Unfamiliar and newly learned face identification: An examination of individual differences.

Kristen A. Baker, MA

Psychology (Behavioural and Cognitive Neuroscience)

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Social Sciences, Brock University
St. Catharines, Ontario

Abstract

Face identification involves two tasks: Recognizing an individual even when their appearance changes, and discriminating them from similar-looking individuals. People vary in the accuracy with which they identify unfamiliar faces. Much of the work investigating individual differences in face identification used tightly controlled stimuli (i.e., focused on discrimination). Few studies have used stimuli that incorporate variability in appearance (e.g., focused on recognition). Despite interest in individual differences, and understanding that recognizing a face across instances poses a difficult challenge, many gaps in the literature remain. These include potential predictors, the reliability and convergent validity of face identification tasks, and whether unfamiliar face identification predicts face learning efficiency.

I examined a potential predictor of face identification—photography experience (Chapter 2). I recruited photography Experts, Hobbyists and Novices to take part in an unfamiliar face identification task. Photography experience was not a significant predictor of sensitivity in unfamiliar face identification. However, it was a predictor of response bias.

I examined the reliability and convergent validity of face identification tasks (Chapter 3). Participants completed four unfamiliar face identification tasks on two days (study 1), or two versions (simultaneous and sequential) of three unfamiliar face identification tasks (study 2). Sensitivity to identity and bias were stable across time and tasks. Response times were fastest on trials that were congruent vs. incongruent with one's bias, providing preliminary evidence that this reflects decision-making processes.

I examined whether unfamiliar face matching predicts face learning efficiency (Chapter 4). Participants completed two unfamiliar face matching tasks and a novel face learning task (which tested recognition four times during learning). Individual differences in the slope of face learning were predicted by unfamiliar face matching ability. These differences appear to be driven by individual differences in recollection, not familiarity.

My dissertation provides insights about individual differences in face identification. Individual differences in sensitivity in unfamiliar face identification were stable across time and tasks. They also predict face learning efficiency. My results suggest that face identification is not just a perceptual problem—it is influenced by decision making and other processes. These results have implications for face identification theories and applied settings.

Key Words: Face identification, face recognition, face learning, within-person variability, individual differences.

Acknowledgements

First, I would like to thank my supervisor, Cathy Mondloch. Your support through my PhD was paramount to my success. I am grateful for your encouragement, words of wisdom and the training that you have provided (e.g., writing, research design)—I am well prepared for my upcoming postdoctoral fellowship and eventual tenure track position. When I have my own lab, I hope that I can be as good of a mentor as you. I would also like to thank my incredibly supportive doctoral Committee members: Karen Arnell and Stephen Emrich. In addition to your thoughtful suggestions, you've always been willing to go way beyond what is expected of committee members by answering many of my questions and providing advice. Your time and effort helped me to grow into a confident researcher.

I would also like to thank past and current members of The Face Perception Lab. Sarah Laurence, you were my first mentor when I joined the lab as an RA and you've never stopped mentoring me. I am incredibly grateful to have you as a mentor, and as a friend. Meryl Ochoa and Mariah Parente, I would like to thank you both for your help with stimuli and coding the Sorting Task data. Vince Stabile, thank you for all of your help in the lab. Mentoring you during your undergrad was incredibly fun and rewarding! Finally, my lab mates (Sophia, Anita, and Claire), your support encouragement and humour through this process has been incredible.

I would also like to take the time to thank my other friends and graduate students in the psychology department at Brock University. Andrew, Holly, Joel, Christine, Kari, Elvira, Jen, Nick and Nadia, you all provided much needed support and humor throughout my entire PhD. I would especially like to thank Brent Pitchford, who provided me with support, care and even cooked for me during the most stressful times (e.g., when I was spending all of my time getting my dissertation on track during the covid-19 pandemic).

This dissertation was only possible because of the support from my family, who have been my cheerleaders through this whole process. My parents have never stopped supporting me, even when I asked them to test out my experiments, when I practiced my presentations in front of them, and when I talked (extensively) about the innerworkings of my research, how to publish journal articles, and academia in general. I would be remiss if I didn't also acknowledge, my Grandma and Papa Baker (who teared up when they found out I was starting my PhD), my Grandma and Papa Peddle (who kept announcing to the world that there was going to be "a doctor in the family"), as well as my Aunts (e.g., Judy, Dale) and Uncles (e.g., Rod, Dean) who never stopped believing in me.

Table of contents

List of Tables

List of Figures

List of Abbreviations

AIFMT	Ambient Image Face Matching Task
ANOVA	Analysis of Variance
CCTV	Closed-circuit television
CFMT	Cambridge face memory test
DCNN	Deep convolutional neural networks
DPSD	Dual process signal detection
FA	False alarm
FRU	Face recognition unit
GFMT	Glasgow face matching test
ID	(i.e., Photographic ID) Identification
PIN	Person identity node
SDT	Signal detection theory
VET	Vanderbilt expertise test

Chapter 1: General Introduction

Modern society functions in a manner that is reliant on face identification. For example, an inability to recognize friends and colleagues could leave one feeling socially isolated. Another example is our use of Government issued ID (e.g., passport, driver's licence, health card). This ID enables millions of Canadians to cross international borders or to purchase age-restricted goods. Face identification is even used in criminal cases, such as in those that use eyewitness testimony or when police use CCTV to catch the perpetrator of a crime. The challenge in each of these cases is to recognize the target identity, even if their appearance changes, while discriminating the target from individuals who look similar. The wide-spread use of face identification in these high-cost situations is likely the result of a conflation of familiar and unfamiliar face identification (for a discussion see Burton, 2013). The ease with which we recognize familiar faces led to the common misconception that we are experts at recognizing all faces, rather than just the faces of people we know.

Historical work into face identification focused primarily on the ability to tell two faces apart (i.e., discrimination). Past research argued that faces are a homogeneous class of stimuli, with each face comprising two eyes above a nose, above a mouth (for a discussion see Maurer et al., 2002; Tanaka & Gordon, 2011). Because of this homogeneity of face stimuli, it was argued that the challenge of face identification was being able to discriminate between two similar-looking faces (for examples see Bruce et al., 1999; Mondloch et al., 2010; Tanaka & Farah, 1993). To assess this fine-tuned discrimination, it made sense to use stimuli that were tightly controlled and manipulated such that there was only a slight change in one aspect of the face, such as feature spacing or shape (for examples see Mondloch et al., 2002; Tanaka & Farah, 1993).

This approach provided many insights into the factors influencing face identification, including the finding that faces are processed holistically (e.g., the features of a face are processed as a gestalt; Maurer et al., 2002) and that experience influences performance (e.g., Le Grand et al. 2001; Rhodes et al., 2006; Sugita, 2008; for a discussion see Valentine, 1991). However, this approach ignored half of the problem of face identification—recognizing a face despite variability in appearance (e.g., changes in hairstyle, make-up, facial hair; for a discussion see Burton, 2013). This insight about the impact of within-person variability in appearance on face identification turned the field of face identification onto its head. When studies incorporate stimuli containing within-person variability in appearance, the dramatic difference between recognition of familiar vs. unfamiliar becomes clear, raising a wealth of questions about how such faces are represented and the process by which an unfamiliar face becomes familiar.

The overarching goal of my dissertation research was to investigate individual differences in face identification. Many studies have focused on the differences between familiar and unfamiliar face identification (e.g., Burton & Jenkins, 2011; Jenkins et al., 2011; Johnston & Edmonds, 2009). I have focused on unfamiliar face identification. I examined predictors of unfamiliar face identification, reliability and convergent validity of unfamiliar face identification, and whether unfamiliar face identification predicts the efficiency with which a newly encountered face is learned.

Familiar and Unfamiliar Face Identification

The distinction between familiar vs. unfamiliar faces was noted by Bruce and Young (1986). In their model, unfamiliar faces are identified based on an abstract, view-independent representation. This representation is limited based on the nature of the exposure to the face (e.g., the particular photograph viewed or the individual's appearance on a single encounter). In contrast, familiar faces have a robust representation, and the recognition of these faces activates both Face Recognition Units (FRU) and Person Identity Nodes (PIN). When a FRU is activated by a close enough match between an instance and the corresponding visual representation of the identity, it then activates the PIN (a system that stores information such as an individual's name, voice). This activation facilitates recall of the identity.

The most compelling evidence for Bruce and Young's distinction of unfamiliar and familiar face identification was published 25 years after their theory was published. Jenkins and colleagues (2011) asked participants to sort 40 ambient face photographs (i.e., images that contain naturally occurring variability in appearance) into piles. Participants were informed that each pile that they made should contain all images of one identity. Participants were unaware that the photographs comprised 20 instances of each of two individuals. Participants overestimated the number of identities that were present in the photos when the faces were unfamiliar ($Median_{\text{piles}} = 7.5$), but performed the task perfectly when the faces were familiar ($Median_{\text{piles}} = 2$). The representation for familiar faces can easily tolerate changes in appearance, and the representation of unfamiliar faces cannot.

This work galvanized more research in this area, much of which focused on between-group comparisons. For example, whereas children aged 5 to 11 tolerate less variability in appearance than do adults when identifying unfamiliar faces, by 6 years of age (but not younger) children perform at an adult-like level for recognizing familiar faces (Laurence & Mondloch, 2016; Matthews et al., 2022). Similarly, whereas adults tolerate less variability in unfamiliar other- vs. own- race faces (Laurence et al., 2016), they perform perfectly when asked to sort images of familiar other-race faces (Zhou & Mondloch, 2016). Such findings suggest that face-specific experience shapes our ability to match identity in images of wholly unfamiliar faces.

Like familiarity with specific identities, and general experience with faces, professional experience with faces might influence face identification performance. Researchers have examined whether adults in some occupations have superior face identification skills. Many studies have found minimal effects of profession on face identification abilities (e.g., notaries: Papesh, 2018; border patrol officers: White et al., 2014). However, occupation does relate to performance on unfamiliar identification tasks for forensic examiners and artists. Forensic examiners, who are trained to use a featural approach when identifying unfamiliar faces, outperform novices on face identification tasks (Towler et al., 2017), an effect that is attributable to the strategy these experts used. When novices are trained to use this same strategy, their performance improves by around 6% (Towler et al., 2017). Artists out-perform novices on the Cambridge Face Perception Task [CFPT], a sequential matching task and on old/new recognition tasks when the same image of a face is presented during the learning and recognition phases (Devue & Barsics, 2016; Hsiao et al. 2021).

Individual differences in face identification

Differences in face identification abilities across occupations points towards the importance of using an individual difference approach for investigating face identification. Investigating face identification using an individual differences approach is an untapped resource that can provide valuable insights about the mechanisms underlying face identification and learning (see White & Burton 2022 for a review). The ability to match identity in unfamiliar faces varies across individuals (for reviews see Bruce et al., 2018; Lander et al., 2018; Wilmer, 2017). Much of the research investigating individual differences in face identification has compared control participants to individuals at the extreme ends of the distribution—individuals with prosopagnosia (those who struggle with identifying even highly familiar faces; for a review see Kress & Daum, 2003) or super-recognizers (those who have exceptional face-identification abilities; see Russell et al., 2009). Less research has investigated individual differences within the typical population. Such research has focused on the extent to which performance is stable across tasks and, to some extent, consistent across time.

Individual differences in sensitivity to identity

Face identification tasks involve many different protocols. Same/different tasks (e.g., Burton et al., 2010, Fysh & Bindemann, 2018; White et al., 2014) require participants to decide whether a pair of images belong to the same person. Lineup tasks (e.g., Bindemann et al., 2012; Megreya & Bindemann, 2013; Megreya & Burton, 2008) require participants to determine which image (if any) in a set of images belongs to a target identity. Sorting tasks require participants to match identity across many variable instances—without knowing the total number of identities present. Tasks also vary in the extent to which their stimuli incorporate natural within-person variability in appearance. For example, whereas match trials of the Glasgow Face Matching Test (GFMT; Burton et al., 2010) comprise two images an individual that were taken minutes apart with different cameras, match trials in the Ambient Image Face Matching Test (AIFMT; Baker et al., in press; Baker & Mondloch, 2022) were taken on different occasions in which the individual’s appearance changed as a function of hair style, make up, facial hair, lighting. Whether individual differences are stable across tasks has received scant attention. If a unitary skill underlies performance across tasks, then individual differences in one task should correlate with individual differences in another, despite variability in task parameters.

Recent studies (Fysh et al., 2020; McCaffery et al., 2018; Stacchi et al., 2020; Verhallen et al., 2017) have examined individual differences in face identification across a battery of tasks. Some, but not all tasks, required participants to match identity in images of wholly unfamiliar faces, although many used the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006). Very few studies incorporated stimuli that contain natural variability in appearance (for examples see Fysh & Bindemann, 2018; McCaffery et al., 2018). Across studies, relationships between tasks varied in strength (R^2s = 0.004 to 0.61). Whereas Stacchi et al. (2020) reported a null relationship between performance on the Sorting Task and same/different tasks, Fysh et al. (2020) reported a significant relationship between the Sorting Task and their measure of face matching. These findings suggest the need to investigate convergent validity across various types of face identification tasks to establish whether they are all capturing the same ability. Several tasks that have been used as measures of individual differences in face identification (e.g., lineup tasks, sorting tasks) have yet to be examined for test-retest reliability (see White & Burton, 2022). For these, there is no estimate of how stable performance is across time and thus, there is no estimate of how accurately these measures capture individual differences.

Individual differences in criterion

Face identification researchers tend to approach face identification as purely a perceptual problem. As a result, they ignore the influence of decision making. Bindemann and Burton (2021) called on face identification researchers to better integrate decision making into models of face identification. One of the ways we can begin to examine decision making in face identification research is by investigating criterion, as it is considered the boundary that one must reach to make a decision (Summerfield and Egner, 2013). Conservative and liberal response bias have their own associated costs. For example, an eyewitness with a liberal response bias is more likely to select an innocent individual from within a lineup (e.g., around 70% of falsely incarcerated individuals were prosecuted from faulty eyewitness testimony; See Innocence

Project, 2022). Conversely, an eyewitness with a conservative response bias is more likely to let a guilty person go free.

Very little is known about individual differences in criterion across face identification tasks. There is some evidence of age-related changes in criterion during face identification (Megreya & Bindemann, 2015), such that liberal response biases tended to increase with age. As well, although oxytocin administration improves recognition in prosopagnosics (Bate et al., 2014), it merely shifts criterion in typical populations (Bate et al., 2015)—and when examined directly, criterion does differ between prosopagnosics and typical controls (White et al., 2017; supplemental material). The recognition memory literature also hints that the individual differences in criterion are consistent across a wide set of different stimulus types (see Kantner & Lindsay, 2012, 2014). Although widely unexplored in the face identification literature, these effects seem to support the idea that there are individual differences in criterion.

Individual differences in face learning

Consistent differences in familiar and unfamiliar face identification led to research investigating how a newly encountered face becomes familiar. Exposure to variability in appearance facilitates face learning (e.g., Andrews et al., 2017; Baker et al., 2017; Dowsett et al., 2016; Matthews & Mondloch, 2018; Menon et al., 2015; Murphy et al., 2015; Ritchie & Burton, 2017). For example, participants perform better at both face naming tasks and face matching tasks when trained on the identities with high vs. low variability images (Ritchie & Burton, 2017). Children also learn to recognize a newly encountered identity better after watching a video filmed across 3 (high variability) vs. 1 (low variability) days (Baker et al., 2017). To date no study has examined individual differences in the ability to capitalize on variability as a newly encountered face becomes familiar.

There is some evidence to suggest that unfamiliar face matching abilities might predict individual differences in the efficiency with which one learns a new face (i.e., slope of face learning; the amount of benefit one gains from additional exposure to variability in appearance). Group differences in face matching correspond to group differences in face learning. Adults from small towns perform more poorly than those from large towns on unfamiliar face identification tasks and on the Cambridge Face Memory Test (Balas & Saville, 2015, 2017); children perform worse on face learning tasks than adults (Baker et al., 2017; Laurence & Mondloch, 2016); and adults perform worse on unfamiliar face matching and learning tasks when tested with inverted as compared to upright faces (Kramer et al., 2017; see Valentine, 1988). To date, no study has examined the extent to which individual differences in matching ability predict the slope of face learning.

There is also evidence that recollection (e.g., explicit memory details) and familiarity (e.g., an implicit, unconscious sense of "knowing") might change during the process of face learning and/or might capture individual differences. For example, individual differences in CFMT performance not only predict familiarity responses, but also predict correct reports of episodic details for familiar faces—a proxy for recollection (Devue et al., 2019). I examined whether individual differences in face matching predict individual changes in two processes that underlie performance in face learning: Recollection vs. familiarity. I predicted that there would be a relationship because the representation of unfamiliar faces might influence ability to build a robust representation of a new face.

Theoretical models

My dissertation research does not explicitly test theories of face identification or theories of decision making. Nonetheless, my research questions were inspired by theory and my data provide novel insights with theoretical implications.

Theories of Face Identification

Bruce and Young. Bruce and Young's (1986; see also Young & Bruce, 2011) model accounts for both familiarity and within-person variability effects. In this model, Face Recognition Units (FRUs), the aspect of the model that comprise our representations of familiar faces, are activated by a recognizable image of an identity. The activation of the FRU then activates the associated Person Identity Node (PIN). Both the FRU and PIN are all-or-none processes, such that if a threshold is reached, a person is recognized. Bruce and Young proposed that the representation of unfamiliar faces is constrained by the nature of the encounter. Only a close enough match to the representation will result in a "match" response. In Bruce & Young's model, then, the transition from an unfamiliar to familiar face involves building a robust representation. To the extent that this transition is influenced by the quality of the unfamiliar face representation (i.e., one that does not contain cues that are not diagnostic of identity), this model might predict a positive correlation between unfamiliar face matching ability and the efficacy of face learning.

Deep convolutional neural network face space. Past conceptualizations of face identification proposed that faces are represented as individual points within a multi-dimensional face space (Valentine, 1991). Recent work using deep convolutional neural networks (DCNNs) redefined and extended this model (O'Toole et al., 2018; Hill et al., 2019; O'Toole & Castillo, 2021). DCNNs are a class of learning algorithm that provide a potential model for how humans learn and represent faces. DCNNs are based on the human visual system (Phillips et al., 2018; Noyes et al., 2021). They make errors when identifying unfamiliar faces, and the number of identities on which the algorithm is trained influences the number of errors made (Blauch et al., 2021; Rosemblaum et al., 2021)—even for learned identities that are disguised (Noyes et al., 2021). Training a DCNN entails exposing the algorithm to many labelled images that incorporate natural variability of many different identities (e.g., hundreds of images each of a thousand different identities). Although DCNNs are trained on specific identities, when the "top layer" is removed, this eliminates all knowledge about specific identities despite leaving the neural architecture intact (i.e., all faces become unfamiliar; O'Toole et al., 2018). These results suggest that representations of unfamiliar faces could influence the efficacy with which a new face is learned.

Theories of Decision Making

Bindemann and Burton (2021) discussed that decisions in face identification might be reached through a gradual process of gathering evidence for a response option. They suggested that the process of aggregating evidence and reaching a decision could occur through counting or reduction strategies. A counting strategy would comprise an internal counter which tracks the amount of evidence (i.e., evidence accumulation) for either response option (e.g., same, different). In unfamiliar face identification tasks, if two photos of the same person are very similar (i.e., taken on the same day) the amount of evidence to suggest that the photos are of the same person would be grater than the evidence that they are different people. Of course, one would choose the option with more supporting evidence than another option. A reduction strategy would entail accounting for the strength/predictability of each individual piece of evidence (i.e., evidence weighting). In unfamiliar face identification tasks, some evidence will come from sources that are more variable than others; this would contribute to the evidence's relative weighting (e.g., if two faces have a similar hair cut and similar earrings, but one face has brown eyes and the other has blue eyes, eye colour would be weighted as having a higher evidentiary value). Bindemann and Burton also discussed how one might shift decision strategies depending on the situation (e.g., if mismatches are more frequent). These hypotheses about how decision making might influence face identification can be expanded further as several decision-making theories exist.

Signal Detection Theory. Signal detection theory is a model of perceptual choice (Summerfield & Egner, 2013). It proposes that dichotomous decisions are dependent on two factors: 1) The likelihood of whether the observation was from a signal or noise distribution, and 2) the set position of a criterion. Criterion can therefore be used as an initial investigation of decision making. However, as criterion does not account for the whole decision-making process and can conflate noise attributed to the stimulus and decisional noise (Mueller & Weidemann, 2008), the results from studies using criterion should be further investigated using other measures of decision making. Investigating criterion in face identification is important as individuals with similar sensitivity to identity can have different patterns of errors because of different placements of their criterion. For example, despite having similar accuracy, forensic examiners and super recognizers differ in their bias (Towler et al., 2021). Super recognizers vs. forensic examiners have a more liberal bias, and when super recognizers do make errors, they are more likely to make high confidence false alarms. Findings such as these can highlight differences in decision making merely by investigating sensitivity and criterion; measuring only accuracy would mask these important individual differences.

Drift diffusion models. Drift diffusion models are an extension of signal detection theory. They propose that three parameters influence decision making: The starting point, drift rate, and criterion (Ratcliff et al., 2016). The starting point is the point at which evidence accumulation starts. The placement of the starting point can be biased towards a particular option. This bias would influence the amount of evidence required to decide on that option. In face matching, this starting point might vary as a function of the anticipated proportion of match vs. mismatch pairs. The drift rate is a measure of the speed at which evidence is accumulated and the evidence quality. In face matching this might have to do with comparing the number of similarities and differences across image pairs (see Bindemann & Burton, 2021)—of course, this would entail weighting the quality of evidence, such that easily changed attributes (e.g., length of hair) are not treated as having the same evidentiary value as attributes that are less variable (e.g., eye colour). The criterion is the amount of evidence that is required to make a decision. In face matching, this threshold might vary as a function of the cost/reward associated with making a response. According to drift diffusion models, evidence and noise are accumulated until they reach a criterion (Ratcliff et al., 2016).

Dual process models. Dual process theories propose that two systems are involved in decision making (Evans & Stanovich, 2013). System one is rapid and automatic. It requires little working memory capacity and attention for a decision to be reached. System two is associated with slow, deliberate processes. It requires working memory capacity and attention for a decision to be reached. When a problem requires an effortful solution, system two inhibits system one. These processes have also been reconceptualized using Fuzzy Trace Theory in which verbatim and gist are the two systems, which are mediated by a third system—inhibition (see Weldon et al., 2013). There is some evidence of the use of dual-process decision making in face identification. Typical observers use a fast holistic process to identify faces (Richler et al., 2009). Forensic examiners, known for their accuracy in face identification, use a slow and deliberate process to make unfamiliar face identification judgements (Towler et al., 2017; for discussions see Kemp et al., 2021; White et al., 2021). When novices are trained to use this process, their performance shows a modest (6%) improvement (Towler et al., 2017). Moreover, when forensic examiners are given a limited time (i.e., 2 s) in which they can indicate their responses, their performance decreases (for a discussion see Kemp et al., 2021). This suggests that individual differences in face identification might reflect differences in the use of system one and system two processing.

Although I cannot distinguish among these models, they highlight the need to consider individual differences and to better integrate decision making in face identification research. In my research, I integrate theories of face identification and decision making, and use an individual differences approach. By using these two theoretical lenses. My thesis speaks to Bindemann and Burton’s (2021) call to integrate decision making into studies of face identification and to White and Burton’s (2022) call to take an individual difference approach.

The current research

White and Burton (2022), discussed the need to use individual difference approaches to further models of face identification. There are many gaps in the literature including: 1) determining significant predictors of face identification, 2) the reliability and convergent validity of many tasks, and 3) whether matching ability predicts face learning efficiency. Here, I sought to fill these gaps in the literature by examining individual differences in face identification.

Chapter 2

Although photography experience has never been examined as a predictor of face matching ability, there are hints in the literature that suggest that it might. For example, artists out-perform novices on the Cambridge Face Perception Test [CFPT], a sequential matching task and on old/new recognition tasks when the same image of a face is presented during the learning and recognition phases (Devue & Barsics, 2016; Hsiao et al. 2021). Whether photography experience has a similar effect is unknown, but an important point to establish. Many training programs in which those who check photographic ID are enrolled incorporate some sections of the training on photography (see Towler et al., 2019)—however, it is not yet understood whether photography experience even influences face identification abilities.

I examined this question in Chapter 2 by recruiting participants with varying levels of photography experience: Experts, Hobbyists and Novices. Participants completed the Ambient Image Face Matching Task—a task that I developed to measure unfamiliar face matching abilities in the context of naturalistic within-person variability in appearance. Participants then answered several questions regarding their photography experience. I showed that although there was no effect of photography experience on sensitivity to identity, there was an effect on criterion. My findings suggest that individuals who are hobbyists, and thus likely to take multiple photos of a few people, become more tolerant of within-person variability without becoming more sensitive to identity. This has implications for training programs which incorporate modules on the way that photography can influence the appearance of a face (Towler et al., 2019).

Chapter 3

There are some hints in the literature suggesting that face identification tasks should be reliable. The CFMT is a widely used measure of face memory for which reliability has been well established. Likewise, the GFMT is a widely used measure of unfamiliar face matching for which reliability has also been established. However, neither of these measures incorporate within-person variability in appearance. As such, they do not capture a challenge in face identification that is experienced in daily life: Recognition of an identity across multiple instances. Moreover, reliability for the above tasks has only been established for accuracy. To date, no studies have reported test-retest reliability of criterion in any task, nor have any studies investigated the convergent validity of criterion across face identification tasks. The failure to investigate the reliability and stability of criterion across time and tasks, has left many unanswered questions about criterion's ability to serve as an individual difference measure. Indeed, if criterion cannot serve as a quality individual difference measure, then it's ability to provide necessary information about decision making and the pattern of responses that participants make, is quite limited.

Chapter 3 comprises two studies in which I investigated individual differences in the ability to match identity in wholly unfamiliar faces. Across studies 1 and 2, I examined the extent to which individual differences in sensitivity and criterion were stable across time and tasks. In study 1, on each of two sessions (approximately a week apart) participants completed a battery of four face identification tasks. I showed that individual differences in sensitivity and bias were stable across time and tasks, and that the number of piles made in the sorting task might more closely resemble a measure of bias than of sensitivity. In study 2, participants completed two versions (a simultaneous version and a sequential version) of each of three face identification tasks within a single session. I showed that the stability of behaviour remains consistent regardless of whether stimuli are presented simultaneously vs. sequentially. These results suggest that individual differences in both sensitivity and bias are stable across time and tasks—even when memory demands are introduced by presenting stimuli sequentially.

Chapter 4

Individual differences on tasks measuring face memory correlate with individual differences on unfamiliar face matching tests (for examples see Fysh et al., 2020; McCaffery et al., 2018; Verhallen et al., 2017) and group differences in unfamiliar face identification correspond to group differences in face learning (e.g., small town effects, developmental effects and inversion effects all influence performance in unfamiliar face matching and memory or learning: Balas & Saville, 2015; Balas & Saville, 2017; Baker et al., 2017; Laurence & Mondloch, 2016; Kramer et al., 2017; see Valentine, 1988). These two findings suggest that individual differences in unfamiliar face matching might predict individual differences in the slope of face learning. Whereas previous tasks have measured the endpoint of face learning, no study has measured performance as learning unfolds. Further, most studies of face learning have measured sensitivity (d′) or provided separate measures of recognition (hits) and discrimination (correct rejections). This approach has provided many insights but does not allow one to examine underlying processes – familiarity vs recollection.

Chapter 4 comprises a study in which I investigated two questions. First, I examined whether individual differences in face matching predict the slope of face learning. Second, I examined whether familiarity and/or recollection (old, new) reflect changes during face learning and whether changes in familiarity and/or recollection vary as a function of matching ability. Participants completed three face identification tasks. Two of these tasks were previously established measures of face matching. The third task was a novel task that I developed to measure the slope of face learning. I showed that individual differences in matching ability predicted the slope of face learning, such that individuals with poor matching abilities benefitted less from variability in appearance than those with good matching abilities. Importantly, the amount of benefit gained from exposure to variability was predicted by matching ability. I also showed that whereas familiarity responses do not change as a function of either individual differences or learning, recollection "old" responses do. Moreover, matching ability influences the recollection new responses differently at different levels of learning. Although poor matchers never improve on this parameter, both mid-level matchers and good matchers do improve during the learning process—however, mid-level matchers take longer to show a benefit. These results suggest that one's representation of newly learned identities changes in a categorical, episodic way. These findings also suggest that unfamiliar face matching predicts differences in the ability to confidently perceive a distractor as new while learning unfolds.

Collectively, these studies suggest that the two components that underlie face identification (sensitivity, criterion) are related across time and tasks—even when the tasks contain different demands such as memory or learning. Sensitivity to identity predicts the slope of face learning, and the benefit one receives from variability for perceptual discrimination (i.e., recollection new). These studies also suggest that just because a variable is not a significant predictor of sensitivity, does not mean that it is not a useful predictor, as it might predict criterion—just as photography experience had. These findings have theoretical and applied significance, which is discussed in the General Discussion.

References

Andrews, S., Burton, A. M., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. *Quarterly Journal of Experimental Psychology*, *70*(8), 1620-1632. https://doi.org/10.1080/17470218.2016.1195851

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19-30. https://doi.org/10.1016/j.cognition.2016.12.012

Baker, K.A. & Mondloch, C. J. (2022). Picture this: Photographers no better than controls for recognizing unfamiliar faces. Perception. 51(8), 591–595. https://doi.org/10.1177/03010066221098727

Baker, K.A., Stabile, V.J., & Mondloch C.J. (In Press). Stable individual differences in unfamiliar face identification: Evidence from perceptual matching and memory tasks. *Cognition*.

Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. *Neuropsychologia*, *69*, 211-217. https://doi.org/10.1016/j.neuropsychologia.2015.02.005

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision research*, *141*, 228-236. https://doi.org/10.1016/j.visres.2016.12.005

Bate, S., Cook, S. J., Duchaine, B., Tree, J. J., Burns, E. J., & Hodgson, T. L. (2014). Intranasal inhalation of oxytocin improves face processing in developmental prosopagnosia. *Cortex*, *50*, 55-63. https://doi.org/10.1016/j.cortex.2013.08.006

Bate, S., Bennetts, R., Parris, B. A., Bindemann, M., Udale, R., & Bussunt, A. (2015). Oxytocin increases bias, but not accuracy, in face recognition line-ups. *Social cognitive and affective neuroscience*, *10*(7), 1010-1014. https://doi.org/10.1093/scan/nsu150

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(2), 96-103. https://doi.org/10.1016/j.jarmac.2012.02.001

Bindemann, M., & Burton, A. M. (2021). Steps towards a cognitive theory of unfamiliar face matching. *Forensic face matching: Research and practice*, 38-61.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2021). Deep learning of shared perceptual representations for familiar and unfamiliar faces: Reply to commentaries. *Cognition*, 208, 104484. https://doi.org/10.1016/j.cognition.2020.104484

Bruce, V., Bindemann, M., & Lander, K. (2018). Individual differences in face perception and person recognition. *Cognitive Research: Principles and Implications*, *3*(1), 1-3. https://doi.org/10.1186/s41235-018-0109-4

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. Journal of Experimental Psychology: Applied, 5(4), 339. https://doi.org/10.1037/1076-898X.5.4.339

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, 77(3), 305-327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1467-1485. https://doi.org/10.1080/17470218.2013.800125

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. The Oxford handbook of face perception, 28, 287-306.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior research methods*, *42*(1), 286-291. https://doi.org/10.3758/BRM.42.1.286

Devue C., Barsics C. (2016). Outlining face processing skills of portrait artists: perceptual experience with faces predicts performance. *Vision Research*, 127, 92–103. https://doi.org/10.1016/j.visres.2016.07.007

Devue, C., Wride, A., & Grimshaw, G. M. (2019). New insights on real-world human face recognition. *Journal of Experimental Psychology: General*, *148*(6), 994. https://doi.org/10.1037/xge0000493

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. https://doi.org/10.1080/17470218.2015.1017513

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

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. *British journal of psychology*, 109(2), 219-231. DOI:10.1111/bjop.12260

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society open science*, 7(9), 200233. https://doi.org/10.1098/rsos.200233

Hill, M. Q., Parde, C. J., Castillo, C. D., Colon, Y. I., Ranjan, R., Chen, J. C., Blanz, V. & O'Toole, A. J. (2019). Deep convolutional neural networks in the face of caricature. Nature Machine Intelligence, 1(11), 522-529. https://doi.org/10.1038/s42256-019-0111-7

Hsiao J. H., An J., Zheng Y., Chan A. B. (2021). Do portrait artists have enhanced face processing abilities? Evidence from hidden Markov modeling of eye movements. Cognition, 211, 104616. https://doi.org/10.1016/j.cognition.2021.104616

Innocence Project (2022). Retrieved from: https://innocenceproject.org/eyewitness-identification-reform/.

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

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. Memory, 17(5), 577-596. https://doi.org/10.1080/09658210902976969

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & cognition*, 40(8), 1163-1177. https://doi.org/10.3758/s13421-012-0226-0

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, *21*(5), 1272-1280. https://doi.org/10.3758/s13423-014-0608-3

Kemp, R. I., Edmond, G., & White, D. (2021). A proposed solution to the problem of identifying people from CCTV and other images. In *Methods, Measures, and Theories in Eyewitness Identification Tasks* (pp. 13-33). Routledge.

Kramer, R. S., Jenkins, R., Young, A. W., & Burton, A. M. (2017). Natural variability is essential to learning new faces. *Visual Cognition*, 25(4-6), 470-476. https://doi.org/10.1080/13506285.2016.1242522

Kress, T., & Daum, I. (2003). Developmental prosopagnosia: A review. *Behavioural neurology*, 14(3, 4), 109-121. https://doi.org/10.1155/2003/520476

Lander, K., Bruce, V., & Bindemann, M. (2018). Use-inspired basic research on individual differences in face identification: Implications for criminal investigation and security. *Cognitive Research: Principles and Implications*, 3(1), 1-13. https://doi.org/10.1186/s41235-018-0115-6

Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, *143*, 123-138. https://doi.org/10.1016/j.jecp.2015.09.030

Le Grand, R., Mondloch, C., Maurer, D. *et al.* Early visual experience and face processing. *Nature* **410**, 890 (2001). https://doi.org/10.1038/35073749

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*(1), 1-15. https://doi.org/10.1186/s41235-018-0112-9

Matthews, C. M., & Mondloch, C. J. (2018). Finding an unfamiliar face in a line-up: Viewing multiple images of the target is beneficial on target-present trials but costly on target-absent trials. *British Journal of Psychology*, 109(4), 758-776. https://doi.org/10.1111/bjop.12301

Matthews, C. M., Mondloch, C. J., Lewis-Dennis, F., & Laurence, S. (2022). Children's ability to recognize their parent's face improves with age. *Journal of Experimental Child Psychology*, *223*, 105480. https://doi.org/10.1016/j.jecp.2022.105480

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, *6*(6), 255-260. https://doi.org/10.1016/S1364-6613(02)01903-4

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology*, *25*(1), 30-37. https://doi.org/10.1080/20445911.2012.739153

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*(1), 5-22. https://doi.org/10.1068/p7825

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. https://doi.org/10.1037/a0013464

Menon, N., White, D., & Kemp, R. I. (2015). Variation in photos of the same face drives improvements in identity verification. Perception, 44(11), 1332-1341. https://doi.org/10.1177/0301006615599902

Mondloch, C. J., Elms, N., Maurer, D., Rhodes, G., Hayward, W. G., Tanaka, J. W., & Zhou, G. (2010). Processes underlying the cross-race effect: An investigation of holistic, featural, and relational processing of own-race versus other-race faces. *Perception*, *39*(8), 1065-1085. https://doi.org/10.1068/p6608

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*(5), 553-566. https://doi.org/10.1068/p333

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic bulletin & review*, *15*(3), 465-494. https://doi.org/10.3758/PBR.15.3.465

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. http://dx.doi.org/10.1037/xhp0000049

Noyes, E., Parde, C. J., Colón, Y. I., Hill, M. Q., Castillo, C. D., Jenkins, R., & O'Toole, A. J. (2021). Seeing through disguise: Getting to know you with a deep convolutional neural network. Cognition, 211, 104611. https://doi.org/10.1016/j.cognition.2021.104611

O'Toole, A. J., & Castillo, C. D. (2021). Face Recognition by Humans and Machines: Three Fundamental Advances from Deep Learning. *Annual Review of Vision Science*, *7*, 543-570. https://doi.org/10.1146/annurev-vision-093019-111701

O’Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in cognitive sciences*, *22*(9), 794-809. https://doi.org/10.1016/j.tics.2018.06.006

Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive Research: Principles and Implications*, *3*(1), 1-9. https://doi.org/10.1186/s41235-018-0110-y

Phillips, P.J., Yates, A.N., Hu, Y., Hahn, C.A., Noyes, E., Jackson, K., Cavazos, J.G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J., Castillo, C.D., Chellappa, R., White, D. &

O'Toole, A.J., (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *PNAS, 115*(24):6171–76. https://doi.org/10.1073/pnas.1721355115

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, *20*(4), 260-281. https://doi.org/10.1016/j.tics.2016.01.007

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. Psychonomic bulletin & review, 13(3), 499-505. https://doi.org/10.3758/BF03193876

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. https://doi.org/10.1016/j.visres.2009.08.025

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897-905. https://doi.org/10.1080/17470218.2015.1136656

Rosemblaum, M., Grosbard, I., Abudarham, N., & Yovel, G. (2021). What type of experience is needed to generate a human-like view-invariant representation of face identity? Evidence from Deep Convolutional Neural Networks. Journal of Vision, 21(9), 2595-2595. https://doi.org/10.1167/jov.21.9.2595

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, *16*(2), 252-257. https://doi.org/10.1016/j.tics.2018.03.002

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive research: principles and implications*, *5*(1), 1-17. https://doi.org/10.1186/s41235-019-0205-0

Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences*, *105*(1), 394-398. https://doi.org/10.1073/pnas.0706079105

Summerfield, C., & Egner, T. (2013). Attention and decision-making. *The Oxford handbook of attention.* DOI: 10.1093/oxfordhb/9780199675111.013.018

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*(2), 225-245. https://doi.org/10.1080/14640749308401045

Tanaka, J. W., & Gordon, I. (2011). Features, configuration, and holistic face processing. *The Oxford handbook of face perception*, 177-194.

Towler, A., Dunn, J. D., Martínez, S. C., Moreton, R., Eklöf, F., Ruifrok, A., ... & White, D. (2021). Diverse routes to expertise in facial recognition.

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. https://doi.org/10.1371/journal.pone.0211037

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47. https://doi.org/10.1037/xap0000108

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

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. https://doi.org/10.1080/14640749108400966

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, *79*(4), 471-491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision research*, *141*, 217-227. https://doi.org/10.1016/j.visres.2016.12.014

Weldon, R. B., Corbin, J. C., & Reyna, V. F. (2013). Gist processing in judgment and decision making: Developmental reversals predicted by fuzzy-trace theory. In *The developmental psychology of reasoning and decision-making* (pp. 36-62). Psychology Press.

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology*, *1*(5), 287-300. https://doi.org/10.1038/s44159-022-00041-3

White D., Kemp R. I., Jenkins R., Matheson M., Burton A. M. (2014). Passport officers' errors in face matching. PLoS One, 9(8), e103510. https://doi.org/10.1371/journal.pone.0103510

White, D., Towler, A., & Kemp, R. (2021). Understanding professional expertise in unfamiliar face matching. *Forensic face matching: Research and practice*, 62-88.

Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science, 26(3*), 225-230. https://doi.org/10.1177/0963721417710693

Zhou, X., & Mondloch, C. J. (2016). Recognizing “Bella Swan” and “Hermione Granger”: No own-race advantage in recognizing photos of famous faces. *Perception*, *45*(12), 1426-1429. https://doi.org/10.1177/0301006616662046

Chapter 2: Photography experience and unfamiliar face matching. [1]

Abstract

With the exception of super recognizers and forensic examiners, people make a surprising number of errors when deciding whether photographs of unfamiliar faces belong to the same person or different people. Training protocols designed to improve professionals' (e.g., passport officers) performance often include photography. We evaluated the influence of life-time photography experience on the ability to distinguish matched versus mismatched face pairs. Expert photographers were not more sensitive to identity than hobbyists or novices—despite specializing in human subjects; Hobbyists were more liberal (more same responses) than Experts. We conclude that photography experience is not a route to expertise.

Introduction

Barkeepers, passport officers, and cashiers are charged with a challenging task: Matching photos of unfamiliar faces or matching a photo to a live person. Within-person variability in appearance (e.g., changes in lighting, expression, make-up) coupled with similarity across faces (e.g., Natalie Portman and Keira Knightley) makes this task error-prone (Burton, 2013). Although trivial in some contexts (even perceived as positive by the minor who successfully uses a sibling's ID), errors can have serious consequences in law and security. The discovery that passport officers are as error-prone as untrained undergraduates (White et al., 2014) sparked interest in improving individuals' ability to distinguish between matched (photos of the same person) versus mismatched (photos of two people) face pairs.

---

[1] This chapter is based on the published article: Baker, K.A. & Mondloch, C. J. (2022). Picture this: Photographers no better than controls for recognizing unfamiliar faces. Perception. *51*(8), 591–595. https://doi.org/10.1177/03010066221098727

Face matching tasks in the lab approximate checking photo-ID in applied settings. Signal detection theory is an ideal estimate of performance. Sensitivity (d′) accounts for both hits and false alarms (i.e., responding same on match and mismatched pairs, respectively) and criterion (c) accounts for response bias (Stanislaw & Todorov, 1999). Professional training protocols have been largely unsuccessful in increasing sensitivity to facial identity (Towler et al., 2019), increasing performance on match or mismatch trials, but not both (i.e., shifting c; Ritchie & Burton, 2017). One component of several training protocols is photography (Towler et al., 2019). The logic is that learning how photographic conditions alter appearance improves performance; photographers can manipulate lighting, viewpoint, and lenses to make someone look like a saint, criminal, or anything in between. This assumption is consistent with evidence that artistic ability is linked to unfamiliar face-matching performance (Devue & Barsics, 2016; Hsaio et al., 2021). This is the first examination of whether photography experience provides effective training by examining whether life-time photography experience influences d′ and c.

Methods

Ninety-five Caucasian participants (Women: $n = 58$; $M_{Age} = 47.84$, $SD_{Age} = 17.02$)[2] completed 80 trials (50% female; 50% match) of the Ambient Image Face Matching Task (AIFMT; Baker et al., under review). In this task, participants were shown pairs of images, presented simultaneously, and were tasked with determining whether the pair of images reflected the same person/different people. Image pairs remained on screen until participants indicated their response. Participants were asked to respond as accurately as possible. The images used in this task were obtained from the Face and Ocular Challenge Series (Phillips & O’Toole, 2014;

---

[2] Two participants did not report their age.

Phillips et al., 2011) and Brock University's Let's Face It database. Images were ambient (incorporating natural variability in appearance) color photographs, photoshopped to have flat gray backgrounds, and cropped to 275 × 295 pix (See Figure 2-1).

After completing the AIFMT, participants indicated their photography experience on a 9-point scale: Novices (1−3; $n = 28$), Hobbyists (4−6; $n = 33$), and Experts (7−9; $n = 33$). Thirty in the Expert category self-identified as professional ($M_{experience} = 18$ years); most specialized in human subjects ($n = 21$). Participants also indicated the extent to which they specialized in human subjects (e.g., faces, portraits) on a 9-point scale. A significant effect of photography group confirmed that Experts specialized in human subjects more than both Hobbyists and Novices, $F(2,91) = 5.12$, $p = 0.008$, $\eta^2 = 0.10$, who did not differ from each other ($p = 0.85$).

Results

A one-way ANOVA showed that d′ did not vary across photography groups, $F(2,91) = 0.23$, $p = 0.79$, $\eta^2 = 0.005$. A Bayesian ANOVA (JASP (2021) default priors) confirmed that these data were more likely to occur with the null hypothesis ($BF_{10} = 0.10$). Consistent with the analyses of d′, Novices ($M = 65.04\%$, $SD = 9.64$), Hobbyists ($M = 66.10\%$, $SD = 8.48$), and Experts ($M = 66.86\%$, SD = 9.01) did not differ in percentage correct, $p = 0.74$. The ANOVA for c was significant, $F(2,91) = 4.33$, $p = 0.02$, $\eta^2 = 0.09$. A Bayesian ANOVA confirmed that these data were more likely to occur with the alternate hypothesis ($BF_{10} = 3.01$). Tukey's HSD revealed that Novices did not differ significantly from the Hobbyists or Experts, $ps > 0.19$. Hobbyists were more liberal (i.e., made more same responses) than Experts, $p = 0.01$ (See Figure 2-2).

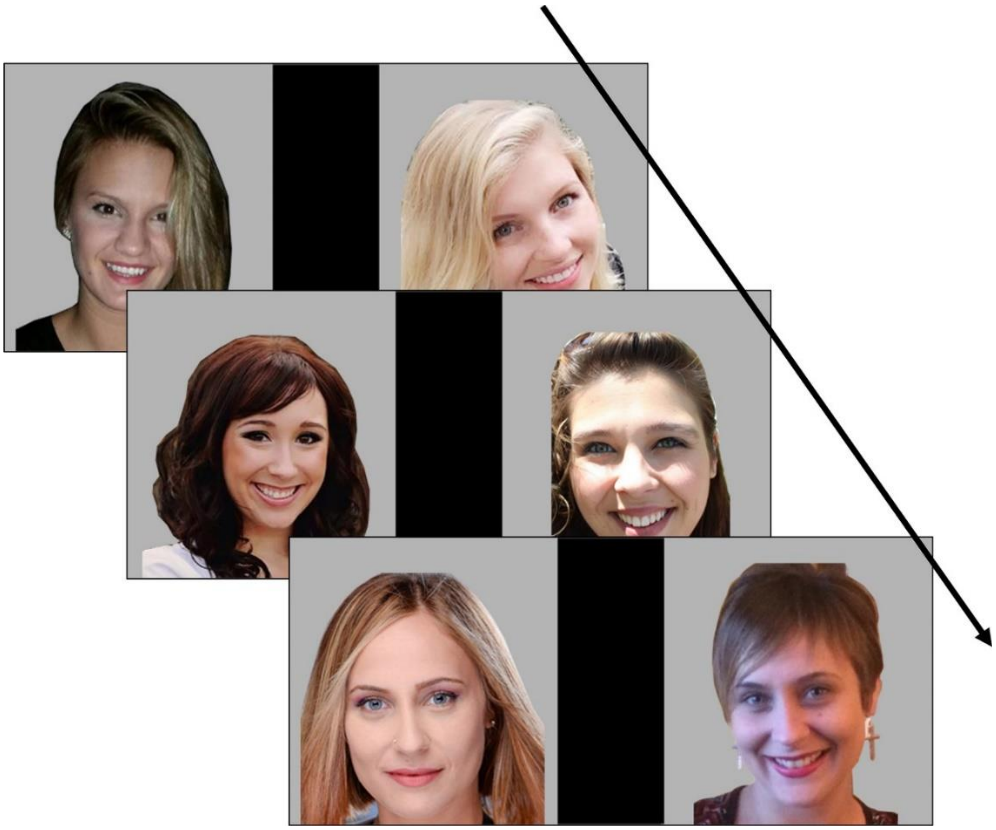

Figure 2-1. Depicts AIFMT trials.

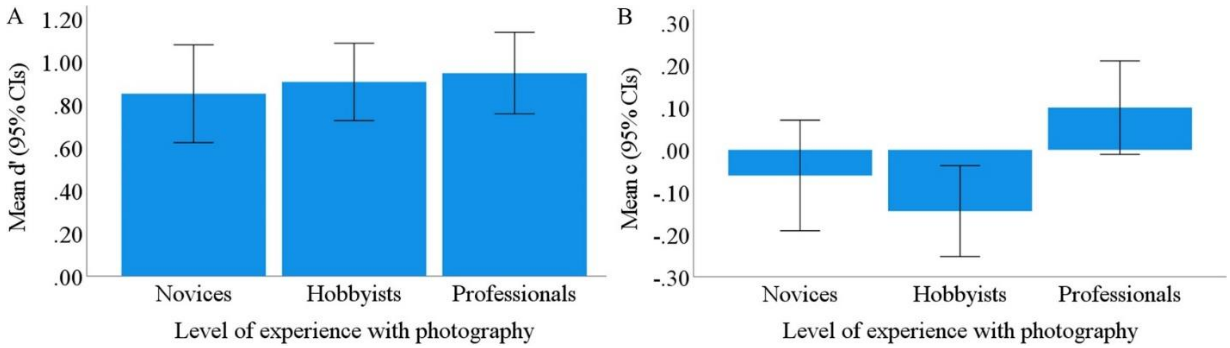

Figure 2-2. Mean d′ (A) and c (B) for Novices, Hobbyists, and Experts.

Discussion

Photography experience changes the type, but not the number of errors. This aligns with evidence that performance is attributable to genetics (Wilmer, 2017) and early life experience (Balas & Saville, 2017), but is not easily shifted by experience beyond that gained by most adults (e.g., passport officers: White et al., 2014). The finding that Hobbyists and Experts differed in response bias is similar to Towler et al.'s (2021) finding that super recognizers do not differ from forensic examiners in accuracy, but are more liberal in response bias. Hobbyists might be more liberal because they take multiple photos of the same people (e.g., friends, family) while manipulating lighting and perspective (see Alenezi & Bindemann, 2013 for evidence that feedback shifts criterion). Experts are more likely to take a few photos of many different people.

Future research should examine why individuals with artistic experience are more sensitive to identity (see Devue & Barsics, 2016; Hsaio et al., 2021) whereas Expert photographers are not. I suggest that the difference lies in process. Whereas artists create faces, painstakingly drawing the details of every feature, photographers' capture face images in an instance. Artists might adopt a feature-based approach and/or a slow and serial process to face identification. Artists' approaches might reflect innate skills that foster their artistic success or training, much like forensic examiners (Towler et al., 2021; White et al., 2015).

In summary, this provides evidence that expert photography experience does not improve sensitivity to identity. I recommend reconsidering the utility of including photography in training courses. I encourage practitioners and researchers to develop models of face identification to account for both sensitivity to identity and response bias. Capitalizing on stable individual differences in sensitivity to identity (e.g., see Baker et al., in press) is one route to improved face identification in applied settings; relying upon photography experience is not. Indeed, the local hobbyist might readily accept a higher number of fraudulent ID cards.

References

Alenezi H. M., Bindemann M. (2013). The effect of feedback on face-matching accuracy. Applied Cognitive Psychology, 27(6), 735–753. https://doi.org/10.1002/acp.2968

Baker K. A., Stabile V. J., Mondloch C. J. (in press). Stable individual differences in unfamiliar face identification: Evidence from perceptual matching and memory tasks.

Balas B., Saville A. (2017). Hometown size affects the processing of naturalistic face variability. Vision Research, 141, 228–236. https://doi.org/10.1016/j.visres.2016.12.005

Burton A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. The Quarterly Journal of Experimental Psychology, 66(8), 1467–1485. https://doi.org/10.1080/17470218.2013.800125

Devue C., Barsics C. (2016). Outlining face processing skills of portrait artists: perceptual experience with faces predicts performance. Vision Research, 127, 92–103. https://doi.org/10.1016/j.visres.2016.07.007

Hsiao J. H., An J., Zheng Y., Chan A. B. (2021). Do portrait artists have enhanced face processing abilities? Evidence from hidden Markov modeling of eye movements. Cognition, 211, 104616. https://doi.org/10.1016/j.cognition.2021.104616

JASP Team (2021). JASP(Version 0.16) [Computer software].

Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. Image and Vision Computing, 32(1), 74&85. https://doi.org/10.1016/j.imavis.2013.12.002

Phillips, P. J., Beveridge, J. R., Draper, B. A., Givens, G., O'Toole, A. J., Bolme, D. S., ... Weimer, S. (2011). An introduction to the good, the bad, & the ugly face recognition challenge problem. In 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG) (pp. 346-353). IEEE. 10.1109/FG.2011.5771424

Ritchie, K. L., Burton, A. M. (2017). Learning faces from variability. Quarterly Journal of Experimental Psychology, 70(5), 897–905. https://doi.org/10.1080/17470218.2015.1136656

Stanislaw, H., Todorov, N. (1999). Calculation of signal detection theory measures. Behavior Research Methods, Instruments, & Computers, 31(1), 137–149. https://doi.org/10.3758/BF03207704

Towler, A., Dunn, J. D., Martínez, S. C., Moreton R., Eklöf F., Ruifrok A., Kemp R., White D. (2021). Diverse routes to expertise in facial recognition. https://doi.org/10.31234/osf.io/fmznh

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. https://doi.org/10.1371/journal.pone.0211037

White, D., Kemp, R. I., Jenkins, R., Matheson, M., Burton, A. M. (2014). Passport officers' errors in face matching. PLoS One, 9(8), e103510. https://doi.org/10.1371/journal.pone.0103510

White, D., Phillips, P. J., Hahn, C. A., Hill, M., O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. Proceedings of the Royal Society B: Biological Sciences, 282(1814), 20151292. https://doi.org/10.1098/rspb.2015.1292

Wilmer, J. B. (2017). Individual differences in face recognition: a decade of discovery. Current Directions in Psychological Science, 26(3), 225–230. https://doi.org/10.1177/0963721417710693

Chapter 3: Unfamiliar face matching abilities over time and across tasks.[3]

Abstract

Matching identity in images of unfamiliar faces is difficult: Images of the same person can look different and images of different people can look similar. Recent studies have capitalized on individual differences in the ability to distinguish match (same ID) vs. mismatch (different IDs) face pairs to inform models of face recognition. We addressed two significant gaps in the literature by examining the stability of individual differences in both sensitivity to identity and response bias. In Study 1, 210 participants completed a battery of four tasks in each of two sessions separated by one week. Tasks varied in protocol (same/different, lineup, sorting) and stimulus characteristics (low vs. high within-person variability in appearance). In Study 2, 148 participants completed a battery of three tasks in a single session. Stimuli were presented simultaneously on some trials and sequentially on others, introducing short-term memory demands. Principal components analysis revealed two components that were stable across time and tasks: sensitivity to identity and bias. Analyses of response times suggest that individual differences in bias reflect decision-making processes. I discuss the implications of these findings in applied settings and for models of face recognition.

Introduction

Our ability to identify faces is challenged every day. We need to recognize colleagues at a conference, even if their appearance has changed; we need to discriminate actors to follow the plot of a movie, even if their appearance is similar. Accurate face identification depends on the ability to recognize a face despite variability in its appearance (*telling faces together*) and

[3] This chapter was based on the article: Baker, K.A., Stabile, V.J., & Mondloch,C.J. (In press). Stable individual differences in unfamiliar face identification: Evidence from simultaneous and sequential matching tasks. Cognition.

discriminating between identities (*telling faces apart*). Earlier work in the face recognition field focused on fine-tuned discrimination of faces (e.g., Mondloch et al., 2002; Tanaka & Corneille, 2007; for a review see Burton, 2013; Maurer et al., 2002)—an approach that provided many insights, but ignored the challenge of telling faces together.

Work on identifying a face despite variability in its appearance was galvanized by a seminal paper by Jenkins and colleagues (2011). Participants were asked to sort 40 ambient face photographs (i.e., images that contain naturally occurring variability in appearance) into piles such that each pile contained all of the images of one identity. Participants were not told that the photographs comprised 20 instances of each of two individuals. This task was performed accurately when participants were familiar with the identities. When unfamiliar, participants overestimated the number of identities that were present in the photos ($Median_{\text{piles}} = 7.5$), but rarely made misidentifications (i.e., put images of two different people in the same pile). Data from other paradigms have confirmed that face identification is error-prone when viewing images of unfamiliar faces (e.g., same/different tasks: Burton et al., 2010, Fysh & Bindemann, 2018; lineup tasks: Bruce et al, 1999; Megreya & Bindemann, 2013; Megreya & Burton, 2008; visual working memory tasks: Lorenc et al., 2014; Ritchie et al., 2021; Zhou et al., 2018; recognition memory tasks: Burton et al., 2010; Estudillo & Bindemann, 2014). Whereas familiar face identification is robust to within-person variability in appearance (e.g., changes in appearance from hairstyle, lighting, angle), unfamiliar face identification is more fragile.

The ability to identify unfamiliar faces varies across face categories. Participants are more accurate if presented own- vs. other-age/race faces and for upright than for inverted faces. These effects have been attributed to experience (e.g., Laurence et al., 2016; Kramer et al., 2017;

Prioetti et al., 2019; Tanaka & Farah, 1993; Yin, 1969; for discussions see Valentine, 1991; Valentine et al., 2016).

Under ideal conditions (e.g., when tested with upright images of own-race, own-age faces), the ability to identify unfamiliar faces varies across individuals (for reviews see Bruce et al., 2018; Lander et al., 2018; Wilmer, 2017; for examples see Fysh et al., 2020; Stacchi et al., 2020). At the extremes are those with prosopagnosia (who struggle with recognizing highly familiar faces; for a review see Kress & Daum, 2003) and super recognizers (individuals with extraordinary face-recognition abilities; Ramon, 2021; see Russell et al., 2009). In between these extremes remains an abundance of variability. Understanding these individual differences is important for understanding expertise and can inform models of face recognition. For example, if good vs. poor performers differ in the information that they use to make same/different judgements, then training could focus on shifting attention towards features used by better performers. Prior to capitalizing on individual differences, it is essential to determine the extent to which individual differences are stable. In the current study I investigated whether individual differences are stable over time, across tasks, and regardless of whether faces are presented simultaneously or sequentially.

Individual differences in sensitivity to identity

Despite recent interest in individual differences in face identification, little is known about how stable they are over time. Establishing stability is vital as between-task relationships are constrained by the reliability of each independent measure (Spearman, 1910; see also Goodhew & Edwards, 2019 and Hedge et al., 2018 for discussions). Researchers often have examined relationships between tasks without providing estimates of reliability for each individual task (for a discussion see Goodhew & Edwards, 2019). For instance, the Sorting Task has been used to investigate both non-manipulable group differences (e.g., age-related changes, other-race effects; Laurence & Mondloch, 2016; Matthews & Mondloch, 2021; Laurence et al., 2016; Zhou & Mondloch, 2016), and in correlational approaches (Fysh et al., 2020; Stacchi et al., 2020). Despite this widespread interest in the Sorting Task as a measure of individual differences, it has never been examined for reliability. Therefore, it is important to establish whether performance on this task is stable over time. In many studies, task reliability is estimated from a single session (e.g., Burton et al., 2010; Robertson et al., 2017; Verhallen et al., 2017). This approach can be problematic; high reliability can be driven by state-like factors that are prone to changes (e.g., emotion, situational factors) rather than stable trait-like factors (e.g., personality, IQ).

Only a few studies provide test-retest reliability estimates (e.g., for the Cambridge Face Memory Test: Murray & Bate, 2020, Wilmer et al., 2010; for the Kent Face Matching Test: Fysh & Bindemann, 2018; for the Yearbook Task: Fysh et al., 2020; for the Glasgow Face Matching Test: Stantic et al. 2021), which is the ideal estimate of performance stability (Goodhew & Edwards, 2019). Evidence from twin studies suggests that individual differences in matching tightly controlled images of unfamiliar faces are heritable and, thus, are likely to be stable over time (e.g., Shakeshaft & Plomin, 2015; Wilmer et al., 2010). Nonetheless, research examining the stability of individual differences in the context of natural within-person variability in appearance is needed.

Another important issue is whether individual differences in face identification are stable across protocols that vary in task demands. Same/different tasks (e.g., Burton et al., 2010; Fysh & Bindemann, 2018; White et al., 2014) require participants to decide whether two images belong to the same person, whereas lineup tasks (e.g., Bindemann et al., 2012a; Bruce et al., 1999; Megreya & Bindemann, 2013; Megreya & Burton, 2008) require participants to determine which image (if any) in a set of images belongs to a target identity. The Sorting Task (Jenkins et al., 2011) is perhaps most challenging because the number of identities present is unknown. Tasks also vary in the extent to which the images incorporate within-person variability in appearance. For example, the Sorting Task contains vast amounts of within-person variability in appearance, the Yearbook Task (Fysh et al., 2020) contains less variability—but images vary in age, and the Glasgow Face Matching Test (GFMT; Burton et al., 2010) contains less variability in appearance. To investigate convergent validity across face identification tasks, and therefore to establish stability of performance across tasks, it is advantageous to include tasks that should capture the same ability but vary in task demands (e.g., the Sorting Task, same/different tasks, lineup tasks). Doing so would enable researchers to separate face identification ability from more task-specific abilities (i.e., the ability to make the "best guess" when two alternatives are provided). Here I aimed to establish the Sorting Task's convergent validity with other measures of sensitivity to identity in unfamiliar faces.

If a unitary skill underlies performance across protocols, then individual differences in one task should correlate with individual differences in another. Three recent studies (Fysh et al., 2020; McCaffery et al., 2018; Stacchi et al., 2020) have examined individual differences in face identification across a battery of tasks. Some, but not all examined identification of wholly unfamiliar faces (e.g., in the absence of learning) —the focus of the current research. Across studies, relationships between tasks varied in strength (R^2s = 0.004 to 0.61). Whereas Stacchi et al. (2020) reported a null relationship between performance on the Sorting Task and same/different tasks (Person Identification Challenge Test, Expertise in Facial Comparison Test), Fysh et al. (2020) reported a significant relationship between the Sorting Task and their measure of face matching (Kent Face Matching Test). This discrepancy could occur because the Sorting Task might have poor reliability or convergent validity. One goal of the current study was to build on this literature to examine the extent to which individual differences in sensitivity to identity are stable over time and across tasks.

To do so, I took advantage of Signal Detection Theory (SDT). I arbitrarily defined the signal as two images belonging to the same identity (i.e., as a matched face pair). Thus, a hit was defined as responding “same” on a matched face trial and a false alarm was defined as responding “same” on a mismatched face trial. Many studies in the field of face recognition report accuracy on matched or mismatched face pairs separately; however, this approach makes it difficult to interpret good performance. For example, good performance on match trials might reflect extraordinary skill in recognizing identity in unfamiliar faces or merely the tendency to respond same (see Stanislaw & Todorov, 1999). This is a problem that SDT was designed to solve, as both criterion and sensitivity influence each response that is made.

Individual differences in response bias

Past studies have focused on individual differences in sensitivity to identity; none have examined individual differences in bias. This is important, given that bias can influence the type of errors that individuals make. To date, no study has examined whether individual differences in response bias are stable over time and across face identification tasks. Despite the importance of response bias, it is underrepresented in classic models of face identification. For example, response bias is not discussed in the Bruce and Young model (1986), nor is it discussed in Valentine’s multi-dimensional face space model (1991; see also Valentine et al., 2016). However, there are hints from the recognition memory literature that we might expect individual differences in response bias to be stable across time (e.g., Kantner & Lindsay, 2012).

Three factors motivated my examination of individual differences in response bias. First, there has been a recent call to better understand decision making in models of unfamiliar face identification (Bindemann & Burton, 2021). Decision making involves multiple processes including evidence accumulation, comparison of evidence to a criterion, and a decision policy (which influences the placement of a criterion) that is used to maximize benefit and reduce costs (for a discussion see Summerfield & Egner, 2013). Even in classic SDT, the selection of one option over another depends on both perceptual evidence (e.g., signal strength) and the position at which a criterion was set (i.e., the decision boundary; see Stanislaw & Todorov, 1999). Analyzing individual differences in criterion will set the groundwork for a better understanding of decision making within the context of unfamiliar face identification. Doing so will help to inform contemporary models of face identification.

Second, evidence from studies investigating the other-race effect suggests that response bias might vary as a function of task demands. On same/different tasks the other-race effect is driven by errors on *mismatch* trials; participants are biased to perceive images of different people as belonging to the same person when viewing other-race face pairs as compared to own-race face pairs (e.g., Meissner & Brigham, 2001). The opposite pattern is observed in sorting tasks; participants are biased to perceive images of the same person as belonging to different people when sorting other-race faces as compared to own-race faces, resulting in their making more piles (Laurence et al., 2016; Zhou & Mondloch, 2016); misidentification errors are low for both own- and other-race faces. Such findings suggest that individual differences in response bias might vary across tasks. They also suggest that the number of piles made in the Sorting Task might measure response bias rather than sensitivity to identity per se. Evidence has been shown for this possibility in recent work by Cavazos et al. (2020). They discussed the importance of how manipulating the decision threshold (which is manually set) in Deep Convolutional Neural Networks (DCNNs) influences performance. When each of four different algorithms used the same threshold for own- and other-race faces, more false alarms were made for other-race faces. Setting different thresholds for own- and other-race faces attenuated the other-race effect in terms of false alarms. If the number of piles made in the Sorting Task does capture response bias rather than sensitivity to identity, this would be extremely relevant for the field; creating more piles is considered to be the primary error in the Sorting Task (e.g., Jenkins et al., 2010; Laurence et al., 2016).

Third, whereas training protocols have not been effective in improving sensitivity to identity, they have altered response bias (e.g., as shown by an increase in both hits and FAs, [e.g., Alenzi & Bindemann, 2013 exp.3; Matthews & Mondloch, 2018], or by only influencing either hits or false alarms, [e.g., Baker & Mondloch, 2019; see also Ritchie et al., 2021]). A better understanding of bias might lead to promising avenues for training. For instance, economic models of decision making suggest that response bias increases in the context of perceptual uncertainty, a condition that defines unfamiliar face identification (see Lynn & Barrett, 2014; Lynn et al., 2015; Summerfield & Tsetsos, 2012). Under conditions of uncertainty, the direction of response bias (liberal [e.g., more match responses] vs. conservative [e.g., more mismatch responses]) is influenced by base rates (e.g., proportion of match vs. mismatch trials) and the costs vs. benefits associated with each response. These claims also align with findings in the face identification field—for example, asymmetric base rates in combination with feedback or similarity of stimuli affect criterion (Papesh et al., 2018; Alenzi & Bindemann, 2013). A better understanding of individual differences in bias could provide a better understanding of how to shift an individual's criterion. A second goal of the current study was to examine the extent to which individual differences in response bias are stable over time and across unfamiliar face identification tasks.

The current study

In Study 1, participants completed an online battery of face identification tasks that differed in protocol and in the extent to which images were tightly controlled vs. ambient: A Sorting Task, the GFMT, a newly developed same/different task that incorporates natural within-person variability in appearance (the Ambient Image Face Matching Task; AIFMT) and a Lineup Task. Approximately one week later, participants completed these tasks a second time with a completely new stimulus set. I conducted principal components analyses (PCA) to determine whether individual differences in sensitivity to identity (d′ or proportion correct) reflect the same underlying ability across tasks, and also to examine whether individual differences in bias (c or proportion of target-absent responses) capture the same response biases across tasks.

This protocol allowed for the investigation of two main questions regarding individual differences in unfamiliar face identification abilities. First, I examined the extent to which performance on unfamiliar face identification tasks is stable over time. If the tasks are quality measures of individual differences, there would be a moderate-to-strong relationship between the ability to distinguish images showing the same person vs. different people in Sessions 1 and 2. I also examined the extent to which performance on unfamiliar face identification is stable across tasks. If sensitivity to identity on each of the tasks loads onto the same component, this would suggest that the ability to distinguish same vs. different identities is stable across tasks that vary in demands and that there is only one underlying face identification ability. Second, I provided the first examination of the extent to which individual differences in response bias are stable across time and tasks. If response bias on each of the tasks loads onto the same component, this would suggest that there is one underlying component for bias in face identification tasks. Examining response bias also allowed us to explore the possibility that the number of piles made in the Sorting Task captures individual differences in bias rather than sensitivity to identity. I did so by determining on which of the two components the number of piles loads.

In Study 2, participants completed two versions (simultaneous and sequential stimulus presentation) of three face identification tasks: The GFMT, AIFMT, and the Lineup Task. This protocol allowed us to assess whether individual differences are stable across simultaneous vs. sequential matching tasks. This protocol also allowed us to use exploratory analyses to further examine bias using response times (RTs).

Study 1

Methods

Participants. Two-hundred fifty-three Caucasian participants completed Session 1; four were excluded because of failing attention checks (see below). Of the remaining 249 participants, 212 completed Session 2; two were excluded from analyses because of failing attention checks. Thus, for analyses only containing Session 1 data, the final sample comprised 249 participants (women: n = 166; Age: M = 22.56, SD = 4.29) and for analyses containing both Sessions 1 and 2, the final sample comprised 210 participants (women: n = 144; Age: M = 22.71, SD = 4.40). Just under half ($n_{Session1}$ = 117; $n_{Session2}$ = 98) of the participants were Brock University students who received 2 hours of research credit for their participation. The remaining participants were recruited online via Prolific (www.prolific.co) and were paid £5 per session for their participation.

Stimuli and tasks. Stimuli and tasks are depicted in Figure 3-1. Images for the GFMT were presented in their original size (image pairs: 1000 x 700 pix) and format (e.g., the images were full-faced, greyscale and the images within the match trials were taken using different cameras). For all other tasks I presented ambient images that contained naturally occurring variability in appearance. Images for the AIFMT were colour photos of white identities that were photoshopped to have a flat grey background and cropped to be 275 x 295 pix. Each identity was only presented in one trial (either match or mismatch). Mismatched pairs were chosen based on physical similarity of the two identities. Images for all other tasks were colour photos, had a roughly frontal view of the model's face, and were cropped to be 125 x 167 pix. Half of the images in each task were male identities. Of the 873 identities in this study (not including attention check trials), only 53 were recognized by at least one participant. Trials in which a participant recognized an identity were not included in the analyses.

Images were obtained from the following databases: The Face and Ocular Challenge Series (Phillips & O'Toole, 2014; Phillips et al., 2011), The Center for Vital Longevity Face Database (Minear & Park, 2004), and Brock University's *Let's face it* database. Images were also obtained from previous publications: Burton et al. (2010), Baker and Mondloch (2019), Matthews and Mondloch (2018), and Dowsett et al. (2016). Additional images were collected using a google search and social media (e.g., Instagram and Twitter). Image searches occurred for identities from the following categories: Celebrities, chefs, or politicians from other countries (e.g., Thea Sofie Loch Naess, Chef Richard McCormick), and minor league athletes. For the Sorting Task, the first 20 (Session 1) or 15 images (Session 2) were gathered from a Google search. The images were required to be an unoccluded, roughly frontal view of the model's face. When obtaining multiple images of the same identity, images were required to have been taken on different days. The faces had to be larger than 125 x167 pix. If there were not enough images that met the qualifications on Google, I selected the first images that met the requirements on social media.

To assess the stability of performance across sessions, I created two versions of each task such that participants completed Version A on the first day of testing and Version B on the second; the two versions differed only in the identities presented. I did so to prevent any possibility of learning the identities via exposure to multiple images, which is a mechanism for face learning (e.g., Andrews et al., 2015; Andrews et al., 2017; Baker et al., 2017; Matthews & Mondloch, 2018; Murphy et al., 2015; Ritchie & Burton, 2017; Burton et al., 2016).

Procedure. Participants were tested online, as per Covid-19 protocols, using Testable (www.testable.org). Prior to starting the tasks, participants were asked to calibrate their screens by using their arrow keys to resize a line to the size of a credit card. This ensured that stimuli were presented to scale given each participant's screen resolution. The task order within each session and the order of the two versions were fixed[4], as is recommended for studies of individual differences (for a brief discussion see Goodhew and Edwards, 2019). Participants completed the four face identification tasks in the following order: The Sorting Task, GFMT, AIFMT, Lineup Task. (A fifth task was included but excluded from analyses because responses were hard to interpret. Only the attention check was retained). Participants were asked to respond as accurately as possible for each task. Feedback was not provided. Five to nine days following Session 1, participants completed Session 2.

[4] Task order is typically fixed, not counter-balanced, in individual difference studies. Task performance can differ because of task order. If task order is counterbalanced scores would then be confounded with this additional variability (Dale & Arnell, 2014). This confound is easily removed by using a fixed task order. Likewise, which stimulus set was used in each session was held constant across participants.

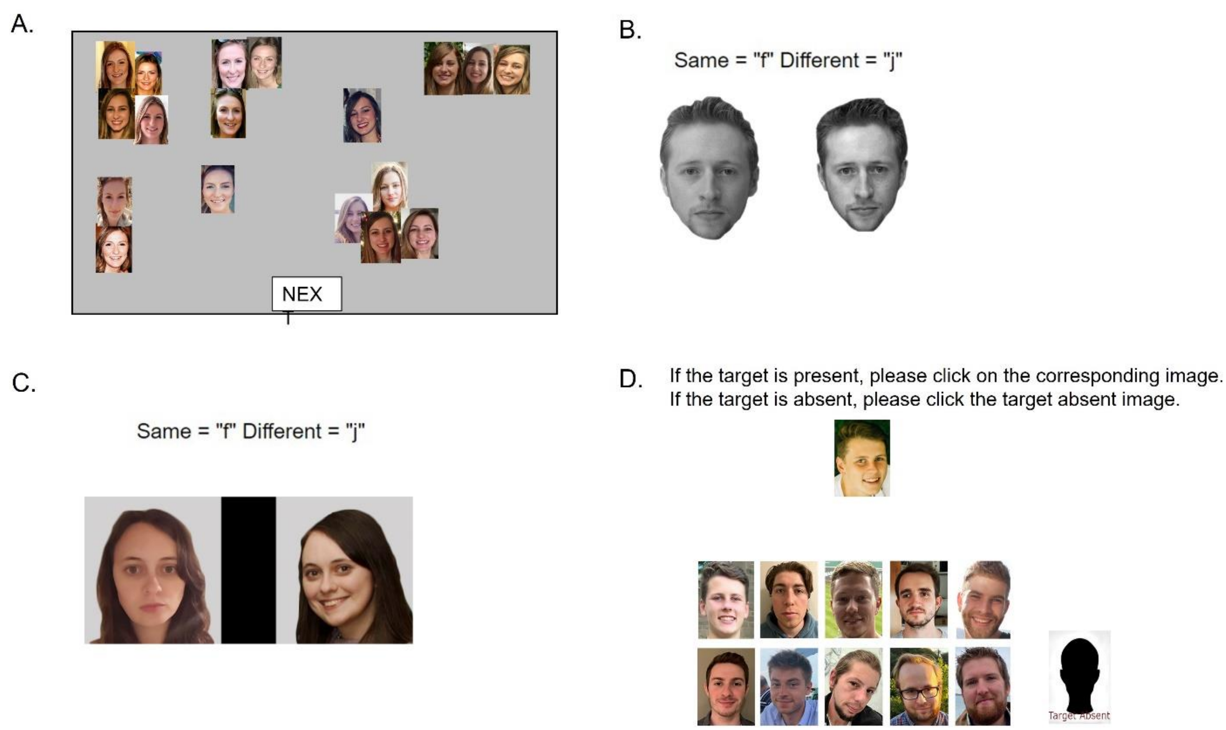

Figure 3-1 depicts examples of A) Sorting Task; B) the Glasgow Face Matching Test; C) the Ambient Image Face Matching Task; and D) the Lineup Task.

Unfamiliar face identification tasks. ***Sorting Task.*** In Session 1, the Sorting Task comprised 40 face photographs of two identities (as per Jenkins et al., 2011): 20 were of Gigi Ravelli and 20 were of Fern Sutherland. In Session 2, the Sorting Task comprised 45 images of three identities (15 images each of Donald Stamper, Andy Murray, and Philipp Boy). The number of identities and images per identity varied across days to reduce potential carryover effects. Although none of the participants recognized the identities from Session 1, $N = 12$ participants recognized one identity (Andy Murray) in Session 2.

To adapt the task for online testing, participants were shown images one at a time in the bottom-centre of their screen. Participants were asked to drag each image (using their mouse or their laptop trackpad) to different positions on the screen to create “piles” such that each pile contained images of only one identity. Participants were told that they could take as much time as they needed to complete the task. Participants were asked to keep track of the number of piles they made, as they would be asked about it later. When completed, participants clicked the "NEXT" button, which brought them to a screen where they indicated how many piles they made.[5]

To recreate the final sorting solution, I used RStudio (RStudio Team, 2020) to determine the location of each image on the screen. I then used Matlab and Psychtoolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) to recreate the final sorting pattern. When participants had overlapping images, I created a video of the images appearing in the screen locations. From these images and videos, three independent raters scored the images and videos for the number of piles and misidentifications. On 13 occasions, one rater did not agree with the other two. To solve these discrepancies, the three raters met to reach a consensus. The number of piles reported by participants and raters was strongly related ($r = 0.90$, $p < 0.001$). I used the sorting solutions provided by the raters because these solutions included information about errors.

GFMT. I used the same trial sets as Towler et al. (2019), such that there were two versions of the GFMT (20 pairs per version). When used as pre- and post-tests in a training study, Towler et al. (2019) reported equivalent accuracy across these versions of the GFMT. In each session, there were 10 match trials (two images of the same person) and 10 mismatch trials (images of two different people). As in Burton et al. (2010), participants were shown face pairs and asked to indicate whether the pair showed the same person or different people. Images remained on screen until participants made a response. No identities were recognized.

[5] Twenty-six participants did not complete the sorting task in at least one session (Session 1, $n = 15$; Session 2, $n = 7$; both sessions, $n = 4$) and were excluded from the associated sorting task analyses.

AIFMT. In each session, participants completed 60 trials (50% match). The testing protocol was identical to the GFMT. Five participants recognized one identity.

Lineup Task. In each session, participants completed 28 trials (50% target present) of a Lineup Task. Each trial comprised a single image of a target identity in the top-centre of the screen and a 10-image array. Participants were asked to select the image of the target within the lineup of 10 images or, if the target was absent, to click the target-absent button on the lower-right of their screen. Images were presented simultaneously and remained on screen until participants responded. Forty-six participants recognized at least one identity; the range of removed trials was 0 to 3. An error in the program allowed for participants to skip a trial by pressing the target identity image. Participants (n = 17) infrequently proceeded to the next trials by doing so; these trials were also removed (range: 1 to 4).

Attention checks. To ensure that participants were attentive during the experiment I included three attention checks per session (1 per each of the last three tasks). The attention checks in Session 1 and Session 2 had different correct responses. *Match* attention checks were easily solved because the same image of the target was presented; *mismatch* attention checks were easily solved because distractors differed from the target identity in age and sex. The main analyses were based on Session 1; participants who failed >1 attention check in Session 1 were excluded from all analyses (n = 4). Test-retest reliability was based on both sessions; I excluded one additional participant who failed one attention check on each of the two sessions and another who failed >1 attention check on Session 2.

Results

Data analysis. Data analyses were performed using SPSS. I used data from Session 2 only to assess the reliability of individual differences. All other analyses were based on Session 1 only. Doing so reduces introducing additional sources of variability (e.g., practice effects, state-based effects) into the data.

Wherever possible, I used signal detection theory to examine performance. For the GFMT and AIFMT I defined a hit as responding *same* on match trials and a false positive as responding *same* on different trials. Hit rates of 1 and false alarm rates of 0 were corrected using the approach suggested by Macmillan and Kaplan (1985); values of 0 are replaced with $0.5/n_{\text{Signal or Noise trials}}$ and values of 1 are replaced with $(n_{\text{Signal or Noise trials}} - 0.5)/n$. For the Lineup Task, in lieu of analyzing d′ and c, I analyzed total proportion correct and proportion of target-absent responses. I did so because there are two types of ‘misses’ when errors are made on target-present trials: misidentifications (selecting a different image) and misses (responding target-absent). Thus, the Lineup Task does not lend itself to analysis via SDT. Nonetheless, total proportion correct is akin to d′; it accounts for all accurate and inaccurate responses. Proportion of target-absent responses is akin to c; it provides a measure of participants’ tendencies to choose one option (target absent) over all others, regardless of whether their response was accurate or inaccurate. I did not calculate sensitivity to identity and bias for the Sorting Task.

Stability across time. I report test-retest data in terms of correlation coefficients, as they allow for the comparison to past research. I also report Intraclass Correlation Coefficients (ICCs; Mcgraw & Wong, 1996) to assess absolute agreement (e.g., a measure of whether participants' scores were exactly the same in sessions 1 and 2; see Hedge et al., 2017). ICCs provide a more conservative measure of stability in performance. Whereas correlation coefficients assess whether relative performance is stable over time, ICCs assess whether absolute performance is stable over time, despite using entirely different stimulus sets in the two sessions and despite any potential practice effects. I do not report ICCs for the Sorting Task because I varied the number of identities between Sessions 1 and 2. In addition to measuring stability across sessions within each task, I also report Cronbach's alpha as a measure of internal consistency. Given that there are only two measures of performance in the Sorting Task, I do not report internal consistency for the Sorting Task.

I discovered stability across days (rs > 0.52) and moderate to high levels (moderate: ICCs ≥ 0.5 - 0.75, high: ICCs ≥ 0.75-0.90; Koo & Li, 2016) of absolute agreement in performance across sessions for each task (see Table 1). Though not included in the table, the dependent variables for the Sorting Task (number of piles and misidentifications made) also showed good reliability (rs > 0.52). I also showed internal consistency, as reported via Cronbach's alpha, that fell within or above the adequate range (Cortina, 1993). Together, the estimates of stability, absolute agreement, and internal consistency suggest that these tasks are suitable measures of individual differences. The estimates of stability and absolute agreement suggest that performance is stable across time.

Table 3-1.

Correlations for test-retest reliability across sensitivity to identity and bias for the Glasgow Face Matching Test, Ambient Image Face Matching Task, and Lineup Task.

Task	Sensitivity to identity	Bias	Cronbach's alpha (Sessions 1-2)
GFMT	0.53 *** (0.67)	0.52 *** (0.66)	0.744
AIFMT	0.66 *** (0.79)	0.67 *** (0.79)	0.813
Lineup	0.60 *** (0.75)	0.73 *** (0.84)	0.768

Note. ICC values are provided in brackets. Internal consistency scores are provided by Cronbach's alpha.

*** $p < 0.001$.

Relationships between tasks. As shown in Table 2, there were significant relationships across the tasks. Across the three tasks where measures of overall sensitivity to identity and bias were calculated, both sensitivity to identity (*r*s > 0.41, *p*s < 0.001) and bias (*r*s > 0.25, *p*s < 0.001) were correlated. The number of piles made in the Sorting Task was related to sensitivity to identity on the two same/different tasks (GFMT and AIFMT) and bias measures for two tasks (GFMT and the Lineup Task). The number of misidentifications made in the Sorting Task was negatively related to all three measures of sensitivity to identity and to the bias measure in the Lineup Task. This high degree of relatedness between tasks, despite using controlled vs. ambient images and different task demands, suggests that the tasks are capturing similar abilities in distinguishing match vs. mismatch face pairs.

Table 3-2.

Correlations between task performance for each of the four tasks.

	GFMT d′	AIFMT d′	Lineup acc.	GFMT c	AIFMT c	Lineup bias	Sorting Piles	Sorting MisIDs
GFMT d′	--	0.54 ***	0.42 ***	- 0.01	0.05	0.06	- 0.17**	- 0.36 ***
AIFMT d′		--	0.53 ***	- 0.01	0.07	0.09	-0.27***	- 0.31 ***
Lineup acc.			--	0.08	0.28 ***	0.40 ***	-0.05	- 0.39 ***
GFMT c				--	0.59 ***	0.30 ***	0.14*	- 0.08
AIFMT c					--	0.40 ***	0.09	- 0.10
Lineup bias						--	0.17**	- 0.15*
Sorting Piles †							--	- 0.06
Sorting MisIDs. †								--

†. Spearman’s rho is reported for relationships with the Sorting Task.

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Underlying face identification abilities. I conducted a Principal Components Analysis (PCA) with an Oblimin rotation. I included standardized scores for sensitivity to identity and bias for the GFMT, AIFMT, and Lineup Task, and the number of piles and misidentifications made in the Sorting Task (the Sorting Task is not ideally suited for measures of sensitivity to identity or bias). The Kaiser-Meyer-Olkin was adequate (KMO = 0.66) and Bartlett's test of sphericity was significant (χ^2 (28) = 428.42, $p < 0.001$), suggesting these data were suitable for PCA. Although Kaiser's criterion suggested a 3-component solution, both the Skree plot and Parallel analysis suggested that there were two components; these two components explained 53.33% of the variance in scores.

As shown in Table 3, sensitivity to identity for the GFMT, AIFMT, and Lineup Task loaded positively on Component 1; misidentification errors in the Sorting Task loaded negatively. Bias for the GFMT, AIFMT, and Lineup Task loaded positively on Component 2, as did number of piles made in the Sorting Task. Thus, Component 1 represents sensitivity to identity and Component 2 bias.

Table 3-3.

Communality and factor loadings for performance on each face task.

	Measure	Communality	Loading	Cronbach's alpha
Component 1	GFMT d′	0.62	0.79	0.30
	AIFMT d′	0.67	0.82	
	Lineup acc	0.68	0.77	
	Sorting Misidentifications	0.32	- 0.56	
Component 2	GFMT c	0.62	0.79	0.59
	AIFMT c	0.72	0.84	
	Lineup bias	0.50	0.64	
	Sorting Piles	0.12	0.33	

Note. The provided loadings are the result of an Oblimin rotation. The internal consistency (Cronbach's alpha) for both components is also included.

Discussion

Study 1 revealed two important findings. First, sensitivity to identity remained stable across time and tasks. The stability estimates for sensitivity to identity are similar to other test-retest reliability estimates in the field (e.g., r = 0.67 to 0.78; for examples see Fysh & Bindemann, 2018; White et al., 2021; Murray & Bate, 2020). The stability of these tasks is consistent with the high heritability of performance on face identification tasks (e.g., Shakeshaft & Plomin, 2015; Wilmer et al., 2010; for a discussion see Wilmer, 2017). Stable individual differences in the ability to distinguish match vs. mismatch face pairs do suggest that there is one underlying ability for identity perception as suggested by Verhallen et al. (2017)—a claim that received additional support from McCaffery et al. (2018).

The second novel finding is that individual differences in response bias are stable across time and tasks. The PCA showed that individual differences in response bias are independent of individual differences in sensitivity to identity, suggesting that individual differences in performance on face identification tasks reflect two underlying components. This finding expands on the findings by Verhallen et al. (2017) and McCaffery et al. (2018) by providing the first examination of individual differences in response bias in face identification tasks. Individual differences in response bias have important implications for applied settings and suggest that decision making is influencing responses, a possibility I explore in Study 2. Though it remains unclear as to whether this underlying bias is face-specific, these findings draw attention to how bias and sensitivity to identity independently influence unfamiliar face identification performance.

Although the number of misidentifications made in the Sorting Task loaded with measures of sensitivity to identity, the number of piles made in the Sorting Task loaded with measures of bias. The number of piles made in the Sorting Task is a metric typically conceptualized as a measure of accuracy (e.g., Balas & Saville, 2017; Jenkins et al., 2011; Laurence et al., 2016). My data suggest this metric might be better conceptualized as response bias. This finding suggests a need to re-conceptualize past findings. For example, differences in the types of errors made for other-race faces in dichotomous-choice tasks (e.g., Meissner & Brigham, 2001) vs. the Sorting Task (Laurence et al., 2016) might reflect the influence of task demands on response bias. Task demands, such as consistency of colour in image pairs, have been shown to affect response biases (Bobak et al., 2019). Likewise, improved performance following exposure to within-person variability in appearance (e.g., Andrews et al., 2015; Ritchie & Burton, 2017) might be driven, at least in part, by a shift in criterion. There are some hints in the literature that suggest that this might be the case. For instance, White et al. (2014) showed that performance improved when participants were matching a target identity to multiple images vs. a single photo—but only on match trials (see also Menon et al., 2015). Dowsett et al., (2016) showed that increased exposure to variability in appearance helped participants select the correct target identity out of set of 30 image. However, Matthews and Mondloch (2018) showed that this increased variability also led participants to select non-target identities during target-absent trials.

All of the tasks in Study 1 involved the simultaneous presentation of stimuli. Simultaneous presentation allows participants to scan back-and-forth between images—a task performed daily by passport officers and clerks selling age-restricted goods. What remains unknown is whether individual differences would be stable regardless of whether stimuli are presented simultaneously or sequentially, with all other task demands remaining consistent. This is important to establish, as many tasks in the field use different procedures. For example, stimuli are presented simultaneously in the GFMT (Burton et al., 2010) and the Oxford Face Matching Test (Stantic et al., 2021), but sequentially in visual working memory tasks (Zhou et al., 2018) and in the recognition memory procedures used in the Cambridge Face Memory Test (Duchaine & Nakayama) and UNSW face test (Dunn et al., 2020), tasks that require participants to maintain the representation in memory for longer. Thus, there is a need in the field to check face identification tasks for convergent validity (White & Burton, 2022).

Evidence to suggest that sensitivity to identity should load together across simultaneous and sequential versions of my tasks is two-fold: Sensitivity to identity for same/different judgements and recognition memory performance are correlated when using tightly controlled face images (e.g., Burton et al., 2010) and matching and face memory are correlated when assessed via different tasks (e.g., McCaffery et al., 2018; Verhallen et al., 2017; Fysh & Bindemann, 2018). Nonetheless, no studies have examined whether individual differences in bias are stable across simultaneous vs. sequential tasks, a question I examined in Study 2.

Study 2

To investigate these questions, participants from Prolific completed three tasks from Study 1 (the GFMT, the AIFMT, and the Lineup Task) in a single session. Trials from Session 1 were unchanged. Trials from Session 2 were converted to sequential presentation, introducing short-term memory demands. I hypothesized that sensitivity to identity (as measured by d′ and proportion correct) would be highly correlated across the two versions of each task. PCA allowed us to examine whether the simultaneous vs. sequential matching versions would load onto the same or different components. If all measures of sensitivity to identity load onto the same component, it would suggest that performance on simultaneous and sequential versions is based on the same ability. However, if sensitivity to identity on the simultaneous and sequential versions loads onto two different components, it would suggest that performance is based on separable abilities. Likewise, I hypothesized that response bias would be highly correlated across the two versions of each task and used PCA to examine whether response bias on the simultaneous vs. sequential versions of the tasks loaded onto the same or different components.

To begin exploring whether the individual differences in response bias reflect decision making, I also analyzed RTs. RTs are a common dependent variable in decision-making tasks and have been used to model various aspects of decision-making processes (e.g., in the drift-diffusion model, for a review see Ratcliff et al., 2016; Serial sampling models, see Summerfield & Egner, 2013). Here, I used RTs to determine whether participants' criterion influenced the speed with which they responded on match vs. mismatch trials in the GFMT and AIFMT. My hypothesis was that RTs would be faster on trials that were congruent vs. incongruent with participants' response bias (i.e., that participants with a liberal bias would respond faster on match vs. mismatch trials, whereas participants with a conservative bias would respond faster on mismatch trials). To confirm that RTs provide insights about decision making, I examined RTs in the Lineup Task. I hypothesized that participants would respond slower on target-absent than target-present trials. Much like self-terminating search in the visual search literature (e.g., Treisman & Gelade, 1980), RTs on target-absent trials should be slower overall than RTs on target-present trials; in target-absent trials one must determine that each photo in the lineup is not the target prior to determining that the target is absent. I did not anticipate an effect of response bias; rather, I predicted a null, or a small correlation between bias and the difference in RTs.

Methods

Participants. One hundred forty-eight Caucasian participants completed Study 2 (women: n = 131; Age: M = 23.42, SD = 5.30). Two additional participants were excluded because of failing attention checks (i.e., failed > 1 attention check). Participants were recruited online via Prolific and were paid £5 for their participation.

Stimuli and tasks. Participants completed three of the four tasks used in Study 1: The GFMT, AIFMT, and Lineup Task. Session 1 stimuli from Study 1 were used to create the simultaneous version of each task and Session 2 stimuli were used to create the sequential version. Identities that were recognized by more than four participants in Study 1 were replaced.

Procedure. Participants were tested online using Testable. Prior to starting the tasks, participants were asked to calibrate their screens by using their arrow keys to resize a line to the size of a credit card. All participants completed the tasks in the same order: The GFMT, AIFMT, Lineup Task. Trials within each task were blocked, such that for each task the simultaneous version preceded the sequential version. Participants were asked to respond as accurately as possible for each task. Feedback was not provided.

The protocols for the simultaneous version of each task were the same as in Study 1 (i.e., images were presented simultaneously until participants indicated their responses). For each trial in the sequential version, the target image (left-most image from each pair in the GFMT and AIFMT) was presented for 1500 ms. After an 800-ms interstimulus interval (ISI), the test image or lineup array was presented and remained on screen until participants indicated their response. For the GFMT and AIFMT participants were instructed to press "f" to indicate that the images were of the same person, and "j" to indicate that the images were of two different people. For the Lineup Task, participants were instructed to select the image that corresponded with the target if they felt the target was present, or to select the "target absent" image, if they felt the target was absent. See Figure 3-2 for a depiction of the tasks.

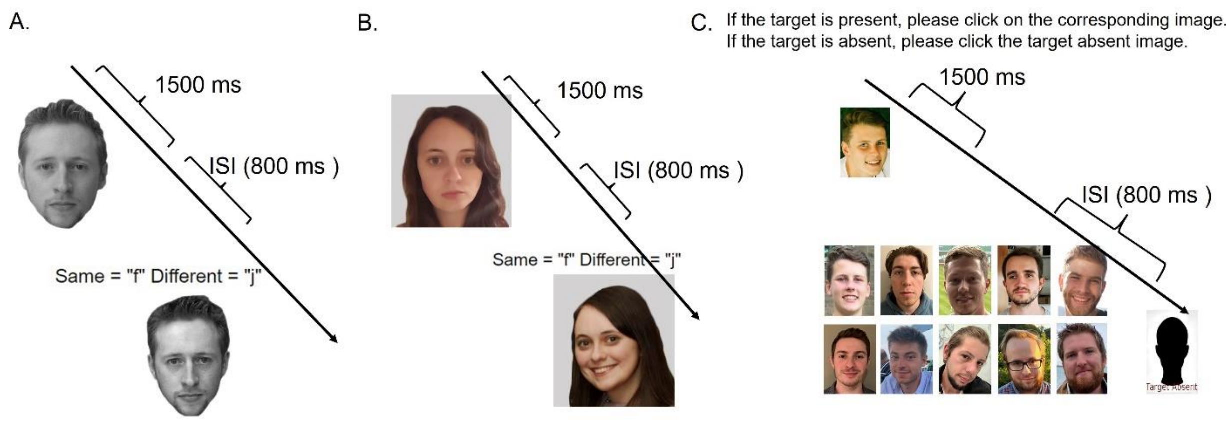

Figure 3-2 depicts examples of the sequential versions of the Glasgow Face Matching Test (A), the Ambient Image Face Matching Task (B), and the Lineup Task (C).

Results

Data analyses. The primary analyses were identical to those in Study 1. To further investigate what is being captured by c, I also investigated RTs. For each task, I calculated each participant's median RT on correct trials separately for match and mismatch trials. I then created a difference score (match RT – mismatch RT) as a measure of participants' tendencies to respond faster to match vs. mismatch trials. To investigate whether participants were faster on trials that were congruent with their bias I examined the correlation between difference scores. A positive correlation (larger difference scores associated with a more liberal response bias) would provide evidence that c provides a measure of decision making. Separate analyses were conducted for the simultaneous and sequential versions of each task[6].

[6] In the RT analyses, n = 9 participants were excluded ($GFMT_{sim.}$: n = 1; $Lineup_{sim.}$: n = 1; $Lineup_{seq.}$: n = 7) as they failed to make any correct responses on the mismatch trials (n = 1), target present trials (n = 6) or target absent trials (n = 2), respectively.

Relationships between tasks. As shown in Table 4, sensitivity to identity (rs > 0.26, ps < 0.002) was related across tasks, regardless of whether the task involved simultaneous or sequential stimulus presentation. Bias showed the same pattern (rs > 0.19, ps < 0.03), with the exception that the correlation between bias in the sequential GFMT and simultaneous Lineup task was not significant ($r = 0.05$, $p = 0.58$).

Table 3-4.

Correlations between performance for simultaneous and sequential versions of each face task.

	GFMT sim. d′	AIFMT sim. d′	LU sim. acc	GFMT seq. d′	AIFMT seq. d′	LU seq. acc	GFMT sim. c	AIFMT sim. c	LU sim. bias	GFMT seq. c	AIFMT seq. c	LU seq. bias
GFMT sim. d′ †	--	0.51***	0.47***	0.43***	0.41***	0.28**	- 0.11	-0.12	- 0.17	-0.16	0.05	0.21**
AIFMT sim. d′		--	0. 48***	0.41***	0.51***	0.33***	0.006	-0.11	-0.03	0.001	-0.03	0.23**
LU sim. acc			--	0.43***	0.63***	0.60***	0.05	0.23**	0.14	0.19*	0.07	0.72***
GFMT seq. d′				--	0.40***	0.27**	0.003	0.07	0.12	-0.01	0.03	0.24**
AIFMT seq. d′					--	0.47***	-0.05	0.04	0.04	0.08	-0.02	0.32***
LU seq. acc						--	0.16*	0.23**	0.24**	0.24**	0.09	0.57***
GFMT sim. c							--	0.43***	0.19*	0.45***	0.24**	0.21*
AIFMT sim. c								--	0.34***	0.39***	0.63***	0.49***
LU sim. bias									--	0.05	0.27***	0.43***
GFMT seq. c										--	0.31***	0.28**
AIFMT seq. c											--	0.33***
LU seq. bias												--

†. Spearman’s correlations reported for the GFMT match d′ due to lack of normality.

* *p< 0.05. ** p < 0.01. ***p < 0.001.*

Underlying face identification abilities. I conducted a PCA with an Oblimin rotation using standardized scores for sensitivity to identity and bias for each version of each task. The Kaiser-Meyer-Olkin was adequate (KMO = 0.73) and Bartlett's test of sphericity was significant (χ^2 (66) = 693.38, $p < 0.001$), suggesting this data was suitable for PCA. Although Kaiser's criterion suggested a 3-component solution, both the Skree plot and Parallel analysis suggested there were only two components; these two components explained 52.30% of the variance. As shown in Table 5, sensitivity to identity for both versions of all three tasks loaded positively on Component 1. Bias for both versions of all three tasks loaded positively on Component 2.

Table 3-5.

Communality and factor loadings for performance on each of the face tasks (simultaneous and sequential versions).

	Measure	Communality	Loading	Cronbach's alpha
Component 1	GFMT simultaneous d′	0.44	0.65	0.81
	AIFMT simultaneous d′	0.57	0.75	
	Lineup simultaneous acc	0.77	0.82	
	GFMT sequential d′	0.39	0.63	
	AIFMT sequential d′	0.61	0.79	
	Lineup sequential acc	0.55	0.70	
Component 2	GFMT simultaneous c	0.39	0.63	0.75
	AIFMT simultaneous c	0.73	0.86	
	Lineup simultaneous bias	0.27	0.50	
	GFMT sequential c	0.39	0.63	
	AIFMT sequential c	0.47	0.69	
	Lineup sequential bias	0.69	0.61	

Note. The loadings are the result of an Oblimin rotation. The internal consistency for both components (Cronbach's alpha) is also included.

RTs and bias. ***Lineup Task.*** As hypothesized, both the simultaneous and sequential tasks showed longer RTs on the target-absent trials ($M_{sim.}$ = 16.58 s, $SD_{sim.}$ = 8.09; $M_{seq.}$ = 6.83, $SD_{seq.}$= 2.65) than on target-present trials ($M_{sim.}$ = 13.62, $SD_{sim.}$ = 8.54; $M_{seq.}$ = 5.73, $SD_{seq.}$= 2.21), *p*s <0.001 , *d*s >0.42, suggesting that RTs are appropriate to investigate these processes. Bias predicted a small, but significant, amount of variance in RT difference scores (present – absent) in the sequential version of the Lineup Task ($r = 0.17$, $p = 0.04$), but not the simultaneous version ($r = 0.13$, $p = 0.13$).

GFMT. c positively predicted significant variance in the difference in RTs between match and mismatch trials in both the simultaneous ($r = 0.38$, $p < 0.001$) and sequential versions of the GFMT ($r = 0.34$, $p < 0.001$). In both cases, as criterion becomes more conservative RTs become longer on match as compared to mismatch trials.

AIFMT. c positively predicted significant variance in the difference in RTs between match and mismatch trials in both the simultaneous ($r = 0.53$, $p < 0.001$) and sequential versions of the GFMT ($r = 0.59$, $p < 0.001$). In both cases, as criterion becomes conservative criterion RTs become longer on match as compared to mismatch trials (See Figure 3-3).

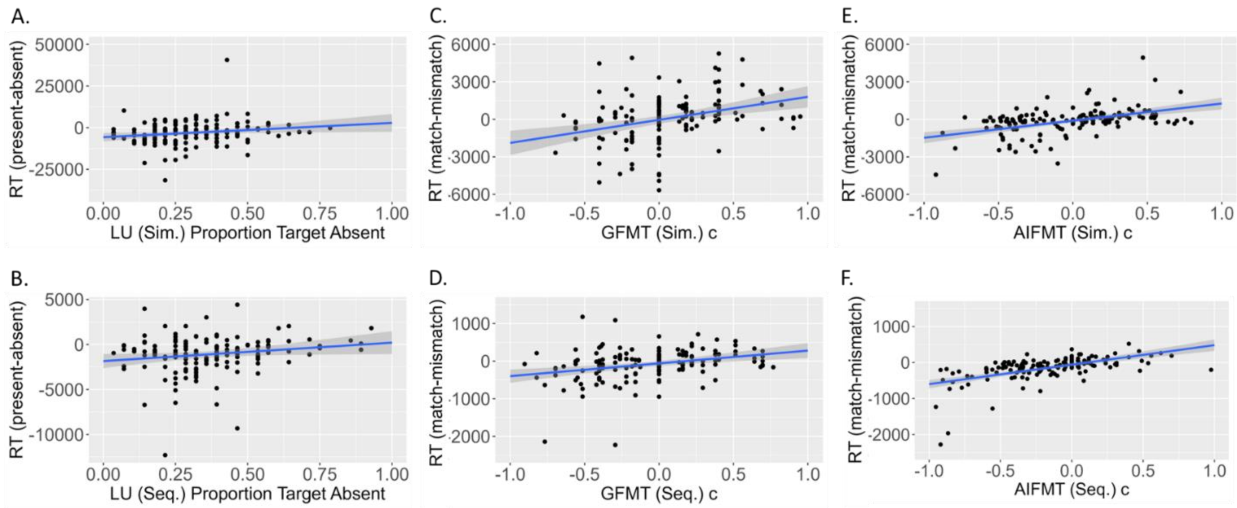

Figure 3-3. Depicts the difference score in RTs for match vs. mismatch trials as a function of response bias in the simultaneous (top row) and sequential (bottom row) of the lineup (A, B), GFMT (C, D), and AIFMT (E,F) tasks.

Discussion

Study 2 yielded three important findings. First, Study 2 confirmed that individual differences in sensitivity to identity are stable across tasks, even when memory demands are introduced by presenting the stimuli sequentially. Many of relationships in Study 2 were the same size as those shown in Study 1, with the only exception being the sequential Lineup Task. Weaker relationships between the sequential Lineup task in Study 2 and the matching tasks might be the result of the complexity of the lineup array as compared to a pair of images in the matching tasks. Recent studies reported that individual differences in matching identity of unfamiliar faces correlate with other aspects of face perception (e.g., holistic processing: Verhallen et al., 2017; accuracy on a face learning task: Fysh & Bindemann, 2018; familiar face identification: McCaffery et al., 2018). My results build upon previous findings and show that measures of sensitivity to identity from three different tasks loaded onto Component 1, regardless of whether stimuli were presented simultaneously or sequentially (i.e., regardless of whether participants could scan between two images or had to hold the first face in short-term memory).

Second, Study 2 provided the first evidence that individual differences in response bias are stable across tasks, even when memory demands are varied by presenting the stimuli sequentially vs. simultaneously. Bias has been ignored in most studies despite it influencing the types of errors (false alarms vs. misses) perceivers with comparable sensitivity to identity will display.

Third, my analyses of RTs strongly suggest that individual differences in response bias are related to decision making. Slower RTs on target-absent vs. target-present trials of the Lineup Task confirm validity in using RTs to examine decision making; in order to reach a target-absent solution, participants must first consider and ultimately decide against each identity in the lineup. On tasks in which there were two alternatives, response bias and trial type interacted such that participants were faster on trial types (i.e., match, mismatch) that were congruent with their biases.

It is striking that the simultaneous and sequential versions of the tasks loaded onto the same, rather than different, components—especially given that participants performed worse on the sequential than the simultaneous tasks. This difference in performance occurred despite sequential versions needing to be recalled only a second later than simultaneous versions. Despite using stimuli from the same database, Burton et al. (2010) found lower, albeit significant, relationships between the GFMT and an old/new face recognition memory task than for the GFMT and the Familiar Figures Matching Test. This discrepancy could be because recognition memory paradigms have a longer delay between targets and recall than in the current study. Whereas tasks with memory demands require perceivers to build a representation in memory, perceptual matching tasks do not (Ritchie et al., 2021). My findings suggest that the ability to identify faces when stimuli are presented simultaneously, a protocol that does not require building a mental representation, predicts the ability to identify faces when short-term memory demands are introduced (i.e., in sequential tasks). This pattern of results suggests that the ability to identify unfamiliar faces in a perceptual matching task may have cascading effects on the ability to learn a new face. I discuss the implications of these findings below.

General Discussion

There are vast individual differences in unfamiliar face identification (for examples see Fysh, 2018; Fysh & Bindemann, 2017; McCaffery et al., 2018; Megreya & Bindemann, 2013; Verhallen et al., 2017; for a review/editorial see Bruce et al., 2018); interest in those individual differences is growing (see Wilmer, 2017). Prior to this study, the degree to which individual differences in sensitivity to identity are stable across time and consistent across tasks remained unclear and no study had investigated the stability of individual differences in response bias on face identification tasks. Here, I report two main findings.

Individual differences in sensitivity to identity are stable across time and tasks. In Study 1, I showed ample evidence of moderate to high stability for sensitivity to identity in face identification tasks. This reliability suggests that the face identification measures reported in Study 1 accurately capture stable individual differences, consistent with previous estimates (e.g., Fysh et al., 2020; Fysh & Bindemann, 2018; White et al., 2021; Murray & Bate, 2020; Burton et al., 2010; Rhodes et al., 2014; Verhallen et al., 2017). PCA revealed a component that reflected sensitivity to identity—suggesting stable individual differences in the ability to distinguish between face pairs that belong to the same person vs. different people. In Study 2, the measures of sensitivity to identity loaded onto the same component in the simultaneous and sequential versions, suggesting comparable individual differences when short-term memory demands are introduced. Thus, I provide further evidence of one underlying face identification ability (*f* ; see McCaffery et al., 2018; Verhallen et al., 2017).

The present studies do not speak to potential sources of individual differences, other research has shown potential underlying mechanisms. In addition to genetics (e.g., Wilmer et al., 2010; for a discussion see Wilmer, 2018), differential experience likely underlies individual differences in sensitivity to identity. Jenkins and colleagues (2018) estimated that on average individuals recognize 5000 faces, but participants ranged from knowing approximately 1000 to 10,000 people. One hypothesis is that individuals who know more faces are more sensitive to identity in unfamiliar faces. This hypothesis is consistent with evidence that individuals from small towns perform more poorly on face memory tasks as compared to individuals from cities (Balas & Saville, 2015). A role for experience is also consistent with age-related improvements in face identification during childhood, attributable to children becoming familiar with more identities (e.g., Megreya & Bindemann, 2015; Neil et al., 2016; Laurence & Mondloch, 2016), and with evidence that adults make more errors when identifying unfamiliar faces from categories with which they lack experience (e.g., other-race and other-age faces, Laurence et al., 2015; Prioetti et al., 2019; Yovel et al., 2012; Meissner & Brigham, 2001). Individual differences in norm-based coding, a process thought to be influenced by experience, may also play a role (Rhodes et al., 2014).

A role for experience is also consistent with evidence from recent models of face learning. DCNNs are a class of learning algorithm built to model the human visual system

(Phillips et al., 2018; Noyes et al., 2021). Like humans, DCNNs make errors when identifying unfamiliar (i.e., untrained) faces, but their accuracy in doing so is dependent on the number of identities on which the algorithm is trained (Blauch et al., 2021; Hill et al., 2019; O'Toole et al., 2018; O'Toole & Castillo, 2021; Rosemblaum et al., 2021).

Individual differences in response bias are stable across time and tasks. Here, I provided the first examination of whether individual differences in response bias on face identification tasks are stable across time. Similar to Lindsay and Kantner (2012), who showed stable individual differences in response bias to word and art stimuli, I showed stable individual differences in response bias on face identification tasks.

Building on past research by Verhallen et al. (2017) and McCaffery et al. (2018) who reported a single ability underlying face recognition, *f*, I showed another component—one that reflected bias. Again, individual differences held across tasks that varied in structure (e.g., procedure, variability in stimuli) and across simultaneous vs. sequential tasks. This finding is novel, as individual differences in bias had yet to be explored. I propose that there might also be one general overarching bias for face identification tasks: f_b.

Another novel finding from the PCA concerns interpretations of the number of piles made in the Sorting Task (Study 1). To the authors' knowledge, this was the first investigation into whether each measure from the Sorting Task (i.e., number of piles, misidentification) loads onto a component with measures of sensitivity to identity or response bias. Although the number of misidentifications made in the Sorting Task loaded onto the sensitivity to identity component, the number of piles made in the Sorting Task loaded onto the bias component. A previous study reported that individuals who were homeschooled made more piles than individuals who attended public school (Short et al., 2017). My findings suggest that this finding might be reinterpreted as evidence that experience influences criterion, such that individuals with less experience are more conservative. This interpretation might also apply to evidence that adults make more piles when sorting unfamiliar other- as compared to own-race faces (Laurence et al., 2016; Zhou & Mondloch, 2016).

What captures the unexplained variance? The estimates provided here (e.g., correlations and PCA) do not account for 100% of the variance in scores. This further demonstrates the need for assessing reliability and convergent validity of face identification tasks—a topic discussed by White and Burton (2022). This is especially important as there is evidence that participants make different responses when tested with the same trials (Bindemann et al., 2012b; Fysh & Bindemann, 2018) and show a choice blindness when asked to justify decisions (Sauerland et al., 2016). Collectively, this is evidence that individual differences will not align perfectly across face-identification tasks (for other examples see Fysh et al., 2020; McCaffery et al., 2018; Stacchi et al., 2020; Verhallen et al., 2017; for a discussion on this effect in Super recognizers, see Ramon et al., 2019). Some of this variance would arguably be error variance, and/or attributable to the ability to perform well under the demands associated with each individual task (i.e., with dichotomous forced-choice tasks vs. lineup tasks). It is likely that the residuals between simultaneous and sequential versions will be accounted for by individual differences in memory.

What other mechanisms might explain the remaining variance in face identification scores? Researchers have investigated the ability of IQ, personality, and self-reported face identification abilities to serve as predictors (Wilhelm et al., 2010; Megreya & Bindemann, 2013; McCaffery et al., 2018; Verhallen et al., 2017; for reviews see White & Burton, 2022 and Wilmer, 2017). It is also possible that individual differences in face identification strategies might explain some of this unexplained variance. For example, Fysh and Bindemann (2022) showed that directing participants' attention to moles improved face identification performance and Towler et al. (2021) showed that novices' performance improves when they implement the same identification strategies as used by Forensic Examiners. Unexplained variance might also be attributed to the fact that our typical experience with faces is multimodal (for a discussion see White & Burton, 2022); the amount of explained variance in scores might increase if additional identity information (e.g., voice, body, gait) was provided to participants.

Are these components face-specific? Although I did not include any non-face tasks (e.g., object recognition), past studies suggest little or no relationship between face identification accuracy and other cognitive skills. Verhallen and colleagues (2017) used a factor analysis to show that although their four face-perception tasks loaded onto a single factor, other skills (e.g., form perception, motion perception, contrast sensitivity) did not load with them. In a recent review, Wilmer (2017) suggested that other skills (e.g., IQ, verbal recognition) were limited in their ability to predict face identification accuracy. The lack of relationship is not surprising because the challenge of unfamiliar face identification extends beyond that required for object matching. Faces are one of the very few socially relevant stimulus categories that need to be recognized at the individual (Tim, Jane) as compared to the categorical (cup, bowl) level—at least among categories that are studied by cognitive psychologists. The challenge of discriminating between identities is compounded by faces being homogeneous on the one hand (all faces have two eyes over a nose and mouth) and displaying more within-individual variability than any other stimulus category on the other hand. In addition to changing with lighting and viewpoint, changes common to all categories, faces show non-rigid changes (e.g., due to changes in expression—the face contains several muscles, many of which belong to a subclass of muscles [i.e., mimetec] that are used in emotional expression [see Marur et al., 2014]). Further, many individuals vary their appearance via aesthetic changes (e.g., make-up, glasses). Thus, I suggest that both the challenge of unfamiliar face identification and individual differences are largely face-specific.

It remains unclear if the component reflecting bias is also face-specific. Given that bias is flexible (i.e., is sensitive to both base rates and costs vs. benefits associated with each response), it might not be. Evidence in line with this hypothesis comes from the recognition memory literature and suggests that individual differences in bias are stable across stimulus types (Kantner & Lindsay, 2014). One conceptualization is that sensitivity to identity is shaped by face-specific experience and reflects the challenge of discriminating within- vs. between-person variability in appearance, whereas bias is best conceived in light of economic models and reflects more general principles of decision making—a fruitful direction for future research.

Implications. ***Theoretical.*** Models of face identification have focused primarily on the ability to distinguish between face pairs showing the same person vs. two different people (i.e., sensitivity to identity; e.g., Bruce & Young, 1986; Valentine, 1991; O'Toole et al., 2018). Advances have also been made for understanding the representations of familiar faces. For example, recent work using similarity ratings reported that familiarity with a face influences representation of within-person variability more so than between-person variability (White et al., 2022).

Classic models of face identification approach the challenge of identifying faces as if it is purely a perceptual problem. Work investigating how context (e.g., Memon & Bruce, 1985; Young et al., 1985) and priming effects (e.g., Brennen & Bruce, 1991; Faulkner et al., 2002; Laurence et al., 2022) influence face identification suggests that multiple cognitive mechanisms are involved. For example, a familiar face is more likely to be recognized when encountered in an expected vs. novel context. Similarly, faces are learned better when they are learned with congruent conceptual information (i.e., person-related labels; e.g., name) than with incongruent conceptual information (e.g., object-related labels; see Schwartz & Yovel, 2016). The current study aligns with Bindemann and Burton (2021) who argued that contemporary models of face identification need to integrate decision making into their frameworks. Though contemporary models using DCNNs have begun to broach the idea of criterion (e.g., Cavazos et al., 2020), more research is needed. For example, researchers might examine the extent to which individual differences in criterion reflect individual differences in perceived similarity (i.e., perception) vs. individual differences in the threshold set for reaching a *same identity* conclusion (i.e., decision making).

Applied. The ability to distinguish between photographs of the same person vs. different people is important in applied settings. My discovery that individual differences in sensitivity to identity are stable over time and across tasks provides a timely answer to the question of whether people who are especially good at face identification in one context (e.g., a lineup task) are also especially good in other contexts (e.g., when judging the validity of photo ID; Kemp et al., 2021). A high-performing passport officer is likely to be good at spotting an unfamiliar face in a crowd.

Though overshadowed by sensitivity to identity (e.g., percent correct, d′), individual differences in criterion also have ramifications for applied settings. Towler et al. (2022) showed that there are different paths to expertise. Forensic examiners and super recognizers had comparable percent correct trials, but differed in criterion. Whereas forensic examiners had a neutral response bias, super recognizers had a strong liberal bias; thus, super recognizers were more likely than forensic examiners to misperceive photos of different people as belonging to the same person. The stable individual differences in criterion observed in the current study extend this pattern to the general population. Passport officers, clerks in stores selling age-restricted goods, and health care workers differ not only in the probability of making an error when checking photo ID, but also in the type of error they are most likely to make.

One goal of applied research is to improve performance of individuals tasked with face identification. To date, no training protocol has substantially improved overall accuracy (for brief reviews see Kemp et al., 2021; White et al., 2021). Most training protocols have improved performance on match *or* mismatch trials (e.g., Alenzi & Bindemann, 2013; Ritchie et al., 2021; for a brief discussion see Ritchie & Burton, 2017). I argue that improved performance on match *or* mismatch trials is best conceptualized as a shift in criterion. Although stable over time and across tasks (current study), this effect of training suggests that response bias might be more flexible (i.e., responsive to training) than sensitivity to identity. This is consistent with Gentry and Bindemann (2019), who argued that providing task examples might help low performers to adopt a task-appropriate criterion.

Integrating response bias into training protocols aligns with economic models of decision making. Economic models suggest that response bias increases in the context of perceptual uncertainty, a condition that defines unfamiliar face identification (see Lynn & Barrett, 2014; Lynn et al., 2015; Summerfield & Tsetsos, 2012). Under conditions of uncertainty, the direction of response bias (liberal vs. conservative) is influenced by base rates (e.g., proportion of match trials) and the costs vs. benefits associated with each response. Such effects provide further evidence that response bias might be flexible.

Protocols designed to influence response bias might be a promising route to minimize costly errors, even if they do not reduce overall errors. Future research should examine shifts in response bias as a function of perceptual uncertainty (e.g., similarity of faces, individual differences in sensitivity to identity) under different task conditions and the extent to which people can adaptively shift their bias based on task demands. Doing so would speak to the recent call to action by Bindemann and Burton (2021) who discussed that decision-making processes must be incorporated into our theories of face identification, as face identification is not a purely perceptual problem.

Conclusion. There is a consensus that most adults are experts with familiar faces, but there exists an ongoing debate as to whether adults are experts with unfamiliar faces (Rossion, 2018; Sunday & Gauthier, 2018; Young & Burton, 2018a, 2018b, 2018c). Regardless of whether adults should be labelled *experts*, it is clear that individuals vary in their ability to distinguish match vs. mismatch images. Whereas past studies showed individual differences can be attributed to experience (e.g., Balas & Saville, 2015; Short et al., 2017) and are heritable (e.g., Wilmer et al., 2010; for a discussion see Wilmer, 2017), I have shown that individual differences in sensitivity to identity and bias are stable over time and tasks. Rather than focusing on *whether* humans' expertise extends to unfamiliar faces, it would be more profitable to focus on individual differences in such expertise. Understanding the causes of individual differences and the strategies used by good vs. poor performers would provide insight into the nature of adults' expertise, informing theoretical models and providing avenues for training protocols in applied settings.

Supplemental material

S1. Descriptive statistics of each test for Study 1 (Session 1), as well as the maximum possible d′ that could be achieved in the task.

Table 3-S6-1

Lineup Task	*M*	*SD*
Total Proportion correct	0.35	0.15
Proportion of Target absent selection	0.32	0.16
Misidentifications	0.35	0.20
Accuracy when target present	0.32	0.16
Accuracy when target absent	0.38	0.23

Table 3-S7-2

GFMT	*M*	*SD*
Hits	0.79	0.20
FAs	0.35	0.18
d′ *	1.73	0.91
C	-0.04	0.48

* max possible d′ = 3.29

Table 3-S8-3

AIFMT	*M*	*SD*
Hits	0.68	0.17
FAs	0.34	0.18
d′ *	0.97	0.66
c	-0.02	0.45

* max possible d′ = 4.26

Table 3-S9-4

Sorting Task	*M*	*SD*
Number of piles	5.05	4.59
Misidentifications	2.50	3.82

S2. Descriptive statistics of each test and the number of participants showing a liberal bias for the GFMT and AIFMT.

Table 3-S10-1

Lineup Task	Simultaneous		Sequential	
	M	*SD*	*M*	*SD*
Total Proportion correct	0.39	0.15	0.31	0.12
Proportion of Target absent selection	0.31	0.15	0.36	0.18
Misidentifications	0.32	0.20	0.46	0.18
Accuracy when target present	0.38	0.18	0.22	0.12
Accuracy when target absent	0.41	0.22	0.39	0.22

Table 3-S11-2

GFMT	Simultaneous		Sequential	
	M	*SD*	*M*	*SD*
Hits	0.83	0.20	0.78	0.16
FAs	0.15	0.18	0.26	0.17
d′	2.16	0.96	1.56	0.73
c*	0.04	0.39	-0.07	0.40

* In the simultaneous version, 51 participants showed a liberal bias. In the sequential version 73 participants showed a liberal bias.

Table 3-S12-3

AIFMT	*Simultaneous*		*Sequential*	
	M	*SD*	*M*	*SD*
Hits	0.71	0.14	0.72	0.14
FAs	0.29	0.14	0.40	0.16
d′	1.31	0.58	0.93	0.49
c*	-0.04	0.39	-0.18	0.38

* In the simultaneous version, 71 participants showed a liberal bias. In the sequential version 92 participants showed a liberal bias.

References

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*(6), 735-753. https://doi.org/10.1002/acp.2968

Andrews, S., Burton, A. M., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. *Quarterly Journal of Experimental Psychology*, *70*(8), 1620-1632. https://doi.org/10.1080/17470218.2016.1195851

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. https://doi.org/10.1080/17470218.2014.1003

Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. *Neuropsychologia*, *69*, 211-217. https://doi.org/10.1016/j.neuropsychologia.2015.02.005

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision research*, *141*, 228-236. https://doi.org/10.1016/j.visres.2016.12.005

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19-30. https://doi.org/10.1016/j.cognition.2016.12.012

Baker, K. A., & Mondloch, C. J. (2019). Two Sides of Face Learning: Improving Between-Identity Discrimination While Tolerating More Within-Person Variability in Appearance. *Perception*, 48(11), 1124-1145. https://doi.org/10.1177/0301006619867862

Bindemann, M., Avetisyan, M., & Rakow, T. (2012b). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied*, *18*(3), 277. https://doi.org/10.1037/a0029635

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012a). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(2), 96-103. https://doi.org/10.1016/j.jarmac.2012.02.001

Bindemann, M., & Burton, A. M. (2021). Steps Towards a Cognitive Theory of Unfamiliar Face Matching. In *Forensic Face Matching*. Oxford University Press. https://doi.org/10.1093/oso/9780198837749.003.0003

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2021). Deep learning of shared perceptual representations for familiar and unfamiliar faces: Reply to commentaries. *Cognition*, 208, 104484. https://doi.org/10.1016/j.cognition.2020.104484

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). A grey area: how does image hue affect unfamiliar face matching?. Cognitive Research: Principles and Implications, 4(1), 1-10. https://doi.org/10.1186/s41235-019-0174-3

Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision 10*:433-436.

Brennen, T., Bruce, B. Context effects in the processing of familiar faces. *Psychological Research* 53, 296–304 (1991). https://doi.org/10.1007/BF00920483

Bruce, V., Bindemann, M., & Lander, K. (2018). Individual differences in face perception and person recognition. *Cognitive Research: Principles and Implications, 3(18).* https://doi.org/10.1186/s41235-018-0109-4

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. Journal of Experimental Psychology: Applied, 5(4), 339. https://doi.org/10.1037/1076-898X.5.4.339

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, 77(3), 305-327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, 66(8), 1467-1485. https://doi.org/10.1080/17470218.2013.800125

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

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, *40*(1), 202-223. https://doi.org/10.1111/cogs.12231

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior research methods*, 42(1), 286-291. https://doi.org/10.3758/BRM.42.1.286

Cavazos, J. G., Phillips, P. J., Castillo, C. D., & O'Toole, A. J. (2020). Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?. *IEEE transactions on biometrics, behavior, and identity science,* 3(1), 101-111. doi: 10.1109/TBIOM.2020.3027269.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98. https://doi.org/10.1037/0021-9010.78.1.98

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. https://doi.org/10.1080/17470218.2015.1017513

Dunn, J. D., Summersby, S., Towler, A., Davis, J. P., & White, D. (2020). UNSW Face Test: A screening tool for super-recognizers. PloS one, 15(11), e0241747. https://doi.org/10.1371/journal.pone.0241747

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *i-Perception, 5(7),* 589-601. https://doi.org/10.1068/i0669

Faulkner, T.F., Rhodes, G., Palermo, R. et al. Recognizing the un-real McCoy: Priming and the modularity of face recognition. *Psychonomic Bulletin & Review* 9, 327–334 (2002). https://doi.org/10.3758/BF03196289

Fysh, M. C. (2018). Individual differences in the detection, matching and memory of faces. *Cognitive Research: Principles and Implications*, *3*(1), 1-12. https://doi.org/10.1186/s41235-018-0111-x

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. *British journal of psychology*, 109(2), 219-231. DOI:10.1111/bjop.12260

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society open science*, 7(9), 200233. https://doi.org/10.1098/rsos.200233

Gentry, N. W., & Bindemann, M. (2019). Examples improve facial identity comparison. *Journal of Applied Research in Memory and Cognition*, *8*(3), 376-385. https://doi.org/10.1016/j.jarmac.2019.06.002

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, 69, 14-25. https://doi.org/10.1016/j.concog.2019.01.008

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior research methods*, 50(3), 1166-1186. https://doi.org/10.3758/s13428-017-0935-1

Hill, M. Q., Parde, C. J., Castillo, C. D., Colon, Y. I., Ranjan, R., Chen, J. C., Blanz, V. & O'Toole, A. J. (2019). Deep convolutional neural networks in the face of caricature. *Nature Machine Intelligence*, *1*(11), 522-529. https://doi.org/10.1038/s42256-019-0111-7

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know?. *Proceedings of the Royal Society B, 285*(1888), 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. https://doi.org/10.1016/j.cognition.2011.08.001

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & cognition*, 40(8), 1163-1177. https://doi.org/10.3758/s13421-012-0226-0

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, *21*(5), 1272-1280. https://doi.org/10.3758/s13423-014-0608-3

Kemp, R. I., Edmond, G., & White, D. (2021). A proposed solution to the problem of identifying people from CCTV and other images. In *Methods, Measures, and Theories in Eyewitness Identification Tasks* (pp. 13-33). Routledge.

Kleiner M., Brainard D. & Pelli D. 2007 “What’s new in Psychtoolbox-3?” Perception 36 ECVP Abstract Supplement.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kramer, R. S., Jenkins, R., Young, A. W., & Burton, A. M. (2017). Natural variability is essential to learning new faces. *Visual Cognition*, 25(4-6), 470-476. https://doi.org/10.1080/13506285.2016.1242522

Kress, T., & Daum, I. (2003). Developmental prosopagnosia: A review. *Behavioural neurology*, 14(3, 4), 109-121. https://doi.org/10.1155/2003/520476

Lander, K., Bruce, V., & Bindemann, M. (2018). Use-inspired basic research on individual differences in face identification: Implications for criminal investigation and security. *Cognitive Research: Principles and Implications*, 3(1), 1-13. https://doi.org/10.1186/s41235-018-0115-6

Laurence, S., Baker, K. A., Proietti, V. M., & Mondloch, C. J. (2022). What happens to our representation of identity as familiar faces age? Evidence from priming and identity aftereffects. British Journal of Psychology. https://doi.org/10.1111/bjop.12560

Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, *143*, 123-138. https://doi.org/10.1016/j.jecp.2015.09.030

Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. *British Journal of Psychology*, 107(2), 374-388. https://doi.org/10.1111/bjop.12147

Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, & Psychophysics*, 76, 1975–1984. https://doi.org/10.3758/s13414-014-0653-z

Lynn, S. K., & Barrett, L. F. (2014). “Utilizing” signal detection theory. *Psychological science*, *25*(9), 1663-1673. https://doi.org/10.1177/0956797614541991

Lynn, S. K., Wormwood, J. B., Barrett, L. F., & Quigley, K. S. (2015). Decision making from economic and signal detection perspectives: development of an integrated framework. *Frontiers in psychology*, *6*, 952. https://doi.org/10.3389/fpsyg.2015.00952

Marur, T., Tuna, Y., & Demirci, S. (2014). Facial anatomy. Clinics in dermatology, 32(1), 14-23. https://doi.org/10.1016/j.clindermatol.2013.05.022

Matthews, C. M., & Mondloch, C. J. (2018). Finding an unfamiliar face in a line-up: Viewing multiple images of the target is beneficial on target-present trials but costly on target-absent trials. *British Journal of Psychology*, *109*(4), 758-776. https://doi.org/10.1111/bjop.12301

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, 6(6), 255-260. https://doi.org/10.1016/S1364-6613(02)01903-4

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*(1), 1-15. https://doi.org/10.1186/s41235-018-0112-9

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30. doi:10.1037/1082-989X.1.1.30

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology*, 25(1), 30-37. https://doi.org/10.1080/20445911.2012.739153

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*(1), 5-22. https://doi.org/10.1068/p7825

Megreya, A. M., Bindemann, M., & Havard, C. (2011). Sex differences in unfamiliar face identification: Evidence from matching tasks. *Acta Psychologica*, *137*(1), 83-89. https://doi.org/10.1016/j.actpsy.2011.03.003

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. https://doi.org/10.1037/a0013464

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7(1), 3. https://doi.org/10.1037/1076-8971.7.1.3

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & cognition*, *33*(5), 783-792. https://doi.org/10.3758/BF03193074

Memon, A., Bruce, V. Context effects in episodic studies of verbal and facial memory: A review. *Current Psychology* 4, 349–369 (1985). https://doi.org/10.1007/BF02686589

Menon, N., White, D., & Kemp, R. I. (2015). Variation in photos of the same face drives improvements in identity verification. Perception, 44(11), 1332-1341. https://doi.org/10.1177/0301006615599902

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630-633. https://doi.org/10.3758/BF03206543

Mondloch, C.J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception*, *31*, 553–566. https://doi.org/10.1068/p3339

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. http://dx.doi.org/10.1037/xhp0000049

Murray, E., & Bate, S. (2020). Diagnosing developmental prosopagnosia: repeat assessment using the Cambridge Face Memory Test. *Royal Society open science*, 7(9), 200884. https://doi.org/10.1098/rsos.200884

Neil, L., Cappagli, G., Karaminis, T., Jenkins, R., & Pellicano, E. (2016). Recognizing the same face in different contexts: Testing within-person face recognition in typical development and in autism. *Journal of Experimental Child Psychology*, *143*, 139-153. https://doi.org/10.1016/j.jecp.2015.09.029

Noyes, E., Parde, C. J., Colón, Y. I., Hill, M. Q., Castillo, C. D., Jenkins, R., & O'Toole, A. J. (2021). Seeing through disguise: Getting to know you with a deep convolutional neural network. *Cognition*, *211*, 104611. https://doi.org/10.1016/j.cognition.2021.104611

O'Toole, A. J., & Castillo, C. D. (2021). Face Recognition by Humans and Machines: Three Fundamental Advances from Deep Learning. *Annual Review of Vision Science*, *7*, 543-570. https://doi.org/10.1146/annurev-vision-093019-111701

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in cognitive sciences*, *22*(9), 794-809. https://doi.org/10.1016/j.tics.2018.06.006

Papesh, M. H., Heisick, L. L., & Warner, K. A. (2018). The persistent low-prevalence effect in unfamiliar face-matching: The roles of feedback and criterion shifting. *Journal of Experimental Psychology: Applied*, 24(3), 416. https://doi.org/10.1037/xap0000156

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and human behavior*, *36*(3), 247. https://doi.org/10.1037/h0093923

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision 10*:437-442.

Phillips, P. J., Beveridge, J. R., Draper, B. A., Givens, G., O'Toole, A. J., Bolme, D. S., ... & Weimer, S. (2011). An introduction to the good, the bad, & the ugly face recognition challenge problem. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 346-353). IEEE. DOI: 10.1109/FG.2011.5771424

Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, *32*(1), 74-85. https://doi.org/10.1016/j.imavis.2013.12.002

Phillips, P.J., Yates, A.N., Hu, Y., Hahn, C.A., Noyes, E., Jackson, K., Cavazos, J.G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J., Castillo, C.D., Chellappa, R., White, D. & O'Toole, A.J., (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *PNAS, 115*(24):6171–76. https://doi.org/10.1073/pnas.1721355115

Proietti, V., Laurence, S., Matthews, C. M., Zhou, X., & Mondloch, C. J. (2019). Attending to identity cues reduces the own-age but not the own-race recognition advantage. *Vision research*, *157*, 184-191. https://doi.org/10.1016/j.visres.2017.11.010

Ramon, M. (2021). Super-Recognizers-a novel diagnostic framework, 70 cases, and guidelines for future work. Neuropsychologia, 158, 107809. https://doi.org/10.1016/j.neuropsychologia.2021.107809

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, 20(4), 260-281. https://doi.org/10.1016/j.tics.2016.01.007

Rhodes, G., Jeffery, L., Taylor, L., Hayward, W. G., & Ewing, L. (2014). Individual differences in adaptive coding of face identity are linked to individual differences in face recognition ability. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 897. https://doi.org/10.1037/a0035939

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897-905. https://doi.org/10.1080/17470218.2015.1136656

Ritchie, K. L., Kramer, R. S., Mileva, M., Sandford, A., & Burton, A. M. (2021). Multiple-image arrays in face matching tasks with and without memory. *Cognition*, *211*, 104632. https://doi.org/10.1016/j.cognition.2021.104632

Robertson, D. J., Jenkins, R., & Burton, A. M. (2017). Face detection dissociates from face identification. *Visual Cognition*, *25*(7-8), 740-748. https://doi.org/10.1080/13506285.2017.1327465

Rosemblaum, M., Grosbard, I., Abudarham, N., & Yovel, G. (2021). What type of experience is needed to generate a human-like view-invariant representation of face identity? Evidence from Deep Convolutional Neural Networks. *Journal of Vision*, *21*(9), 2595-2595. https://doi.org/10.1167/jov.21.9.2595

Rossion, B. (2018). Humans are visual experts at unfamiliar face recognition. *Trends in cognitive sciences*, *22*(6), 471-472.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, *16*(2), 252-257. https://doi.org/10.1016/j.tics.2018.03.002

Sauerland, M., Sagana, A., Siegmann, K., Heiligers, D., Merckelbach, H., & Jenkins, R. (2016). These two are different. Yes, they're the same: Choice blindness for facial identity. *Consciousness and Cognition*, *40*, 93-104. https://doi.org/10.1016/j.concog.2016.01.003

Schwartz, L., & Yovel, G. (2016). The roles of perceptual and conceptual information in face recognition. Journal of Experimental Psychology: General, 145(11), 1493. https://doi.org/10.1037/xge0000220

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences*, *112*(41), 12887-12892. https://doi.org/10.1073/pnas.1421881112

Short, L. A., Balas, B., & Wilson, C. (2017). The effect of educational environment on identity recognition and perceptions of within-person variability. *Visual Cognition*, *25*(9-10), 940-948. https://doi.org/10.1080/13506285.2017.1360974

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 1904-1920, 3(3), 271-295. https://doi.org/10.1111/j.2044-8295.1910.tb00206.x

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive research: principles and implications*, *5*(1), 1-17. https://doi.org/10.1186/s41235-019-0205-0

Stantic, M., Brewer, R., Duchaine, B., Banissy, M. J., Bate, S., Susilo, T., Catmur, C., & Bird, G. (2021). The Oxford Face Matching Test: A non-biased test of the full range of individual differences in face perception. Behavior research methods, 1-16. https://doi.org/10.3758/s13428-021-01609-2

Stanislaw, H., Todorov, N. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers* **31**, 137–149 (1999). https://doi.org/10.3758/BF03207704

Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: neural and computational mechanisms. *Frontiers in neuroscience*, *6*, 70. https://doi.org/10.3389/fnins.2012.00070

Summerfield, C., & Egner, T. (2013). Attention and decision-making. *The Oxford handbook of attention.* DOI: 10.1093/oxfordhb/9780199675111.013.018

Sunday, M. A., & Gauthier, I. (2018). Face expertise for unfamiliar faces: A commentary on Young and Burton's "Are We Face Experts? *Journal of Expertise*, *1*, 35-41

Tanaka, J. W., & Corneille, O. (2007). Typicality effects in face and object perception: Further evidence for the attractor field model. *Perception & psychophysics*, *69*(4), 619-627. https://doi.org/10.3758/BF03193919

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*(2), 225-245. https://doi.org/10.1080/14640749308401045

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. https://doi.org/10.1371/journal.pone.0211037

Towler, A., Dunn, J. D., Martínez, S. C., Moreton, R., Eklöf, F., Ruifrok, A., Kemp, R. & White, D. (2021). Diverse routes to expertise in facial recognition. doi: 10.31234/osf.io/fmznh

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97-136. https://doi.org/10.1016/0010-0285(80)90005-5

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. https://doi.org/10.1080/14640749108400966

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, *69*(10), 1996-2019. https://doi.org/10.1080/17470218.2014.990392

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision research, 141*, 217-227. https://doi.org/10.1016/j.visres.2016.12.014

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. Nature Reviews Psychology, 1(5), 287-300. https://doi.org/10.1038/s44159-022-00041-3

White, D., Guilbert, D., Varela, V. P., Jenkins, R., & Burton, A. M. (2021). GFMT2: A psychometric measure of face matching ability. *Behavior Research Methods*, 1-9. https://doi.org/10.3758/s13428-021-01638-x

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS one, 9*(8), e103510. https://doi.org/10.1371/journal.pone.0103510

White, D., Towler, A., & Kemp, R. (2021). Understanding professional expertise in unfamiliar face matching. *Forensic Face Matching: Research and Practice*, 62.

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces—One element of social cognition. *Journal of personality and social psychology*, *99*(3), 530. https://doi.org/10.1037/a0019972

Wilmer, J. B. (2017). Individual differences in face recognition: A decade of discovery. *Current Directions in Psychological Science*, *26(3*), 225-230. https://doi.org/10.1177/0963721417710693

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., ... & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of sciences*, *107*(11), 5238-5241. https://doi.org/10.1073/pnas.0913053107

Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, *81*(1), 141. https://doi.org/10.1037/h0027474

Young, A. W., & Burton, A. M. (2018). Are we face experts?. *Trends in cognitive sciences*, *22*(2), 100-110. https://doi.org/10.1016/j.tics.2017.11.007

Young, A. W., & Burton, A. M. (2018). What we see in unfamiliar faces: a response to Rossion. *Trends in cognitive sciences*, 472-473. https://doi.org/10.1016/j.tics.2018.03.008

Young, A. W., & Burton, A. M. (2018). The limits of expertise in face recognition: Response to Mackenzie Sunday and Isabel Gauthier: Face expertise for unfamiliar faces: A commentary on Young and Burton's "Are we Face Experts?. *Journal of Expertise*, 130-135.

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. https://doi.org/10.1111/j.2044-8295.1985.tb01972.x

Yovel, G., Halsband, K., Pelleg, M., Farkash, N., Gal, B., & Goshen-Gottstein, Y. (2012). Can massive but passive exposure to faces contribute to face recognition abilities?. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 285. https://doi.org/10.1037/a0027077

Zhou, X., & Mondloch, C. J. (2016). Recognizing “Bella Swan” and “Hermione Granger”: No Own-Race Advantage in Recognizing Photos of Famous Faces. State and Local Government Review, 45(12), 66–71. https://doi.org/10.1177/0160323X11400211

Zhou, X., Mondloch, C. J., & Emrich, S. M. (2018). Encoding differences affect the number and precision of own-race versus other-race faces stored in visual working memory. *Attention, Perception, & Psychophysics*, *80*(3), 702-712. https://doi.org/10.3758/s13414-017-1467-6

Chapter 4: Unfamiliar face matching and face learning efficiency.[7]

Abstract

I have provided the first examination of individual differences in the efficiency of face learning. Investigating individual differences in face learning can illuminate potential mechanisms and provide greater understanding of why certain individuals might be more efficient face learners. Participants completed two unfamiliar face matching tasks and a learning task in which learning was assessed after viewing 1, 3, 6, and 9 images of to-be-learned identities. Individual differences in the slope of face learning (i.e., increases in sensitivity to identity) were predicted by the ability to discriminate between matched (same-identity) vs. mismatched (different-identity) pairs of wholly unfamiliar faces. A Dual Process Signal Detection model showed that three parameters increased with learning: Familiarity (an unconscious type of memory that varies in strength), Recollection-Old (conscious recognition of a learned identity), and Recollection-New (conscious/confident rejection of novel identities). Good (vs. poor) matchers had higher Recollection-Old scores throughout learning and showed a steeper increase in Recollection-New. We conclude that good matchers are better able to capitalize on exposure to within-person variability in appearance, an effect that is attributable to their conscious memory for both learned and novel faces. These results have applied implications and will inform contemporary and traditional models of face identification.

Introduction

Adults have a remarkable ability to recognize familiar faces despite within-person variability in appearance (e.g., changes in appearance resulting from lighting, hairstyle,

[7] This chapter is based on the submitted article: Baker, K.A. & Mondloch, C. J. (Under review). Unfamiliar face matching ability predicts the slope of face learning. Scientific Reports.

expression, viewpoint, health). Familiar faces are recognized even when image quality is poor (Burton et al., 1999; Bindemann et al., 2013), faces are far away (Noyes & Jenkins, 2017), disguised (Noyes & Jenkins, 2019), or are of a different race than the perceiver (Laurence et al. 2016). In contrast, matching identity of unfamiliar faces is error prone even when adults view high-quality photos of upright own-race faces taken just moments apart (Fysh & Bindemann, 2018; Megreya & Bindemann, 2013; Megreya & Burton,2008; Lorenc et al., 2014; Ritchie et al., 2021; Burton et al., 2010). Poor performance when viewing unfamiliar faces reflects the inherent challenge of face identification: Images of different faces can be very similar, and different images of the same face can vary in appearance.

The stark difference between recognition of unfamiliar vs. familiar faces, combined with the fact that every familiar face was once unfamiliar, has raised a critical question: How does a newly encountered face become familiar? Exposure to within-person variability in appearance is key (Andrews et al., 2017; Baker et al., 2017; Baker & Mondloch, 2019; Burton et al., 2016; Menon et al., 2015; Murphy et al., 2015; Zhou et al., 2022). Viewing multiple high-variability images of a target identity (i.e., ambient images taken on different days) improves performance in lineup tasks (Dowsett et al., 2016; Matthews & Mondloch, 2018), same/different tasks, name verification tasks (Ritchie & Burton, 2017), and old/new face identification tasks in which test stimuli comprise novel images of a learned identity (Baker et al., 2017; Murphy et al., 2015). Semantic associations also play a role. Recognition improves after social (*how trustworthy is this person?*) relative to perceptual (*how round is this face*) judgements—a difference associated with increased activity in social processing regions (e.g., dorsal medial prefrontal cortex; Shoham et al., 2022). Neural signatures of face learning have also been identified. The N250, an event related potential component that reflects recognition, is more negative for new images of learned identities than for images of wholly unfamiliar identities following an implicit learning paradigm (Andrews et al., 2017), and for personally familiar vs. celebrity faces (Wiese et al., 2019). The Sustained Familiarity Effect, an event related potential occurring at approximately 400-600 ms, tracks level of familiarity with an identity (Li et al., 2022).

In the current study I provide the first examination of individual differences in the efficiency with which a newly learned face becomes familiar (i.e., the slope of face learning). The individual differences approach in face perception can provide valuable insights about the perceptual and cognitive mechanisms underlying face identification but is a largely untapped resource (White & Burton, 2022). Individual differences in unfamiliar face identification are reliable (White & Burton, 2022; Baker et al., in press; Fysh et al., 2020) and are consistent across tasks that vary in the type of judgment being made (e.g., same/different vs. detecting a target in a lineup), the amount of within-person variability in appearance, and sequential versus simultaneous stimulus presentation (Baker et al., in press). To date, no study has examined whether individual differences in unfamiliar face matching predict improvement in performance as additional images are presented (i.e., the ability to capitalize on variability as a newly encountered face becomes familiar).

I predicted that individuals who are better at matching identity in wholly unfamiliar faces would have a steeper face learning slope based on several findings in the literature. First, group differences in unfamiliar face identification correspond to group differences in face learning. Adults from small towns perform worse when sorting images of unfamiliar faces than adults from large cities (Balas & Saville, 2015); children perform worse on unfamiliar face identification tasks than adults (Laurence & Mondloch, 2016); and adults perform worse when tested with inverted or other-race faces as compared to upright or own-race faces (Laurence et al., 2016; Valentine, 1988). These same group differences are seen in face learning and memory tasks. Adults from small towns perform more poorly on the Cambridge Face Memory Test than those from larger towns (Balas & Saville, 2017); children learn less efficiently than adults (Baker et al., 2017); and adults learn inverted or other-race faces less efficiently than upright or own-race faces (Kramer et al., 2017; Zhou et al., 2018). Likewise, individual differences on tasks measuring face memory (e.g., Models of Memory Test: Fysh et al., 2020, Cambridge Face Memory Test [CFMT]: Duchaine & Nakayama, 2006) correlate with individual differences on unfamiliar face matching tests (Fysh et al., 2020; McCaffery et al., 2018; Verhallen et al., 2017; Stacchi et al., 2020). Such tasks measure performance at only one point during face learning; I examined performance as learning unfolded.

The primary purpose of the current study was to test the hypothesis that individual differences in unfamiliar face matching predict individual differences in the slope of face learning, such that good matchers benefit more from viewing multiple images of a target identity than do poor matchers. The second purpose was to examine whether the contribution of two processes known to underlie memory (recollection and familiarity) differs between good vs. poor matchers. To address this question, I analyzed the data using the dual process model of signal detection (DPSD).

Dual Process Signal Detection

There is evidence that two distinct processes (recollection and familiarity) influence recall of an event, such that explicit memory details (recollection) and an implicit sense of familiarity inform one’s responses (Yonelinas, 2002; Yonelinas et al., 2010; Yonelinas, 1994; Yonelinas, 1999). The DPSD model proposes that recognition is influenced by three parameters: Recollection old (Ro), recollection new (Rn), and familiarity (Fam). The Ro parameter is a threshold process reflecting high-confidence hits. It is an episodic process in which high-quality information about the recollected stimulus is retrieved (e.g., “She played Black Widow in the Avengers movie”). The Rn parameter is a threshold process reflecting high-confidence correct rejections, sometimes considered to be more perceptual in nature (Aly & Yonelinas, 2012). The Fam parameter is a graded signal process that occurs at all levels of confidence, but in the absence of conscious recollection (e.g., “I think I have seen her before but can’t remember where”; Yonelinas, 2002; Yonelinas et al., 2010; Yonelinas, 1994; Yonelinas, 1999).

There are hints in the literature that suggest that the extent to which the Ro, Rn and/or F parameters change during the process of face learning varies. First, poor performance on the CFMT is associated with more inaccurate “familiar” responses and less recall of accurate semantic detail about a target (akin to recollection) when participants are asked to recognize actors from a familiar TV series (Devue et al., 2019). Second, groups that perform poorly on face recognition tasks (e.g., older adults, adults tested with other-race faces) make fewer recollection-based responses than better performing groups (i.e., young adults, adults tested with own-race faces; Bartlett et al., 1989; Bartlett & Fulton, 1991; Edmonds et al., 2012; Matthews & Mondloch 2021; Semplonius & Mondloch, 2015; Koen & Yonelinas 2016). These results suggest that individual differences in the slope of face learning and/or face matching might be associated with individual differences in recollection and/or familiarity. As binary responses (*same/different*) are not ideal for distinguishing responses based on recollection vs. familiarity, I asked participants in the face learning task to report their confidence on a scale from 1 (certain that the image belongs to the target) to 6 (certain that the image does not belong to the target).

The current study

Participants completed a battery of three face identification tasks. Two tasks (The Glasgow Face Matching Test [GFMT]: Burton et al., 2010; The Ambient Image Face Matching Task [AIFMT]: Baker et al., in press; Baker & Mondloch, 2020) measured face matching in wholly unfamiliar faces. I developed a novel face learning task to measure the slope of face learning as participants were exposed to an increasing number of new images (1, 3, 6, and then 9 images) of a previously unfamiliar face. See figure 4-1 for a depiction of the tasks and the Methods section below for further details.

My primary goal was to provide the first examination of the extent to which individual differences in face-matching abilities (using average d′) predict the slope of face learning (using d′ across four phases [1, 3, 6, and 9 images] in the learning task). I examined whether face matching ability (e.g., poor, mid-level, and good) interacted with learning phase such that good matchers benefitted more from the additional variability provided at each step of learning. I predicted that there would be significant interactions, such that poor matchers would benefit less from viewing more images than good matchers at each level of learning.

To examine how one’s representation changes during face learning, I estimated three parameters in the DPSD (Fam, Ro, and Rn) using the frequency of confidence ratings in each phase of the face learning task. Doing so allowed us to examine the extent to which exposure to multiple images of the target identity leads to changes in Ro (recollection of the target), changes in Rn (certainty that a face was not previously seen), and/or changes in familiarity. I predicted that good vs. poor matchers would differ in how these processes change during learning.

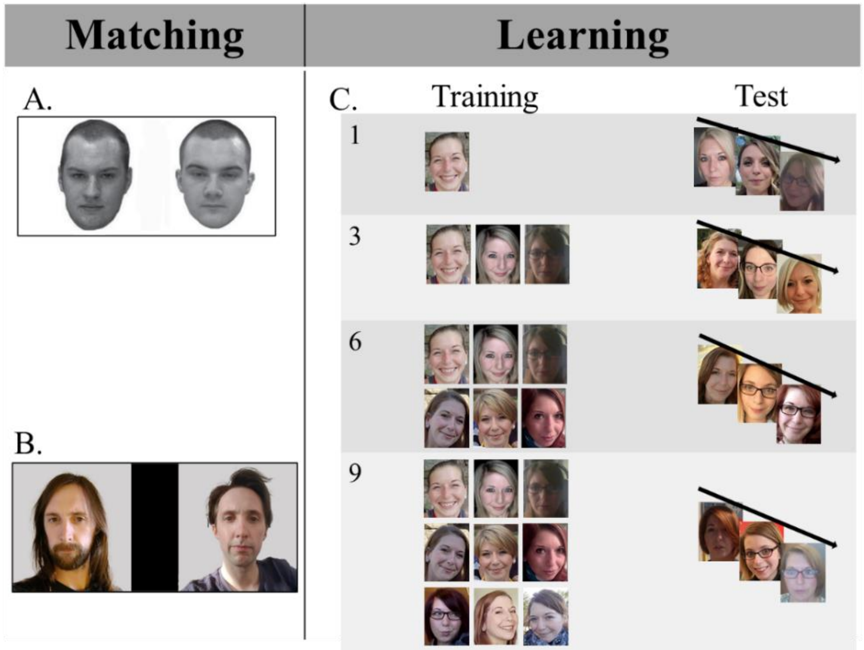

Figure 4-1 depicts the GFMT (A), AIFMT (B), and learning task (C). Images in 4-1B and 4-1C are for illustrative purposes and are not photos of individuals used in the experiment. I obtained informed consent for open-access publication from these models. Figure 4-1A was used in a prior publication (Burton et al., 2010); permission to use this figure has been requested. Participants first completed the GFMT (n_{trials} = 40; 50% match). They were instructed to press "f" if both images showed the same person and "j" if the images showed different people. Next, participants completed the AIFMT (n_{trials} = 80; 50% match); participants indicated their responses on a scale from 1 (certain that the images belong to the same person) to 6 (certain that the images belong to different individuals). Selecting 1, 2 or 3 was considered a match response and selecting 4, 5 or 6 was considered a mismatch response. Finally, participants completed the face learning task. Participants viewed 1, 3, 6, and then 9 images of a target. Their ability to recognize novel images of the learned identity, when intermixed with images of a similar distractor, was assessed after each of the learning phases. Selecting 1, 2 or 3 was considered an old response and selecting 4, 5 or 6 was considered a new response. Participants completed this four times, once for each of four identities (n_{test} trials = 40).

Data Analysis and Results

Data analysis

All statistical analyses were performed using SPSS. Bonferroni corrections were used to control for multiple comparisons in both post-hoc and simple effect analyses. For all tasks signal detection theory was used to measure sensitivity (d′). I arbitrarily defined a hit as responding *same* on match trials (GFMT, AIFMT) or *old* when the target identity was presented in the learning task. Hit rates of 1 were replaced with ($n_{\text{Signal or Noise trials}}$ -0.5)/n and false alarm rates of 0 were replaced with 0.5/$n_{\text{Signal or Noise trials}}$ (Macmillan & Kaplan, 1985).

Face learning slope. I analyzed whether matching ability predicted face learning in three ways. First, I calculated mean d′ across the two unfamiliar face matching tasks (weighted equally). I used a tertial split on the mean d′ to create three groups that differed in sensitivity (poor matchers, mid-level matchers and good matchers). I then conducted a 3 (matching ability: poor, mid-level, good) x 4 (learning phase: 1, 3, 6, and 9 images) mixed factorial Analysis of Variance (ANOVA) with d′ on the learning task as the dependent variable. A significant interaction would provide evidence that the slope of learning varies with matching ability.

Second, I further probed these findings to investigate whether the slope from one learning phase to another was predicted by average matching d′. To obtain an estimate of the slope, I created residual scores using a regression in which performance in each learning phase was predicted by the preceding phase (e.g., using performance on the 1-image phase to predict performance on the 3-image phase). Measures of slope were obtained for the 3-, 6-, and 9-image phases. These measures of slope were correlated with the continuous measure of average matching ability.

Third, I confirmed my findings by calculating a difference score (9 image d′ – 1 image d′) for each participant—an analysis that focuses on the start and end point of face learning, consistent with previous research. These results were analyzed using a one-way ANOVA.

Dual Process Signal Detection. To examine the extent to which improvements during face learning reflected changes in familiarity or recollection (old, new), receiver operating characteristics for the learning task were conducted using the DPSD model. Parameters (Fam, Ro, Rn) were estimated using the frequency that participants reported each level of confidence across trial types (target, distractor) and image learning phases using the default settings of the ROC Toolbox in Matlab (Koen et al., 2017).

Investigations of scatterplots revealed several outliers in the Fam parameter, therefore we used an iterative outlier process (i.e., all participants who scored ± 3 SD away from the mean in any condition were removed until there were no more outliers within the data). For consistency we also used this same process for the Ro and Rn parameters. Thus, Fam analyses are conducted with n = 94 participants, Ro analyses are conducted with n = 129 participants, and Rn analyses are conducted with n = 149 participants. Less strict outlier removal strategies yield the same results.

For each parameter I conducted a 3 (matching group; low, mid-level, high) x 4 (learning phase; 1, 3, 6, and 9 images) mixed ANOVA. Significant interactions were followed up with analyses of simple effects in which I analyzed the effect of learning phase at each level of matching ability.

Results

Face learning slope. Matching ability influenced learning across the four phases. The main effects of learning phase ($p < 0.001$, $\eta_p^2 = 0.69$) and matching ability ($p < 0.001$, $\eta_p^2 = 0.31$)

were qualified by a significant interaction, $F(3,459) = 2.57$, $p = 0.02$, $\eta_p^2 = 0.03$. Simple effects revealed that d′ was higher in the 3-image phase than the 1-image phase for all groups (see Figure 4-2A for a depiction of the means, *ps* < 0.001), but was not higher in the 6-image phase than in the 3-image phase for any groups, (*ps* > 0.10). d′ was higher in the 9-image phase than the 3-image phases for all groups (*p*s < 0.002). However, d′ was higher in the 9- than the 6-image phases for mid-level and good matchers (*p*s < 0.02), but not for poor matchers ($p = 0.28$).

Although the previous analyses tell us *whether* the groups learned at each phase, they do not tell us whether matching ability predicted the amount of learning at each phase. This effect is observed using residual scores (See Figure 4-2 B-D). Matching ability predicted the slope for the 3- ($r = 0.32$, $p < 0.001$) and 9-image phases ($r = 0.29$, $p < 0.001$), such that the benefit of viewing additional images was positively correlated with matching ability. This effect was not observed for the 6-image phase ($r = 0.14$, $p = 0.07$).

The overall slope of learning (from 1 to 9 images) varied between the groups. The ANOVA revealed a main effect of group, $F(2, 153) = 5.66$, $p = 0.004$, $\eta_p^2 = 0.07$. Post-hoc analyses revealed that poor matchers showed less improvement ($m = 1.23$, $SD = 0.62$) than both mid-level ($m = 1.55$, $SD = 1.62$) and good matchers ($m = 1.62$, $SD = 0.65$), *p*s < 0.04. Mid-level and good matchers did not differ, $p > 0.99$. In short, these findings suggest that poor matchers benefit less from exposure to variability in appearance than do mid-level and good matchers.

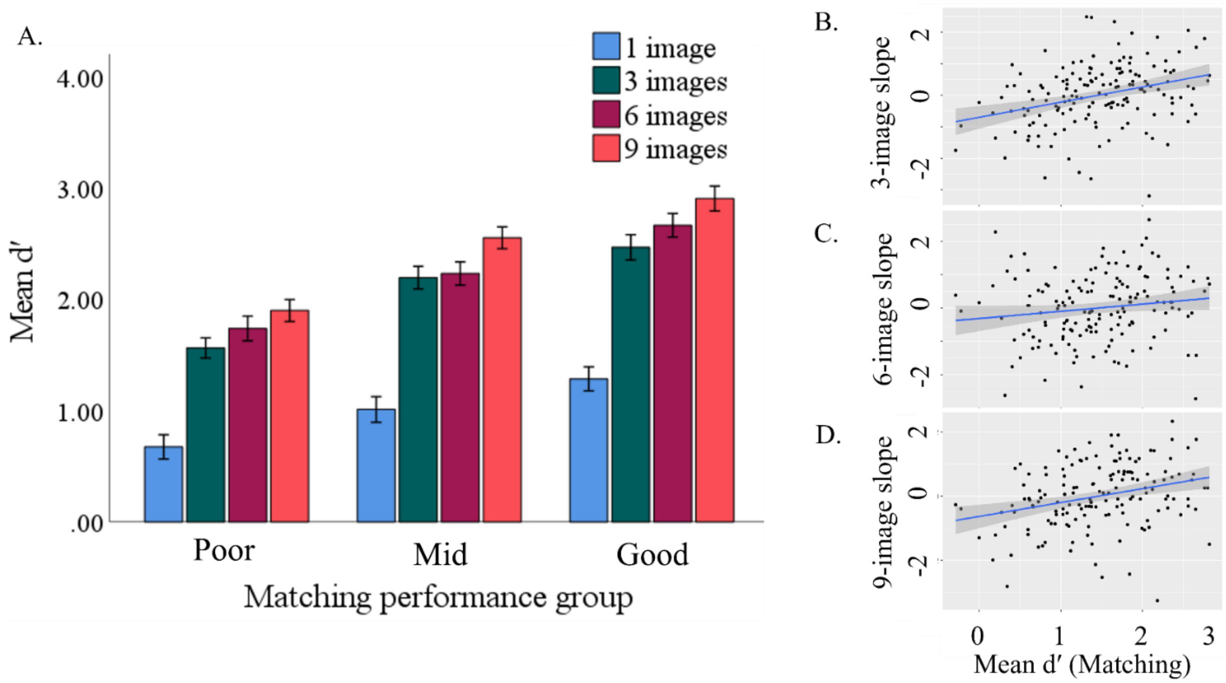

Figure 4-2. Figure 4-2A depicts mean d′ in each phase of the learning condition as a function of unfamiliar face matching ability. Error bars corrected for within-subject error. Figures 4-2 B-D reflect the correlations between average matching ability (d′) and the slope for the 3-, 6- and 9-image phases, respectively. Error bars all reflect 95% CIs.

Dual Process Signal Detection. *Ro.* The main effect of learning phase for the Ro parameter was significant, $F(2.51, 328.83) = 61.25$, $p < 0.001$, $\eta_p^2 = 0.32$. Post-hoc analyses revealed that Ro was smaller in the 1-image phase ($m = 0.04$, $SD = 0.06$) than all other phases, *p*s<0.001. Ro in the 9-image phase ($m = 0.31$, $SD = 0.25$) was significantly larger than in the 3- ($m = 0.18$, $SD = 0.19$) and 6-image phases ($m = 0.21$, $SD = 0.22$), *p*s < 0.001. No other comparisons differed, *p*s>0.59. The main effect of matching ability was also significant, $F(2, 131) = 5.73$, $p = 0.004$, $\eta_p^2 = 0.08$. Post-hoc analyses revealed Ro was smaller in poor matchers ($m = 0.14$, $SD = 0.11$) than good matchers ($m = 0.20$, $SD = 0.15$), $p = 0.001$. Mid-level matchers

(m = 0.23, SD = 0.13) did not differ from either group, $ps > 0.09$. These effects were not qualified by a significant interaction, $F(5.02, 328.83) = 2.15$, $p = 0.06$, $\eta_p^2 = 0.03$, suggesting that changes in high confidence hits did not vary with matching ability. See Figure 3.

Rn. The main effects of learning phase, $F(3, 453) = 24.07$, $p < 0.001$, $\eta_p^2 = 0.14$, and matching ability, $F(2, 153) = 11.41$, $p < 0.001$, $\eta_p^2 = 0.13$, were qualified by a significant interaction, $F(6, 459) = 2.55$, $p = 0.02$, $\eta_p^2 = 0.03$. Simple effects analyses revealed that poor matchers showed no improvement in the Rn parameter across the 1- (m = 0.14, SD = 0.19), 3- (m = 0.17, SD = 0.17), 6- (m = 0.21, SD = 0.23), and 9- image phases (m = 0.26, SD = 0.25), $ps >$ 0.11. Mid-level matchers only showed a significant improvement in Rn in the 9-image phase; their Rn was significantly larger in the 9-image phase (m = 0.43; SD = 0.31) than in the 1- (m = 0.21; SD = 0.23), 3- (m = 0.22; SD = 0.26) and 6-image phases (m = 0.28; SD = 0.31), $ps < 0.03$. No other differences were significant. Good matchers showed immediate improvement in Rn; their Rn was significantly smaller in the 1-image phase (m = 0.18; SD = 0.24) than all other phases, $ps < 0.001$. Rn in the 9-image phase (m = 0.51; SD = 0.34) was significantly larger than in the 3-image phase (m = 0.37; SD = 0.33), $p < 0.001$. Rn in the 6-image phase (m = 0.38; SD = 0.34) did not differ from either the 3- or 9-image phases, $ps > 0.07$. Thus, training improved the ability of mid-level and good matchers, but not poor matchers, to perceive a distractor as being new. Mid-level matchers required more exposure to variability than the good matchers to do so.

Fam. The main effect of learning phase [$F(2.68, 257.49) = 77.44$, $p < 0.001$, $\eta_p^2 = 0.44$] was significant. Post-hoc analyses revealed that Fam was smaller in the 1-image phase (m = 0.48, SD = 0.54) than the 3- (m = 1.45, SD = 0.73) 6- (m = 1.53, SD = 0.80) and 9- (m = 1.57, SD = 0.80) image phase, $ps < 0.001$. No other conditions differed, $ps > 0.22$. Matching ability [$F(2,96)$ =

2.20, $p = 0.12$, $\eta_p^2 = 0.04$] and the learning phase by matching ability interaction [$F(5.36, 257.49) = 0.62$, $p = 0.70$, $\eta_p^2 = 0.01$] failed to reach significance.

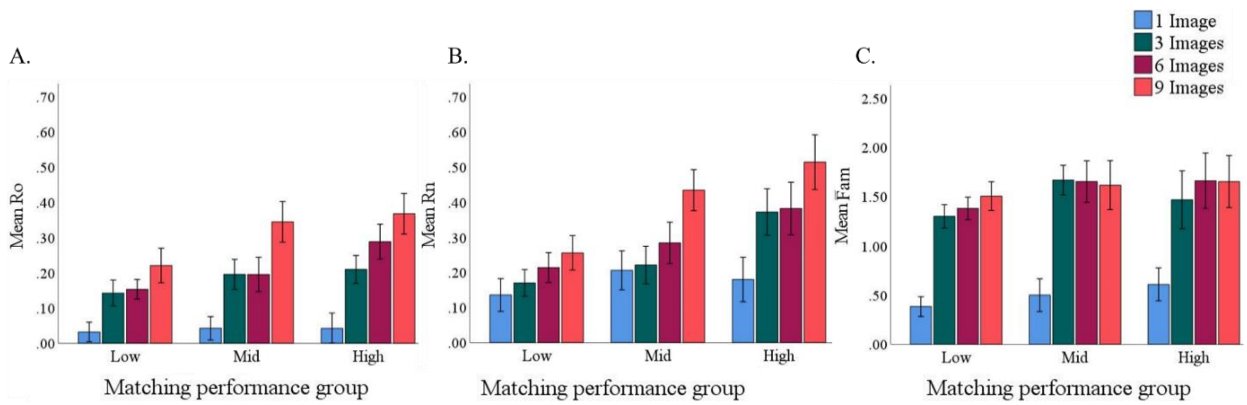

Figure 4-3. Figure 4-3 A-C depicts mean Ro, Rn and Fam in each phase of the learning condition as a function of unfamiliar face matching ability, respectively. Error bars corrected for within-subject error. Error bars reflect 95% CIs.

In short, providing perceivers with additional images led to changes in recollection, a pattern that varied as a function of matching ability. Ro varied with matching ability (good > poor) but, increased comparably for all groups. Rn also increased with the number of presented images, but only for mid-level and good matchers, with mid-level matchers requiring more variability (9 images) than good matchers (3 images) to show a benefit. Although Fam increased with the number of presented images, it did not vary across groups.

Discussion

Little is known about the process by which a newly encountered face becomes familiar. Individual differences have tremendous potential for illuminating our understanding of face learning, but have been largely neglected to date (White & Burton, 2022). I capitalized on an individual difference approach to address two key questions. 1) Do individual differences in face-matching ability predict the slope of face learning? 2) Are individual differences in learning best captured by recollection or familiarity parameters?

Do individual differences in matching ability predict the slope of face learning?

Despite evidence suggesting that the slope of face learning might be related to unfamiliar face matching abilities (Baker et al., 2017; Balas et al., 2015, 2017; Fysh et al., 2020; Kramer et al., 2017; McCaffery et al., 2018; Valentine, 1988; Verhallen et al., 2017), no studies had directly examined this question. Here I provide three pieces of evidence that strongly suggest that matching ability is related to the slope of face learning. First, whereas all groups were more sensitive to identity after viewing three images vs. one image, only mid-level and good matchers were more sensitive to identity after viewing nine vs. six images (i.e., refined their representation with these additional images). Second, all perceivers did not benefit equally from exposure to multiple images. Individuals with the best matching abilities showed the steepest slopes (calculated using residuals) in the 3- and 9-image phases. Third, poor matchers showed less improvement after viewing nine vs. one image than did mid-level and good matchers. Collectively, these findings suggest that good matchers are better able to capitalize on exposure to within-person variability in appearance when learning a newly encountered face.

My findings inform our understanding of face identification and corresponding models and suggest that good vs. poor matchers are interacting with faces differently. One possibility is that good and poor matchers are attending to different information when identifying faces and/or weighting information differently while making their decision—possibilities that were explored in a recent chapter (Bindemann & Burton, 2021). For example, facial moles can be diagnostic of identity and when individuals are instructed to use moles for unfamiliar face identification, their performance improves (Fysh & Bindemann, 2022). Forensic examiners, who show greater accuracy than novices, use a featural approach to examine unfamiliar faces. When novices are trained to use this approach, their performance improves (Towler et al., 2017). These studies suggest that the information to which one attends contributes to unfamiliar face matching—a pattern that my data suggest may extend to face learning. One published study runs counter to this explanation. That study reported that developmental prosopagnosics and super recognizers are both more sensitive to critical features (features that lead to the perception of a different identity when altered) than non-critical features (Abudarham et al., 2021). However, the stimuli used in that study were tightly controlled images in which one or more features were systematically altered, not ambient images where appearance varies naturally. Future studies could examine whether individual differences in face identification reflect differences in attending to and/or weighting cues using faces that capture idiosyncratic variability in appearance. Future studies could also examine whether attending to these cues improves the slope of face learning.

The present data are consistent with evidence from Deep Convolutional Neural Networks (DCNNs)—state-of-the-art algorithms informed by the primate visual system (O'Toole et al., 2018). A recent study varied the number of identities on which a DCNN was trained and the number of images of a to-be-tested identity on which a DCNN was trained (Rosemblaum et al., 2021). DCNNs that were trained on fewer identities (i.e., had a sparse representation of identity) required more exposure to variability (i.e., more images) to learn new identities than DCNNs that were trained on many different identities. This is consistent with the present proposal that a perceiver's matching ability (a proxy for their representation of unfamiliar faces) constrains the efficiency of face learning. The present findings also relate to classic models of face perception (Bruce & Young, 1986). Whereas representations of unfamiliar faces are thought to be pictorial and therefore intolerant to variability, representations of familiar faces are robust and tolerant to variability. Poor matchers might struggle to transition from a pictorial to a robust representation for newly learned faces (e.g., they might include pictorial cues within their representation of newly learned identities). The links between individual differences in unfamiliar face matching and the slope of face learning provide a promising avenue for understanding the relationship between, and transition from, unfamiliar to familiar face recognition (White & Burton, 2022; Lander et al., 2018; Bindemann & Hole, 2020; Bindemann & Johnston, 2017).

Are individual differences in learning best captured by the recollection parameters?

Two processes (recollection and familiarity) influence the recall of an event, such that explicit memory details (recollection) and an implicit sense of familiarity inform one's responses (Yonelinas, 2002; Yonelinas et al., 2010; Yonelinas, 1994; Yonelinas, 1999). No studies had investigated how these parameters change during the process of face learning following exposure to variability in appearance. Using the DPSD model, found no evidence that Fam varies across groups that differ in matching ability or that individual differences in Fam during learning covary with matching ability. This is consistent with evidence that there are no age-related changes in Fam despite young adults showing more sensitivity than older adults when tested with newly learned faces (Bartlett et al., 1989; Bartlett & Fulton, 1991; Matthews & Mondloch 2021; Koen & Yonelinas, 2016). These results suggest that, although familiarity increases with learning, individual differences reflect recollection.

Ro and Rn did change with learning and the pattern of change varied with matching ability. Ro increased as participants learned a newly encountered face, suggesting that recognition was changing in a precise, episodic-like manner for all participants. Ro was higher in good matchers than poor matchers overall, suggesting that better matchers tend to recognize individuals in an episodic manner more than poor matchers. The increase in Ro across learning phase did not vary across groups, suggesting that exposure to variability in appearance influenced this process similarly across all levels of matching ability. In contrast, Rn increased during learning but only in good and mid-level matchers—the same groups that benefitted from 9 vs. 6 images and who showed more overall improvement during learning. These findings are consistent with recent findings for individual differences in episodic recall and familiarity responses for familiar faces. Individuals who had better accuracy in the CFMT provided more responses containing accurate semantic information of the target identities (comparable to Ro), made fewer false alarms and were less likely to rely on feelings of familiarity than poor matchers (i.e., relied on both Ro and Rn; Devue et al., 2019). These results show a role for recollection during face learning. This suggests that efficient learning reflects both recognition of an identity across variable instances and discrimination of the newly learned face from similar-looking distractors.

My examination of how face representations change during the process of learning adds to the literature investigating face learning (Andrews et al., 2017; Baker et al., 2017; Burton et al., 2016; Menon et al., 2015; Murphy et al., 2015; Dowsett et al., 2016; Matthews & Mondloch, 2018; Ritchie & Burton, 2017). My findings suggest that one’s representation of newly learned identities changes in a categorical, episodic way. This corresponds with previous findings where individuals best recognize actors from instances that they had previously seen (Carbon, 2008) and rate previously seen photos as having a better likeness (Ritchie et al., 2018). It also relates to findings that exposure to idiosyncratic variability facilitates face learning (Burton et al., 2016; Dowsett et al., 2016; Matthews & Mondloch, 2018). As exposure to variability in appearance facilitates face learning, and episodic memories are specific to the individual events (or faces) that created them, it makes sense that representations are being changed in an episodic manner (Ro) allowing confident rejections of similar looking identities (Rn).

These findings for Ro, Rn and Fam provide insight into group differences. My results suggest that individuals who are poor at unfamiliar face matching also show fewer high confidence recollection responses. This then leads to the hypothesis that poor matchers rely less on recollection than good matchers. An extension of this hypothesis would be a prediction that groups that perform worse on unfamiliar face matching tasks might also rely less on recollection when learning a new face than do groups that perform well. For example, older adults might rely less on recollection than young adults during their recall of faces (Koen & Yonelinas, 2016) because of their relatively poor face matching abilities (Matthews & Mondloch, 2021). Likewise, individuals might rely less on recollection when recognizing newly learned other- vs. own-race faces (Semplonius & Mondloch, 2015), because of poor face matching abilities (Laurence et al., 2016; Zhou et al. 2022). To the extent that a poor ability to match unfamiliar faces is attributable to a less well refined representation of an identity, my data suggest that poor matching ability might impair the ability to build precise, episodic representations—a hypothesis that could be tested with computer models.

On the surface, my data contrast with a recent examination of the effect of face familiarity (i.e., familiar vs. unfamiliar faces) with an identity on perceivers' similarity judgements of within- vs. between-identity face pairs (White et al., 2022)—at least when tested with highly familiar faces. Familiar faces showed increased similarity ratings of within-identity pairs more than it decreased similarity ratings of between-identity pairs—an effect that was most pronounced for super recognizers. Such findings suggest that Ro might play a larger role than Rn during face learning and better differentiate good vs. poor performers. My data suggest a key role for Rn. The reported increased similarity responses might be partially attributable to all images of a highly familiar face activating the same representation. In classic models of familiar face recognition, Person Identity Nodes (PINs) are theorized to be the backbone of person recognition (Bruce & Young, 1986). PINs comprise identity-specific information, regardless of modality (e.g., an individual's voice, their face). A person is recognized when the information reaches a threshold. Increased similarity ratings for within-identity pairs might reflect that both images activated the same PIN, emphasizing the role of telling faces together. Future studies could investigate this hypothesis by integrating newly learned faces into White and colleagues' design and examining the overlap between similarity ratings and confidence judgements.

Implications for Applied Settings.

These findings are significant for applied work. There are many applied contexts in which face learning is relied upon, such as missing person searches, eyewitness testimony, and when using CCTV to monitor individuals for illegal activities. My data suggest that the duration of exposure and the amount of variability to which an eyewitness is exposed will have differential benefits for good vs. poor matchers. The processes underlying face identification in applied settings will vary as a function of unfamiliar face matching abilities. Although good and poor matchers might show greater confidence in recognizing the perpetrator of a crime after exposure to variability, only good matchers are likely to show greater confidence for rejecting a similar looking distractor.

These results also speak to discussions on face training protocols. I showed that Rn (high confidence correct rejections) can improve with exposure to variability—but is dependent on matching ability. Group-level analyses have shown that training improved either performance on match or mismatch trials, but rarely both (Ritchie et al., 2021; Ritchie & Burton, 2017; Alenezi

& Bindemann, 2013). A potential explanation is that only those with high matching abilities show an immediate benefit of variability in Rn and those with poor matching abilities do not appear to get any benefit at all. Using group level statistics and averaging across individual differences may reduce our ability to show an effect of training on d′, at least in good matchers.

Conclusion

Models of face identification (Bruce & Young, 1986; Valentine, 1991; O’Toole et al., 2018) have theorized on the mechanisms underlying face learning and identification. Investigating individual differences can help to further face identification theories (White & Burton, 2022). Using an individual difference approach, I showed that 1) individual differences in matching ability predict the slope of face learning. 2) Individual differences in face learning are best captured by recollection. These findings can help to inform models of face identification. As we understand little about the transition from, and relationship between unfamiliar and familiar face recognition (White & Burton, 2022; Lander et al., 2018; Bindemann & Hole, 2020; Bindemann & Johnston, 2017), the approach used here provides a fruitful avenue for future research.

Methods

Participants

One-hundred fifty-six Caucasian participants (women: n = 133; Age: M = 21.62, SD = 3.79) completed a battery of three face identification tasks. This sample size provides enough power to detect moderate effects. Three additional participants were excluded because of failing attention checks (see below). Just over half (n = 84) of the participants were Brock University students who received research credit for their participation. The remaining participants were recruited online via Prolific (www.prolific.co) and were paid £10 for their participation. Each participant gave informed consent. All participants completed the tasks in a fixed order—a key component of individual differences designs (Dale & Arnell, 2015; Goodhew & Edwards, 2019).

Stimuli and tasks

Images in the GFMT (half male) were greyscale (1000 x 700 pix), full-faced and shown on a white background (Burton et al., 2010). Images within match trials were taken using different cameras. Images in the AIFMT (half male) were colour photographs (652 x 260 pix) shown on a grey background (Baker et al., in press; Baker & Mondloch, 2022). Images were taken from a variety of sources and included natural variability in appearance. All images in the AIFMT had a roughly frontal view of each model's face. Images for the learning task (half male) were colour photographs (125 x167 pix) and were presented with the original background in each image. Learned identities were individuals with whom Canadian participants were expected to be unfamiliar (Fern Sutherland, Philipp Boy, Gigi Ravelli and Donald Stamper); distractor identities for each target were chosen based on physical similarity. Images for the learning task were gathered from Brock University's Let's face it database, a google search and social media (e.g., Instagram, Twitter). The photographs of each model depicted a roughly frontal view of the face. I presented 33 images of each to-be-learned identity (9 training, 20 test, 4 attention checks) and 20 images of each distractor. All images were of White faces.

Procedure

All procedures received clearance from the Social Sciences Research Ethics Board of Brock University. I carried out the procedures in accordance with the guidelines specified by the Canadian tri-council and the ethics review board at Brock University.

Participants were tested online using Testable (www.testable.org). After providing informed consent, all participants completed three tasks. The order of the tasks and the order in which stimuli were presented within the learning phase of the learning task were both fixed. Doing so ensures that individual differences are not confounded with variance that arises from any effect of task or stimulus order on performance (Dale & Arnell, 2015; Goodhew & Edwards, 2019).

GFMT. Participants completed all 40 trials (50% match) of the shortened version of the GFMT[11]. Participants were shown face pairs and asked to indicate whether each pair showed the same person (by pressing “f”) or different people (by pressing “j”). Images remained on screen until participants made a response.

AIFMT. Participants completed 80 trials (50% match) of the AIFMT (Baker et al., in press; Baker & Mondloch, 2022). Participants were presented with pairs of images and were asked to determine whether the image pair was of the same person or different people. For the purposes of another study, participants indicated their responses on a scale from 1 (certain that the images belong to the same person) to 6 (certain that the images belong to different individuals). In this manuscript selecting 1, 2 or 3 was considered a match response and selecting 4, 5 or 6 was considered a mismatch response. One participant recognized one identity in the AIFMT. This trial was removed from the analyses.

Face learning task. Participants learned each of four identities, one at a time. Participants’ ability to recognize a learned identity was assessed after viewing 1, 3, 6, and then 9 images; new images were presented during each test phase. Each assessment included five novel images of the to-be-learned identity, five images of a similar looking distractor, and one attention check (see below). Collapsing across all four identities, participants completed 40 test trials (50% target) at each level of training (1, 3, 6, and 9 images). All images in each learning set were presented simultaneously for 20 s and were removed from the screen when the test stimuli were presented. Each test stimulus remained on screen until participants pressed a response key. Participants indicated whether each test image belonged to the target identity using a scale from 1 (certain that the images belong to the target) to 6 (certain that the images do not belong to the target). To create binary responses, I treated 1, 2, or 3 as match responses and 4, 5, or 6 as mismatch responses. The order in which to-be-learned identities and individual learning images were presented was fixed.

Attention checks. To ensure that participants were attentive during the experiment I included 18 total attention checks. The attention checks in the AIFMT (n = 2) comprised match and mismatches. Match attention checks were easily solved because the same image of the target was presented; mismatch attention checks were easily solved because distractors differed from the target identity in age and sex. The attention checks in the learning task (n = 16; 4 per identity, with 1 per phase) comprised the first image of each identity on which participants were trained. Participants who failed >75% of the attention checks were excluded from all analyses (n = 3).

Data availability

The datasets collected and analysed during the current study are available in the OSF repository, https://osf.io/3crmy/?view_only=056a559ae5f54089821731731c5a7c36.

References

Abudarham, N., Bate, S., Duchaine, B., & Yovel, G. (2021). Developmental prosopagnosics and super recognizers rely on the same facial features used by individuals with normal face recognition abilities for face identification. *Neuropsychologia, 160*, 107963. https://doi.org/10.1016/j.neuropsychologia.2021.107963

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*(6), 735-753. https://doi.org/10.1002/acp.2968

Aly, M., & Yonelinas, A. P. (2012). Bridging consciousness and cognition in memory and perception: evidence for both state and strength processes. *PLoS one*, *7*(1), e30231. https://doi.org/10.1371/journal.pone.0030231

Andrews, S., Burton, A. M., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. Quarterly Journal of Experimental Psychology, 70(8), 1620-1632. https://doi.org/10.1080/17470218.2016.1195851

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. Cognition, 161, 19-30. https://doi.org/10.1016/j.cognition.2016.12.012

Baker, K. A., & Mondloch, C. J. (2019). Two Sides of Face Learning: Improving Between-Identity Discrimination While Tolerating More Within-Person Variability in Appearance. Perception, 48(11), 1124-1145. https://doi.org/10.1111/bjop.12560

Baker, K. A., & Mondloch, C. J. (2022). Picture this: Photographers no better than controls for recognizing unfamiliar faces. Perception. https://doi.org/10.1177/03010066221098727

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

Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. Neuropsychologia, 69, 211-217. https://doi.org/10.1016/j.neuropsychologia.2015.02.005

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision Research an International Journal in Visual Science*, 141, 228–236. https://doi.org/10.1016/j.visres.2016.12.005

Bartlett, J. C., & Fulton, A. (1991). Familiarity and recognition of faces in old age. *Memory & Cognition*, *19*(3), 229-238. https://doi.org/10.3758/BF03211147

Bartlett, J. C., Leslie, J. E., Tubbs, A., & Fulton, A. (1989). Aging and memory for pictures of faces. *Psychology and aging*, *4*(3), 276. https://doi.org/10.1037/0882-7974.4.3.276

Bindemann, M., Attard, J., Leach, A. M. Y., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar-face matching. *Applied Cognitive Psychology*, *27*(6), 707-717. https://doi.org/10.1002/acp.2970

Bindemann, M., & Burton, M. (2021). Steps towards a cognitive theory of unfamiliar face matching. *Forensic face matching: Research and practice*, 38-61.

Bindemann, M., & Hole, G. J. (2020). Understanding face identification through within-person variability in appearance: Introduction to a virtual special issue. *Quarterly Journal of Experimental Psychology*, *73*(12), NP1-NP8. https://doi.org/10.1177/1747021820959068

Bindemann, M., & Johnston, R. A. (2017). Understanding how unfamiliar faces become familiar: Introduction to a special issue on face learning. *Quarterly Journal of Experimental Psychology*, *70*(5), 859-862. https://doi.org/10.1080/17470218.2016.1267235

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, *77*(3), 305-327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. Cognitive Science, 40(1), 202-223. https://doi.org/10.1111/cogs.12231

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior research methods*, *42*(1), 286-291. https://doi.org/10.3758/BRM.42.1.286

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science,* 10(3), 243-248. https://doi.org/10.1111/1467-9280.00144

Carbon, C. C. (2008). Famous faces as icons. The illusion of being an expert in the recognition of famous faces. *Perception, 37*(5), 801-806. https://doi.org/10.1068/p5789

Dale, G., & Arnell, K. M. (2015). Multiple measures of dispositional global/local bias predict attentional blink magnitude. *Psychological Research*, *79*(4), 534-547. https://doi.org/10.1007/s00426-014-0591-3

Devue, C., Wride, A., & Grimshaw, G. M. (2019). New insights on real-world human face recognition. *Journal of Experimental Psychology: General*, *148*(6), 994. https://doi.org/10.1037/xge0000493

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. https://doi.org/10.1080/17470218.2015.1017513

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

Edmonds, E. C., Glisky, E. L., Bartlett, J. C., & Rapcsak, S. Z. (2012). Cognitive mechanisms of false facial recognition in older adults. *Psychology and Aging*, *27*(1), 54. https://doi.org/10.1037/a0024582

Fysh, M. C., & Bindemann, M. (2022). Molistic processing in facial image comparison. *Applied Cognitive Psychology*, *36*(4), 830-841. https://doi.org/10.1002/acp.3975

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. British journal of psychology, 109(2), 219-231. DOI:10.1111/bjop.12260

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society open science*, 7(9), 200233. https://doi.org/10.1098/rsos.200233

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, 69, 14-25. https://doi.org/10.1016/j.concog.2019.01.008

Koen, J. D., Barrett, F. S., Harlow, I. M., & Yonelinas, A. P. (2017). The ROC Toolbox: A toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behavior research methods*, *49*(4), 1399-1406. https://doi.org/10.3758/s13428-016-0796-z

Koen, J. D., & Yonelinas, A. P. (2016). Recollection, not familiarity, decreases in healthy ageing: Converging evidence from four estimation methods. *Memory*, *24*(1), 75-88. https://doi.org/10.1080/09658211.2014.985590

Kramer, R. S., Jenkins, R., Young, A. W., & Burton, A. M. (2017). Natural variability is essential to learning new faces. *Visual Cognition*, *25*(4-6), 470-476. https://doi.org/10.1080/13506285.2016.1242522

Lander, K., Bruce, V., & Bindemann, M. (2018). Use-inspired basic research on individual differences in face identification: Implications for criminal investigation and security. *Cognitive Research: Principles and Implications*, 3(1), 1-13. https://doi.org/10.1186/s41235-018-0115-6

Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, 143, 123-138. https://doi.org/10.1016/j.jecp.2015.09.030

Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. *British Journal of Psychology*, 107(2), 374-388. https://doi.org/10.1111/bjop.12147

Li, C., Burton, A. M., Ambrus, G. G., & Kovács, G. (2022). A neural measure of the degree of face familiarity. *cortex*, *155*, 1-12. https://doi.org/10.1016/j.cortex.2022.06.012

Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, & Psychophysics*, 76, 1975–1984. https://doi.org/10.3758/s13414-014-0653-z

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, *98*(1), 185. https://doi.org/10.1037/0033-2909.98.1.185

Matthews, C. M., & Mondloch, C. J. (2018). Finding an unfamiliar face in a line-up: Viewing multiple images of the target is beneficial on target-present trials but costly on target-absent trials. *British Journal of Psychology*, 109(4), 758-776. https://doi.org/10.1111/bjop.12301

Matthews, C. M., & Mondloch, C. J. (2021). Learning and recognizing facial identity in variable images: New insights from older adults. *Visual Cognition*, *29*(10), 708-731. https://doi.org/10.1080/13506285.2021.2002994

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*(1), 1-15. https://doi.org/10.1186/s41235-018-0112-9

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal of Cognitive Psychology*, 25(1), 30-37. https://doi.org/10.1080/20445911.2012.739153

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. https://doi.org/10.1037/a0013464

Menon, N., White, D., & Kemp, R. I. (2015). Identity-level representations affect unfamiliar face matching performance in sequential but not simultaneous tasks. The Quarterly

Journal of Experimental Psychology, 68(9), 1777-1793. https://doi.org/10.1080/17470218.2014.990468

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. http://dx.doi.org/10.1037/xhp0000049

Noyes, E., & Jenkins, R. (2017). Camera-to-subject distance affects face configuration and perceived identity. *Cognition*, 165, 97-104. https://doi.org/10.1016/j.cognition.2017.05.012

Noyes, E., & Jenkins, R. (2019). Deliberate disguise in face identification. *Journal of Experimental Psychology: Applied*, *25*(2), 280. https://doi.org/10.1037/xap0000213

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in cognitive sciences*, *22*(9), 794-809. https://doi.org/10.1016/j.tics.2018.06.006

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. Quarterly Journal of Experimental Psychology, 70(5), 897-905. https://doi.org/10.1080/17470218.2015.1136656

Ritchie, K. L., Kramer, R. S., & Burton, A. M. (2018). What makes a face photo a 'good likeness'?. *Cognition*, *170,* 1-8. https://doi.org/10.1016/j.cognition.2017.09.001

Ritchie, K. L., Kramer, R. S., Mileva, M., Sandford, A., & Burton, A. M. (2021). Multiple-image arrays in face matching tasks with and without memory. *Cognition*, *211*, 104632. https://doi.org/10.1016/j.cognition.2021.104632

Rosemblaum, M., Grosbard, I., Abudarham, N., & Yovel, G. (2021). What type of experience is needed to generate a human-like view-invariant representation of face identity? Evidence from Deep Convolutional Neural Networks. *Journal of Vision*, *21*(9), 2595-2595. https://doi.org/10.1167/jov.21.9.2595

Semplonius, T., & Mondloch, C. J. (2015). Attentional biases and recognition accuracy: What happens when multiple own-and other-race faces are encountered simultaneously?. *Perception*, *44*(1), 52-70. https://doi.org/10.1068/p7892

Shoham, A., Kliger, L., & Yovel, G. (2022). Learning faces as concepts improves face recognition by engaging the social brain network. *Social cognitive and affective neuroscience*, *17*(3), 290-299. https://doi.org/10.1093/scan/nsab096

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive research: principles and implications*, *5*(1), 1-17. https://doi.org/10.1186/s41235-019-0205-0

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47. https://doi.org/10.1037/xap0000108

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology*, *1*(5), 287-300. https://doi.org/10.1038/s44159-022-00041-3

White, D., Wayne, T., & Varela, V. P. (2022). Partitioning natural face image variability emphasises within-identity over between-identity representation for understanding accurate recognition. *Cognition, 219*, 104966. https://doi.org/10.1016/j.cognition.2021.104966

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A robust neural index of high face familiarity. *Psychological science*, *30*(2), 261-272. https://doi.org/10.1177/0956797618813572

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. https://doi.org/10.1080/14640749108400966

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, *79*(4), 471-491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision research*, *141*, 217-227. https://doi.org/10.1016/j.visres.2016.12.014

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of experimental psychology: Learning, memory, and cognition*, *20*(6), 1341. https://doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characterstics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1415. https://doi.org/10.1037/0278-7393.25.6.1415

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, *46*(3), 441-517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178-1194. https://doi.org/10.1002/hipo.20864

Zhou, X., Matthews, C. M., Baker, K. A., & Mondloch, C. J. (2018). Becoming familiar with a newly encountered face: Evidence of an own-race advantage. *Perception*, *47*(8), 807-820. https://doi.org/10.1177/0301006618783915

Zhou, X., Mondloch, C. J., Chien, S. H. L., & Moulson, M. C. (2022). Multi-cultural cities reduce disadvantages in recognizing naturalistic images of other-race faces: evidence from a novel face learning task. *Scientific Reports*, *12*(1), 1-14.

Zhou, X., Vyas, S., Ning, J., & Moulson, M. C. (2022). Naturalistic Face Learning in Infants and Adults. *Psychological Science*, *33*(1), 135-151. https://doi.org/10.1177/09567976211030630

Chapter 5: General Discussion

Face identification is important for both our social and societal functions. Socially, face identification is often performed when we need to recognize our friends and family. Even when watching a movie, we must recognize an actor to follow the plot. Although these examples might seem like easy tasks, that is merely because we are performing them with familiar faces. Societally, face identification is often performed by matching identity in unfamiliar faces. This is a task that is performed in instances such as checking the ID of individuals who are crossing the border, purchasing age-restricted goods, or taking a university exam. Despite having never met the person holding an ID card, they must be recognized even when their appearance changes and discriminated from individuals who look similar. Whereas it is easy to recognize familiar faces despite these challenges, recognizing unfamiliar faces is much more difficult. Indeed, it is even difficult to identify unfamiliar faces in laboratory experiments that use photographs that were taken within the same photography session (for examples see Burton et al., 2010; Benton et al., 1994; Duchaine & Nakayama, 2006).

The overarching goal of my dissertation was to study individual differences in unfamiliar face matching and face learning. By doing so, my dissertation speaks to White and Burton's (2022) call for face identification researchers to use individual difference approaches to drive theory and models of face identification. Identifying faces requires both the ability to recognize an individual across variable instances (i.e., telling together) and the ability to discriminate them from individuals who look similar (i.e., telling apart). For this reason, many face identification studies separately analyze match vs. mismatch trials. This way of thinking makes sense for applied settings as it enables researchers to separate two types of errors (*misses* and *false alarms* [FAs]) which have different consequences (e.g., in eyewitness testimony settings: a miss could result in allowing a criminal to go free and a FA could result in sending an innocent person to jail). This way of thinking has allowed researchers to show that training typically only affects match or mismatch accuracy, but rarely both (e.g., Alenzi & Bindemann, 2013; Ritchie et al., 2021; for a brief discussion see Ritchie & Burton, 2017), and has led to reports that there is no mirror effect in unfamiliar face identification (Megreya & Burton, 2007).

However, performance on match and mismatch trials is influenced by both sensitivity and bias. A perfect score on match or on mismatched trials could reflect sensitivity or a strong bias towards one of the response options. This problem is what signal detection theory (SDT) was designed to address. In my dissertation, I took a SDT approach to examine individual differences in unfamiliar face identification and face learning. In using this approach, I was able to focus on both sensitivity to identity and on bias. Where that was not possible (e.g., the lineup task) I created a proxy measure. This was an important first step to addressing the role of decision making in face identification—a step that speaks to the recent call to action by Bindemann and Burton (2021). I also went beyond the traditional approach by analyzing the data from the learning task using the dual-process signal detection (DPSD) model. This allowed me to look at two processes that underly memory and influence responses: Recollection (old, new) and familiarity.

Three key findings

By using these novel approaches, my dissertation package reflects three key lessons for the face identification field. These findings show the stability of individual differences in sensitivity to identity, that unfamiliar face matching predicts face learning efficiency, and that face identification is not just a perceptual problem. Although my thesis was not designed to test models of face identification or decision making, my research questions were inspired by these models. Thus, following the discussion of each of these three findings, I have reflected on their theoretical significance. As these findings also have applied significance, I have also discussed it below.

Sensitivity to identity is stable across times and tasks

I showed that individual differences in sensitivity to identity in unfamiliar face identification tasks are stable across time and tasks (Chapter 3). Prior to this study, very little was known about the reliability of performance on face identification tasks over time. Several measures of face identification had not even been investigated for their reliability (for a discussion see White & Burton, 2022). Understanding the reliability of a measure is important because it is the upper threshold of that measures' ability to correlate with another measure (White & Burton); reliability also reflects the measures' quality as an individual difference measure (Goodhew & Edwards, 2019). Moreover, prior to this study, most of what we knew about the convergent validity of face identification measures was primarily based on relationships between face tasks that incorporate limited variability in appearance (e.g., Cambridge Face Memory Test [CFMT] and Glasgow Face Matching Test [GFMT]; for examples see, McCaffery et al., 2018; Verhallen et al., 2017). As our field employs many different types of tasks, assessing the convergent validity across multiple face identification tasks is important. In fact, by doing so, I showed that it is the number of misidentification errors made during the Sorting Task that falls on the PCA component that reflects sensitivity to identity (Chapter 3). This is important because researchers often conceptualize it as measure of accuracy (for examples see Jenkins et al., 2011; Laurence et al., 2016; Zhou & Mondloch, 2016).

This consistency of performance among face identification measures (i.e., multiple face identification tasks) is stronger than the consistency of performance among measures identification that span across modalities (i.e., tasks that measure other visual stimulus categories including faces; see Verhallen et al., 2017; McCaffery et al., 2018). For example, despite sharing very similar task demands, the CFMT and Cambridge Car Memory Test share 13.7% of variance (Dennett et al., 2012). In Chapter 3, I used tasks with different task demands and found stronger relationships. In fact, component 1 accounted for 29-31% of variance. Collectively, these findings can speak to the question of whether these individual differences in performance are face specific. As there is more shared variance between face identification tasks than between tasks with different stimulus categories, it suggests that these individual differences are face specific, at least in part. They are likely not entirely face specific, as there is still leftover variance that is not accounted for by face identification performance (see future directions).

Theoretical implications. Establishing convergent validity across tasks that vary in demands has important theoretical implications for the field and speaks to the classic models of face identification. Individual differences in face identification can be conceptualized in light of Valentine's (1991) multi-dimensional face space model (see also Valentine et al., 2016). This model discusses differences in recognition as a function of the face being recognized (e.g., atypicality, face category [for a discussion see O'Toole et al., 2000]); my dissertation draws attention to differences in recognition across perceivers. This idea is reminiscent of experiential effects that are accounted for by Valentine's model. Valentine's model predicts that children would be less sensitive to the individual dimensions of face space than adults, and that adults would be less sensitive for other- vs. own-race face—effects that correspond to poorer accuracy (e.g., Megreya & Bindemann, 2015; Neil et al., 2016; Laurence & Mondloch, 2016; Laurence et al., 2016; for a discussion see Valentine et al., 2016). According to this model, the individual differences in face identification ability shown in my dissertation would be driven, at least in part, by individual differences in experience.

Although Bruce and Young (1986) did not speak to individual differences in performance in their model, they did suggest the need to better understand them in their review of their model (Young & Bruce, 2011). My findings speak to this and suggest that individual differences in face identification ability greatly influence whether a face is accurately identified. Although my data cannot point to where in the model these individual differences are occurring, it is likely that they could occur at many different points within the Bruce and Young model. For example, they might even occur at the formation of a pictorial representation of an identity. This would likely have downstream effects for learning and familiar face identification. On the other hand, these individual differences might occur in the other cognitive systems—such as in attention or even decision-making systems. Clearly, individual difference research might be able to further refine this model of face identification, just as articulated by White and Burton (2022).

Applied implications. The stability of sensitivity to identity over time and across tasks suggests that people who are especially good at face identification in one context are also especially good in other contexts. An accurate eyewitness would also be accurate on the GFMT. Sixty-nine percent of wrongful convictions occur because of misidentifications in eyewitness testimony (see The Innocence Project). My research suggests that an eyewitnesses' testimony could be evaluated based on their unfamiliar face identification ability. For example, they could complete an unfamiliar face identification task after their selection of an individual from a lineup. This could have important implications for the use of eyewitness testimony in court hearings. Putting an eyewitness' testimony in the context of their face identification ability could help to reduce the number of wrongful convictions in the court system.

Unfamiliar face matching predicts the slope of face learning

I showed that the ability to match identity in wholly unfamiliar faces predicts slope of learning (Chapter 4). As surprising little is known about face learning, this is an incredibly important finding. Most studies that have examined face learning only measure the final learning outcome (for examples see Baker et al., 2017; Baker & Mondloch, 2019; Ritchie & Burton, 2017; Murphy et al., 2015). To examine learning as it unfolds, I developed a novel task in which participants were tested on their recognition of an identity four times during the process of face learning. I showed that the slope of face learning not only varies across people, but it is also related to unfamiliar face matching ability.

This is consistent with hints in the literature that suggest that group differences in unfamiliar face matching (i.e., effects attributable to differences in experience; for examples see Laurence et al., 2016; Prioetti et al., 2019; Tanaka & Farah, 1993; Valentine, 1988) correspond with differences in face learning (for examples see Baker et al., 2017; Zhou et al., 2018; Kramer et al., 2017). It is also consistent with individual differences in unfamiliar face matching and face memory (e.g., the CFMT; Models of memory test; for examples see McCaffery et al., 2018; Verhallen et al., 2017; Fysh et al., 2020; Stacchi et al., 2020). Here, I added to this literature by being the first to look at performance as learning unfolds.

Theoretical Implications. This finding has important implications for models of face identification. It relates to Bruce and Young's (1986) model. Bruce and Young explicitly incorporate face familiarity and within-person variability in their model. Differences in sensitivity result from differences in representations for unfamiliar (image-based; pictorial codes) and familiar faces (robust representations; person identity nodes [PINs]). My finding that matching ability predicts the slope of face learning suggests that an initial view-dependent representation (i.e., from the first encounter with a face) shapes the development of a view-independent representation and face recognition unit (FRU; i.e., by attending to the identity-diagnostic cues). A person is recognized when the FRU activates the PIN and therefore has access to identity-specific information. An unrepresentative FRU would reduce the probability that a view-independent representation will activate the PIN. This would have cascading effects on one's representation of an identity. The new instance would not be included in the representation of that identity, resulting in the representation of an identity remaining the same and not becoming stronger or more robust with exposure to new instances.

Although Valentine's (1991) multi-dimensional face space cannot account for face learning, the DCNN face space model can account for these effects (for a discussion see O'Toole et al., 2018). DCNNs are a class of algorithm that is trained to model the human visual system (e.g., Phillips et al., 2018; Noyes et al., 2021) and are currently helping to refine discussion about models of face representation. In a recent study, Rosemblaum et al. (2021) showed that the number of identities on which a DCNN is trained influences the number of images on which it needs to be trained to learn a new identity. This could suggest that representations of familiar faces (how many we know) shapes representation of unfamiliar faces, which bootstraps face learning (see future research). Another possibility is that the number of different faces encountered (including unfamiliar faces) could shape the representation of unfamiliar faces (see future research). These two possibilities are not mutually exclusive.

Applied Implications. I showed that there are individual differences in the amount of benefit that individuals will gain from learning a new face via exposure to variability in appearance. In the context of eyewitness testimony, eyewitnesses can vary in their face identification ability. Eyewitnesses with good unfamiliar face identification abilities would need less exposure to the criminal's appearance to be accurate than those with poorer matching abilities. Estimating an eyewitness' unfamiliar face identification ability and having an idea about the context of the crime (e.g., how long were they witnessing the crime, were they able to see the criminal's face from many angles), would help to put their testimony in context and could help to reduce wrongful convictions.

In the context of police procedure, this finding can also help to improve wanted criminal searches. Whereas one officer might show great improvement after being shown multiple images of the criminal, another officer might show less improvement and might need exposure to more images. Police could improve their chances at locating wanted criminals by estimating officers' unfamiliar face identification ability and using these estimates to guide the number of photos to which one needs to be exposed when tasked with finding wanted criminals.

Face identification is not merely a perceptual problem

In my dissertation I have shown that face identification is not purely a perceptual problem. First, I showed that individual differences in criterion are reliable across time and tasks, and are independent of sensitivity (Chapter 3). Prior to this study, no studies in the face identification field had examined the stability of bias. Establishing the reliability and convergent validity of bias could be seen as the first step towards Bindemann and Burton's (2021) call to better integrate decision making into theories of face identification, especially as reliability and convergent validity speak to the quality of individual difference measures (White and Burton, 2022; Goodhew et al., 2019). This process also allowed me to establish that the number of piles made in the Sorting Task is related to bias, not sensitivity (Chapter 3). This is important as the number of piles made in the Sorting Task is often conceptualized as the primary measure of accuracy (e.g., Jenkins et al., 2011). Therefore, checking the convergent validity of performance across multiple face identification tasks allowed me to understand that it is unlikely that performance on the Sorting Task is merely perceptual discrimination and recognition, but actually involves other processes.

Second, I showed that response times on correct trials varied based on whether the trial was congruent with one's bias (Chapter 3). This is important because it suggests that individual differences in criterion are capturing decision making (see theoretical implications). Similarly, photography experience predicted bias, not sensitivity (Chapter 2). Prior to this study, photography experience had never been investigated as a potential predictor of face identification performance. These two findings provide hints to the decision-making processes that are involved in face identification. In SDT, criterion is a decision boundary placed between the distributions of evidence for each alternative (Summerfield & Egner, 2014). The decision that one makes depends on the placement of criterion. For either response option the response that is made is determined by whether there is enough signal, or evidence, to reach the criterion (Summerfield & Egner, 2014; Stanislaw & Todorov, 1999). These findings suggest that participants need less evidence, or perhaps take less time accumulating evidence, when the accurate response for a trial is aligned with their threshold, and that the placement of this threshold is likely influenced by experience—at least with photography. These findings suggest that it is likely that face identification responses are not entirely based on perception.

Third, individual differences in recollection captured individual differences in face learning. Recall that recollection old reflects high confidence hits and recollection new reflects high confidence correct rejections. Not only do these findings suggest that the responses that are made in face identification tasks incorporate episodic memory, but that the decisions that one makes in face identification tasks are informed by recollection or feelings of familiarity. Thus, this suggests that face identification is not only a perceptual process. This is consistent with findings that face memory predicts episodic recall (a proxy for recollection) of familiar faces (Devue et al., 2019), and that group level differences in unfamiliar face matching (e.g., Matthews & Mondloch, 2021; Zhou et al., 2022) correspond to fewer recollection responses (e.g., Bartlett & Fulton, 1991) and episodic recall (Semplonius & Mondloch, 2015). However, I was the first to examine this during the process of learning following exposure to variability in appearance.

Collectively, these findings suggest that face identification is not a purely perceptual ability, but rather it is also influenced by other factors such as decision making. This evidence has made clear that to fully understand individual differences in face identification we cannot just focus on accuracy or d′; we need to also measure RTs and bias. This evidence also suggests that we need to treat interpretations of Hits and FAs carefully—as any changes might be driven by sensitivity, bias or both.

Theoretical Implications. These data speak to classic and contemporary models of face identification. While Bruce and Young (1986) alluded to decisions having a role in face identification, by including an “other cognitive processes” section in their model, DCNNs consider decision making more explicitly and have recently started to manipulate criterion. Cavazos et al. (2020) discussed how manipulating a manually set decision threshold in DCNNs influences performance. Fixed thresholds for own- and other-race faces result in fewer FAs for Caucasian than East Asian faces. Setting a separate threshold for own- and other-race faces produces similar levels of FAs for both Caucasian and East Asian faces. This finding further establishes that criterion influences the errors made, and hints towards a role for decision making influencing performance on face identification tasks.

The findings in this section also address a call to better integrate decision making into theories of face identification (see Bindemann and Burton, 2021). Doing so makes sense as in every behavioural task that is used to measure face identification, participants are asked to make a decision. For instance, in a same/different task participants are asked to decide whether two images belong to the same person or different people. Indeed, different observers make different decisions to the same faces (Bindemann et al., 2012a; Fysh & Bindemann, 2018; Stantic et al., 2021), and often the same observers make different decisions to the same faces at different times (Alenezi & Bindemann, 2013; Alenezi et al., 2015; Bindemann et al., 2012b). Contextual manipulations in face identification tasks also change participants' decisions, even when stimuli are held constant (Fysh & Bindemann, 2017; Papesh et al., 2018). These patterns should simply not arise if identification errors only reflect the available visual information in faces, and therefore point to inconsistencies in how visual information is used to reach an identification decision. By investigating bias and RTs as an initial glance into decision making, I have added to this growing body of literature.

Recall that there are two typical models of decision making, dual process and drift diffusion. Proponents of dual-process theories propose that two systems are involved in decision making (Evans & Stanovich, 2013). System one is rapid and automatic and System two is slow and deliberate. Although dual-process theories of decision making cannot speak to my findings that participants respond fastest to trials that are congruent with their bias, or to the differences found in criterion for Expert vs. Hobbyist photographers, it can account for the differences shown in DPSD. In the dual-process memory literature, familiarity is akin to system one and recollection is akin to system two (e.g., Joordens & Hockley, 2000; see Ozubko, 2007). Although there were no differences in familiarity, there were group differences in recollection new responses. This suggests that poor matchers are engaging less in this system two process overall than mid-level and good matchers—despite benefitting from variability, just as mid-level and good matchers had. Moreover, the interaction for recollect old suggests that whereas the poor matchers are not ever engaging in this system two process, the mid-level matchers do so eventually, and the good matchers do so immediately.

Proponents of drift diffusion models argue that, until a decision threshold is reached, evidence and noise are accumulated (Ratcliff et al., 2016). The subsequent decisions and their accuracy are influenced by three parameters. One parameter measures the point at which evidence accumulation starts (starting point). A second parameter measures the speed of evidence accumulation and evidence quality (i.e., drift rate). This can be biased towards one of the response options, reducing the distance between starting point and criterion for that option. Bindemann and Burton (2021) spoke about something similar in their review. They referred to it as a counting strategy (i.e., counting and comparing the number of similarities and differences across image pairs). A third parameter measures the amount of evidence that one requires to make a decision (i.e., criterion; decision threshold). Response speed is influenced by the above parameters. It is also influenced by non-decision components (e.g., speed of initiating a response). The drift diffusion model easily accounts for the differences in RTs on trials that were congruent vs. incongruent with participants' bias. Both accuracy and response speed contribute to the drift rate parameter. The amount of evidence required for the drift rate is influenced by the starting point or the decision threshold. It is likely that participants' criterion on trials that are congruent with their bias are placed such that they need less evidence to make that response. It also can account for differences in criterion for hobbyist vs. expert photographers. This too speaks to differences in evidence accumulation—the expert photographers have their criterion placed such that they need more evidence to make a “same” response. Although it is possible that the differences in recollection (old and new) occurred because good and mid-level matchers were better able to extract quality evidence for the drift rate parameter than the poor matchers, these data cannot speak to this hypothesis. It is possible that drift-diffusion models cannot account for the DPSD findings.

Applied Implications. In many applied situations, some errors are more costly than others (e.g., Selling alcohol to a minor is a more costly error than refusing to sell alcohol to an appropriately aged patron. This is because it is more costly for the business to lose its liquor licence than it is to lose a customer). My finding that criterion is stable across time and tasks has important implications for applied settings, especially as one’s criterion influences the types of error that they make. For example, despite having similar accuracy, super recognizers and forensic examiners have different biases (Towler et al., 2021), such that super recognizers have a stronger “same-person” bias than forensic examiners. Indeed, when super recognizers and forensic examiners do make errors, super recognizers more often make high confidence FAs than do the forensic examiners (Towler et al.).

These findings also have implications for training protocols. Many studies have shown that face identification training does not extend beyond the identities on which an individual was trained (e.g., Matthews & Mondloch, 2018; Dowsett et al., 2016). When training does improve performance, the benefit is only modest (for an example see Towler et al., 2021). The very fact that face identification does not rely entirely on perception, points to a possible reason as to why sensitivity is difficult to train. Nonetheless, the fact that face identification is not purely a perceptual problem also points to a probable solution for how to improve face identification ability: We need to attempt to train these other factors influencing face identification. Success in doing so would benefit all interactions in which checking photographic ID is required. For example, I showed that criterion placement is stable across time and tasks, this would suggest that ones' typical criterion placement can be estimated. Other studies have also shown that individuals can shift their criterion flexibly (Kantner & Lindsay, 2014). Thus, individuals checking photographic ID (e.g., passport officers, bartenders) who are aware of their criterion placement, might be able to adjust their criterion to make fewer of their more frequent errors or reduce the type of error that is most costly.

Future Directions

Two key questions are raised by this work. The first is what is driving individual difference in performance. It is possible that individual differences in meta-cognition could be playing a role. Meta-cognition is often thought to reflect three core components: Meta-cognitive knowledge (i.e., knowing about ones' cognitive processes, such as their own vs. other's abilities), meta-cognitive strategies (i.e., use of strategies to control cognition), meta-cognitive experiences (knowledge/judgements of task performance; for discussions see Norman et al., 2019; Norman, 2020). Being aware of one's performance and cognition can allow for on-line adjustments in strategies to better performance.

Self-reported insight into face identification abilities and performance on face identification tasks are typically only moderately related. The shared variance ranges between 1.7% to 9% (e.g., Bobak et al., 2018; Matsuyoshi & Watanabe, 2021; Palermo et al., 2017; for a brief discussion see White & Burton, 2022), although two studies have reported higher estimates (11% to 66%; Livingston & Shah, 2017; Shah et al., 2015). Insight has also been measured using participants' ratings of confidence. Recent data shows that the relationship between accuracy and confidence is strongest for the highest performing individuals (Grabman & Dodson, 2020; Zhou

& Jenkins, 2020)—at least on a trial-by-trial basis (Kramer et al., 2021). Although I did not measure participants' predicted accuracy, my data suggest that the inability to predict cannot be attributed to performance being unstable.

Bias has not been discussed in the context of metacognition in face identification tasks. However, it is possible that shifting one's criterion might work as a metacognitive strategy. For example, if one were to determine that their performance was sub-optimal and reasoned that they were mostly wrong when making *match* responses, they would be able to shift their criterion such that they need more evidence to make a match response. There is evidence from economic models to suggest that criterion can shift flexibly; individuals can shift their criterion in response to unequal base rates, costs/benefits, and stimulus difficulty (Kantner & Lindsay, 2014). Moreover, the ability to shift criterion has been shown to vary between individuals (for a discussion see Miller & Kanter, 2020). In recent work on which I am a co-author, we showed that the shift of criterion in response to base rate or cost manipulations is related to accuracy (Stabile et al., in prep). However, it is unclear if this effect would remain in a neutral environment.

The second question raised by work is what mediates the relation between unfamiliar face identification and face learning. This relationship could be mediated by domain-general and/or domain-specific mechanisms. Importantly, these are not mutually exclusive. First, a potential mediator could be low-level or general (object) visual processing. This would be consistent with Gauthier and colleagues (2014), who showed that the relationship between face identification (measured via the CFMT) and object processing across multiple categories (measured via the Vanderbilt Expertise Test [VET]) depends on experience with objects (measured via self report). Other studies have investigated potential relationships between face and word processing. For example, dyslexic readers tend to perform more poorly on the CFMT and when recognizing novel images of objects than typical readers (Sigurdardottir et al., 2015). Reading ability also relates to configural processing of faces (Jozranjbar et al., 2021). These studies suggest that it is plausible that some of the variability in matching and learning might be mediated by domain general processing.

A second potential mediator could be the number of unfamiliar faces encountered (a domain-specific account). A finding such as this would be consistent with hometown size effects (i.e., small vs. large), where in those from small towns show poorer face identification abilities (Balas & Saville, 2015, 2017). It is well understood that adults perform more poorly on face identification tasks that use faces from categories with which they lack experience (e.g., Laurence et al., 2016; Prioetti et al., 2019; Tanaka & Farah, 1993; Valentine, 1988). This effect might be driven by the number of faces typically encountered. Evidence for this possibility is hinted by Zhou and colleagues (2022) in their recent paper. Individuals from more ethnically homogenous cities (e.g., St. Catharines, Canada) show a larger other-race effect than those from more ethnically heterogeneous cities (e.g., Toronto, Canada). These studies suggest that it is plausible that some of the variability in matching and learning might be mediated by the number of unfamiliar faces encountered.

A third potential moderator is the number of known faces (a domain-specific account). Individual differences in sensitivity to identity in wholly unfamiliar faces is partially attributable to individual differences in face-specific experience. On average, adults recognize approximately 5000 faces, however this estimate varies widely (i.e., 1000 to 10,000 faces) between individuals (Jenkins et al., 2018). Even infants facial experience is tuned to lots of experience with a few faces. Although most of the time is spent with the three most frequent faces, the number of faces infants have experience with ranges from 2 to 20 (Jayaraman et al., 2015; Sugden & Moulson, 2019). Just as in humans, DCNNs that are trained on fewer identities are less accurate in identifying unfamiliar faces than DCNNs that are trained with more identities (Blauch et al., 2021; Hill et al., 2019; O'Toole et al., 2018; O'Toole & Castillo, 2021). Moreover, DCNN trained with more identities need fewer images to learn new identities than those that are trained with less (Rosemblaum et al., 2021). Collectively, these studies would suggest that it is plausible that some of the variability in matching and learning might be mediated by the number of familiar faces that are known.

Conclusions

In using SDT, DPSD and an individual difference approach to examine face identification, I was able to provide key insights for theory and our understanding of unfamiliar face identification and face learning. First, the stability of individual differences in sensitivity to identity are stable across time and tasks. Second, unfamiliar face matching predicts face learning efficiency. Third, face identification is not just a perceptual problem. In using these approaches, I spoke to recent calls to drive theory by examining individual differences in face identification (White & Burton, 2022) and to incorporate decision making in face identification theories (Bindemann & Burton, 2021). These findings speak to classic and contemporary models of face identification and also have implications for unfamiliar face identification and for face learning in applied settings.

References

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology*, *27*(6), 735-753. https://doi.org/10.1002/acp.2968

Alenezi, H. M., Bindemann, M., Fysh, M. C., & Johnston, R. A. (2015). Face matching in a long task: Enforced rest and desk-switching cannot maintain identification accuracy. *PeerJ*, *3*, e1184. https://doi.org/10.7717/peerj.1184

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19-30. https://doi.org/10.1016/j.cognition.2016.12.012

Baker, K. A., & Mondloch, C. J. (2019). Two Sides of Face Learning: Improving Between-Identity Discrimination While Tolerating More Within-Person Variability in Appearance. *Perception*, 48(11), 1124-1145. https://doi.org/10.1177/0301006619867862

Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. *Neuropsychologia*, *69*, 211-217. https://doi.org/10.1016/j.neuropsychologia.2015.02.005

Balas, B., & Saville, A. (2017). Hometown size affects the processing of naturalistic face variability. *Vision research*, *141*, 228-236. https://doi.org/10.1016/j.visres.2016.12.005

Benton, A. L., Sivan, A. B., Hamsher, K. D., Varney, N. R., & Spreen, O. (1994). *Contributions to neuropsychological assessment: A clinical manual*. Oxford University Press, USA.

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012a). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(2), 96-103. https://doi.org/10.1016/j.jarmac.2012.02.001

Bindemann, M., Avetisyan, M., & Rakow, T. (2012b). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied*, *18*(3), 277. https://doi.org/10.1037/a0029635

Bindemann, M., & Burton, A. M. (2021). Steps towards a cognitive theory of unfamiliar face matching. *Forensic face matching: Research and practice*, 38-61.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2021). Deep learning of shared perceptual representations for familiar and unfamiliar faces: Reply to commentaries. *Cognition*, 208, 104484. https://doi.org/10.1016/j.cognition.2020.104484

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. https://doi.org/10.1177/1747021818776145

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology*, 77(3), 305-327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior research methods*, *42*(1), 286-291. https://doi.org/10.3758/BRM.42.1.286

Cavazos, J. G., Phillips, P. J., Castillo, C. D., & O'Toole, A. J. (2020). Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?. *IEEE transactions on biometrics, behavior, and identity science,* 3(1), 101-111. doi: 10.1109/TBIOM.2020.3027269.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior research methods*, *44*(2), 587-605. https://doi.org/10.3758/s13428-011-0160-2

Devue, C., Wride, A., & Grimshaw, G. M. (2019). New insights on real-world human face recognition. *Journal of Experimental Psychology: General*, 148(6), 994. https://doi.org/10.1037/xge0000493

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. https://doi.org/10.1080/17470218.2015.1017513

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

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, *8*(3), 223-241. https://doi.org/10.1177/174569161246068

Fysh, M. C., & Bindemann, M. (2018). The Kent face matching test. *British journal of psychology*, 109(2), 219-231. DOI:10.1111/bjop.12260

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: implications for theory, research and personnel selection. *Royal Society open science*, 7(9), 200233. https://doi.org/10.1098/rsos.200233

Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Van Gulick, A. E. (2014). Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of vision*, *14*(8), 7-7. doi:https://doi.org/10.1167/14.8.7

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, 69, 14-25. https://doi.org/10.1016/j.concog.2019.01.008

Grabman, J. H., & Dodson, C. S. (2020). Stark individual differences: Face recognition ability influences the relationship between confidence and accuracy in a recognition test of Game of Thrones actors. *Journal of Applied Research in Memory and Cognition*, *9*(2), 254-269. https://doi.org/10.1016/j.jarmac.2020.02.007

Hill, M. Q., Parde, C. J., Castillo, C. D., Colon, Y. I., Ranjan, R., Chen, J. C., Blanz, V. & O'Toole, A. J. (2019). Deep convolutional neural networks in the face of caricature. Nature Machine Intelligence, 1(11), 522-529. https://doi.org/10.1038/s42256-019-0111-7

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PloS one*, *10*(5), e0123780. https://doi.org/10.1371/journal.pone.0123780

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know?. *Proceedings of the Royal Society B, 285*(1888), 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. https://doi.org/10.1016/j.cognition.2011.08.001

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: when old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1534. https://doi.org/10.1037/0278-7393.26.6.1534

Jozranjbar, B., Kristjánsson, Á., & Sigurdardottir, H. M. (2021). Featural and configural processing of faces and houses in matched dyslexic and typical readers. *Neuropsychologia*, *162*, 108059. https://doi.org/10.1016/j.neuropsychologia.2021.108059

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, *21*(5), 1272-1280. https://doi.org/10.3758/s13423-014-0608-3

Kramer, R. S., Jenkins, R., Young, A. W., & Burton, A. M. (2017). Natural variability is essential to learning new faces. *Visual Cognition*, 25(4-6), 470-476. https://doi.org/10.1080/13506285.2016.1242522

Kramer, R. S., Gous, G., Mireku, M. O., & Ward, R. (2022). Metacognition during unfamiliar face matching. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12553

Laurence, S., & Mondloch, C. J. (2016). That's my teacher! Children's ability to recognize personally familiar and unfamiliar faces improves with age. *Journal of Experimental Child Psychology*, *143*, 123-138. https://doi.org/10.1016/j.jecp.2015.09.030

Laurence, S., Zhou, X., & Mondloch, C. J. (2016). The flip side of the other-race coin: They all look different to me. British Journal of Psychology, 107(2), 374-388. https://doi.org/10.1111/bjop.12147

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*(1), 1-15. https://doi.org/10.1186/s41235-018-0112-9

Matsuyoshi, D., & Watanabe, K. (2021). People have modest, not good, insight into their face recognition ability: a comparison between self-report questionnaires. *Psychological Research*, *85*(4), 1713-1723. https://doi.org/10.1007/s00426-020-01355-8

Matthews, C. M., & Mondloch, C. J. (2018). Finding an unfamiliar face in a line-up: Viewing multiple images of the target is beneficial on target-present trials but costly on target-absent trials. *British Journal of Psychology*, 109(4), 758-776. https://doi.org/10.1111/bjop.12301

Matthews, C. M., & Mondloch, C. J. (2021). Learning and recognizing facial identity in variable images: New insights from older adults. *Visual Cognition*, *29*(10), 708-731. https://doi.org/10.1080/13506285.2021.2002994

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*(1), 5-22. https://doi.org/10.1068/p7825

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. http://dx.doi.org/10.1037/xhp0000049

O'Toole, A. J., & Castillo, C. D. (2021). Face Recognition by Humans and Machines: Three Fundamental Advances from Deep Learning. *Annual Review of Vision Science*, *7*, 543-570. https://doi.org/10.1146/annurev-vision-093019-111701

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in cognitive sciences*, *22*(9), 794-809. https://doi.org/10.1016/j.tics.2018.06.006

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, *20*(4), 260-281. https://doi.org/10.1016/j.tics.2016.01.007

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897-905. https://doi.org/10.1080/17470218.2015.1136656

Rosemblaum, M., Grosbard, I., Abudarham, N., & Yovel, G. (2021). What type of experience is needed to generate a human-like view-invariant representation of face identity? Evidence from Deep Convolutional Neural Networks. Journal of Vision, 21(9), 2595-2595. https://doi.org/10.1167/jov.21.9.2595

Livingston, L. A., & Shah, P. (2018). People with and without prosopagnosia have insight into their face recognition ability. *Quarterly Journal of Experimental Psychology*, *71*(5), 1260-1262. https://doi.org/10.1080/17470218.2017.1310911

Stacchi, L., Huguenin-Elie, E., Caldara, R., & Ramon, M. (2020). Normative data for two challenging tests of face matching under ecological conditions. *Cognitive research: principles and implications*, *5*(1), 1-17. https://doi.org/10.1186/s41235-019-0205-0

Sugden, N. A., & Moulson, M. C. (2019). These are the people in your neighbourhood: Consistency and persistence in infants' exposure to caregivers', relatives', and strangers' faces across contexts. *Vision Research*, *157*, 230-241. https://doi.org/10.1016/j.visres.2018.09.005

Summerfield, C., & Egner, T. (2013). Attention and decision-making. *The Oxford handbook of attention.* DOI: 10.1093/oxfordhb/9780199675111.013.018

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly journal of experimental psychology*, *46*(2), 225-245. https://doi.org/10.1080/14640749308401045

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

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. https://doi.org/10.1080/14640749108400966

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, *79*(4), 471-491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision research*, *141*, 217-227. https://doi.org/10.1016/j.visres.2016.12.014

White, D., & Burton, A. M. (2022). Individual differences and the multidimensional nature of face perception. *Nature Reviews Psychology*, *1*(5), 287-300. https://doi.org/10.1038/s44159-022-00041-3

Zhou, X., & Mondloch, C. J. (2016). Recognizing “Bella Swan” and “Hermione Granger”: No own-race advantage in recognizing photos of famous faces. *Perception*, *45*(12), 1426-1429. https://doi.org/10.1177/0301006616662046