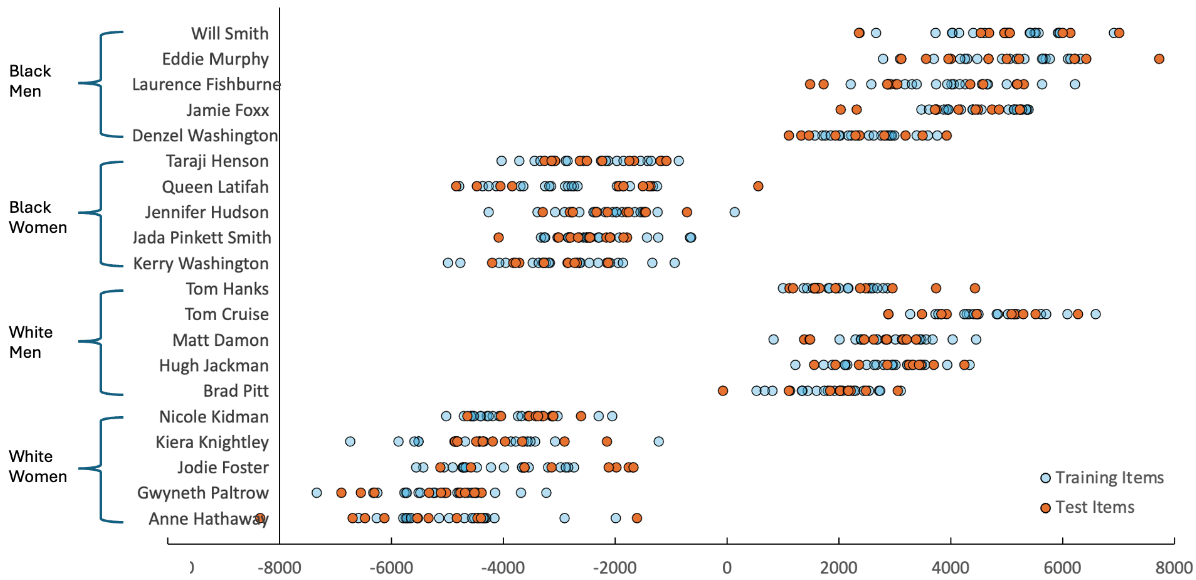

Fig 4. Values on derived dimension for a system trained to identify 20 individuals. Although the data is not labelled by sex, the model discovers this as an efficient classifier.

The computation of sex as an emergent property of identity, is particularly interesting, because perception of sex based on facial appearance has been difficult to understand. It has been known for many years that sex judgements are very fast and highly accurate, even in the absence of cultural adornments (Bruce et al., 1993; Martin & Macrae, 2007). There are many physical cues for sex, including 3d shape, coloration and brightness – but all these differences are subtle and show highly overlapping distributions, none of which could be used in isolation to deliver a fast "read-off" of sex (Brown & Perrett, 1993; Bruce et al., 1993; Burton et al., 1993; Hoss et al., 2005). The observation that sex classification simply emerges as a consequence of identification, and from a very small number of faces, shifts the focus away from an elusive perfect cue. Again, instead of seeking the "essence" of a perceptual representation, we find that the statistical process of telling faces apart and together delivers performance consistent with human behaviour.

Idiosyncrasy – a challenge for the future

A statistical approach to familiar face recognition alerts us to a fundamental difficulty for research in this area. The issue of *idiosyncrasy* arises in at least three different contexts. First, we have seen above that faces themselves vary in idiosyncratic ways. Second, there are large individual differences in face recognition ability. Third, we all know different people – individuals' personal history of exposure to faces is unique to them. So, the issue of idiosyncrasy arises for the *stimulus*, for the *capability* of the perceiver, but also for the *experience* of the viewer. I have discussed the first of these issues in detail above, but will finish this paper with a brief consideration of the other two sources of idiosyncrasy.

The study of individual differences in face recognition has become increasingly important recently (Wilmer, 2017; Yovel et al., 2014) and we have reviewed this field extensively elsewhere (White & Burton, 2022). It is now clear that these differences are large and stable (Baker et al., 2022), and there is strong evidence for a significant genetic contribution (Shakeshaft & Plomin, 2015; Zhu et al., 2010). However, almost all tests of these differences use unfamiliar faces. There is a general view in the literature that some people are better at recognising *familiar* faces than others, and there are certainly differences in people's self-reported abilities, though whether these self-reports correspond to actual ability remains a topic of debate (Bobak et al., 2019; G. Rhodes et al., 2016).

Overall, the nature of individual differences for familiar faces remains very poorly understood. One reason for this is a failure in the literature to acknowledge the profound differences between familiar and unfamiliar face processing – discussed at the start of this paper – and manifesting here as an assumption that patterns found with any face stimuli will generalise to all, including those that are familiar. However, there is also a more fundamental technical reason: it is very hard to develop a test of familiar face recognition. In almost all existing tests, participants are shown images of celebrities (perhaps difficult photos distorted in some way, or chosen to be untypical), and asked to identify them (Herzmann et al., 2008; McCaffery et al., 2018; Russell et al., 2009; Wilmer et al., 2012). The problem here is that failure to recognise a face could be due to poor ability, but could equally well indicate that the viewer simply does not know the person shown.

This leads us to the final issue of idiosyncrasy – each individual's unique experience with faces. As we discussed above, early theoretical models of face recognition took their inspiration from models of word recognition. But one substantial difference between the study of language (in lexical terms) and the study of faces, is that we can expect a very large shared vocabulary between all users of a language – in the order of tens of thousands of words. This is not the case for faces. In our study of how many faces people know (Jenkins et al., 2018), we were surprised by the small intersection of famous people known by different individuals, with pairs of individual participants sometimes recognising very few faces in common. Participants in that study were all undergraduates, and the problem becomes more extreme when we consider the different faces known by different sections of society, for example people of different ages, and with different professional and leisure interests. Furthermore, this fractionation of known faces is becoming more extreme as mass media, delivered by the internet, increasingly targets specific groups. There was a peak time in global media when relatively few channels *broad*cast to everyone – that time has now passed.

A possible response to this diversity in face knowledge is to suggest that one's specific experience is unimportant, it is sufficient that all participants in a psychology experiment will have seen many thousands of faces over their lives. However, this is inconsistent with evidence from elsewhere in the face literature. Most famously, the other-race effect (Malpass & Kravitz, 1969) is commonly interpreted as an effect of *exposure* to a specific set of people through life, shaping one's "face-space" through that exposure (O'Toole et al., 1994; Valentine, 1991). The phenomenon is very well known and routinely replicated, despite being an order of magnitude smaller than the effect of familiarity (Zhou et al., 2021).

Other findings also invoke the idea that one's individual exposure affects subsequent recognition substantially. For example, there have been suggestions that growing up in high- or low-density geographical areas influences one's subsequent ability to learn faces (Balas & Saville, 2017; Sunday et al., 2019). There are also own-age and own-gender biases reported which are normally explained in terms of exposure (Herlitz & Lovén, 2014; M. G. Rhodes & Anastasi, 2012). These effects all imply that we should be taking idiosyncratic exposure patterns seriously in face research – and that implies that we should probably consider an alternative to using standard face sets of "famous" faces as our stimuli.

A more recent approach has been to tailor stimuli to individual participants. For example, in a series of studies of familiar face recognition, led by Holger Wiese, we have asked participants to provide photos of their friends or family, of themselves, and also to list celebrities who they know well through the media (Wiese et al., 2022, 2023; Wiese, Ingram, et al., 2019; Wiese, Tüttenberg, et al., 2019). In this way, we can develop bespoke stimuli for each experimental condition for each participant. This approach is time-consuming and brings its own design problems. However, despite the difficulties this entails, we have argued that it is a worthwhile approach for tackling some issues.

Conclusions

I have argued for the critical importance of familiarity in understanding face recognition. First, a working assumption that all faces are processed in the same way seems unwarranted. In particular, for the purposes of recognition/identification, we are experts only in *some* faces, i.e. those we know (Young & Burton, 2018). Second, we need to understand the nature of the representations underlying our facility with familiar faces. I have argued here that most models assume, often implicitly, an essentialist stance to representation, whereas a statistical approach more naturally captures the range of observed behaviour. While some progress is being made in this regard, much remains to be done, and particularly in the understanding of sampling. We simply do not know the pattern of exposure we have to faces, day to day, and this will be a critical determinant of our recognition behaviours. Third, I have argued that idiosyncrasy is a key part of face recognition, and that this is studied very little. Idiosyncrasy manifests in the stimulus (people's faces vary idiosyncratically), the viewer's abilities (individual differences in perceptual skill) and the viewer's experience. For certain problems, our group-based approach to experimentation is not adequate to understand the phenomenon of face recognition.

An argument for the importance of idiosyncratic familiarity in face recognition suggests a similar issue in other within-class recognition domains. For example, we have analogous problems to solve when recognising our own car, pet or suitcase. While I have outlined the necessary requirements for a familiar face representation above, I have not addressed the more general phenomenon of recognition, as it applies to any stimulus. In recent work (Wiese et al., 2023), again using individually-tailored stimuli, we have examined ERP responses to familiar faces, pets, shoes and dwellings. Perhaps surprisingly, there are very similar effects of familiarity (i.e. differences between familiar and unfamiliar items) for all of these stimulus items. Common effects appear at relatively late processing stages, while earlier effects are quite divergent - presumably reflecting visual processing of these very different types of stimuli. While this work is in its early days, it suggests a way forward for understanding the nature of familiarity itself.

In the end, it seems likely that converging evidence will be most compelling in understanding face recognition. In just the same way that the Bruce and Young model (Bruce & Young, 1986) drew on multiple sources (experimentation, naturalistic observation, neuropsychological case studies etc), modern theorising is likely to be advanced further by diverse evidence. Individual studies rarely solve a problem conclusively – instead, methodological diversity is a strength. In this claim, I am following Sir Frederic Bartlett who, in his celebrated book *Remembering*, wrote: “While experimental psychology has to arrange external conditions with an eye to uniformity and control, the experimenter should never hesitate to break this external uniformity in the interests of stability of response.” (Bartlett, 1932/1995).

References

Baker, K. A., Stabile, V. J., & Mondloch, C. J. (2022). Stable individual differences in unfamiliar face identification: Evidence from simultaneous and sequential matching tasks. *Cognition*, *232*, 105333.

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision Research*, *141*, 228–236.

Bartlett, F. C. (1995). *Remembering*. Cambridge University Press. (Original work published 1932)

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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 711–720.

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly Journal of Experimental Psychology*, *72*(4), 872–881.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329.

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 M.D. Vernon Memorial Lecture. *Quarterly Journal of Experimental Psychology*, *47A*(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., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology. Applied*, *5*(4), 339–360.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology. Applied*, *7*(3), 207–218.

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

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *259*(1355), 137–143.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*(2), 153–176.

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

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

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., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), 482–496.

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

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

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

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.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *The Behavioral and Brain Sciences*, *16*(4), 681–694.

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.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107.

Fysh, M. C., & Bindemann, M. (2018). The Kent Face Matching Test. *British Journal of Psychology* , *109*(2), 219–231.

Galper, R. E., & Hochberg, J. (1971). Recognition memory for photographs of faces. *The American Journal of Psychology*, *84*(3), 351–354.

Geng, X., Zhou, Z.-H., & Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(12), 2234–2240.

Gobbini, M. I., & Haxby, J. V. (2006). Neural response to the visual familiarity of faces. *Brain Research Bulletin*, *71*(1–3), 76–82.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*(1), 32–41.

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage*, *22*(4), 1628–1635.

Goldstein, A. G., & Chance, J. E. (1970). Visual recognition memory for complex configurations. *Perception & Psychophysics*, *9*(2B), 237–241.

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

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.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology*, *15*(4), 445–464.

Herlitz, A., & Lovén, J. (2014). Sex differences and the own-gender bias in face recognition: A meta-analytic review. *Visual Cognition*, *21*(9–10), 1306–1336.

Herzmann, G., Danthiir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods*, *40*(3), 840–857.

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

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.

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings of the Royal Society B: Biological Sciences*, *285*(20181319). https://doi.org/10.1098/rspb.2018.1319

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

Kapsetaki, M. E., & Zeki, S. (2022). Human faces and face-like stimuli are more memorable. *PsyCh Journal*, *11*(5), 715–719.

Kemp, R. I., 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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(1), 103–108.

Kovács, G. (2020). Getting to Know Someone: Familiarity, Person Recognition, and Identification in the Human Brain. *Journal of Cognitive Neuroscience*, *32*(12), 2205–2225.

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46–58.

Kramer, R. S. S., Young, A. W., Day, M. G., & Burton, A. M. (2017). Robust Social Categorization Emerges From Learning the Identities of Very Few Faces. *Psychological Review*, *124*(2), 115–129.

Leippe, M. R., Wells, G. L., & Ostrom, T. M. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. *The Journal of Applied Psychology*, *63*(3), 345–351.

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *The Journal of Neuroscience*, *20*(2), 878–886.

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.

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law: An Official Law Review of the University of Arizona College of Law and the University of Miami School of Law*, *7*(1), 134–152.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334.

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.

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(21), 1–15.

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.

Mileva, M., Young, A. W., Jenkins, R., & Burton, A. M. (2020). Facial identity across the lifespan. *Cognitive Psychology*, *116*(101260). https://doi.org/10.1016/j.cogpsych.2019.101260

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178.

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar Variance Supports Robust Learning of Facial Identity. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(3), 577–581.

Noad, K. N., & Andrews, T. J. (2024). The importance of conceptual knowledge when becoming familiar with faces during naturalistic viewing. *Cortex*, *177*, 290–301.

Noad, K. N., Watson, D. M., & Andrews, T. J. (2024). Familiarity enhances functional connectivity between visual and nonvisual regions of the brain during natural viewing. *Cerebral Cortex*, *34*(7). https://doi.org/10.1093/cercor/bhae285

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*(2), 208–224.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research*, *38*(15), 2351–2363.

Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive Research: Principles and Implications*, *3*(1), 19.

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

Phillips, P. J., & White, D. (2025). The state of modelling face processing in humans with deep learning. *British Journal of Psychology (London, England: 1953)*. https://doi.org/10.1111/bjop.12794

Ramon, M., & Gobbini, M. I. (2018). Familiarity matters: A review on prioritized processing of personally familiar faces. *Visual Cognition*, *26*(3), 179–195.

Read, J. D., Vokey, J. R., & Hammersley, R. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy-confidence relationship. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *16*(5), 870–882.

Rhodes, G., Al-Janabi, S., Rossion, B., Malaspina, M., Daini, R., Taylor, L. C., Hall, B., Tez, T., Palermo, R., Irons, J., Laguesse, R., Albonico, A., McKone, E., & Rivolta, D. (2016). Do People Have Insight into their Face Recognition Abilities? *The Quarterly Journal of Experimental Psychology*, *70*(2), 218–233.

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological Bulletin*, *138*(1), 146–174.

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.

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology* , *70*(5), 897–905.

Ritchie, K. L., Smith, F. G., Jenkins, R., Bindemann, M., White, D., & Burton, A. M. (2015). Viewers base estimates of face matching accuracy on their own familiarity : Explaining the photo-ID paradox. *Cognition*, *141*, 161–169.

Robertson, D. J., & Burton, A. M. (2021). Checking ID-cards for the sale of restricted goods: Age decisions bias face decisions. *Applied Cognitive Psychology*, *35*, 71–81.

Russell, R., Duchaine, B. C., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257.

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, *14*(3), 398–409.

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences*, *112*(41), 201421881.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74.

Sunday, M. A., Patel, P. A., Dodd, M. D., & Gauthier, I. (2019). Gender and hometown population density interact to predict face recognition ability. *Vision Research*, *163*, 14–23.

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. 177–194). Oxford University Press.

Towler, A., Kemp, R. I., Burton, A. M., Dunn, J. D., Wayne, T., Moreton, R., & White, D. (2019). Do professional facial image comparison training courses work? *PloS One*, *14(2)*(e0211037), 1–17.

Turk, M. A., & Pentland, A. P. (1991). Face recognition using eigenfaces. *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference On*, 586–591.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204.

Wells, G. L., Lindsay, R. C., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *The Journal of Applied Psychology*, *64*(4), 440–448.

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology*, *1*(5), 287–300.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport Officers' Errors in Face Matching. *PloS One*, *9*(8), e103510.

Wiese, H., Hobden, G., Siilbek, E., Martignac, V., Flack, T. R., Ritchie, K. L., Young, A. W., & Burton, A. M. (2022). Familiarity is familiarity is familiarity: Event-related brain potentials reveal qualitatively similar representations of personally familiar and famous faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(8), 1144–1164.

Wiese, H., Ingram, B. T., Elley, M. L., Tüttenberg, S. C., Burton, A. M., & Young, A. W. (2019). Later but not early stages of familiar face recognition depend strongly on attentional resources: Evidence from event-related brain potentials. *Cortex*, *120*, 147–158.

Wiese, H., Popova, T., Lidborg, L. H., & Burton, A. M. (2024). The temporal dynamics of familiar face recognition: Event-related brain potentials reveal the efficient activation of facial identity representations. *International Journal of Psychophysiology*, *204*(112423), 112423.

Wiese, H., Schipper, M., Popova, T., Burton, A. M., & Young, A. W. (2023). Personal familiarity of faces, animals, objects, and scenes: Distinct perceptual and overlapping conceptual representations. *Cognition*, *241*, 105625.

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y. X., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A robust neural index of high face familiarity. *Psychological Science*, *30*(2), 261–272.

Wilmer, J. B. (2017). Individual Differences in Face Recognition: A Decade of Discovery. *Current Directions in Psychological Science*, *26*(3), 225–230.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology*, *29*(5–6), 360–392.

Young, A. W., & Burton, A. M. (2018). Are We Face Experts? *Trends in Cognitive Sciences*, *22*(2), 100–110.

Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology* , *76*(4), 495–523.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14*(6), 737–746.

Yovel, G., Wilmer, J. B., & Duchaine, B. C. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, *8*(August), 1–9.

Zhou, X., Burton, A. M., & Jenkins, R. (2021). Two factors in face recognition: Whether you know the person's face and whether you share the person's race. *Perception*, *50*(6), 524–539.

Zhou, X., & Jenkins, R. (2022). Face-evoked thoughts. *Cognition*, *218*(104955). https://doi.org/10.1016/j.cognition.2021.104955

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the Specific Cognitive Ability of Face Perception. *Current Biology*, *20*(2), 137–142.