Face processing in photographic identity documents

Xinran Feng

Doctor of Philosophy

University of York

Psychology

August 2021

Abstract

Photographic identity documents (ID) are widely used to prove the bearer's identity. In classic laboratory experiments, face images are mostly presented in isolation against a white background, while real life photo-ID checks normally compare a face embedded in a document with the holder. Researchers have begun to ask whether this additional document context might affect face matching. Recent research shows that in face matching tasks, embedding faces into passports introduces a response bias, such that viewers are more likely to accept two pictures as showing the same person. The experiments in this thesis examine the cause of this bias, and whether it generalises to other face processing tasks. In the first three experimental chapters (Chapter 2, 3 and 4), the bias is replicated using various identity documents (passports, driving licences, and student-ID) and in different face matching conditions (e.g. varying mismatch prevalence and task difficulty). Results show that the bias does not rely on perceived authority of the ID or the isolated processing of document elements. Instead, it seems to occur only in the presence of both card background and personal information, which converge in photo-ID. The document-induced bias is specific to unfamiliar faces, and occurs in face matching tasks where the documents themselves are task-irrelevant. Chapters 5 and 6 examine the locus of the document bias, testing both encoding and decisional processes. The effect of documents on memory tasks and first impressions is also examined. The results show that the document-bias seems primarily to affect decision-stage processes. The theoretical and practical implications of these findings are discussed.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Mike Burton, for offering me the opportunity to complete my PhD. Thank you for all your guidance, patience, inspiration, kindness and massive support. I have never thought I would be so lucky to have you lead me through my PhD. You always support me, encourage me to become better. You give me strength and confidence to face difficulties. Without your help, I have no idea what life would be like in those hard days.

Thank you to my TAP members and everyone in Face Lab for their insightful comments and suggestions, especially Professor Rob Jenkins, for your warmth, thoughtfulness, and all your greetings on our traditional days.

Thank you to Mila, for your assistance every time I was in trouble. You are such an angel in my life making me feel like family. Thank you to Gloria and Novia, for your kind help, support and friendship that have made my life in York a wonderful time.

Finally, I would like to thank all my friends and family, especially Yingnan, for your unconditional support, love and belief in me.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Experiment 2 and 3 from Chapter 2 and Experiment 5 from Chapter 3 have been published in Perception. Feng, X., & Burton, A. M. (2019). Identity documents bias face matching. *Perception, 48*(12), 1163-1174.

Experiment 6, 7, 9 and 10 from Chapter 3 have been published in Quarterly Journal of Experimental Psychology. Feng, X., & Burton, A. M. (2021). Understanding the document bias in face matching. *Quarterly Journal of Experimental Psychology*, 17470218211017902.

Chapter 1 - General Introduction

1.1 Introduction

Consider the following two scenarios: an international student has landed at the airport and waits to pass through the border to start his new college life; a terrorist pretending to be someone else tries to enter the country using a fake ID. Distinguishing between these two cases for the border control officers is extremely important for the safety of the country.

Photographic IDs are widely used to prove the identity of the bearer. We use our ID documents in daily life when purchasing alcohol or getting through airport security. It would bring us huge problems for personal and property safety if our ID documents were stolen and falsely used. However, this is a real threat, for example, a stolen passport was found on or near the body of a dead suicide bomber from the terror attacks on Paris in 2015 (Kingsley, 2015). Terrorists may try to mimic a person in documents in order to pass border control if they use a fake passport showing someone with similar appearance. Therefore, for all those involved in identity-checking, it is essential to identify fraudulent ID documents to protect the safety of the owners and the country. This implicates a key perceptual ability to correctly match the photographic-ID with the bearer.

When officers check passports or cashiers verify the information on ID cards, they are carrying out unfamiliar face matching. It is easy for us to recognize people we are familiar with, even in various viewpoints, lightings, etc (see Johnston & Edmonds, 2009). However, ID-checkers are mostly dealing with unfamiliar faces that they are trying to identify. Apart from the familiarity difficulty, photo cards like driving licences or passports are usually renewed every 10 years, which means there will probably be a time gap between the photo and the holder presenting the card. Sometimes, we think our own ID photos do not resemble ourselves, and we have large advantages over unfamiliar perceivers. Small changes like hairstyle, the lens, lightning or make-up can make two images of an unfamiliar face seem quite different. On the other hand, different people may seem similar to perceivers who are not familiar with them because of the same hairstyle or angles when taking the picture. These variations make matching unfamiliar faces a really hard task (Bruce, Henderson, Greenwood, Hancock, Burton, & Miller, 1999; Henderson, Bruce, & Burton, 2001; Megreya, Bindemann, Havard, & Burton, 2012; Megreya & Burton, 2006, 2008).

This error-prone nature of face matching raises great practical issues, which are not widely understood: people tend to believe that professional officers would do a better job on this than the general public (Ritchie, Smith, Jenkins, Bindemann, White, & Burton, 2015), because of the officers' experience of ID-checking. However, there are large individual differences in people's ability to match faces, even within professionals. Surprisingly, one's ability in matching faces is not improved by the length of one's professional experience (Papesh, 2018; White, Kemp, Jenkins, Matheson, & Burton, 2014; Wirth & Carbon, 2017). This is an issue that needs to be addressed. If officers or facial examiners show similar error rates to normal people, there are clear implications for security.

A further difficulty for face matching is that faces are not always encountered in ideal conditions typical of lab studies (i.e., a cropped face image against a white background) in real-life conditions. We meet people on the road and see faces in posters, social media, and ID cards. So, will there be any differences between processing faces in context and processing isolated faces presented in laboratories (see Figure 1.1)? To establish the characteristics of real face processing, it is important to examine the processing of faces across different contexts, and especially in documents. The investigation of complex contexts where faces are encountered in real life may help us build a better understanding of face processing, and expand and enrich our theoretical framework.

Figure 1.1 Examples of faces in different contexts: (a) plain faces used in Bruce et al. (1999)'s experiments, (b) a poster of Sherlock TV Series, (c) an example of a student-ID card from Global Edulink

This chapter gives an overview of current literature on the perception of faces in documents. I will start by introducing general face processing, including face memory, face matching and first impressions of faces. Then I will review processing faces within specific contexts, especially in identity documents. Finally, I will present the wider use of face processing in applied conditions and with professional viewers.

1.2 Face perception

We may ask many questions when encountering faces: Do I know this person? How can I interact with him? Is he trustworthy? These questions reflect how people learn, perceive, recognize and evaluate faces, which is of great importance and has attracted attention for decades. Faces are special visual stimuli that are different from ordinary objects. They convey a great deal of information, both identity information and social information, which can be detected and interpreted even with a simple glance.

A large amount of face perception research relies on standard and highly controlled images, which aim to control low-level factors to make higher-level processes more detectable. This is an efficient way to directly explore how the human visual system processes the information in faces. However, this may not reflect the way we deal with faces in real visual environments (Bindemann & Hole, 2020). Because of this, more recent studies have embraced the surroundings of faces. For example, ambient images (Jenkins, White, Van Montfort, & Burton, 2011), which are drawn from the surrounding environment, raise the importance of context when faces are processed. The context may provide more and help us understand how the human visual system extracts and integrates multiple sources of information from faces and the environment.

In this section, I will review three key aspects of face processing: remembering faces, matching faces, and evaluating faces. Before that, I will introduce familiarity first, which plays an important role when it comes to face identity.

Familiarity

In general, we are considered as experts at perceiving and recognizing faces (Young & Burton, 2018). We can easily recognize our friends and family members in non-optimal conditions, such as at night, from a long-distance or an old photograph. However, the incredible ability to recognize faces only applies to familiar faces. The recognition of unfamiliar faces is far from perfect and we may make mistakes all the time.

Unlike familiar faces, those faces we are not familiar with are quite poorly recognized (e.g., Bruce, 1982; Ellis, Shepherd, & Davis, 1979; Klatzky & Forrest, 1984). In one key study, people could almost perfectly recognize familiar faces that they had previously seen in poor-quality CCTV video clips, but they could remember and recognize unfamiliar faces barely better than chance (Burton, Wilson, Cowan, & Bruce, 1999).

Without memory load, the difference between familiar and unfamiliar faces is also found in face matching tasks (e.g., Ritchie et al., 2015; Young, Hay, McWeeny, Flude, & Ellis, 1985). Ritchie et al. (2015) asked participants to do a simple same/different identity judgement on pairs of familiar faces (celebrities) and unfamiliar faces. Compared to familiar faces (overall accuracy 85%), the performance on unfamiliar faces (overall accuracy 70%) was much lower.

Poor performance was also found in unfamiliar faces with video clips or multiple images (Bruce, Henderson, Newman, & Burton, 2001). Participants were told to match either a video clip, three still images, or a single still image to a colour photograph presenting either the same or a different person. Half of the participants were familiar with all these faces while another half were totally unfamiliar. Results showed that personally familiar faces were more highly recognized than unfamiliar faces.

The large body of evidence for the difference between familiar and unfamiliar faces suggests that there may be different processes underlying the perception of familiar and unfamiliar faces. For example, the internal features of a face (eyes, nose, and mouth area) seem to be more important than the external features for recognizing familiar faces (face shape, hair; Ellis et al., 1979). Clutterbuck and Johnston (2002) asked participants to match an intact image of unfamiliar, moderately familiar, or highly familiar faces with internal (eyes, nose, mouth) or external features (hairline, chin, ears). They found that only for internal feature matching, familiar faces were faster matched than moderately familiar faces and unfamiliar faces. More errors were found when participants matched unfamiliar faces than familiar faces.

The processing of familiar and unfamiliar faces may be independent of one another. It has been suggested that unfamiliar face processing is more like image-bound processing (Hancock, Bruce, & Burton, 2000). Recognition for unfamiliar faces is highly image-specific and quite vulnerable to changes like viewpoint, lighting, expression, or time (see Johnston & Edmonds, 2009). Even small changes can make two images of the same person look dissimilar for unfamiliar perceivers. Jenkins and colleagues (2011) gave participants 40 unfamiliar face images, with a realistic range of variability, to sort into different identity piles. Participants were not aware that these images actually depicted only two individuals. Results showed the median number of identities participants perceived was 7.5 different identities, at a range of 3 to 16, much more than 2 identities. When the same image set was shown to observers who were familiar with the two individuals, almost all of them performing perfectly (Median 2; Range 2–5). This comparison indicated unfamiliar perceivers made more errors and sorted wrongly with face images in high variability. The large variation in photos depicting the same identity seems to suggest that there are multiple different people present, in the view of unfamiliar perceivers.

Megreya and Burton (2006) set out to establish which cognitive processes correlate with unfamiliar face processing. They asked participants to complete several cognitive tasks (e.g., perceptual speed, visual short-term memory, figures matching) as well as familiar face processing tasks. Results showed a correlation between unfamiliar face processing and figure matching, but unfamiliar face processing did not correlate with familiar face processing. This finding shows unfamiliar faces seem to be treated as images or visual patterns like objects rather than an expertise-domain, which may result in the independence of the processsing of familiar and unfamiliar faces.

There is further evidence shown by neuropsychological studies. Early research on patients also shows two components, familiar and unfamiliar face perception (e.g., Benton, 1980). Warrington and James (1967) reported that there is a dissociation between disorders of familiar and unfamiliar face recognition. Studies on prosopagnosia found some patients could match unfamiliar faces while had difficulty recognizing familiar faces (e.g., Malone, Morris, Kay, & Levin, 1982), also some patients preserved the ability to recognize familiar faces while showed deficits in matching unfamiliar faces (e.g., Bauer, 1984).

Thus, familiarity is an essential factor when we are considering face recognition because familiar and unfamiliar faces are processed in different ways.

Face memory

How people remember and recognize faces has been a popular research topic for decades. One important indicator of the ability to recognize faces is face memory. It has been widely studied to demonstrate how well we identify a face. A traditional paradigm for testing face memory (old/new task) consists of two phases: a learning phase and a test phase (e.g., Lamont, Stewart-Williams, & Podd, 2005; Wilson, See, Bernstein, Hugenberg, & Chartier, 2014). In the learning phase, participants will try to remember a set of faces. Then, they will be shown more faces in a sequence to pick out the learned faces (i.e. "old" faces) and the new faces. Corresponding tests have also been built to discriminate face recognition abilities among people. For example, the Cambridge Face Memory Test (CFMT, Duchaine & Nakayama, 2006) is widely used to assess impairments in face processing (Bobak, Parris, Gregory, Bennetts, & Bate, 2017), as well as selecting the super-recognisers (Russell, Duchaine, & Nakayama, 2009) or classifying prosopagnosia (Bowles et al., 2009). In the CFMT test, participants need to memorize 6 model faces in three recognition stages: recognition of the same image learned, recognition of a new image of the learned identity, recognition of a new image of the learned identity with noise. This reduces the memory load from the traditional paradigm but introduces more variability.

Face memory research also has shown great implications for eyewitness testimony, which is often considered to be highly credible in court (Fitzgerald, Price, & Valentine, 2018; Wells, 1993). It usually requires witnesses to select the person they probably have seen in an earlier setting, such as in a robbery scene. This testimony is undoubtedly important, but it is also extremely error-prone, in both field (e.g., Vredeveldt, Charman, den Blanken, & Hooydonk, 2018) and laboratory studies (e.g., Palmer, Brewer, & Weber, 2010; Pryke, Lindsay, Dysart, & Dupuis, 2004). It has been suggested that the problem of eyewitness identification appears not only because of the difficulty of an instant memory for encountered unfamiliar faces, but also involves difficulties in the initial encoding of these faces (e.g., Megreya & Burton, 2008). For example, Smith, Wilford, Quigley-McBride, and Wells (2019) showed an impairment in the encoding phase weakened final line-up recognition. Participants were presented with a culprit-present video either in high-resolution or low-resolution (impaired condition). The video was a scene of people waiting in line for a check-in at an airport. After viewing the video, they were shown a line-up with 6 people, and asked to spot the person in the video who switched a bag. Participants were not aware of the purpose of the task before watching the video. The degraded video decreased the performance of identifying the target (78.3% correct to 29% correct) in the lineup that followed, along with a large increase in 'identifying' innocent people (from 1.7% to 33.8%).

This difficulty reflects the error-prone nature of face memory. Face recognition memory is easily affected by the intrinsic facial characteristics (i.e., the face itself) such as distinctiveness (Newell, Chiroro, & Valentine, 1999; Shepherd, Gibling, & Ellis, 1991), race (Meissner & Brigham, 2001; Zhao, Hayward, & Bülthoff, 2014), and social traits (Mattarozzi, Todorov, & Codispoti, 2015). It also can be influenced by extrinsic factors, such as encoding time (Bornstein, Deffenbacher, Penrod, & McGorty, 2012; Reynolds, & Pezdek, 1992), background (Rainis, 2001), and verbal overshadowing (Lloyd-Jones & Brown, 2008; Macrae & Lewis, 2002). For example, MacLin, MacLin, and Malpass (2001) recruited Hispanic participants and asked them to do a face recognition task on both Hispanic and Black faces. Results showed that participants recognized Hispanic faces better than Black faces, and the performance for both race faces decreased when the encoding time shortened from 5 seconds to 0.5 seconds.

Face matching

Face matching is a widely used method to assess the ability of face perception, without memory. If memorizing faces is hard for unfamiliar viewers, it seems to be easier when the memory load is removed. However, it has been widely demonstrated that matching faces can be difficult (Henderson et al., 2001; Kramer, Mohamed, &

Hardy, 2019; Megreya, Sandford, & Burton, 2013; White, Kemp, Jenkins, & Burton, 2014). Bruce et al. (1999) asked participants to match a target face to an array of 10 faces, one of which might be the target (see Figure 1.2). With no time pressure, the error rate was up to 30% in both target-present and target-absent trials. The poor performance persists in 1 to 6 matching arrays (e.g., Rubínová, Fitzgerald, Juncu, Ribbers, Hope, & Sauer, 2020), even in 1 to 1 pairwise comparisons (e.g., Kramer et al., 2019; White, Burton, Jenkins, & Kemp, 2014).

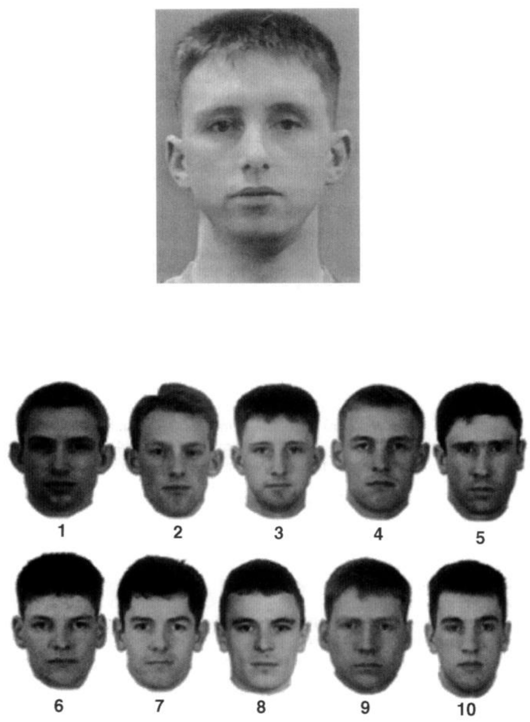

Figure 1.2 Example of a 1-to-10 matching array from the study of Bruce et al. (1999). The person on the top is the target, which may or may not be one of the 10 numbered faces below.

To assess people's ability of matching faces, there are several matching tasks developed. A widely used unfamiliar face matching task is the Glasgow Face Matching Test (GFMT; Burton, White, & McNeill, 2010). In this task, participants see several pairs of faces. They are asked to decide if the two face images are of the same person or different people. The stimuli are all frontal-oriented faces with a neutral expression, in high-quality resolution (see Figure 1.3). Despite providing optimized viewing conditions, the mean performance is about 81.3%, with a range of 51%-100%.

Figure 1.3 Example test items from the Glasgow Face Matching Test (Burton et al., 2010). (A) Mismatching pair. (B) Matching pair.

While in real-life matching, the two matching face pairs are not always in perfect views. A newer task, the Kent Face Matching Test (KFMT; Fysh & Bindemann, 2018a), aims to provide a more applied aspect of face matching. The two matching faces are a personal photo uploaded by the student-volunteer and a high-quality portrait, which are taken three months apart (see Figure 1.4). It provides more variability in the matching task and the overall mean performance is 66%, poorer than GFMT (80%). The KFMT provides more challenging conditions, and more closely matches settings like passport control. However, the test still represents more optimal conditions than real-life checking, where photos may be taken years apart and watermarks may obscure the faces. It reveals the difficulty when people match photos of unfamiliar faces.

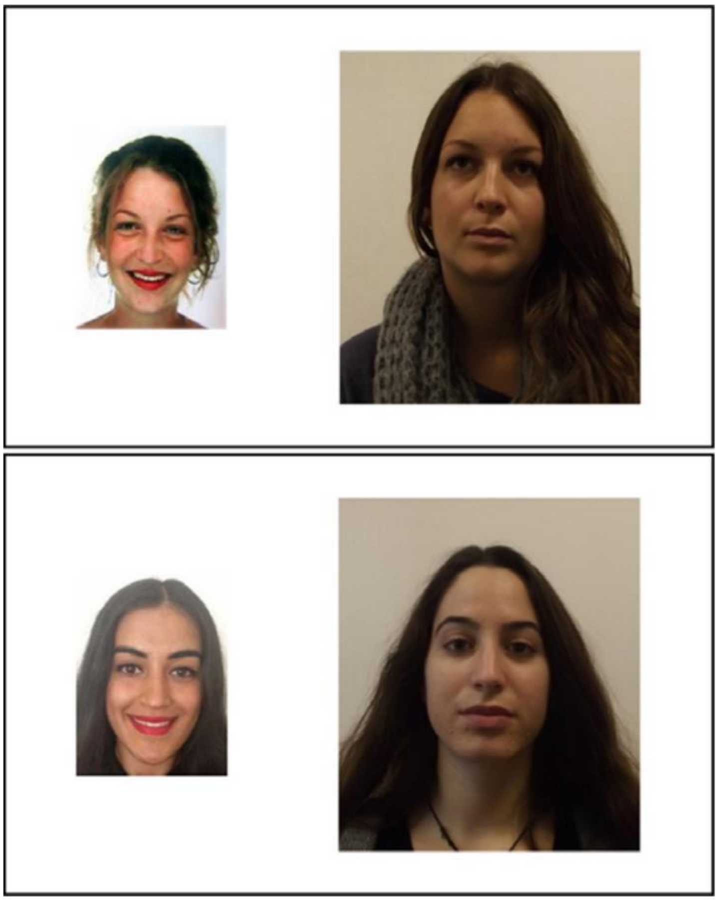

Figure 1.4 Example match (top row) and mismatch (bottom row) pairs from the Kent Face Matching Test (Fysh & Bindemann, 2018a).

Unfamiliar face matching is considered to be highly image-bound (Hancock et al., 2000), because even low-level image differences (e.g., lighting, image quality, etc.) affects performances. We will expand this later in the chapter. Due to the image variability and within-person variability (e.g., expressions, viewpoints), the same person can look quite different, and different people can look incredibly similar (Jenkins et al., 2011). This increases the difficulty of matching identities. In a pairwise matching task, people may fail to realize that two photos depict the same person (a failure of same detection), or they may think that two photos depict the same person when they actually do not (a failure of difference detection).

First impressions

Apart from the identity that we can extract from faces, there is other information, such as age, gender, emotion, and social attributes. Without judging who the person is, people can form stable first impressions within a very short time (Bar, Neta, & Linz, 2006; Todorov, Loehr, & Oosterhof, 2010; Willis & Todorov, 2006), for example, whether the person is trustworthy or not. These first impression judgements show high agreement between different perceivers (Willis & Todorov, 2006; Cogsdill, Todorov, Spelke, & Banaji, 2014).

Although first impressions show high agreement, this does not mean the impressions are correct (Hassin & Trope, 2000; Todorov, 2008). There are no right or wrong responses for social evaluations, because the impressions for one single identity vary a lot using different photos of the same person (e.g., Mileva, Young, Kramer, Burton, 2019; White, Sutherland, & Burton, 2017). Todorov and Porter (2014) showed viewers five front-facing images of 20 individuals and asked them to rate eight traits (attractiveness, competence, etc.) for each image (e.g., "How attractive is this person?") on a 9-point scale. Results showed that different images of the same individual can lead to different impressions.

How were the main traits of first impressions chosen? Oosterhof and Todorov (2008) identified two dimensions underlying first impressions: trustworthiness and dominance. They first asked participants to write unconstrained descriptions of a set of faces, then these descriptions were sorted into trait dimensions. Another group of participants then were asked to rate each face on these traits, like aggressiveness, caring, sociability, etc. After collection of the ratings, they submitted these judgments to a principal components analysis (PCA). The first principal component (PC) accounted for 63.3% of the variance and the second PC accounted for 18.3% of the variance of the mean trait judgments. The judgment of trustworthiness was closest in space to the first PC, and the judgment of dominance was closest to the second PC (see Figure 1.5), so, trustworthiness and dominance could be regarded as the fundamental dimensions of social face evaluation. This structure is also consistent with other models of social perception, like warmth and competence of social cognition (Fiske, Cuddy, & Glick, 2007).

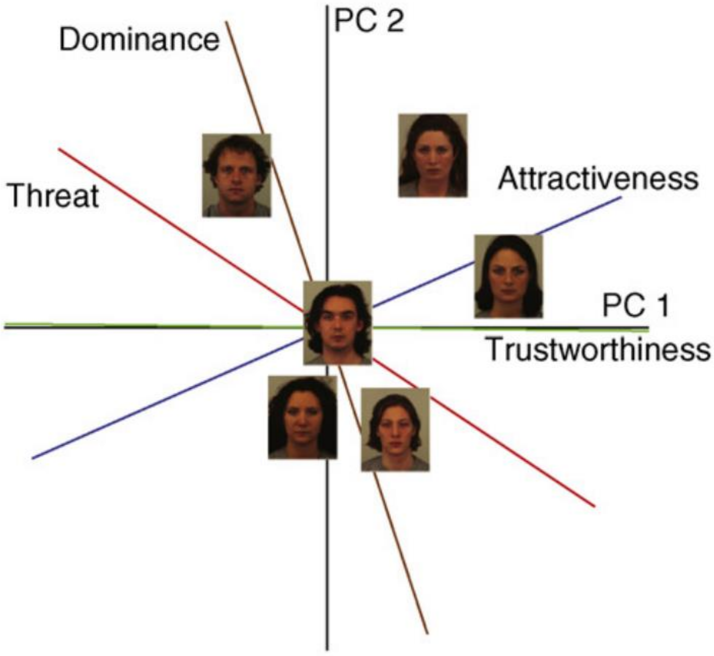

Figure 1.5 The solution of principal components (PC) analysis of face evaluation from Todorov, Said, Engell, and Oosterhof (2008). The PC 1 could be interpreted as valence/trustworthiness evaluation, and the PC 2 could be interpreted as power/dominance evaluation.

These two dimensions were constructed using face images that were tightly controlled, homogenous stimuli, such as neutral faces from the Karolinska directed emotional faces database (KDEF, Lundqvist, Flykt, & Öhman, 1998) and computer-generated faces (Oosterhof & Todorov, 2008). When more variable images are included, the model changes a bit, as "ambient images" (Jenkins et al., 2011) introduce more natural face variability. A third factor, "youthful-attractiveness" emerges (Sutherland, Oldmeadow, Santos, Towler, Burt, & Young, 2013). The new face evaluation model demonstrates real-world variability better, and these three traits (attractiveness, trustworthiness, and dominance) contribute a lot to social behaviour and decisions.

First impressions have consequences in the real world. Attractive people are more likely to be considered as having more positive characteristics, such as being more intelligent (Zebrowitz & Rhodes, 2004) and successful (Eagly, Ashmore, Makhijani, & Longo, 1991). This relates to a halo effect that "what is beautiful is good" (Dion, Berscheid, & Walster, 1972; Little, Burt, & Perrett, 2006). Being attractive, leads to people receiving more help than unattractive peers (see Patzer, 2012) and experiencing less punishment (Stewart, 1980).

Similarly, trustworthiness can affect court decisions (Blair, Judd, & Chapleau, 2004; Wilson & Rule, 2016). When viewers were presented with a severe crime (a murder) and a minor crime (a fraud or a theft), if the defendants' faces seem quite trustworthy, then in severe crimes, viewers required more evidence to give a guilty verdict; while for untrustworthy faces, viewers needed less evidence and were more confident to give guilty decisions (Porter, ten Brinke, & Gustaw, 2010). One account of these findings is the Dangerous Decisions Theory (DDT) (Porter & ten Brinke, 2009), suggesting that "the reading of defendant's face and emotional expressions play a major role in initiating a series of 'dangerous' decisions concerning his/her credibility". Indeed, trustworthiness, as well as dominance has a huge influence on our decisions in important domains. For example, dominance can predict voting decisions (Ballew & Todorov, 2007); more dominant people receive larger salaries (Rule & Ambady, 2008).

1.3. Factors affecting face matching

As we discussed above, unfamiliar face matching is a hard but widely used task in real life (Bruce et al., 1999; Henderson et al., 2001; Megreya et al., 2012). Because of its image-bound nature, even a small change can affect the recognition of an unfamiliar face. The difficulty of face matching emerges with both the properties of images and the face itself (e.g., expressions) (Burton, 2013). For example, early research on whether expressions affected face memory showed that if the viewpoint or the expression of unfamiliar faces changed between the learning and test phase, participants performed only 61% correctly compared to 90% for unchanged images (Bruce, 1982). The same results were replicated in a matching task conducted by Bruce et al. (1999), i.e., performance was best when the viewpoint and expression matched, and reduced with viewpoint changes.

When it comes to applied conditions, most of the photos in identity documents such as passports and driving licences have standard requirements. For example, one cannot wear sunglasses, hats, face masks, or other accessories that may cover parts of the face. The regulations do help to present a clear frontal view of your face. However, some of the photos on documents are not in high quality, or have a long time gap between the day that photos were taken and the present, making matching difficult. For example, driving licences are typically only renewed every ten years. These factors can all contribute to an even harder situation for matching an unfamiliar person to his/her photo. In this section, we will address the factors that will be experienced in applied settings.

First, the resolution or quality of images matters. Bindemann, Attard, Leach, and Johnston (2013) compared matching a high-resolution photograph to a pixelated image with matching two high-resolution photographs. Results showed when participants saw two high-resolution photos, the accuracy was nearly 90% but significantly lower when one of the match photos was pixelated (less than 70%). This demonstrated that the image quality had a strong effect on unfamiliar face matching (Kramer et al., 2019).

The time gap between two matching images could also result in large errors (Alenezi & Bindemann, 2013; Alenezi, Bindemann, Fysh, & Johnston, 2015). When taken only minutes apart, with the same lighting, the overall performance of matching two face images is around 80% (the GFMT stimuli, Burton et al., 2010), while the accuracy reduces when two images were taken with a larger time gap: days, months or even years apart (Bahrick, Bahrick, & Wittlinger, 1975; Bruck, Cavanagh, & Ceci, 1991; Davis & Valentine, 2009; Fysh & Bindemann, 2018a). Megreya and colleagues (2013) tested participants with images of unfamiliar faces taken months apart and those taken on the same day. Compared to those face images taken on the same day, more errors were found (58.6% accurate when taken months apart and 79% when taken on the same day). The decrease was similar in a 1-in-10 matching task or an easier pairwise face matching task.

The proportion of mismatch pairs is also influential. In the real world, matches are thought to be much more prevalent than mismatches. For example, the officers at border control may not suppose many of the passports they check are fraudulent ones. Actually, the different prevalence of mismatch photos does affect people's matching performance. Papesh and Goldinger (2014) asked participants to match a face to a face enclosed in a driving licence in either a low-prevalence condition (10% trials were mismatches) or a high-prevalence condition (50% trials were mismatches). Results showed a higher miss rate on mismatches when the prevalence of mismatches was low. This effect has been demonstrated as a low-prevalence effect (LPE) in visual search tasks so that people tend to miss a target if the target is infrequently presented (Godwin, Menneer, Cave, Thaibsyah, & Donnelly, 2015; Wolfe, Horowitz, & Kenner, 2005). However, some show the opposite results. Bindemann, Avetisyan, and Blackwell (2010) compared a high mismatch prevalence condition (50% match, 50% mismatch) with a low mismatch prevalence one (98% match, 2% mismatch) in a face matching task. They found viewers showed better detection of mismatch photos in the low prevalence condition than in the high prevalence one, but had more errors with match trials in the low prevalence condition. Despite the inconsistency, it is clear that the frequency of mismatch can affect performance (Growns & Kukucka, 2021).

The last factor I would like to stress is the time pressure (Bindemann, Fysh, Cross, & Watts, 2016; Fysh & Bindemann, 2017; Özbek & Bindemann, 2011). Laboratory matching tasks usually use unrestricted presentation time (e.g., GFMT, KFMT), while in real-world settings, face matching often occurs under time pressure, such as passport-matching when border control officers have to work under a time limit. Wirth and Carbon (2017) assigned participants into three groups: the control group with no time limit; the local time-limit group where each face was presented for 5.81 seconds and needed response; the global time-limit group where all faces were needed to be checked within 25 minutes. Results showed that for sensitivity, it decreased significantly from the control group to the global time-limit group, while the local time-limit group lay in between of these groups. Specifically, the time pressure affected mismatches significantly, but did not affect performance on matching pairs.

There are other factors affecting face matching performance, such as disguises (sunglasses, hats, etc) on the face (Graham & Ritchie, 2019; Kramer & Ritchie, 2016), image sizes (Lee, Matsumiya, & Wilson, 2006), image colours (Bobak, Mileva, & Hancock, 2019), etc. As discussed above, the variability of faces increases the difficulty of matching faces, and because of this, it is important to consider face matching within an applied context.

1.4 Face processing in document contexts

Context affects the perception of words (e.g., Brierley, Medford, Shaw, & David, 2007; Liu, Hu, & Peng, 2012), objects (see Bar, 2004 for a review), as well as faces. Most of the research work we have reviewed above uses plain, single, static face photos, while in real-life conditions, faces appear in diverse forms. For example, we meet real people on the street, at workplaces; we see dynamic faces in the movies; we see posters and selfies on social media, etc. These are not plain faces at all. Research using dynamic faces or ambient photos has shown different results compared with plain faces (e.g., Rubenstein, 2005; Sutherland et al., 2013). Our interest here is about faces in documents, especially identity cards, such as passports, or other documents which include identity information and a frame background (e.g., a rectangle colour frame of a UK driving licence, see Figure 1.6). In this section, we are going to review some research using the ID-card context when considering face processing. Because normally an ID document contains identity information and card background, we include these in the following sections.

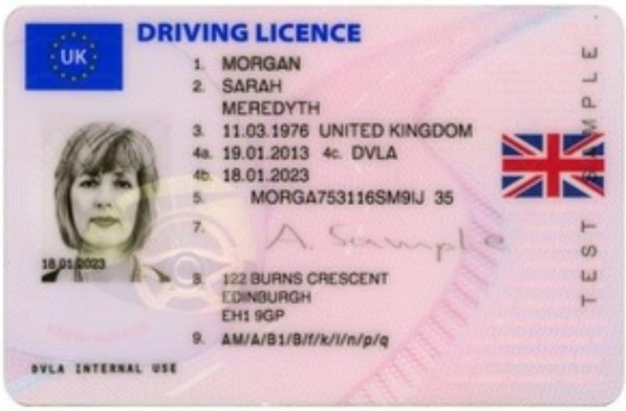

Figure 1.6 A sample of the UK driving licence provided (www.gov.uk).

Colour background and scenes

The perception of a face seems to be easily affected by a local visual context. For example, the perceived trait associated with a face can be influenced by a context as simple as a geometric shape. Toet and Tak (2013) asked participants to rate the perceived dominance of several neutral faces that were presented overlaid on downward- or upward-pointing triangles (see Figure 1.7). As downward-pointing triangles have been found to convey threat and negative valance (Watson, Blagrove, Evans, & Moore, 2012), results showed the neutral faces were rated as more dominant on downward-pointing triangle background compared to upward-pointing ones.

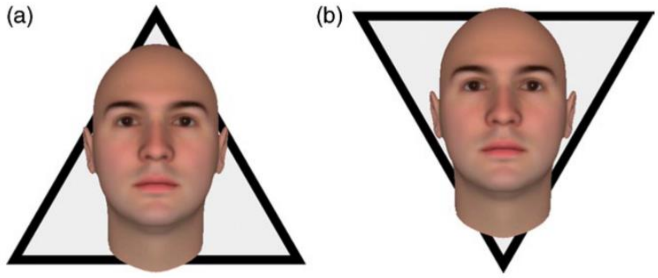

Figure 1.7 Neutral face with an (a) upward-pointing and (b) downward-pointing triangle in the background (Toet & Tak, 2013)

Koji and Fernandes (2010) also provided positive, negative, and neutral scenes as background, presented in colour. The neutral faces on top of each scene were presented in black and white. Participants needed to rate the faces which were shown on a scene on a scale of -3 (negative) to 3 (positive). They found that faces were rated as more positive when positioned in a positive context than in a neutral or negative context; faces presented in a negative context were rated more negative compared to neutral or positive context. This showed people were biased by the scene in the same direction when they rated the valence of the face, even though the scene was task-irrelevant.

There are other studies on the effect of scenes and background related to facial emotions: the valence of the background modulates face processing (e.g., Barrett, Mesquita, & Gendron, 2011; Rainis, 2001; Van den Stock & de Gelder, 2014; Van den Stock, Vandenbulcke, Sinke, Goebel, & de Gelder, 2014;). In addition to these studies using affective backgrounds, Deffler, Brown, and Marsh (2015) showed participants famous places (landmarks), novel places, or neutral (colour) background (see Figure 1.8) paired with novel faces and asked them to rate the familiarity of each face on a scale of 1 (very unfamiliar) to 6 (very familiar). Although the faces were all novel to participants, they still responded that faces were more familiar on the landmark backgrounds compared with novel places and colour backgrounds.

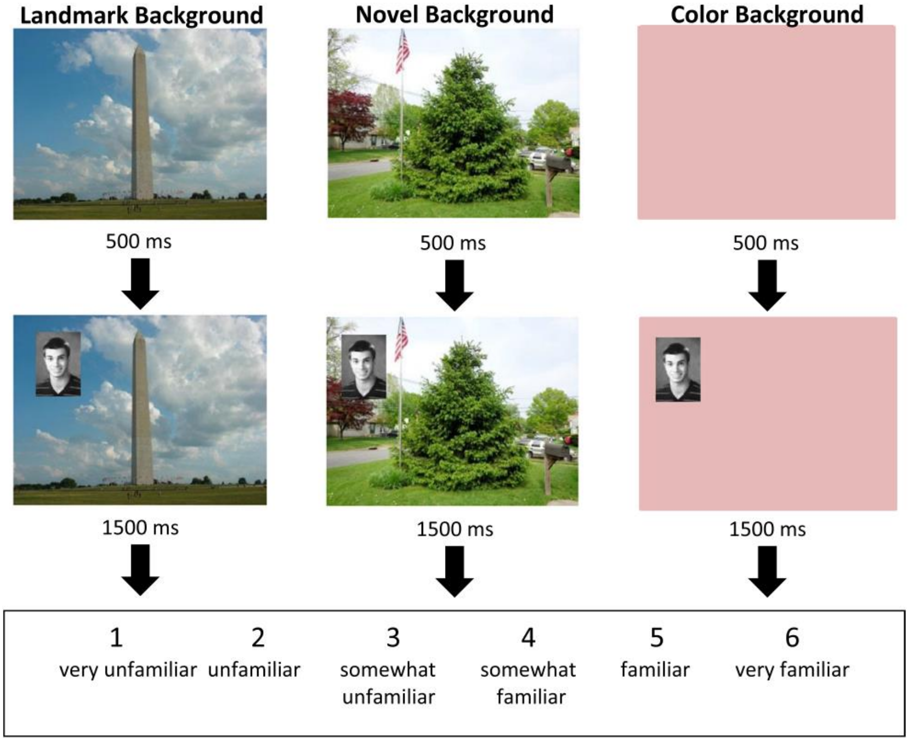

Figure 1.8 Presentations of a novel face with different backgrounds from Deffler et al. (2015)'s study.

Pure text and words

Without scenes or images (Shriver, Young, Hugenberg, Bernstein, & Lanter, 2008), presenting pure text or words with faces can also affect face processing. These studies are normally related to emotions (e.g., Falvello, Vinson, Ferrari, & Todorov, 2015; Mattarozzi, Colonnello, Russo, & Todorov, 2019; McCrackin, Lee, Itier, & Fernandes, 2021), social categorizations (e.g., Bernstein, Young, & Hugenberg, 2007; Bhardwaj & Hole, 2020; Fuller, Majolo, Flack, & Ritchie, 2021), and social strategies (e.g., mate-choice copying, Eva & Wood, 2006).

Mattarozzi and colleagues (2019) presented viewers with several faces to remember, each face was presented either alone or with a behavioural description (positive, neutral, or negative). They found that when the faces were studied with behavioural descriptions, recognition memory was better than for faces learned without context; more importantly, if the descriptions were salient behaviours (positive "He volunteered to stay late to help a co-worker" or negative "She insulted a stranger by making a racial slur"), recognition performance was better than for faces with either no context or neutral descriptions – and this difference persisted even after a week delay. Note that the neutral descriptions also affected the performance in some way. There is similar research finding that person-related information affects face recognition (e.g., Klatzky, Martin, & Kane, 1982). Kerr and Winograd (1982) required participants to memorize faces with zero, one, two, or three descriptive phrases, such as "he's a vegetarian", "he smokes cigars". Results showed that participants recognized faces learned with information better than learned faces alone. Also, words related to groups can affect the recognition of faces, such as the own-group bias in face recognition: faces that were labelled as in-group are better recognized than those labelled as out-group (Bernstein et al., 2007).

Another series of studies showed that the perception of faces and words interacts with each other. Stenberg, Wiking, and Dahl (1998) showed participants several emotional faces (angry, happy, and neutral). On each face, a word displayed in a grey rectangle was positioned in the center of the face, lower part of the nose. Participants' task was to only classify each word as positive or negative, by ignoring the faces behind. Results showed that happy faces speeded the processing of positive words, but slowed that of negative ones. On the other way round, when the words (happy, angry, and blank) were served as distractors, the determination of face expressions (happy or angry) was also be influenced by the congruency of valence (Anes & Kruer, 2004).

Identity documents

The majority of laboratory tasks use isolated face images, but in real-life settings, it is more common to match a person to a contextual document such as photo-IDs or passports. Both in security-critical settings like officers checking in border control and in daily commerce, such as age-verification for purchasing alcohol, unfamiliar face matching in documents plays an important role.

It is widely believed that the inclusion of photographs on ID cards may reduce fraudulent use. In order to test the utility of photo cards, Kemp, Towell, and Pike (1997) tested six experienced supermarket cashiers in a real supermarket. Shoppers showed the cashiers their purchase cards with photographs, then the cashiers decided whether the card matched the carrier. The overall accuracy was 67.4%, and the cashiers falsely accepted more than 50% of the fraudulent cards, even though they were aware that at least some of the cards would not match the carrier. Bindemann and Sandford (2011) showed participants a photo-ID card image, and asked them to select one target from 30 single faces to match the person shown in the photo-ID. There were three cards in total. They found only 38% of participants matched correctly to all provided photo-ID images, and none of the individual IDs produced high recognition accuracy (67% in ID1, 46% in ID2, 58% in ID3). There are also studies using driving licences (Papesh, 2018; Robertson & Burton, 2021), passports (Meissner, Susa, & Ross, 2013; Robertson, Kramer, & Burton, 2017; Wirth & Carbon, 2017), posters (Mileva & Burton, 2019; O'Brien & Thorley, 2020) and newspapers (Galli, Feurra, & Viggiano, 2006; Mattarozzi et al., 2015) as stimuli.

These findings suggest that matching faces with photographs embedded in documents may not be as easy as we thought. However, none of these studies directly compared matching isolated faces with matching one of the faces embedded in a document. McCaffery and Burton (2016) did compare matching isolated face pairs and faces embedded in passport frames. They found that if one of the faces was in a passport frame, participants made more “same” responses to that face pair. In other words, the accuracy went up for passport matching on match trials, but went down on mismatch trials. They also asked participants to check the biographical information such as gender and date of birth when the passport was presented. Results showed that the bias in face matching was not affected by the validity of information, but participants were more likely to make errors on information checking if the faces did not match.

Robertson and Burton (2021) set a scene of selling alcohol to customers. In a critical trial, participants saw a face image and a face embedded in a driving licence (see Figure 1.9). They should only "sell" alcohol when the age shown in the driving licence was above 18 and when the two faces matched. Results showed large errors (50%) when age requirements were met but faces did not match. Similar errors were found even using the PASS+ card ("Proof of Age Standards Scheme" set up in the UK) instead of driving licence, where no age calculation is needed.

Figure 1.9 An example trial in the study of Robertson and Burton (2021), in which participants were required to match the faces as well as verifying the age shown on the driving licence.

The interaction between data checking and face matching tasks indicates the practical difficulty of processing faces within the document contexts, which also suggests that the processing of matching faces in context seems to be different from pure isolated face matching and needs our attention.

1.5 Applied settings

Faces are important for identification. As reviewed above, face processing and recognition are not as easy as generally thought (e.g., Ritchie et al., 2015). This poses a significant problem in applied settings that depends on accurate face matching or recognition to verify the identity of unfamiliar people. From daily scenarios of checking ID cards to more important ones such as eyewitness testimony, looking for missing people, and searching for criminals in surveillance video, applied issues in face recognition continue to be critical in society.

The poor performance of face matching is normally observed on plain faces, mostly using face photos. People may believe it should be easier to decide whether a photograph matches a live person compared with two photos (Ritchie, Mireku, & Kramer, 2020), which is the "live superiority hypothesis" in eyewitness testimony (Fitzgerald et al., 2018). This hypothesis suggests that live presentation of police line-ups yields the best eyewitness identification outcomes than using image line-ups. However, results are quite mixed (see Price, Harvey, Anderson, Chadwick, & Fitzgerald, 2019; Rubínová et al., 2020). Even reducing the number of line-ups into one, i.e., a pairwise matching, matching a photograph to a live person standing in front of you remains poor (e.g., Davis & Valentine, 2009; Kemp et al., 1997; Megreya & Burton, 2008; Ritchie et al., 2020).

Visual search

Because of the poor performance of eyewitnesses (e.g., Pryke et a.., 2004; Rubínová et al., 2020), closed-circuit television (CCTV) surveillance is widely used in public spaces. Unlike the matching contexts described before, searching for people in crowds represents a less controlled and more complex scenario (Kramer, Hardy, & Ritchie, 2020). Multiple distractors and an uncontrolled environment make visual search a hard task, such as looking for missing persons (O'Brien & Thorley, 2020) or criminals at the airport.

Davis, Forrest, Treml, and Jansari (2018) created the Spot the Face in a Crowd Test (SFCT). They recorded an 18-minute video and split it into 11 clips of 21-second videos. In these videos, actors and bystanders were walking through the field of the camera view. Participants were instructed to review the video footage and searched for two, four, or eight “missing people”, the target actors. For each target, participants were provided four images with high variability. As the number of targets that they needed to find increased, the overall performance decreased. Fortunately, police experts performed better than novice participants.

Mileva and Burton (2019) used real CCTV footage depicting a large city transport hub, and asked viewers to search for target individuals. The targets were either presented using one, three, or sixteen images. Results showed that participants who saw three images had better performance than participants who saw a single image, but more than three images did not help (There was no difference between seeing three or sixteen images). They also showed that using video clips of the targets did not enhance searching performance compared with static images, but they did find the quality of stimuli (high quality vs standard quality) helpful.

In these studies, participants could pause, rewind, and replay the CCTV footage to familiarize and to double-check their choice. There was no time constraint at all. But in real-time searching, as you walk into a station to look for missing people, it is impossible to ask the crowd to walk back. This one-chance search resulted in large errors (only 39% accuracy, Kramer et al., 2020), which suggests the level of searching difficulty in real-life conditions.

Professional populations

There are huge individual differences among people when it comes to face recognition (Bruce, Bindemann, & Lander, 2018; McCaffery, Robertson, Young, &

Burton, 2018). For example, subjects' performance which ranged from 50%-100% accuracy was found in many studies (e.g., studies using GFMT, KFMT).

There are groups of professionals who are always doing face matching and recognition, such as border inspectors, officers in police stations. People naturally feel and hold the concept that professionals (especially passport officers) outperform naïve populations (Ritchie et al., 2015). But, is this the truth - do professionals perform well in recognition of faces?

Face matching results from expert observers, like passport officers, show that the performance is far from perfect, and often no better than untrained student participants (Towler et al., 2019; White, Kemp, Jenkins, Matheson, et al., 2014). White and colleagues used a live photo-to-person matching task with professional passport officers. The officers accepted 14% of fraudulent IDs, also 6% of correct IDs were rejected wrongly. This brings to mind Kemp et al's. (1997) finding that cashiers also performed more errors than expected. However, it should be noted that there are huge individual differences. Some officers did perform highly accurately, but there was no relationship between the length of time employed in a professional role and face identification accuracy (e.g., Weatherford, Roberson, & Erickson, 2021; White, Dunn, Schmid, & Kemp, 2015).

There is a smaller set of highly trained staff who are called facial examiners. They are taught to compare faces by detailed features and search for similarities and differences between facial images (White, Phillips, Hahn, Hill, & O'Toole, 2015). Facial examiners rated higher levels of feature usefulness and showed less face-inverted effect than students, which may be the reason that they were more accurate than normal people on face matching tasks (Towler, White, & Kemp, 2017).

Super recognizers (SR) are those who show excellent performance in unfamiliar face recognition regardless of experience. Russell and colleagues (2009) tested four SRs on two different face recognition tests, the Before They Were Famous (BTWF) test and the Cambridge Face Memory Test (CFMT, Duchaine & Nakayama, 2006). The BTWF test contains photos when famous people were in their childhood, and the CFMT involves photos in different views, lighting conditions, and photographs with noise. These SRs scored significantly better in all the tasks than the control group. Apart from face memory, SRs are exceptional at unfamiliar face matching, using either pairwise matching (Bobak, Dowsett, & Bate, 2016), 1-to-10 matching, or still-to-video matching (Bobak, Hancock, & Bate, 2016)

Despite the performance, expert populations are also vulnerable to cognitive biases (Dror, 2020; Kukucka, 2018), such as the confirmation bias (Kassin, Dror, & Kukucka, 2013) and the low-prevalence effect (Nakashima, Kobayashi, Maeda, Yoshikawa, & Yokosawa, 2013; Weatherford et al., 2021). Super recognizers also showed more response bias in matching tasks compared with typical people (Bobak, Dowsett, et al., 2016).

Comparing humans with computer algorithms

The importance of identity judgements has been reviewed many times (Towler, Kemp, & White, 2017). Apart from the situations we mentioned above, there are actually situations that face recognition is made by algorithms or computer systems. In applied settings, especially nowadays, the judgements of identities are done by a collaboration of algorithms and humans. For example, in the UK's airports, there are electronic passport gates installed. These gates employ algorithms that compare the holder's live face with the photo stored in his/her passport (Fysh & Bindemann, 2018b).

Unlike human responses of "same person" or "different people", the face recognition algorithms return a similarity score that serves to index the likelihood of "same person" or "different people" (Cavazos, Phillips, Castillo, & O'Toole, 2020). The higher a similarity score is, the more likely that the two images show the same individual (O'Toole, Abdi, Jiang, & Phillips, 2007). For example, the algorithm reported by White, Dunn, and colleagues (2015), took a target face into the system, searched a stored image database (1.6 million police mugshot images of adults), and returned the eight highest similar images in this database (see Figure 1.10).

Figure 1.10 An example of a one-to-eight face matching array (White, Dunn, et al., 2015). The target face was shown on the top, and the eight faces were selected by the proprietary face recognition software.

The best face recognition algorithms have now achieved perfect or near-perfect performance in benchmark tests (Phillips et al., 2009), and outperform novice participants in most challenging conditions (O’Toole, An, Dunlop, & Natu, 2012; Phillips et al., 2011). Although algorithms failed to perform better than humans in some cases (Phillips et al, 2018; Rice, Phillips, Natu, An, & O'Toole, 2013), they are developing with time (Cavazos et al., 2020).

When asking human also to rate similarities of face pairs, fusing algorithms and humans’ scores result in nearly perfect performance (O'Toole et al., 2007). It has also been found that single forensic facial examiners fused with the best algorithm were more accurate than the combination of two examiners (Phillips et al, 2018). Although algorithms and humans rely on different information and strategy to determine identity, the collaboration between humans and computers offers benefits to face identification accuracy in important applications.

1.6 Overview of the current work

The aim of this thesis is to investigate whether the perception of faces in documents differs from plain faces, as normally used in lab-based experiments. As reviewed previously, context affects face processing. Matching unfamiliar faces in a passport is as difficult as matching them isolated, but a bias for viewers to give more 'same person' responses has been reported (McCaffery & Burton, 2016). This bias could be very important practically, because it increases the risk that those ID checkers will accept more fraudulent documents. It is therefore important to investigate the underlying causes of this bias, as well as trying to establish its generality.

In the first experiment chapter (Chapter 2), we try to replicate the bias effect and test whether the effect found in passports exists for other kinds of documents. A passport is an official document with high authority as identity proof. The bias effect may be due to the authority that a passport carries, so we introduced various documents (driving licence, student-ID, and posters) with different authorities to test this possibility.

Chapter 3 focuses on investigating more of the factors of the bias effect, mostly on the visual properties of a document. We separate the information and background in an identity card to see if the bias is based on either of the two components. Then we manipulate the information and background to investigate more about the reason for the bias. In Chapter 4, we try to vary the faces and tasks, by taking the familiarity and difficulty of faces, the frequency, and the task demands into consideration.

The last two chapters (Chapter 5 and Chapter 6) are designed to test if the bias effect found in unfamiliar face matching generalises to face memory and first impressions, which is to say, if a document frame will affect other aspects of face processing.

Taken together, the experiments demonstrate that processing a face embedded in a document is different from processing an isolated face. In face matching, viewers will be biased by any readable card, but only on criterion, not overall accuracy. The bias effect exists for unfamiliar faces, and only when identity decisions are needed. As we will describe, the presence of a document affects identity decisions, but appears not to affect first impressions or recognition memory.

Chapter 2 – Unfamiliar face matching in various documents

2.1 Introduction

Photographic IDs are widely used in everyday life. People need to prove their identity by providing an acceptable ID document, such as passport, driving licence, or staff card to get through security checks or purchase restricted goods. In these situations, the ID-checking staff need to match an ID document to the holder, i.e. to do an unfamiliar face matching task.

It is surprising that people's performance on such a common task is highly error-prone (Bruce et al., 1999; Davis & Valentine, 2009; Megreya & Burton, 2006, 2008; Ritchie et al., 2015), even for professionals like cashiers and passport officers (Phillips et al., 2018; White, Kemp, Jenkins, Matheson, et al., 2014). The poor performance persists from 1-to-10 matching (e.g., Bruce et al., 1999) or 1-to-1 pairwise matching (e.g., Megreya & Burton, 2006), and in both photo-photo (e.g., Henderson et al.,2001; Susa, Gause, & Dessenberger, 2019) and photo-person (e.g., Kemp et al., 1997; Megreya & Burton, 2008; White, Kemp, Jenkins, Matheson, et al., 2014) matching tasks.

However, most people are unaware of such poor performance levels when matching unfamiliar faces (Zhou & Jenkins, 2020). They may usually underestimate the error rate because of their excellence at recognizing familiar faces, even in non-optimal conditions (e.g., Burton et al., 1999). It has also been found that the performance on face matching shows a graded decrease in accuracy as familiarity drops (Clutterbuck & Johnston, 2002). So, when it comes to unfamiliar faces, performance is far from perfect (Bruce et al., 2001; Ritchie et al., 2015).

Although there is extensive research on unfamiliar face matching, most studies use laboratory face images that are normally cropped to show a full face, without specific context. Reflecting real-life conditions, when we engage in a matching task, it usually happens with an ID document. So, instead of using isolated plain faces as stimuli in unfamiliar matching studies, it may be more practical for experiments to embed the faces in daily surroundings.

Some previous researchers have done matching tasks using cards rather than matching isolated images (e.g., Kemp et al., 1997; Kramer et al., 2019), but the error rate was similar as that found in isolated face matching tasks. For example, Meissner et al. (2013) showed participants a face image and a scanned passport simultaneously, and asked them to decide whether the two images showed the same person or different people. The error rate was about 17 - 28% across all three experiments. Similar results were found using staff/student photo-IDs in university (Bindemann & Sandford, 2011). To mimic real-life settings, Kemp and colleagues (1997) asked real cashiers to do photo-person matching, using a mocked-up purchase card with the bearer's photo on it. The results were no better than photo-photo matching by naïve participants.

These studies reported consistently poor performance when using documents. However, some recent studies have reported a clearer comparison between matching isolated faces and faces in documents. McCaffery and Burton (2016) directly compared matching of isolated face pairs with matching when one of the photos was embedded in a passport frame. Consistent with the studies above, they found the same overall error rate for matching faces that were isolated or in passport frames. However, by considering the match and mismatch trials separately they observed a difference in the bias between conditions. Across three experiments, McCaffery and Burton found a consistent effect that viewers would make more "same person" responses to the face pair if one of the faces was embedded in a passport frame. This means viewers would accept 'fraudulent' matches more often when the faces are presented in passports than isolated. The results they reported raised a serious problem that the direction of bias is forensically disadvantageous – faces are thought more likely to depict the same person when one of the images appears within a passport, so fraudulent use of a passport becomes harder to detect, and more fraudulent passports might be accepted.

McCaffery and Burton (2016) also asked participants to check the biographical information on the passport while doing the face matching task. The information was designed to be valid or invalid associated with the face. For example, using nouns as forename (e.g. ‘Fork’), wrong-gender forename, or presenting a wrong date of birth (e.g., 30 February) were all invalid passports. They found an interference in this dual-task processing, but the tendency of accepting the passport existed in both valid and invalid passport conditions. McCaffery and Burton suggested that this bias effect seems to be due to the pure presence of the frame, which affected the viewer’s perception of identity, whether the frame was task-irrelevant or not.

It seems that the ‘same person’ bias may be related to the ‘forensic confirmation bias’, in which the forensic context can affect image comparison or memory (Kassin et al, 2013). The bias may also be based on prior experience: passports carry high authority, because they are issued by the government and are regarded as powerful proof of identity. It is hard to make a fraudulent one, and the consequences of detection would be severe. It is possible that people hold this idea and therefore show a tendency to accept the holder as genuine because of this authority property.

As McCaffery and Burton (2016) showed, the presence of a passport frame affected viewers’ decisions. It is possible that the high authority property of a passport makes people respond ‘same person’ more often, by comparison to a condition in which no passport context exists. In this chapter, we aim to replicate this passport bias effect they found, and then to investigate its causes. In the following experiments, we first test the hypothesis that this bias arises because of the authority, by varying the authority that a document carries. We introduce documents that do not carry the same authority to see if authority matters: student-ID cards as low authoritative document compared with passports and driving licences (Experiment 1-3). Then we test posters as a different layout from commonly seen identity documents but carrying high authority (Experiment 4). We find that authority is not a sufficient explanation for the observed bias, but the layout of a document may affect this bias.

2.2 Experiment 1

Introduction

McCaffery and Burton (2016) found that matching a face to a passport biases viewers' response, by comparison to matching isolated face images. A passport is used to verify the bearer's citizenship and issued by the government so that it is widely accepted as a valid identity proof. Also, it is hard and risky to make a fraudulent one. So, the bias may exist because of the authority that a passport carries, i.e., participants tend to accept a face pair representing the same person when one image appears in a passport. To test the effect of authority, we introduce a student-ID frame for face matching in this experiment. A student-ID is only locally-significant and carries less authority compared to passports. It is less formal and unlikely to be accepted as a proof of identity when purchasing age-restricted goods. It will be interesting to see the comparison between the two documents.

In the first experiment, we would like to see if embedding one of the faces in documents which carry different authority could affect face matching. Here we compare viewers' matching performance on isolated plain faces and faces embedded in two different ID documents: passport and student-ID (see Figure 2.1).

We are aiming to first replicate the passport effect reported by McCaffery and Burton (2016). If we observe the effect, we then will see whether the bias effect is due to authority. If it is, then we may expect a larger effect for passport and a considerably smaller effect or no effect with the student-ID. If authority does not make a difference, i.e., the effect is present across any document, then we expect no difference between passport and student-ID, and viewers should give more 'same' responses to any face pair involving a document. The third possibility is that the authority does not matter, and neither of the documents shows a bias effect compared to isolated face pairs. As the overall performance is not affected by a passport frame (McCaffery & Burton, 2016), we anticipate the overall error rates to be similar across conditions.

Method

Participants

Thirty female students (aged from 18 to 27, mean age = 20.2) from the University of York participated for course credit. All reported normal or corrected to normal vision. A power analysis using GPower (Erdfelder et al., 1996) indicated that a sample of 30 participants would be needed to detect an effect of size $\eta_p{}^2 = .2$, with 90% power using a within-subjects ANOVA and alpha at .05. Each participant completed a consent form before the experiment.

Stimuli

Sixty face pairs were chosen randomly from the Glasgow Face Matching Test (GFMT; Burton et al., 2010) and used as stimuli. Face images are all in good quality, depicted at the same size and all front-facing. For this experiment, we constructed three formats for each face pair: plain faces, faces with a passport frame, and faces with a student-ID frame (see Figure 2.1). Passport frames were based on genuine documents, but all information (e.g. names, dates of birth) showed fictitious details. The student-ID frame was constructed to emulate a typical university card, but was from a fictitious university (with which our participants could not be familiar). For the frame conditions, we only embedded the right face image in a face pair in these frames. The distance between faces in each pair was constant across all three conditions.

Figure 2.1 Face pairs from GFMT in three presentations for Experiment 1. Each of these examples shows different identities.

Design and Procedure

This experiment was run using a lab-based computer and presented using PsychoPy software (Peirce, 2007). We used a one-factor (presentation type: plain faces, faces in passport, and faces in student-ID) within-subjects design. Each participant performed three blocks (plain faces, passport frame, and student-ID frame) during the task. They saw 20 face pairs in each block, so there were 60 face pairs in total. In each block, half of the face pairs were images of the same individual and half showed different individuals. The face images were counterbalanced across conditions. The order of the three blocks was counterbalanced across participants. Participants were asked to indicate if the faces shown in each pair were the same individual or different individuals by pressing corresponding keys on a keyboard (‘F’ for ‘same person’ and ‘J’ for ‘different people’). Each face pair was displayed until a response was made.

Results

Table 2.1 Means (standard deviations in brackets) of matching performance in Experiment 1 to 3.

Experiments	Presentation type			
	Plain	Passport	Student-ID	Driving licence
Experiment 1	0.92 (0.10)	0.91 (0.10)	0.90 (0.13)	/
Experiment 2	0.73 (0.16)	0.73 (0.16)	0.75 (0.17)	/
Experiment 3	0.73 (0.14)	/	0.73 (0.14)	0.73 (0.15)

The overall accuracy of all conditions was shown in Table 2.1. Because we focused on the *bias* of responding ‘same person’, we calculated the sensitivity (*d’*) and criterion (C) for matching decisions (see Appendix 1 for the detailed calculation of *d’* and C). For these purposes, we count participants’ ‘match’ responses as corresponding to *hits* when the two photos show the same person, and *false positives* when they show different people. We eliminated scores (*d’* and C) that were three standard deviations away from the mean across all experiments.

Repeated-measures ANOVA showed no significant effects of presentation type on *d’* (F(2,58) = 0.81, p>.05, $\eta_p{}^2$ = .03) or on C (F(2,58)=0.65, p>.05, $\eta_p{}^2$=.02) (see Figure 2.2). A by-image analysis showed the same absence of difference between conditions (details can be found in Appendix 2).

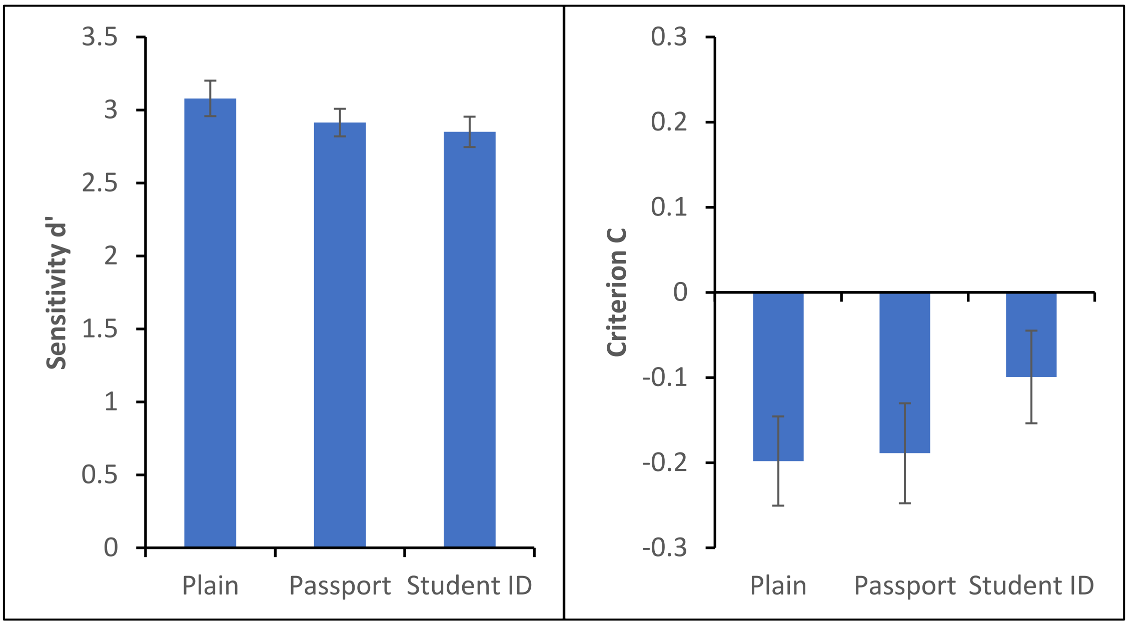

Figure 2.2 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 1. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

From these results, we did not find any differences among the three presentation formats, which means the presence of frames did not make participants give more 'same' responses compared to plain faces. This is inconsistent with McCaffery and Burton's (2016) data. Considering we selected the original face pairs from GFMT, as did McCaffery and Burton, there are two possible reasons explaining this inconsistency. One possibility is that there is actually no such bias effect for passports, and this bias cannot be replicated. The bias effect is perhaps a weak one, and is not large enough to appear with every experiment. Another possibility is that there is something problematic with our manipulation. Looking back to the data, the *d'*s across three conditions were extremely high, which means participants show nearly perfect discrimination for same and different face pairs. The overall accuracy was also quite high (90%), while in McCaffery and Burton's, the overall accuracy was around 80%. This may indicate a ceiling effect in our experiment. Participants performed with very high accuracy in all three conditions, which makes the additional manipulation weak for detecting the effect. The GFMT has been widely used and our participants might have seen some of the exemplars in papers or in other experiments in the University of York. Also, it has been claimed that the GFMT provides optimized conditions (i.e. similarity in pose and format between face pairs), which might result in a ceiling effect and obscure other effects (Fysh & Bindemann, 2018a), like the bias effect in our experiment. Thus, it could make a difference if we use a harder task to test this bias effect.

The Kent Face Matching Test (KFMT; Fysh & Bindemann, 2018a) provides a more challenging task of face matching, with an average population performance of 66% compared to 80% in the GFMT. In the KFMT, each face pair comprises two photos: a high-resolution portrait image and a student-ID photograph provided by the target himself/herself. The ID photos were not constrained by expression or devices, while the portraits showed a neutral expression. These two photos were taken at least three months apart and are presented in different sizes. These aspects of the KFMT therefore encapsulate some of the difficulty in performing face matching in real settings. This database seems more appropriate for our study.

So, in Experiment 2, we decided to repeat the experimental manipulation, using the stimuli from the KFMT instead of the GFMT. If we find the same pattern as Experiment 1, then we should conclude that embedding a face to a document will not bias viewers' responses. If not, we may find the bias effect is not restricted to one particular face set. Our other hypotheses about the possibility of the influence of authority are the same as those in Experiment 1.

2.3 Experiment 2

Introduction

In this experiment, we use the same manipulation of images as that in Experiment 1 (i.e. plain faces, faces embedded in passport and faces embedded in student-ID).

However, all face images are chosen from the KFMT instead of the GFMT. First, we aim to replicate the results of McCaffery and Burton using a different face set. Second, we aim to establish if the high authority carried by passports accounts for any bias effect, such that passports make people accept pairs as showing the same person more often than when viewing isolated images.

Method

Participants

Fifty-six students (51 females, aged from 18 to 32, mean age = 20.1) from the University of York participated for course credit or a small amount of money. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli

The sixty face pairs (30 matches and 30 mismatches) we used for stimuli were randomly chosen from the KFMT instead of GFMT. The KFMT faces are all university students and images are in good quality, taken with different cameras, three months apart, depicting two different sizes and all front-facing. Apart from that, the creation of three formats was the same as that in Experiment 1 (see Figure 2.3). Notably, the smaller picture of the KFMT-pair was used in conditions with a frame.

Figure 2.3 Face pairs from KFMT in three presentations for Experiment 2. Each of these examples shows different identities.

Design and Procedure

The design and procedure were identical to those in Experiment 1. Each participant performed three face matching blocks as illustrated in Figure 2.3 (plain faces, passport frame, student-ID frame). Participants saw 20 faces (half match) in each block, with the order of blocks counterbalanced across the study.

Results

The overall accuracy was shown in Table 2.1, with an average performance of 73% correct in all three conditions. Figure 2.4 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of condition on *d'* ($F(2,110) = 1.30$, $p > .05$, $\eta_p^2 = .02$), but a significant effect on C ($F(2,110)=7.56$, $p<.001$, $\eta_p^2=.12$). Pairwise comparisons (Tukey HSD) showed that C was reliably larger for plain photos than both passport frames ($p<.05$) and student-ID frames ($p<.05$), but these did not differ significantly (HSD=.119; $F_{crit}(1,110) = 5.65$: *F*(plain vs. passport = 14.01; *F*(plain vs. student-ID) = 7.74; *F*(student-ID vs. passport) = 0.92)). A by-image analysis shows similar results, and can be found in Appendix 3.

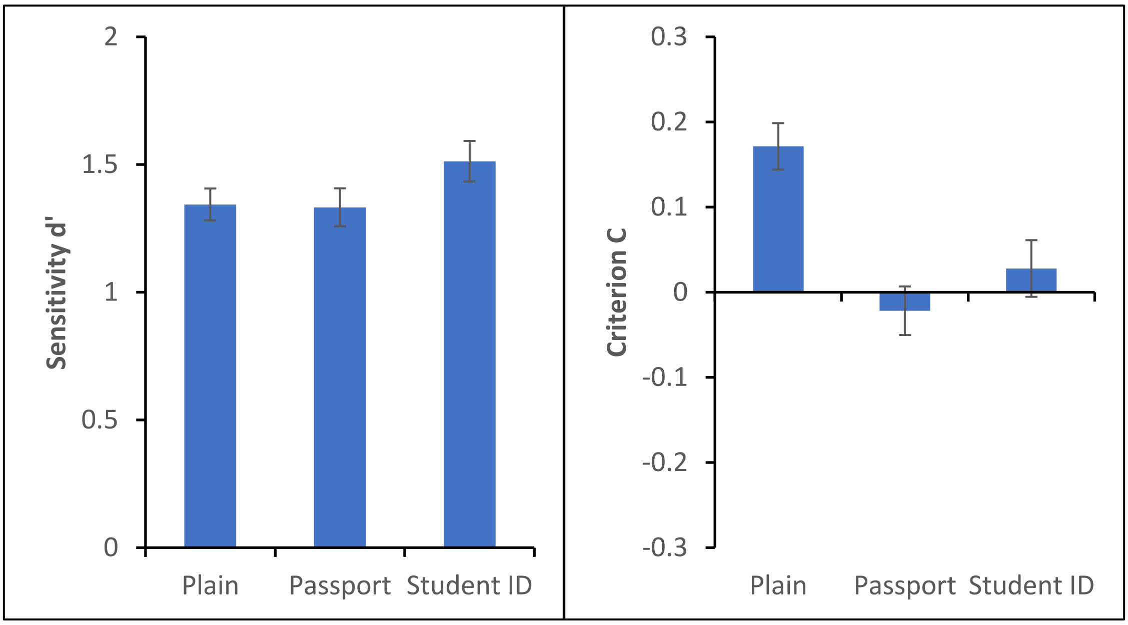

Figure 2.4 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 2. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

Experiment 2 shows that embedding a face into a passport or a student-ID affects viewers' matching decisions but not matching performance. The range of accuracy was 72-75% across three conditions. Compared with Experiment 1 (89-92%), the overall accuracy and *d'* were much lower in the present experiment, giving confidence that the results are no longer affected by ceiling performance, indicating the performance in Experiment 1 was too high to find other effects. In the following experiments, we will continue to use the faces selected from the KFMT.

When the overall levels of accuracy are down below ceiling, the basic effect with passports does replicate. Participants were more likely to make 'same person' judgements when one face was embedded in a document frame, consistent with McCaffery and Burton (2016). The bias effect found in the present experiment indicates that the effect is not restricted to a particular face set (KFMT in the current study, GFMT in McCaffery and Burton's).

Results show that both passports and student-IDs resulted in significant effects on matching response bias, and the two document types did not differ significantly. This result provides evidence for the hypothesis that the bias effect arises simply from embedding photos in ID, and is not strongly affected by the authority of that ID. Whether the ID document carries low or high authority, embedding one of the faces into a document does appear to bias people's choice.

Although passports are known to our participants, they may not be carried and seen in everyday life. Passports appear more at airports and in more official situations. The fictitious student-ID is also not familiar with our participant pool. So, will the exposure of a document lead to such bias? Will it be different when participants view more commonly used documents? It would be interesting to see whether this bias generalises to more commonly used ID documents specific to the participant pool, like driving licences and student-IDs from the host university. Therefore, in the next experiment, we are going to create a driving licence frame and a student-ID frame derived from the genuine ID documents which participants are quite familiar with.

2.4 Experiment 3

Introduction

The previous experiment demonstrates the bias effect found by McCaffery and Burton (2016) is not restricted to a certain face database or to passports only. For example, a fictitious student-ID can also bias viewers' choice. Would such a bias exist more generally for the ID documents we use nearly every day? In this experiment, we turn to more commonly used documents to see if the bias can be generalised to other ID cards. We choose the driving licence from the UK and the student-ID from the host university, which are highly familiar to the participant pool, university students (see Figure 2.5). They use driving licences as a proof of age to buy alcohol or entering nightclubs; they also use their student-ID to receive parcel mail or gain access to library facilities.

Driving licences carry more authoritative weight than student-IDs. Because driving licences are also issued by the government as with passports, except that driving licences do not carry international significance. We then are still able to include two documents that carry different authority as in previous experiments.

If the bias effect also exists in highly familiar documents, the driving licence and the student-ID would both induce that effect, which means viewers will give more 'same' responses to these two IDs. If neither of the IDs used in this experiment shows the bias, possibly familiarity or the specific layout plays a role in the bias effect. If authority is a key factor, then we might expect to see a larger effect for driving licences than student-ID.

Method

Participants

Twenty-nine students (26 females, aged from 18 to 24, mean age = 19.5) from the University of York participated for course credit or a small amount of money. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli

Face pairs were the same as in Experiment 2. For this experiment, we constructed three formats for each face pair: plain faces, faces with a driving licence frame, and faces with a student-ID frame (see Figure 2.5). Driving licence frames were based on genuine current UK documents, but the information they contained (e.g. names, dates of birth) was edited to show fictitious details. The student-ID frame was based on the University of York student-ID card. This card carries most of the key information on its reverse side, and so we included this in the display (see Figure 2.5). All the identifying information was fictitious. The distance between faces in each pair was constant across all three conditions.

Figure 2.5 Face pairs from KFMT in three presentations for Experiment 3. Each of these examples shows different identities.

Design and Procedure

The design and procedure were identical to Experiment 2. Each participant performed three face matching blocks as illustrated in Figure 2.5 (plain faces, driving licence frame, student-ID frame). Participants saw 20 faces in each block, with the order of blocks counterbalanced across the study.

Results

The overall accuracy for each condition was shown in Table 2.1. Figure 2.6 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of presentation type on *d'* ($F(2,56) = 0.01$, $p > .05$, $\eta_p^2 < .01$), but a significant effect on C ($F(2,56) = 5.39$, $p < .01$, $\eta_p^2 = .16$). Pairwise comparisons (Tukey HSD) showed that C was reliably larger for plain photos than driving licence frames ($p < .05$), but no other comparisons were significant (HSD = .167; $F_{crit}(1,56) = 5.80$: *F*(plain vs. driving licence = 10.26; *F*(plain vs. student-ID) = 4.97; *F*(student-ID vs. driving licence)= 0.95)).

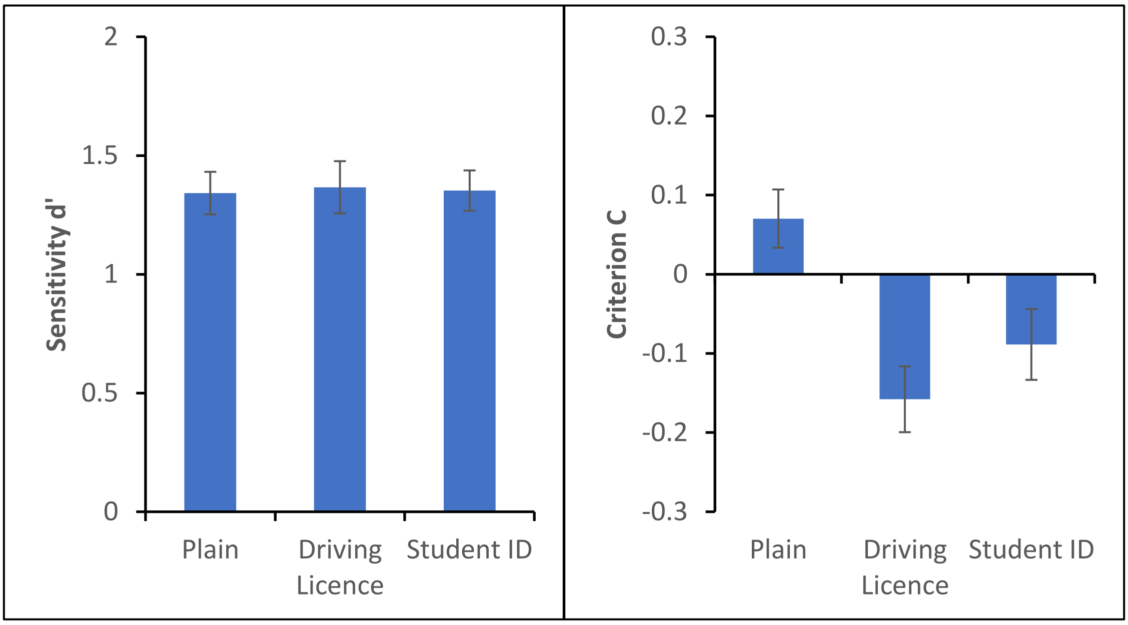

Figure 2.6 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 3. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

From the results, we found that people gave more 'same' responses when they were matching face pairs embedded in a driving licence, in the same way as has been shown previously for passports (McCaffery & Burton, 2016; Experiment 2) and a fictitious student-ID (Experiment 2). The results for student-ID here were less clear. While the difference between plain photos and student-IDs approaches significance, there is very little support for a difference between the two document types.

These data show again that McCaffery and Burton's (2016) document-induced matching bias is not limited to passports. Although there is a clear effect with passports and a fictitious student-ID (Experiment 2), driving licences and student-ID here did not differ significantly in this experiment, and neither was there a significant difference between student-ID and isolated faces. There is a possibility that the failure of detecting the effect of student-ID in this experiment is due to a different layout around the face (see Figure 2.5). The document frames we use for passport and driving licence show a rectangle colour background, with information next to the face. Although the fictitious student-ID shows a white background, the textual information and the face are visually integrated. For our new student-ID in this experiment, the textual information lies below the face and there is more white space around the face image on student cards. This may lead participants to ignore the frame and focus entirely on the faces, so showing no significant difference with plain faces or a driving licence frame. This is a hint that the layout of a document frame may be important to the bias effect. It is possible that the cards produce a bias because of a layout that leads to interference between text and image. This would be consistent with results from McCaffery and Burton's (2016) studies, in which there was evidence of interference between processing the textual information and the face images.

The experiments above use identity documents (passport, driving licence, student-ID) with similar layouts, which are commonly used to prove the holder's identity and are easy to read. In the next experiment, we consider a different situation where both face and information are shown in one document, but the document is larger and contains rich information: a poster. ID cards contain only identification information such as name, date of birth. The key information and card frame are placed close to the face image. But posters provide more additional information and the frame or background is much bigger. So, in the next experiment, we introduce a poster frame as a different type of document which has a different layout but also includes both face and information to test if viewers will be biased by this kind of document. If the bias effect exists, this may indicate the layout of how information and faces are organized is not a key factor of this bias; If we fail to find a bias, then the layout may be critical to the bias.

2.5 Experiment 4

Introduction

Previous experiments show that common identity documents can bias people's choices. In this experiment, we would like to examine a different type of document – a poster that is displayed by authority institutions such as the UK Metropolitan Police. We embedded one of the faces into a poster: either a 'wanted-person poster' or a 'missing-person poster', mimicking the documents provided on national security websites. We created posters based on Mileva and Burton (2019), who used this type of document to provide contextual information about the targets to improve face search performance. These posters contained a large amount of information. In addition to the personal information (name, age, sex, height, etc.) next to the face, there was also a paragraph of further information (see Figure 2.7). In the *wanted* condition, the paragraph presented information about the crime for which the person was wanted; in the *missing* condition, the summary contained information about where the person was last seen.

Figure 2.7 Examples of wanted and missing posters used in Experiment 4.

The posters convey richer semantic context and show a different background than the ID cards we used before: more distance between text and image. This brings a possibility that the cards give a bias because of a specific layout that leads to interference between text and image. So, if we get an effect using posters here, it will suggest a more semantic basis for the effect – information provided affects matching decisions. Whereas if we only get a bias effect with cards, it will suggest an interference between the specific card context and the image. We chose to use two different types of poster in order to manipulate the valence of information about the person. A *wanted* poster reflects negatively on the person shown, whereas a *missing* poster does not imply any wrongdoing. If the basis for interference involves semantic processing, then this valence manipulation might highlight differences between the two types of poster.

For this experiment, we broaden our participant pool, running it remotely, online using Prolific (https://www.prolific.co/), which is an online survey website giving access to a more diverse population. Although the new participant pool contains a wider age and occupations, we continued to limit participation to native English speakers in the UK to make sure they all understand the information on posters.

In this experiment, we are going to test if the bias effect can only be found with a certain layout, and to retest if this kind of document posted by the authorities can bias people's choice. If the posters bias people in the same way as ID cards do, then we will expect to see a difference between the two posters and plain faces.

Method

Participants

This study was run online. One hundred and eighteen registered members (62 females, aged from 18 to 46, mean age = 28.0) from Prolific (https://www.prolific.co/) participated for a small amount of money. All reported normal or corrected to normal vision. Each participant completed an online consent form before the experiment.

Stimuli, Design and Procedure

Face pairs were identical to Experiment 2. We created two poster frames following those used in Mileva and Burton (2019), a wanted poster and a missing poster. There was a physical description of the person and a statement of the background to each case in the poster. All the information was fictitious and the case descriptions were derived from the Metropolitan Police and City of London Police websites. The design and procedure were identical to those in Experiment 2. Each participant performed three face matching blocks as illustrated in Figure 2.8 (plain faces, wanted poster, missing poster). Participants saw 20 faces in each block, with the order of blocks counterbalanced across the study.

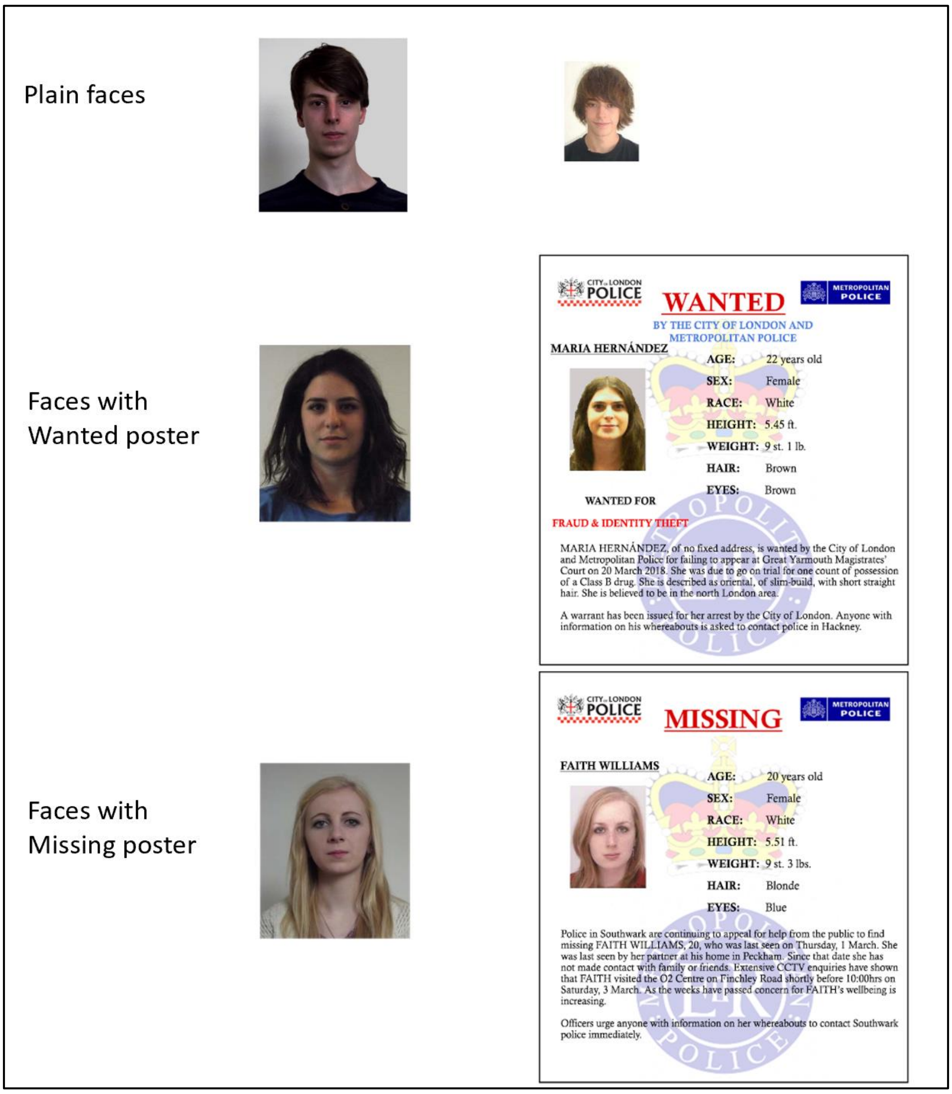

Figure 2.8 Face pairs from three presentation conditions in Experiment 4. Each of these examples shows different identities. Details of the example posters are shown in Figure 2.7.

Results

The overall accuracy for each condition is 75.3%, 75.2%, and 77.3% for plain faces, missing poster, and wanted poster, respectively. Figure 2.9 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of presentation type on *d'* ($F(2,234) = 1.67$, $p>.05$, $\eta_p{}^2 = .01$), and no significant effect on C ($F(2,234) = 2.70$, $p > .05$, $\eta_p{}^2 = .02$).

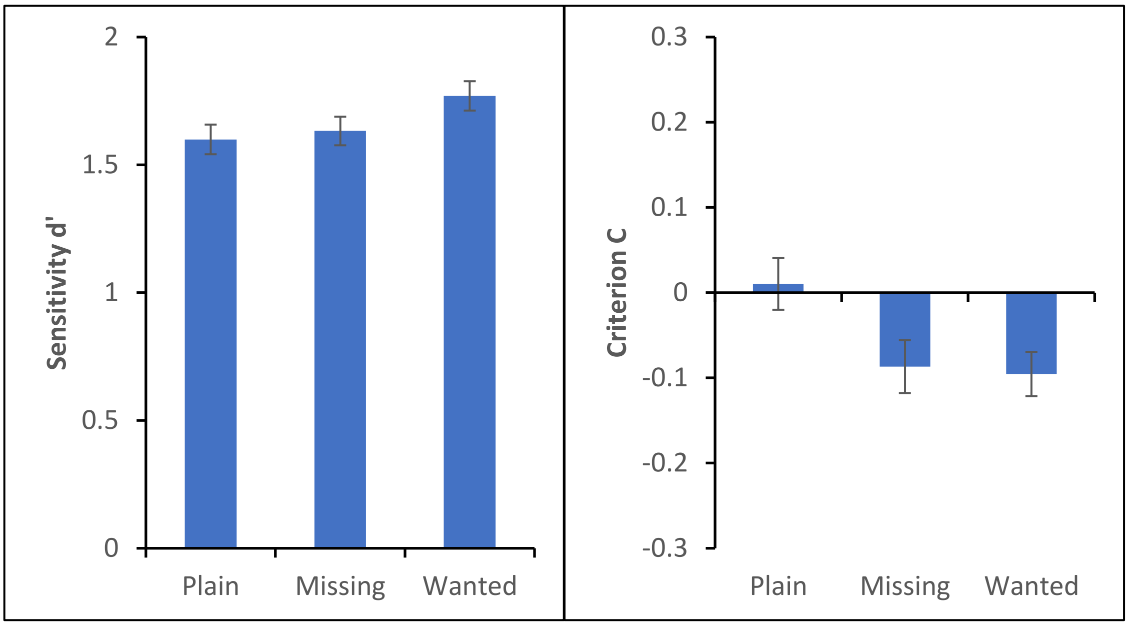

Figure 2.9 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 4. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

Results show no significant difference on overall performance. Neither a missing nor a wanted poster benefit the matching performance. This result is consistent with Mileva and Burton (2019), where adding more contextual information does not improve performance on a search task. The results also show that matching a face pair embedded in a poster does not affect people's response bias. We can see from the graph that there is an observed tendency that participants show lower criterion C, i.e. saying more 'same', in two poster conditions, but the effect does not reach significance.

Here we failed to find a bias using a poster frame, one possible reason was that we provided too much information, participants may just ignore all the information and the background together, because there are too many distractions. This again suggests that authority is not the key factor for the cause of the bias effect. The posters contain authoritative symbols at the top and watermarks behind, but there is no biasing effect found. Another possibility is the layout of a document. The layout of posters may be hard for participants to integrate the face and the frame, whereas the card frame in previous experiments may give rise to a spatial layout that encourages interference between words and images. A commonly seen rectangle card may give viewers' a feeling of matching 'a card', not matching 'a face with some text and background'. It is possible that the experience from real life matters in this task.

2.6 General discussion

Is there any difference between matching a face to a document and matching two isolated faces? Across several experiments, we replicated and extended McCaffery and Burton's (2016) observation that embedding photos in passport frames does not influence the overall accuracy, but affects matching judgements. Viewers were biased by passports to respond 'same person' more when seeing one of the faces embedded in a document. In this chapter, we took the first step to investigate the possible reasons and to see if this bias is replicable. Our results showed that the bias effect could generalise to various documents and face sets, and was not affected by the apparent authority or official nature of the documents.

First, we consider the manipulation of authority. Experiments 1 and 2 introduced a student-ID card with low authority, but participants showed bias in both passport and student-ID conditions. Experiment 4 presented authorized posters, in which the 'wanted' poster is normally published by the police and is a serious document. All posters had titles, symbols and watermarks that indicated the government institution, which carries high authority. However, we failed to find a bias towards more 'same person' with posters. So, the authority does not provide a satisfactory explanation for the observed bias.

Consider the manipulations where we did not find the bias effect: student-ID showing both sides in Experiment 3 and posters in Experiment 4. These two frames show different layouts from the ones that induced bias. There is much white space around the face, and the majority of the text lies below the face. Viewers may just focus on the face alone rather than focus on 'the document' as a whole. This indicates that maybe the ID-like layout is important to the bias. This also suggests the bias effect may arise from the superficial aspect of the stimulus, such that the visual properties of an ID-like frame may affect people's responses. We will investigate this in later experiments.

Second, the bias effect does not rely on whether the document has been commonly seen or not. We used passports and driving licences that were derived from genuine documents. The student-ID card, however, is made up in Experiment 2, and participants were still biased by it. When we used a university student-ID that was actually used in real life by our participants, the bias did not show up.

In fact, some of the face images embedded in documents here would not meet the rules of an acceptable photo when people apply for a UK passport. It seems that image quality was irrelevant to this bias effect. It has been found that there was no difference in performance with matching a face to a driving licence or a passport photograph (Kramer et al., 2019). The performance was still within the range of typical accuracies reported by previous face matching studies. So, different types of face stimuli and documents do explain the bias effect. However, the layout of the document seems to be of importance to this bias.

It should be noted that the presence of ID documents is always task-irrelevant in these experiments, and yet its presence affects behaviour. Participants know that they are taking part in a psychology experiment, and they know that their match/mismatch judgements have no implications at all for the people involved. Despite this, it is possible that their previous experience automatically affects their behaviour, or there was an interference between processing the document and processing the faces. However, note that the effect is specific to bias. None of the experiments above produces any effect in overall performance, as one might observe in typical perceptual interference effects (e.g., Pashler, 1994). Since the task is self-paced, we do not know whether participants actively process the frame or not. But as we suppose the layout of the document matters, it gives us a hint that participants may take the face and that frame as an integral whole.

To summarise, these experiments suggest that presentation within documents is always different from isolated presentation: matching a face to a document biases viewer's choice. The bias effect is quite solid and can generalise to different ID documents (passports, driving licence, and student-ID). In addition, the effect is not due to the authority that a document carries, but may simply be due to embedding photos in IDs.

If it is not because of high authority, then why does a document frame bias people's choice? Another possibility is that the visual properties, like the layout, of that ID document affects people to say more 'same person'. The document frame embraces the face image as part of the document, which leads to visual processing to the other parts of the document and that may lead to a systematic bias. In the following chapter, we turn to manipulate the visual properties of the documents to investigate the underlying cause of the bias effect.

Chapter 3 – Investigating the factors of the bias effect

3.1 Introduction

In the previous chapter, we found matching a face photo to a face embedded in a document biased people's choice: under these conditions, they tend to give more 'same person' responses than they would to isolated face pairs. The effect of the document is not restricted to those with high authority: cards like passports (McCaffery & Burton, 2016), driving licences, and the normal student-ID cards all show this effect. Although we demonstrated the bias is replicable and can generalise to various documents, the underlying cause of the bias remains unknown. In this chapter, we continue investigating the cause of this bias.

As we have excluded the influence of authority, another possible factor derived from previous experiments is the layout or the visual properties of the document display. Although the bias seems to widely exist through our previous experiments, there are conditions in which it did not show up. In those experiments where we did not observe the bias, the layouts of the frames were slightly different from the normal identity document. So, perhaps the layout of the document plays an important role in the bias, which is interesting because the layout of the document is an additional or task-irrelevant element when people do face matching tasks. We do not know if our participants take the frame into account, but the inconsistent results indicate how the document frame presented matters.

Considering the visual properties of a document frame, there were two components: text or information on the card, and the card background. It is unclear which component contributes to the bias or whether they combine. Perhaps visual processing of the faces embedded in documents leads to qualitatively different processing from the cropped isolated photos of a face which are typically used in lab experiments. For example, the bias effect may reflect a more general contextual level of processing, involving viewers implicitly processing the entire ID setting. Priming tasks show clear semantic processing of faces (e.g., Boehm, Klostermann, Sommer, & Paller, 2006), and semantic contexts are known to influence face processing and recognition (Koji & Fernandes, 2010; Rainis, 2001; Shriver et al., 2008).

There are also good reasons to make a hypothesis that the effect arises from the interference between processing the faces and processing the information carried on an ID card. Faces have been demonstrated to interfere with other perceptual tasks (Jenkins, Lavie, & Driver, 2003; Lavie, Ro, & Russell, 2003) and textual labels such as names and occupations can interfere with face classification tasks (Young, Ellis, Flude, McWeeny, & Hay, 1986). There are also studies showing picture-word interference when the picture was spatially adjacent to a word (MacLeod, 1991). So, the information on card documents may interact with the faces when carrying out such matching tasks.

In this chapter, we are going to explore more about the nature of this effect, specifically asking if some visual properties of the display bias viewers' choice. In a series of studies, we manipulate the visual properties of the document, by separating the authoritative context of the card from the personal information it contains (Experiments 5 and 6). Having established that card context is critical to elicit the effect (Experiments 7 and 8), we then manipulate the readability of biographical information by rendering it in a script unknown to the viewers or by blurring (Experiments 9 and 10). Taken together, the results point to multiple sources of the observed bias, depending on both card and linguistic context.

3.2 Experiment 5

Introduction

In this experiment, we compare matching isolated faces to faces embedded in documents that are ID-like (i.e., cards) but carry no information. We compared blank frames derived from driving licences, which have overall shape and colour similarity to the stimuli used in Experiment 3 in Chapter 2, but which have all the identifying information removed (see Figure 3.1 below). We do not know whether our participants would notice that this frame is a blank driving licence (and so may be able to invest it with implicit authority). For that reason, we also introduced a condition with a novel colour, not used in U.K. driving licences. Any effect of these frames on matching bias would support the idea that visual context plays an important role in the effect.

Figure 3.1 The top image shows a face pair embedded in a driving licence used in Experiment 3 in Chapter 2. The bottom image shows a face pair embedded in a blank card preserving the background of a driving licence in the present experiment. Each of these examples shows ‘different’ identities.

Method

Participants

Thirty students (27 females, aged from 18 to 32 years, mean age = 20.6) from the University of York participated for course credit. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

60 face pairs, as used in Experiment 2 in Chapter 2, were selected from the KFMT. We used the pure background of a driving licence as the original colour frame and created a green version of that frame with photo editing software.

The design and procedure were identical to previous experiments. Participants completed three face matching blocks (plain faces, original-colour frame, and green frame), as illustrated in Figure 3.2. There were 20 face pairs per block (10 matches and 10 mismatches), and pairs were counterbalanced across the experiment such that each appeared equally often in the plain and card-embedded conditions. Participants' task on each trial was to indicate whether the face pair showed the same person or different people by pressing corresponding keys on a keyboard. Each face pair was displayed until a response was made. The order of blocks was counterbalanced across participants.

Figure 3.2 Face pairs from three presentation conditions in Experiment 5. Each of these examples shows 'different' identities.

Results

Figure 3.3 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of condition on *d'* (F(2,58) = 1.35, p>.05, η_p^2 = .01), and no significant effect on C (F(2,58) = 0.64, p>.05, η_p^2 = .02).

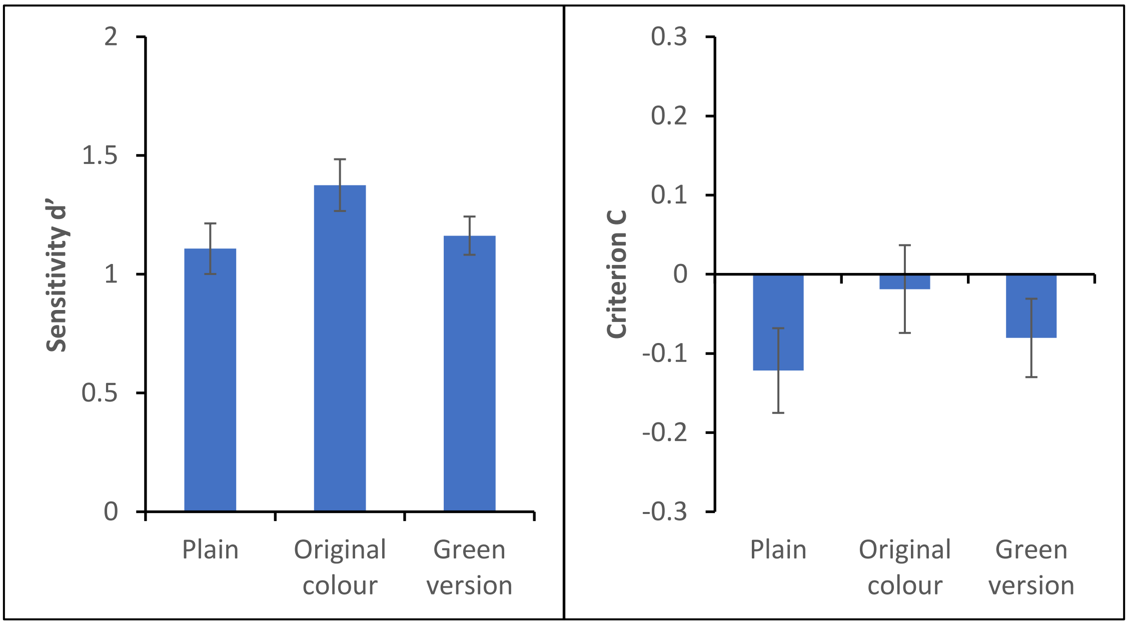

Figure 3.3 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 5. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

This experiment showed no effect of blank cards on matching response bias. Without any ID information, participants' tendency to make "same person" judgements remains constant across conditions. This result ties the phenomenon clearly to the information carried in ID documents, even though this information is task-irrelevant. Furthermore, the original colour of the driving licence does not carry any effect by comparison to a different colouring, suggesting that participants are not implicitly treating this frame as a genuine driving licence.

From this experiment, we find embedding a face to a single background, i.e. the blank card, cannot bias decisions. This suggests the information is of great importance for the bias, which raises the possibility that the biasing effect of card context is driven primarily by the text it contains. In the next experiment, we eliminate the card background, presenting only the text. If the bias is observed in that condition, then it may need to be re-cast as a picture-word interference effect, rather than an effect tied to the social use of ID, as has been previously suggested.

3.3 Experiment 6

Introduction

Experiment 3 in Chapter 2 shows embedding one of the faces into a driving licence biases viewers' choice, while embedding one of the faces into a blank driving licence did not show such an effect (Experiment 5). In this experiment, we compared isolated face matching to a presentation in which biographical information (name, date of birth, and address) is presented alongside a face, but not within a card context (see Figure 3.4). As in previous experiments, this information was task-irrelevant, and participants were simply asked to indicate whether two face images showed the same person or different people. If the bias, now reported across many ID contexts, is induced by fundamental picture-word interference, then we would expect to observe it in this presentation. If not, there will be no difference between the two conditions.

Method

Participants

Twenty-eight students (25 females, aged from 18 to 31, mean age = 20.9) from the University of York participated for course credit or a small amount of money. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

Face pairs were identical to those used in Experiment 5. Two conditions were constructed: plain faces and faces alongside text (see Figure 3.4). For the text condition, we remove the card background and only present the key information (name, date of birth, and address) next to the face, which was the same text extracted from the stimuli used in a driving licence frame in previous experiment.

This experiment employed a within-subjects design. Each participant performed two face matching blocks as illustrated in Figure 3.4 (plain faces, faces with text). They saw 30 face pairs (15 matches and 15 mismatches) in each block, and pairs were counterbalanced across the experiment. Participants' task was to decide if the face pair they saw showed the same person or different people. Each face pair was displayed until a response was made. The order of the two blocks was counterbalanced across participants.

Figure 3.4 Face pairs from two presentation conditions in Experiment 6. Each of these examples shows 'different' identities.

Results

Figure 3.5 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of presentation type on *d'* ($F(1,27) = 0.23$, $p > .05$, $\eta_p^2 = .01$), and no significant effect on C ($F(1,27) = 0.03$, $p > .05$, $\eta_p^2 < .01$).

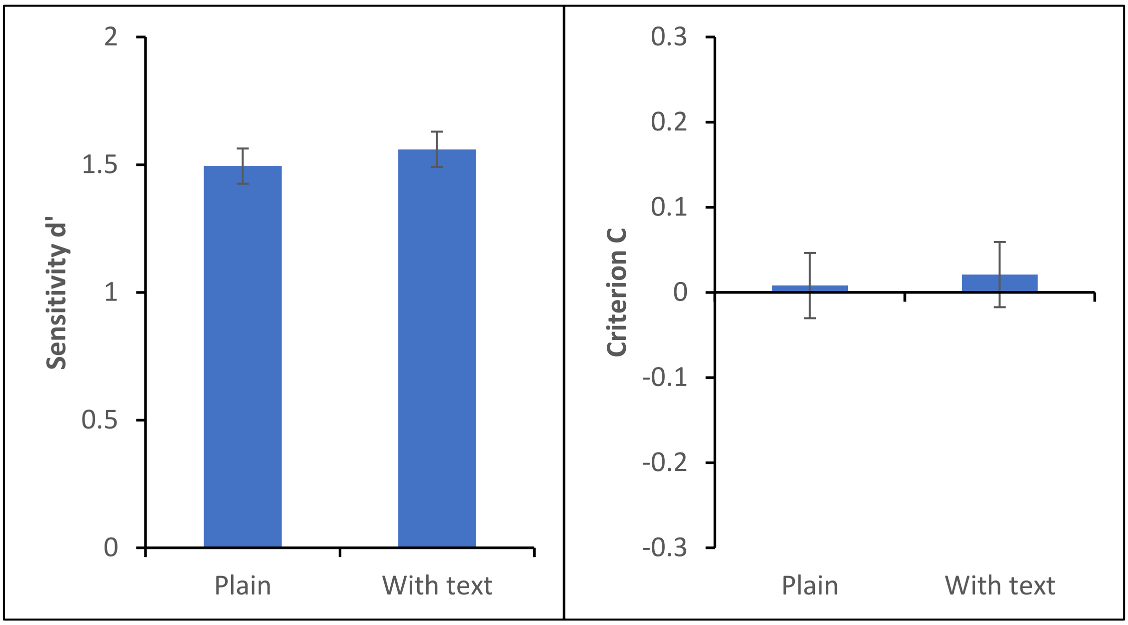

Figure 3.5 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 6. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

The results showed no effect of adjacent biographical text on the bias effect in face matching, which seems to rule out any simple explanation based entirely on textual interference on face matching. Combined with the last experiment, neither the background nor the information can bias people's choice on their own, suggesting that the overall card context is critical to understanding this effect.

It is possible that the card background alone will make the face 'stand out', and the text alone has no visual relation with the face. In this case, viewers probably will not be influenced by the background or the text. They may not treat the face and the background/text as a whole, but as separated, so that the presentation of these items has no effect on biasing responses. The bias effect seems to rely on both components of an ID document.

Therefore, in the next experiment, we use a very simple card frame, without any official marks of symbols, which combines the background (in Experiment 5) and the text (in Experiment 6). If viewers are biased by the new card, it will suggest both the information and the card background are necessary to induce a bias. If not, maybe some more sophisticated elements on a genuine card, such as the peripheral marks of authority, is more important than the visual elements.

3.4 Experiment 7

Introduction

From the two experiments above, we found that faces in a blank card or faces with simple text did not lead to a bias effect. This suggests the two components of a card context may need to be presented at the same time. To test this hypothesis, we examined the effect of minimal ID-like context on face matching. Card frames were created containing personal information (name, date of birth and address), but no further cues about the purpose of the card. This simple layout, illustrated in Figure 3.6, preserves many of the features of a standard ID, but does not convey any information at all about its nature, i.e. whether or not it carries official status.

It will also be a good comparison that enables us to retest the bias effect. The comparison of performance on standard isolated face matching and the minimal card context will establish a baseline ID-effect, independently of expectations induced by specific contexts, such as passports, driving licences or workplace IDs. If the simple card leads to a bias, then we will find participants make more ‘same person’ responses when the card frame presented.

Method

Participants

Thirty-two students (27 females, aged from 18 to 24, mean age = 19.0) from the University of York participated for course credit. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

Face pairs were identical to Experiment 5. We created a simplified card condition, using the green version as a background and added the same information (name, date of birth, and address), in the same relative position as in Experiment 6. The design and procedure were identical to those in Experiment 6. Each participant performed two face matching blocks as illustrated in Figure 3.6 (plain faces, simple card). Participants saw 30 face pairs (15 matches and 15 mismatches) in each block, with the order of blocks counterbalanced across the study.

Figure 3.6 Face pairs from two presentation conditions in Experiment 7. Each of these examples shows 'different' identities.

Results

Figure 3.7 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of condition on *d'* (F(1,31) = 2.30, $p > .05$, $\eta_p^2 = .07$), but a significant effect on C (F(1,31) = 6.89, $p < .01$, $\eta_p^2 = .18$). Participants had a more negative criterion when they saw faces embedded in card frames (M = -0.06) than when they saw faces alone (M = 0.15).

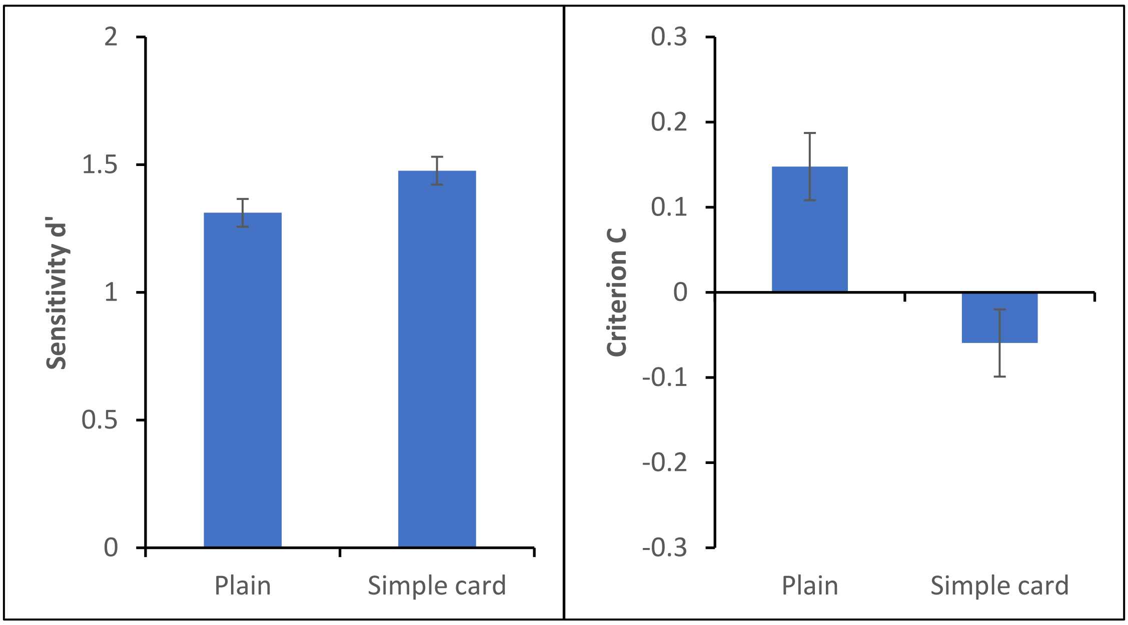

Figure 3.7 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 7. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

Our results show if we have both card components, i.e. information and card background, shown to viewers, they will be biased to make more ‘same person’ judgements than they did when viewing plain, isolated faces. That is to say, a simple card leads to a bias. This echoes previous studies using more sophisticated ID documents such as passports and driving licences. The fact that this bias is elicited by so simple a presentation seems to suggest a rather minimal role for the perceived authority of a card. The simple card did not carry any symbols or titles for participants to indicate the official nature of an ID, but only name, date and address. This provides more evidence that the document frame does not bias people by its authority.

Combined with Experiment 5 and 6, the integration of background, text and face seems to be the key characteristic of the bias effect. People were not biased by single information or by a frame background, which suggests that they did not treat faces with background or faces with text as one single, integrated visual object. While documents are different from simply ‘sticking’ faces onto cards, viewers treat faces in documents differently from plain faces.

As the bias needs both a colour background and some written text, and seems to involve the integration of text, background and faces, will any colour background that can ‘group’ the faces and the text together result in the effect? For example, if the background is not ID-like, but a random shape, will the bias effect exist?

3.5 Experiment 8

Introduction

From the experiments above, we find that the information and the background are both important for the bias effect. It appears that it is the visual grouping of face, information and background that biases people’s choice. Viewers appear to form a Gestalt processing so that they perform differently with isolated faces. In order to test this possibility, we will use a circle background that is different from an ID card, which is usually constructed in a rectangle shape. We would like to see if the visual grouping of face, along with text and background will bias people in the same way as we found before.

In this experiment, we create a circle ‘card’, which is different from the normal ID cards that are usually seen. Using this manipulation, we are able to see if the ‘card-like’ layout is necessary to the bias effect, or it is just a group of background, information and face. We also included the faces with text condition, to make an obvious comparison with the circle background (see Figure 3.8). If it is the group processing of the three elements that induce the bias effect, we will expect no bias in the faces with text condition, but a bias in the circle card, compared with isolated faces. If the card-like layout matters, we will see no bias in these manipulations.

Method

Participants

This study was run online. One hundred and twelve members (79 females, aged from 18 to 50, mean age = 31.2) from Prolific (https://www.prolific.co/) participated for an amount of money. All reported normal or corrected to normal vision. Each participant completed an online consent form before the experiment.

Stimuli, Design and Procedure

Face pairs were identical to Experiment 5. We created a new circle card (see Figure 3.8), using a purple circle as a background and added personal information the same as that in Experiment 6. This colour was selected from a base colour in the driving licence. The design and procedure were identical to those in Experiment 5. Each participant performed three face matching blocks as illustrated in Figure 3.8 (plain faces, faces with text, faces with circle cards). Participants saw 20 face pairs (10 matches and 10 mismatches) in each block, with the order of blocks counterbalanced across the study.

Figure 3.8 Face pairs from three presentation conditions in Experiment 8. Each of these examples shows ‘different’ identities.

Results and discussion

Figure 3.9 shows sensitivity (*d’*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of condition on *d’* ($F(2,222) = 0.24$, $p > .05$, $\eta_p{}^2 < .01$) or on C ($F(2,222) = 1.37$, $p > .05$, $\eta_p{}^2 = .01$).

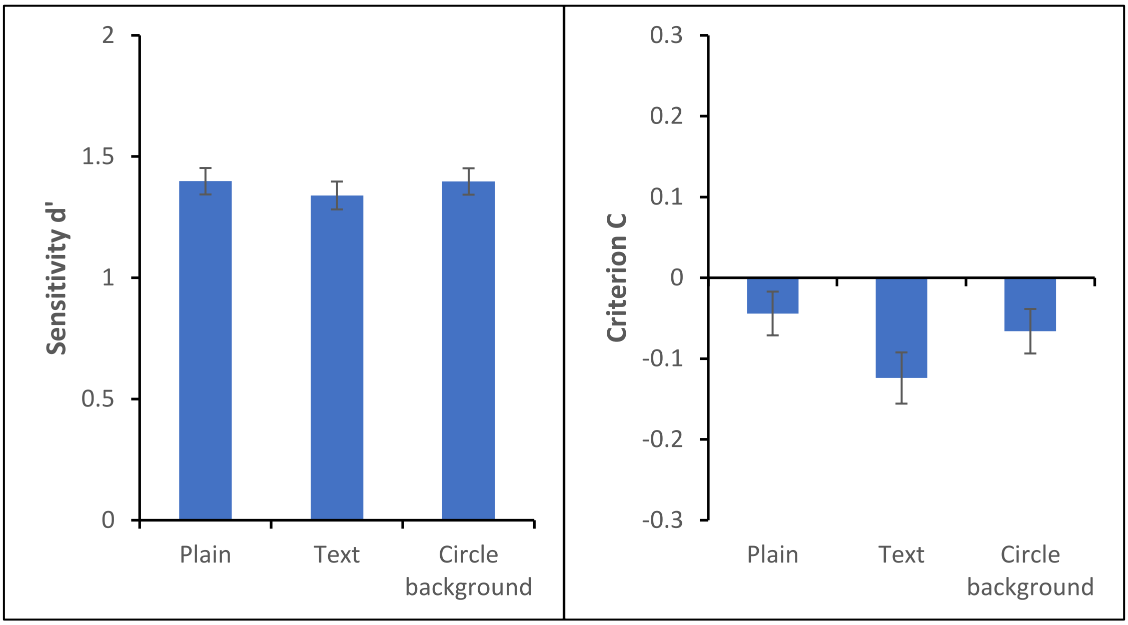

Figure 3.9 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 8. Error bars represent within-subjects standard error (Cousineau, 2005).

Our results show that even though we have both elements, i.e. information and card background, shown to viewers, they were not biased to make more 'same person' responses compared to isolated faces. The only thing we changed from a simple card is the shape of the background, but the effect disappeared (see Figure 3.6 for a comparison). So, the bias is not simply reliant on the perceptual process of grouping the components, it is also related to an expectation of normally seen ID documents. This rules out the hypothesis that a Gestalt processing of face, text and background together is sufficient to produce the bias to say "same person".

Combined with the last three experiments, it suggests that the effect of visual context on face matching relies on a complex combination of visual display characteristics, incorporating both biographical text and an implied ID card context. The effect of (biographical) text does not therefore appear to be automatic, but somehow facilitated by the surrounding context. In the next two experiments, we explore this relationship further by using ID cards with embedded text that is unreadable or irrelevant to participants due to it being rendered in an unfamiliar script (Experiment 9), blurred or semantically inappropriate to ID (Experiment 10).

3.6 Experiment 9

Introduction

The presence of information and card frame are both important to the bias effect. To investigate this effect further, we manipulate the information presented in documents. In this experiment, we present a card with the accompanying biographical text in Bulgarian – a language unfamiliar to the participants. This information was presented either in the Bulgarian alphabet, rendering it literally unreadable by the participants, or transliterated into Roman script, rendering it readable, but mostly meaningless to participants (see Figure 3.10).

If the effect of context on face matching is carried mainly by the visually ID-like frame (i.e. face, information, and card background), irrespective of content, then it should be observed for both card conditions. However, if the source of the effect relies on processing the meaning of the biographical information, then it should not be observed at all in this experiment.

Method

Participants

Twenty-nine students (26 females, aged from 18 to 26, mean age = 20.3) from the University of York participated for course credit or an amount of money. According to the manipulation of this experiment, we only recruited participants who do not read Bulgarian or Russian (as this uses a related alphabet). All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

Face pairs were identical to Experiment 5. Biographical information for ‘Bulgarian’ cards was created using the template of the driving licence, and the background was coloured green to remove the indication of a driving licence. The information included the card bearer’s name, address and signature, along with various official designation numbers relating to the licence. The information was constructed with the help of a Bulgarian national and combined common forenames and surnames along with plausible addresses. For the Roman-script versions, names and addresses were transliterated, for example, ‘Стоева Петя’ to ‘Stoeva Petya’ (Figure 3.10). While this renders them readable to participants, the names and addresses were nevertheless unfamiliar.

The design and procedure were similar to Experiment 5. They completed three blocks of trials as illustrated in Figure 3.10: plain isolated faces, faces with cards in Bulgarian, and faces with cards in Bulgarian rendered in Roman script. Each block comprised 20 face pairs (half matching), and the order of blocks was counterbalanced across the experiment.

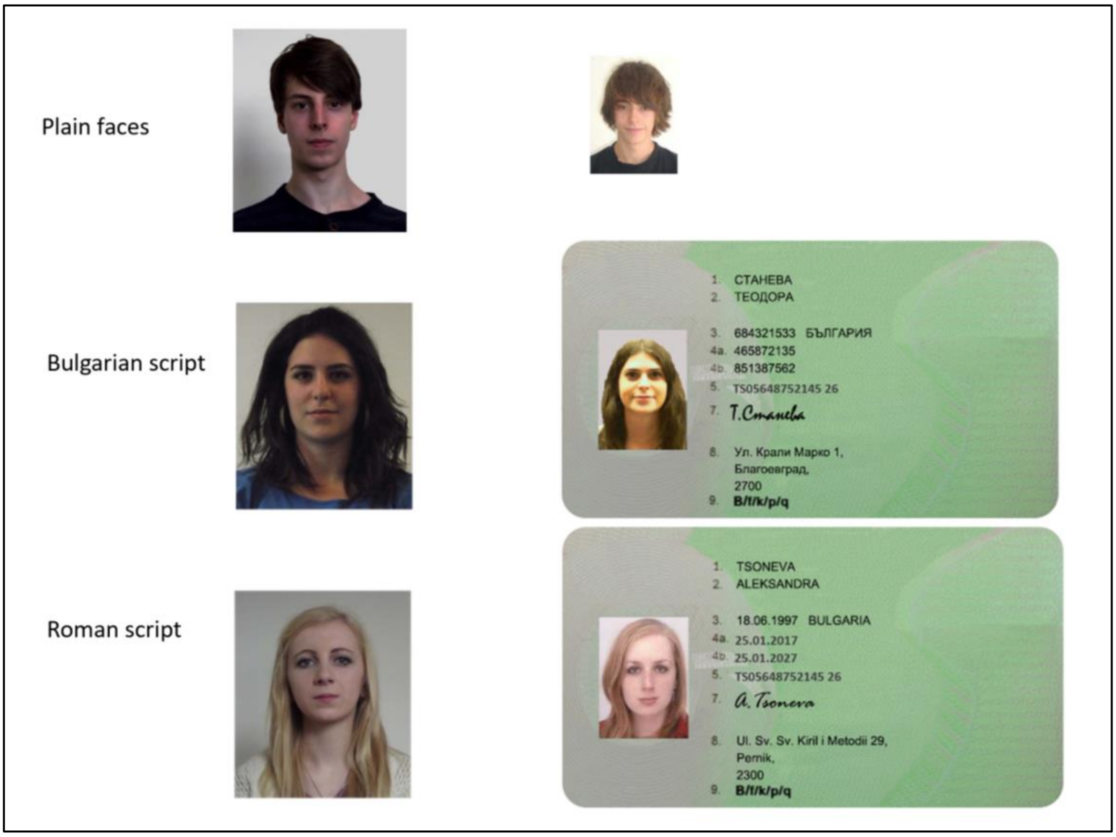

Figure 3.10 Face pairs from three presentation conditions in Experiment 9. Each of these examples shows 'different' identities.

Results

Figure 3.11 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of presentation type on *d'* ($F(2,56) = 0.54$, $p>.05$, $\eta_p{}^2 = .02$), and no significant effect on C ($F(2,56) = 0.23$, $p > .05$, $\eta_p{}^2 = .01$).

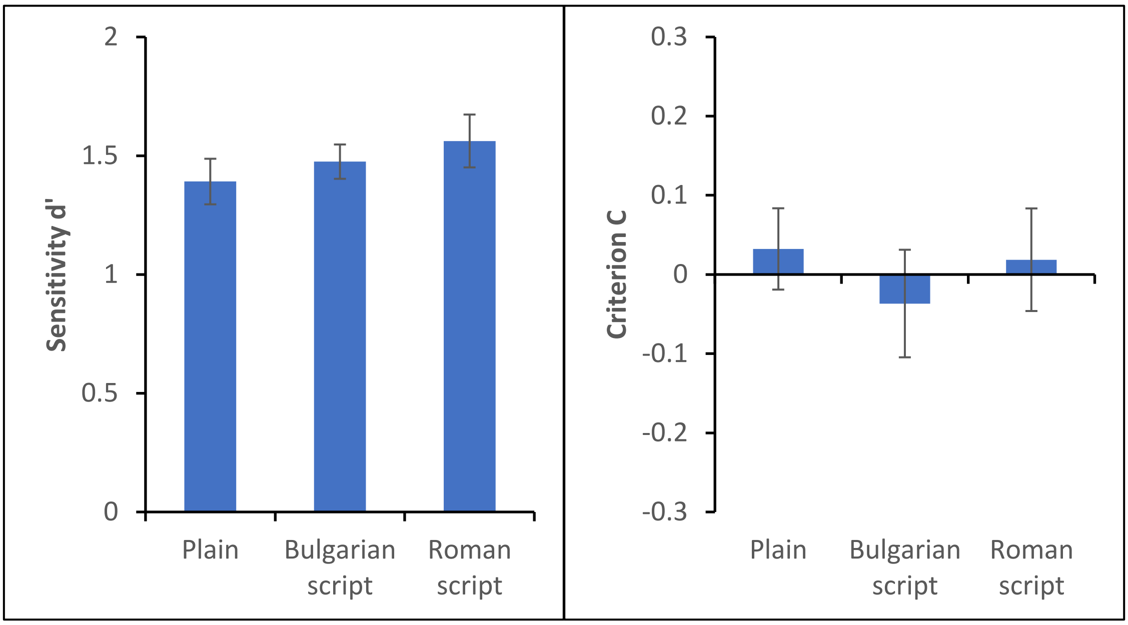

Figure 3.11 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 9. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

This study shows that if people cannot read the information on cards, they are not biased in their responses. It seems then, that the implied nature of the ID is not sufficient to produce an effect on face matching. Instead, it seems necessary that both card-context and meaningful text are necessary to produce this effect.

Note that the text rendered in the Roman script was still insufficient to produce an effect, even though it was readable, the information about names and addresses still relates to Bulgaria, which is unfamiliar with our participants. This gives us a hint that the text information on cards cannot be anything random, but should be at least something understandable or easily processed by the participants. In the next experiment, we invert this relationship by including understandable but irrelevant text on cards (i.e. non-biographical English words).

3.7 Experiment 10

Introduction

The studies presented so far appear to demonstrate that, to have an effect on face matching, it is necessary to present understandable information within a card context. In this experiment, we introduced two new conditions (see Figure 3.12). First, we constructed cards with readable, meaningful text, but this textual information was inappropriate to an ID card, simply comprising English nouns. Second, we presented a card with the full, appropriate information, but in which the text was blurred. It was therefore clear that the card contained information, however, it was not possible to read that information. Here we would like to establish whether only relevant text can influence matching, and whether cards with 'implied' biographical information are sufficient to elicit an effect.

Method

Participants

Thirty-six students (29 females, aged from 18 to 32, mean age = 20.3) from the University of York participated for course credit or an amount of money. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

Face pairs were identical to Experiment 5. The new cards were designed using the driving licence template (see Figure 3.1). For the 'readable cards' condition, arbitrary nouns replaced the licence-holders forename, surname and address. For the 'blurred' condition, the previous driving licences in Experiment 3 were used (Figure 3.1) but the textual part of the card was blurred to a level that preserved word shape but eliminated readability (see Figure 3.12).

Design and procedure were the same as Experiment 5. Participants were asked to make matching decisions (same/different) to pairs of faces. They completed three blocks of trials as illustrated in Figure 3.12: plain isolated faces, readable cards, and blurred cards. Each block comprised 20 face pairs (half matching), and the order of blocks was counterbalanced across the experiment. Face pairs were also counter-balanced, such that across the experiment, each pair occurred equally often in each condition.

Figure 3.12 Face pairs from three presentation conditions in Experiment 10. Each of these examples shows 'different' identities.

Results

Figure 3.13 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of presentation type on *d'*

($F(2,70) = 0.17$, $p>.05$, $\eta_p{}^2 < .01$), but a significant effect on C ($F(2,70)=5.65$, $p<.01$, $\eta_p{}^2=.14$). Pairwise comparisons (Tukey HSD) showed that C was reliably larger for plain photos than readable cards ($p<.05$), and larger for blurred photos than readable cards, but no significant difference between plain and blurred ones (HSD = =.225; *F*crit(1,70) = 5.74: *F*(plain vs blurred) = 0.09; *F*(plain vs readable) = 8.56; *F*(blurred vs readable) = 8.38).

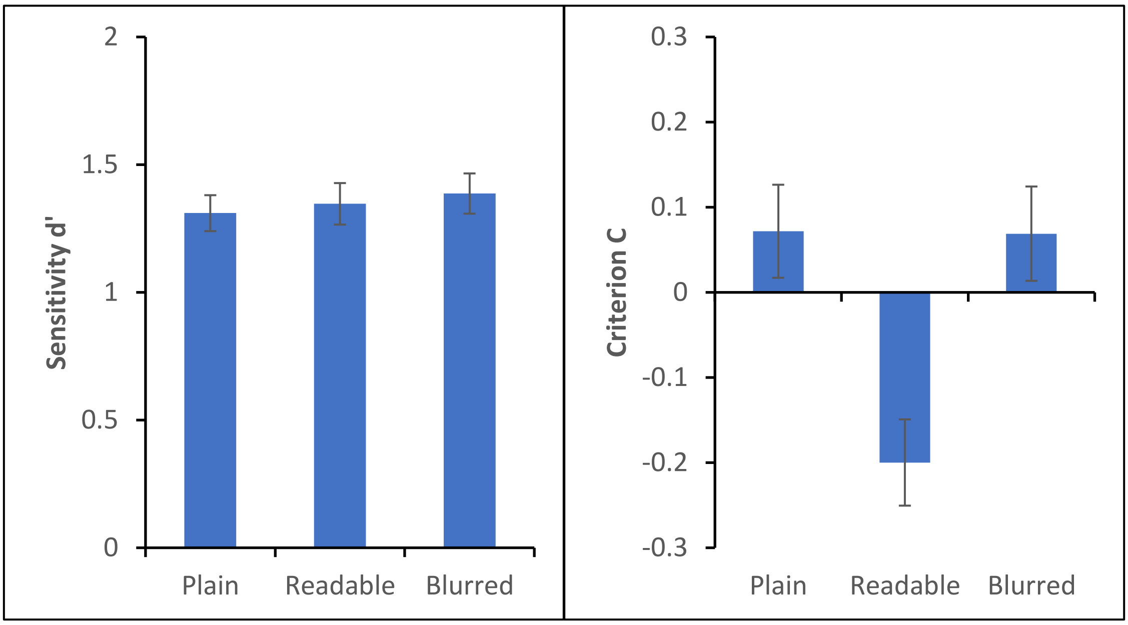

Figure 3.13 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 10. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

These results show that participants made more 'same person' responses to a face pair when one of the faces was embedded in a readable card, compared with a card in which they cannot see the information clearly or with plain isolated faces. The significant difference between the readable card and the blurred condition is quite interesting: the appearance was exactly the same, and the only difference is whether the information is readable – which is task-irrelevant anyway. Note that in all the experiments we have conducted so far, the card context (biographical information and background) is task-irrelevant. This would suggest automatic processing or influence of proximate visual information. However, this is somewhat challenged by Experiment 6 and 9, showing no effect of adjacent text which is outside a card context and no effect of non-understandable information within a card context.

This biasing effect with readable words replicates the findings with simple ID cards in Experiment 7 and those reported in previous studies with more formal ID (Experiments 2 and 3; McCaffery & Burton, 2016). What differentiates this finding from previous studies is that the information on these cards is entirely irrelevant to the ID. Indeed, the arbitrary nouns used are somewhat bizarre in an ID context. Nevertheless, they appear to influence the face matching task.

Faces are suggested to be automatically processed at a semantic level (e.g. Boehm et al., 2006). Even though we do not know whether participants actively read and process the readable information carried on the IDs in these experiments, they did change their criterion when they saw faces in readable cards. It is possible that the processing of documents interferes with the processing of faces, especially when they compare these faces to make an identity judgement.

3.8 General discussion

In this chapter, we tried to establish the reason why people are biased by a document frame when doing face-matching tasks. We demonstrated that a blank card and the adjacent text without a card context could not elicit the bias (Experiments 5 and 6), suggesting neither of these factors was sufficient to account for the effect. But the addition of a very simple card context does (Experiment 7), and it has to be card-like in shape (Experiment 8). Furthermore, ‘interfering’ text needs to be comprehensible but not semantically relevant to elicit this bias (Experiments 9 and 10).

These experiments add additional evidence that the bias effect is not attributed to the apparent authority of official ID documents, which has been demonstrated in Chapter 2. In addition to using different types of documents carrying various levels of authority, there were no official symbols on the cards used here, and they were not related to any familiar type of identity documents: minimal ID-like card contexts (Experiment 7) were sufficient to induce the bias.

Whether the experience or expectation of seeing photo-ID induces the bias effect remains unclear. On the one hand, if one's personal experience of ID cards matters, then viewers seeing unreadable cards should be biased, however, they were not (Experiments 9 and 10). On the other hand, we cannot exclude expectation or experience totally: viewers are used to seeing ID cards, and it is ID-like frames that induce bias (Experiment 8).

Our results suggest a possible explanation of the bias: interference between processing faces and processing the card context, because the card context is critical to the bias (Experiments 5-8) and the information on cards has to be readable (Experiments 9 and 10). However, it is hard to reconcile the observed matching bias with explanations based on a simple interference from the irrelevant text. In typical interference tasks, distractor items are designed to be response-congruent or incongruent. Under those conditions, the literature contains many examples of interference between face and text processing (e.g. Jenkins et al., 2003; Stenberg et al., 1998; Young et al., 1986). However, in our studies, 'interfering' textual information was either consistent with an ID (Experiment 7) or irrelevant to it (Experiment 10). In conditions where text was clearly present, but unreadable (Experiments 9 and 10), no bias was observed. It therefore appears that face matching is somehow biased, in part, by the deployment of resources diverted into task-irrelevant reading, rather than by any semantic processing of the text. Although we cannot be sure whether viewers actually 'read' the information or not, the different results between readable and unreadable cards do suggest the processing, intentionally or not, affects face matching.

Perhaps the most puzzling aspect of these results is the fact that the 'same person' bias does not appear in Experiment 6 and 8, in which faces are presented alongside biographical text, but without an ID card context. Combined with Experiment 7 and 8, these results do suggest that the card frame sets up some expectation in the viewer, such that information on the card is then processed. This gives us a possible answer about the nature of the card contexts: it is not due to the integration of information and background together as a single Gestalt (a perceptual explanation), because the circle background in Experiment 8 does not produce the bias effect. Instead, the result may rely on stimuli that somehow induce an expectation in viewers based on the social use of an ID card (a more social explanation). But in Experiment 9, we used an ID card, nevertheless, when people cannot read the card, they are not biased. This may suggest the perception and expectation relating to an ID document interact with each other in some way.

The bias effect is similar to the Stroop effect (Stroop, 1935) to some extent, in which the distractor word diverted the processing of the required task (naming the colour). In our tasks, face matching is subject to interference by the presence of card context, which is also task-irrelevant. However, the Stroop effect or any other interference effect mostly reflects on reaction times. In the matching tasks here we have never required participants to react as quickly as possible, making it inappropriate to calculate the reaction time here. Maybe in future studies, reaction times should be included in the analysis.

In conclusion, the 'document bias', while frequently demonstrated, resists simple explanation. It appears to rely on the convergence of different stimulus characteristics – most notably those typically found on photo-ID. It also serves as a reminder about the generalisation of simple effects. There is now a large literature on face matching which almost all employs isolated face images, and almost all appeals to the relevance of applied problems. It appears that there is a systematic difference between simple experimental face matching and real-world matching using documents.

Chapter 4 – Varying faces and tasks: what causes the ‘same person’ matching bias?

4.1 Introduction

In previous chapters, we have manipulated stimuli in order to try to understand why matching a face in a document biases viewers to make more ‘same person’ responses. At the very start of the experimental programme, we proposed that the authority or the visual properties of a passport may account for such bias. However, we found that providing readable information in an ID-like frame can induce the bias, regardless of authority. The face set we used was constant across those experiments (Experiment 2 - 10), and the basic experimental settings were quite similar – all face matching tasks. In this chapter, we do not manipulate the card frames themselves, but turn to other factors, such as faces and tasks, that may affect, extend or explain the bias in a different way.

Previous experiments were focused on unfamiliar faces, and it is unknown if a document would bias familiar face matching. Familiar faces are processed differently from unfamiliar ones, and are more accurately processed than unfamiliar faces (see Johnston & Edmonds, 2009, for a review). As a result, embedding a familiar face into a document may affect criterion differently. Furthermore, we have everyday experience of seeing the ID of familiar people, for example, we often have the experience that our friend’s ID photo does not look very much like them, but still accept it as a genuine match. So, perhaps the bias may derive from the experience we have of using photo-IDs. Kramer et al. (2019) have pointed out that people are familiar with the notion that their official photos of ID cards may not look very like them. In this situation, we notice the difference between ‘me’ and the photo, but we will still accept the photo as us, i.e., to respond the two faces match. Generalising this experience to novel faces may support the bias observed here. As this experience is usually related to familiar people, we may observe a bias using familiar faces, too.

Considering the experience of seeing photo-IDs, it is also true that the probability of encountering fake IDs is undoubtedly low. While most studies and previous experiments use equal proportions of match and mismatch trials (50% - 50%). It has been found that the low prevalence of targets influences the ability to detect them, from visual search tasks (Godwin et al., 2015; Rich, Kunar, Van Wert, Hidalgo-Sotelo, Horowitz, & Wolfe, 2008) to more applied situations such as airport baggage screening (Wolfe, Brunelli, Rubinstein, & Horowitz, 2013; Wolfe & Wert, 2010), and also on matching faces (Bindemann et al., 2010; Papesh & Goldinger, 2014). Viewers tend to miss more targets (i.e., to detect a mismatch face pair) when these occur with low prevalence, compared to high prevalence conditions. These experiments compare different proportions of mismatches, and here we ask whether the document bias still exists after increasing match trials (i.e., low-prevalence of mismatches). The document bias effect we found before was based on 50% of mismatch trials, where participants give more 'same person' responses when one of the faces is embedded in a document. Whether the document bias generalises to more realistic settings is of great importance, so we will retest the bias effect using a low-prevalence of mismatches in the following experiments.

Apart from the frequency, another change on the matching task we are going to make is the task itself. Participants in the face matching task are asked to make judgements of same person or different people all the time, which includes identity decisions. If we show participants the same stimuli, but change the task to ask them to rate the similarity of each face pair instead of judging identity, will we still find a bias induced by an additional document? Using this different task also checks the possibility that the change of perceived similarity between two faces results in the bias. People may tend to think the two faces look more similar when one of the faces is in a document context, so that they alter their response to 'same person'. Research has found that similar faces were matched reliably faster than dissimilar faces (Johnston & Barry, 2001). So, the document may make the faces look more similar, then viewers tend to make quicker decisions, which results in giving more 'same person' responses. This manipulation is important to adding explanations to previous results. If people do find faces in documents look more similar than isolated faces, this will suggest the document biases faces on the *perceptual* level, rather than on the stage of making decisions.

In this chapter, we are going to ask if this bias effect can extend to various face sets and experimental settings. We first test different levels of familiarity by asking if there is any document biasing effect for known celebrity faces (Experiment 11) and highly variable unfamiliar faces (Experiment 12). Next, we set a low frequency of mismatch face pairs to mimic the low proportion of fraudulent documents in real-life settings to see if this bias is replicable (Experiment 13). Finally, we change the task from deciding identity to rating the similarity between faces, to see if the bias still exists when performing a different task (Experiment 14). We will gain a broader view of the bias induced by a document.

4.2 Experiment 11

Introduction

People have more accurate performance when recognizing and matching familiar faces compared with unfamiliar faces. It has been suggested that familiar faces are processed differently from unfamiliar ones (see Johnston & Edmonds, 2009, for a review). Previous experiments and research found the document-induced bias exists with unfamiliar faces, but whether people will show the same pattern when matching familiar faces remains unknown. As the expectations of accepting a document derive from the experience of seeing our friends' documents, i.e., a familiar person, it is worth testing if the document bias will show up in familiar face matching tasks.

In this experiment, we use celebrities as familiar faces to test the bias effect on face matching. Will viewers adopt a more relaxed matching criterion for familiar celebrities? To test this, we paired images of celebrities with images of their 'lookalikes' – i.e. people who have established similarity to the celebrity. This makes the matching task non-trivial. Furthermore, because celebrities are well-known people and from different nationalities, it would be inappropriate to use a passport or driving licence frame. For that reason, we created a fictitious visitor ID card (see Figure 4.1), intended to imply that the celebrity has been invited to visit the BBC (British Broadcasting Corporation). The card frame contains personal information (name and profession), with a 'BBC' symbol at the top-right corner and a 'visitor' status at the bottom-right corner. This card, illustrated in Figure 4.1, shows a similar layout with a driving licence or student-ID: rectangular colour background, and some personal information. If people are biased by the card frame, we will find a difference in the matching criterion between faces in documents and faces presented isolated.

Method

Participants

This study was run online. Thirty-six students (32 females, aged from 18 to 22, mean age = 18.8) from the University of York participated for course credits. All reported normal or corrected to normal vision. Each participant completed an online consent form before the experiment.

Stimuli

Sixty well-known celebrities were chosen based on a local database of celebrities provided by university students. For each celebrity, we created a match pair and a mismatch pair. For a match pair, two images of the celebrity were downloaded from the internet. For a mismatch pair, one celebrity photo and a lookalike photo were used. We searched the internet for the lookalikes, mostly taken from celebrity lookalike websites, and we downloaded three images for further selection. To make the final stimulus choice for mistmatch pairs, we asked two people, close to student age, to select the most similar lookalike for each.

In a mismatch pair, the celebrity face image was always on the right and the lookalike image was always on the left. In a match pair, the same celebrity face image appeared on the right, with another different image of the same celebrity shown on the left. All faces were front-facing. For each face pair, there were two formats: plain faces and faces with card (see Figure 4.1 for illustration). We only embedded the face image on the right into a card frame, and this right image was always the celebrity. This means there is no condition of a card saying, for example, 'Will Smith' but showing another person, to make sure the card information is always correct and there is no conflict with deciding whether the two faces show the same person. As in all previous experiments, the card context was task-irrelevant, as participants were only asked to judge whether the faces matched.

The card background was derived from a BBC staff card. We kept the card background and the 'BBC' symbol to indicate this was a BBC visitor card and included the name and profession. The name and profession were correct with this celebrity but the number below was fictitious.

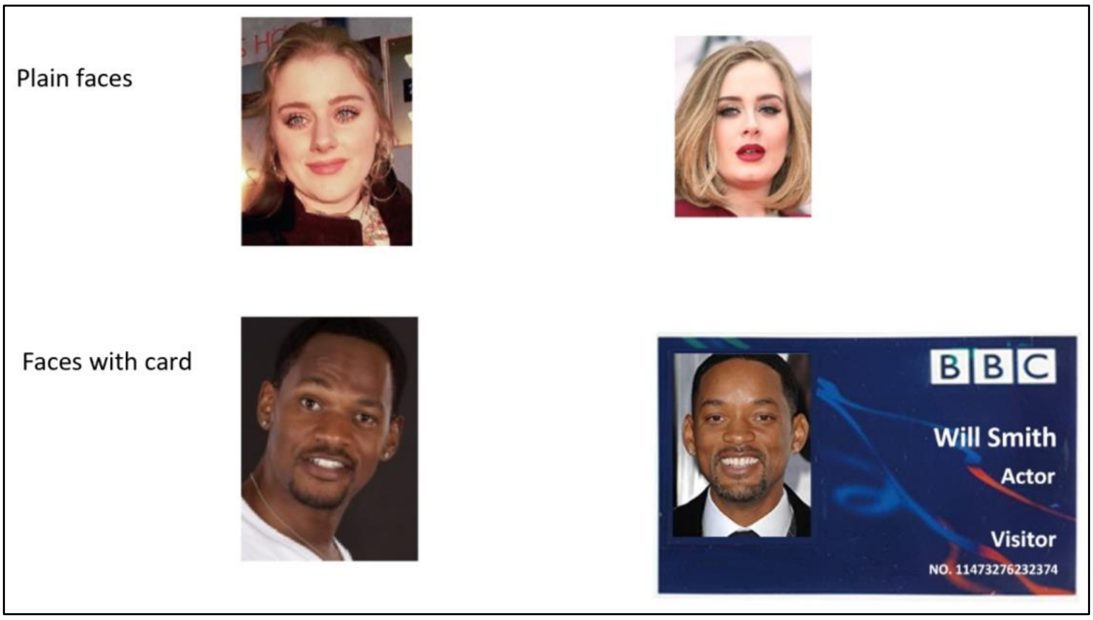

Figure 4.1 Example face pairs from two presentation conditions in Experiment 11. Each of these examples shows different identities. The face on the right for each pair

always shows the celebrity, i.e. Adele on the first row and Will Smith on the second row. The faces on the left are their lookalikes. In this example, the correct answer for each pair should be ‘different people’.

Design and Procedure

In a within-subjects design, all participants completed two face matching blocks, as illustrated in Figure 4.1: plain faces and faces embedded in a card. For each block, 30 face pairs (half matches and half mismatches) were shown, making 60 face pairs in total. The mismatch stimuli were rotated around participants, so that all face pairs appeared equally often in the plain and card conditions. Participants saw each celebrity once, either in plain or card condition. Participants’ task on each trial was to indicate whether the face pair showed the same person or different people by pressing corresponding keys on a keyboard. Each face pair was displayed until a response was made. The order of blocks was counterbalanced across participants.

Results and discussion

The overall accuracy for plain faces was 77.8% (responses for ‘same person’:91.3%, for ‘different people’:64.3%), and for faces in cards was 78.4% (responses for ‘same person’:89.2%, for ‘different people’:67.6%). Figure 4.2 shows sensitivity (*d’*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of condition on *d’* ($F(1,35) = 0.90$, $p > .05$, $\eta_p{}^2 = .02$) or on C ($F(1,35) = 0.58$, $p > .05$, $\eta_p{}^2 = .02$).

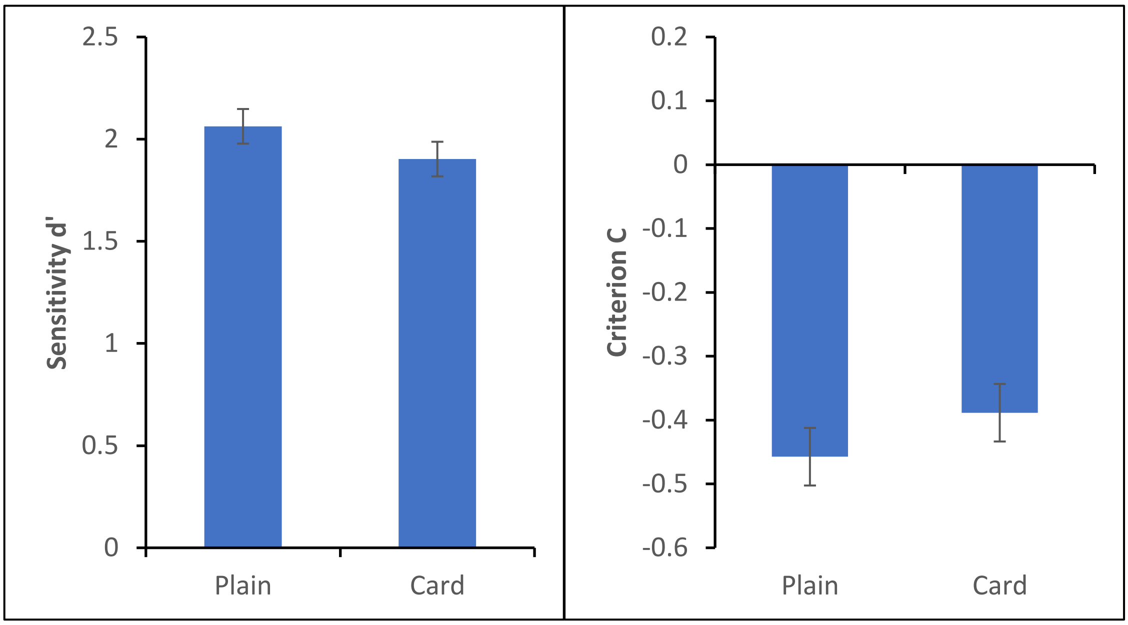

Figure 4.2 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 11. Error bars represent within-subjects standard error (Cousineau, 2005).

We found no difference between plain faces and faces in cards, indicating there was no bias effect of embedding a face into a card frame compared with isolated faces. This result shows that the document frame does not induce the bias effect for familiar faces, in contrast to the previous unfamiliar faces experiments. This is surprising because we made the hypothesis based on the experience or expectation of ourselves or friends' identity documents, and on previous results that the bias effect needs an ID-like frame. The absence of observed bias in familiar faces suggests that the effects reported in previous chapters are not based on expectation.

Although there is no difference between the two conditions, it should be noted that the criterion in both conditions is quite negative, which means there is a general large bias of responding 'same person' for familiar faces compared to unfamiliar faces in previous experiments. Due to individual differences, there are variations among experiments. However, in the present experiment, the basic criterion for plain faces was much more negative compared with previous experiments (e.g., -0.46 present vs. -0.13

in Experiment 5). Nonetheless, it was clear that card context did not affect this differentially.

From the accuracy and the sensitivity in the results, this was a hard task for participants to distinguish the foil. Normally, the performance on matching familiar faces has been found to be more accurate than matching unfamiliar faces, and is often nearly perfect (e.g., Megreya & Burton, 2006). However, the performance in this experiment was far from perfect (78% overall accuracy for both conditions), and only slightly higher than previous experiments using unfamiliar faces (for example, 72% overall accuracy in Experiment 2). Separating the 'same person' and 'different people' responses in plain faces condition, the accuracy for 'same' is 91%, but that for 'different' is only 64%. It is possible that the lookalikes we found for each celebrity are of too great a similarity – after all, these were chosen from look-alike sites. This seems to have resulted in an overall bias of responding 'same person' which could mask any effect caused by a document.

It has been shown that internal features (eyes, nose, mouth) are more important for recognizing familiar faces, while for unfamiliar face processing, the external features (hairline, chin, ears) are as important as the internal features (e.g., Clutterbuck & Johnston, 2002). This brings a possibility that people may extend the external features to the additional document, which raises the interaction of processing faces and documents. But for familiar faces, the internal features weigh more in decisions than external features, so again this may contribute to the fact that the additional document did not affect viewers' decisions on matching.

Nevertheless, we failed to find a bias of documents containing familiar faces. This suggests the bias effect found in unfamiliar faces may not derive from experience or expectation of familiar faces. The inconsistent results of unfamiliar and familiar faces suggest the bias relates more to the visual processing of faces and documents rather than the experience of seeing familiar people's ID cards.

As the criterion in this experiment was quite negative in plain faces, the document bias effect may be hindered by the basic criterion. So, in the next experiment, we consider a basic condition that leads viewers to make more ‘different people’ responses. The Models Face Matching Test (MFMT, Dowsett & Burton, 2015) is designed to be a difficult task (66% accuracy in Bobak et al., 2019), in which the images capture large differences in clothes, hairstyles, lighting conditions, and cameras across identities. These unconstrained images give a more conservative criterion for MFMT compared with other standard face matching tests (e.g. GFMT, Glasgow Face Matching Test; Burton et al., 2010), which means viewers typically give more ‘different people’ responses (Bobak, Dowsett, et al., 2016). In the next experiment, we are going to use this face set (MFMT) to see whether there will be a biasing effect of documents on a face set which itself tends to encourage more ‘different people’ responses.

4.3 Experiment 12

Introduction

In this experiment, we are going to use a different face set - the Models Face Matching Test (MFMT, Dowsett & Burton, 2015). This test contains 90 difficult face pairs that were pre-rated. The faces are all professional models but the images show quite different lightings, hairstyles, etc. This task has been designed to be difficult, and it applies in practical settings, such as looking for or examining super-recognisers (Bobak, Dowsett, et al., 2016). Unlike the face sets we have used before (GFMT or KFMT), the model face images embrace a lot more variation that may reflect real life – i.e. a large variation between ourselves and the photo-IDs.

We are going to make the same manipulation as previous experiments: comparing performance on matching isolated faces with that on matching one of the faces embedded in a document. If we observe the bias in the faces with document condition, then this will confirm that the effect generalises to different face sets beyond those taken in labs, under constrained cameras. The second aim is to see if the failure to observe a document bias on familiar faces (Experiment 11) depends on familiarity or photo variability.

Method

Participants

This study was run online. Thirty-six students (35 females, aged from 18 to 31, mean age = 19.2) from the University of York participated for course credits. All reported normal or corrected to normal vision. Each participant completed an online consent form before the experiment.

Stimuli

Sixty face pairs were chosen from the MFMT (Models Face Matching Test, Dowsett & Burton, 2015) to match the number of faces used in previous experiments. All the face pairs were male models. Half of them presented the same person and the other half presented different people. We created two conditions: plain faces and faces with a passport frame (see Figure 4.3). All the information on passports was fictitious.

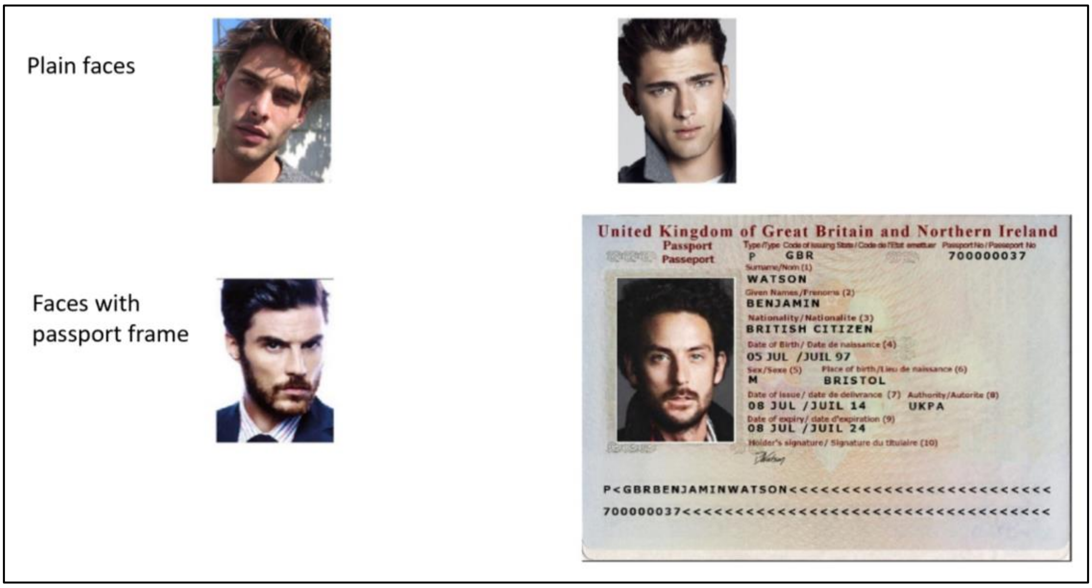

Figure 4.3 Example face pairs from two presentation conditions in Experiment 12. Each of these examples shows different identities.

Design and Procedure

The design and procedure were the same as those in Experiment 11. Each participant performed two face matching blocks as illustrated in Figure 4.3 (plain faces, faces in passport). Participants saw 30 face pairs (15 matches and 15 mismatches) in each block, 60 face pairs in total, with the order of blocks counterbalanced across the study. Participants' task was to decide whether the face pair showed the same person or different people by pressing corresponding keys on a keyboard. Each face pair was displayed until a response was made.

Results

The overall accuracy for plain faces is 71.4% (responses for 'same person':58.9%, for 'different people':83.9%), and for faces in passport is 73.6% (responses for 'same person':66.1%, for 'different people':81.1%). Figure 4.4 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of condition on *d'* ($F(1,35) = 0.12$, $p > .05$, $\eta_p^2 < .01$), but a significant effect on C ($F(1,35) = 13.31$, $p < .001$, $\eta_p^2 = .28$). Mean criterion was positive in both cases, but less so when participants saw faces embedded in passport frames ($M = 0.26$) than when they saw faces alone ($M = 0.47$).

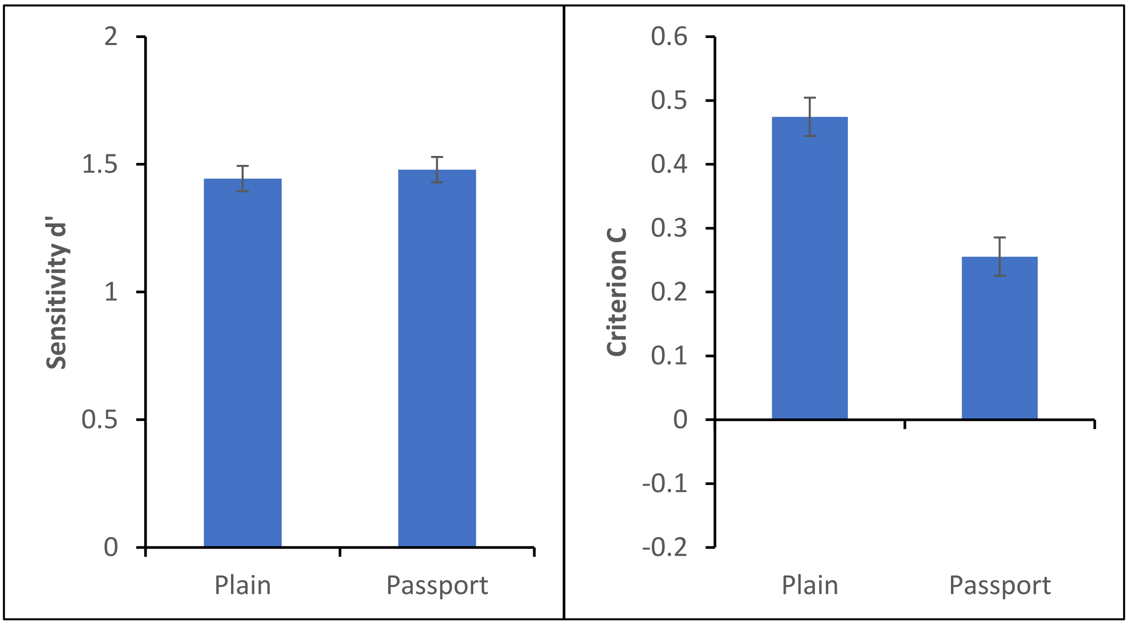

Figure 4.4 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 12. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

These results showed participants were biased to make more 'same person' responses to the face pairs embedded in a passport frame than isolated plain faces, which was consistent with McCaffery and Burton's (2016) and with previous experiments. The bias effect is quite stable with different face sets, even with photos that are not commonly seen in an ID document, i.e. those that do not conform to the normal requirement of a full-face view and neutral expression. This may suggest that the bias effect is only related to the document frame itself but not to the photo image inside.

The model faces induced a strong bias for saying ‘different people’, for the criterion C for plain faces was much more positive than those in our earlier experiments. This can be seen in the accuracy data as well, which shows a shift in the accuracy pattern compared with Experiment 11. This confirms our prediction that this is a difficult set, with an inherent bias to reject matching pairs. Although the values of criterion for plain faces vary across experiments (from Mean = -0.20 to Mean = 0.22), they are all much smaller than the one in the present experiment (Mean = 0.47). The consistent bias found using unfamiliar faces suggests that the effect is induced by a readable card, regardless of baseline criterion.

This result may indicate that the bias effect is specific to unfamiliar faces rather than familiar faces. The face set variation is not a key factor that affects the bias. What is interesting, is that the face images we embedded in a passport are unlike the normal ID-photos, but even in this circumstance, viewers are biased by the presence of a passport. It seems that the bias effect generalises more widely than we had previously hypothesised.

In each of the experiments reported so far, the probability of showing faces depicting the same person was always equivalent to the probability of that depicting different people, which is a 50-50 chance. However, everyday usage of photo-ID is very different – most of the time people present true ID, and fraudulent presentations are quite rare. Might this affect the viewers, leading them to assume a high proportion of matches, even when our experimental manipulations do not follow the proportions generally experienced? In the next experiment, we are going to introduce a low proportion of mismatches to mimic a more real-life condition. We would like to see if the document bias exists in such a situation.

4.4 Experiment 13

Introduction

Officers dealing with photo-ID matching typically encounter an unbalanced proportion of matches and mismatches. The probability of fraudulent IDs (i.e., a mismatch with the holder) is quite low. There is a low-prevalence effect (LPE) frequently seen in visual search tasks, where participants are less likely to detect a target than when prevalence is higher (e.g., targets presented in 1% or 10% of the trials compared with 50% of trials, Wolfe et al., 2005). While the LPE is well demonstrated with visual search tasks where the targets are letters and objects, there have been controversial results on whether the LPE exists in face matching tasks (Bindemann et al., 2010; Papesh & Goldinger, 2014). Despite this inconsistency, the frequency of mismatches does seem to affect performance and criterion in some way. So, the document bias we found previously may be different or may not exist in settings with low-prevalence of mismatches.

Previous research has found that in low-prevalence condition (i.e., mismatches are rare), participants tend to adopt a more conservative criterion than in high-prevalence conditions. When mismatches are rare, but participants have not been warned about this, they may sometimes make errors on a high proportion of 'match' trials because they expect to make similar numbers of 'match' and 'mismatch' responses (Bindemann et al., 2010). In this low-prevalence condition, whether presenting an additional document frame affects criterion (as we found in previous experiments) is worth investigating, because the low-prevalence condition is more real-life related. If under the condition that the fraudulent IDs are presented rarely, we still find a bias effect of accepting 'same person' more, then this will alert passport officers that they should be aware of such bias.

Therefore, in this experiment, we would like to examine the bias effect in a more practical condition by using a low mismatch prevalence (10% mismatches and 90%

matches, Papesh & Goldinger, 2014), to test if viewers will still give more 'same person' responses when seeing passports than isolated faces.

Here, we will only include one low-prevalence condition, unlike previous research using both low- and high-prevalence conditions. We consider the following reasons: first, we focus on whether the bias induced by a document exists in a more realistic setting rather than whether the bias is affected by frequency in general. Second, we have tested the bias effect using the 50% - 50% proportion many times, and have found consistent results. It will be redundant to include this condition again in the present experiment. Third, it has been shown that presenting a high-prevalence block affects the criterion of the following low-prevalence block (Wolfe et al., 2013). To exclude interference, we chose not to expose participants to a high-prevalence condition, but retain the same low-prevalence conditions throughout the experiment.

Method

Participants

This study was run online. Seventy-two registered members (39 females, aged from 18 to 35, mean age = 25.5) from Prolific (https://www.prolific.co/) participated for a small amount of money. All reported normal or corrected to normal vision. Each participant completed an online consent form before the experiment.

Stimuli

One hundred and twenty face pairs were randomly chosen from the KFMT (60 face pairs) and GFMT (60 face pairs). We doubled the number of face pairs, compared to most of the experiments above, so that the low prevalence mismatches (10%), would nevertheless give enough trials in total. For each face pair, we created two conditions: plain faces and faces in a passport (see Figure 4.5). The face images from two different face sets were adjusted to similar sizes. The distance between faces in each pair was constant across all conditions.

Figure 4.5 Example face pairs from the KFMT of two presentation conditions in Experiment 13. Face pairs from the GFMT were manipulated in the same conditions. Each of these examples shows different identities

Design and Procedure

In a within-subjects design, all participants completed two face matching blocks, as illustrated in Figure 4.5: plain faces and faces embedded in passports. In order to preserve a mismatch frequency of 10%, for each participant, 12 mismatch pairs (half from KFMT) and 108 match pairs (half from KFMT) were presented. Therefore, for each block, there were 60 face pairs (54 matches and 6 mismatches, with half from KFMT and half from GFMT). Within each block, the images were randomly presented. The mismatch stimuli were rotated around participants, so that all face pairs appeared equally often in the plain and passport conditions. Participants' task on each trial was to indicate whether the face pair showed the same person or different people by pressing corresponding keys on a keyboard. Each face pair was displayed until a response was made. The order of blocks was counterbalanced across participants.

Results

Figure 4.6 shows sensitivity (*d'*) and criterion (C) for matching decisions. Repeated-measures ANOVA showed no significant effect of condition on *d'* ($F(1,71) = 0.08$, $p > .05$, $\eta_p^2 < .01$), but a significant effect on C ($F(1,71) = 4.09$, $p < .01$, $\eta_p^2 = .05$). Mean criterion was positive in both cases, but less so when participants saw faces embedded in passport frames ($M = 0.04$) than when they saw faces alone ($M = 0.17$).

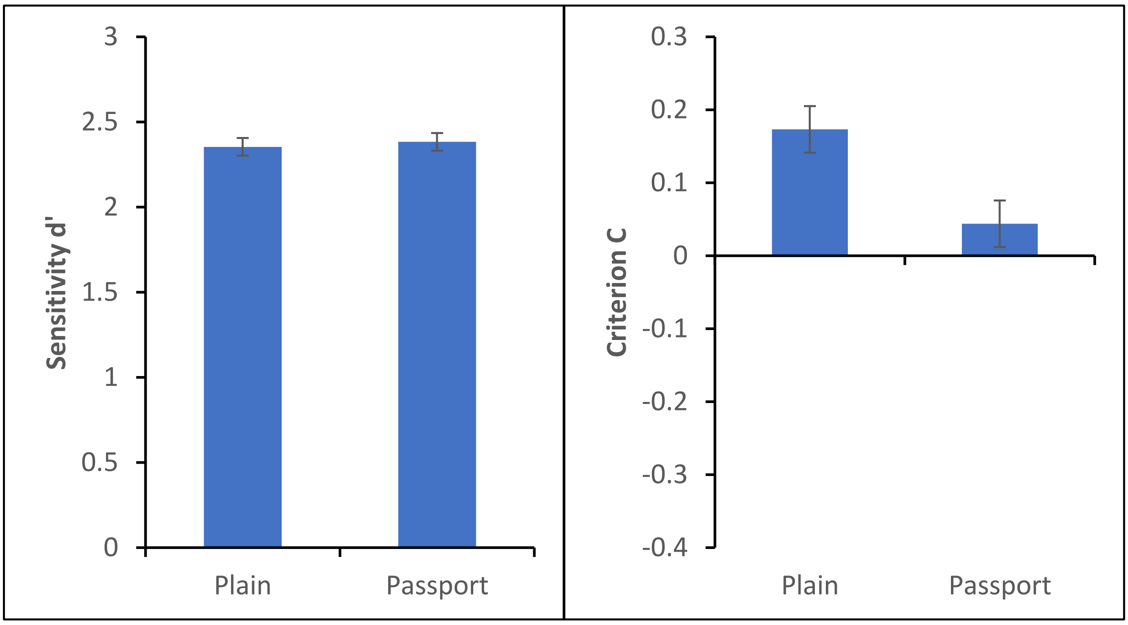

Figure 4.6 Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 13. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

These results show participants responded ‘same person’ more often when one of the faces was embedded in a passport compared to isolated faces, even in a low-prevalence of mismatches (10% of the trials). This replicates the bias effect found in previous experiments, but using a different frequency of mismatch pairs. The overall sensitivity still shows no difference between plain faces and faces in passports, which is consistent with previous experiments that the additional frame does not affect the performance, but induces a bias.

In this experiment, we decreased the proportion of mismatch face pairs in the task to mimic a real-life setting more closely where fraudulent documents are rarely encountered. Although the frequency is still much bigger (10%) than it actually is in real life (less than 1%, HM Passport Office, 2014), it takes a step towards a reflection of reality. Even with this low-prevalence of mismatches, participants still tended to respond more ‘same’ when they saw faces in passports, which somehow reflected the experience we had – the fraudulent IDs are rare. This brings back the concern of officers who deal with matching photo-IDs: a bias to accept pairs with ID cards, which is the forensically worrying mistake when checking documents such as passports or driving licences. This also highlights the importance of investigating the cause of the bias.

We have conducted many experiments on manipulating the visual properties of frames to examine the explanation of perception interference (Experiments in Chapter 2 and 3); we also tried to embrace more real-life experiences to examine the possibility of experience and expectations (Experiment 11-13). Next, we consider whether the bias may be specifically related to the task, i.e., to decide if two images show the same person or different people. In the next experiment, we would like to test the basis of the bias: if we change the task, will the bias still exist?

4.5 Experiment 14

Introduction

In real life, we sometimes feel the photos in ID documents do not look like us, due to the lighting, camera, time gap, etc. Even though the ID-photo is considered as a poor likeness, we do recognize the photo belongs to ourselves. This often happens with familiar faces such as our friends or family members. But for unfamiliar faces, we do not have a previous representation (Bruce & Young, 1986), and so we cannot decide if an image is a ‘good likeness’ or not (Ritchie, Kramer, & Burton, 2018), because we do not know the person. We can only say if the two images look similar when it comes to unfamiliar faces. Deciding whether two face images show the same person or not is a different process from evaluating the similarity of two face images. In previous experiments, we always asked participants to make an identity judgement of two face images - would we observe different results if we asked them to just rate the similarity between the two faces without explicit decisions on identity?

Similarity ratings between faces have often been used in the comparison between automatic face recognition algorithms and human performance on face matching tasks, where the algorithm returns similarity scores on pairs of images (e.g., O’Toole et al., 2007). For face recognition algorithms, the higher a similarity score is, the more likely that the two images show the same individual. Does human performance reflect the same pattern?

When providing three sets of face pairs (good, moderate, and poor similarity) based on the similarity ratings generated by an algorithm, the matching performance of human participants degraded from the good to the moderate, then to the poor similarity set (O’Toole et al., 2012). It has also been found that the accuracy of matching was the highest when the target face and the array of faces were visually similar (Sandford & Ritchie, 2021). So, it seems that the similarity of a face pair relates to the performance of accurate matching.

In this experiment, we are going to present participants with pairs of isolated faces and pairs with one face in a passport, and ask them to rate how similar they think the two faces are. The biggest difference between the present experiment and the previous ones is the response. Collecting responses on similarity ratings allows us to separate the representation and the decision in some way. If the bias is derived from a change in perceived similarity, then we will expect to see a higher similarity rating between faces in passports compared to pairs of plain faces. If the basis of the bias relies on the responding process, then we will expect no differences between the two conditions. Also, whether there is a correlation between the similarity rating and the accuracy responses will be of great interest, in order to establish the relationship between these tasks.

Method

Participants

Twenty-eight students (26 females, aged from 18 to 33, mean age = 20.5) from the University of York participated for course credit. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

Sixty face pairs from the KFMT were used, which were the same as those in Experiment 13. We used a 2 (presentation type: plain faces, passports) × 2 (match type: same, different) within-subjects design. Participants performed two blocks (plain, passport, see Figure 4.7). In each block, they saw 30 face pairs (15 matches and 15 mismatches), with the order of blocks counterbalanced across the study.

In each trial, participants saw a face pair with a Likert scale below. Instead of making a same/different decision, participants were asked to rate the similarity between two face images on a scale of 1 (Extremely dissimilar) to 9 (Extremely similar). They did not need to make judgements on whether the faces shown in pairs were the same person or different people, so there were no right or wrong answers. Each face pair was displayed until a response was made.

Figure 4.7 Face pairs from two presentation conditions in Experiment 14. Each of these examples shows ‘different’ identities.

Results

Figure 4.8 shows the average similarity rating scores for each condition. Repeated-measures ANOVA showed no significant effect of presentation type ($F(1,27) = 2.72$, $p>.05$, $\eta_p{}^2 = .09$) or the interaction between presentation type and match type ($F(1,27) = 0.08$, $p>.05$, $\eta_p{}^2 < .01$). Only a significant difference was found in match type ($F(1,27) = 325.09$, $p<.001$, $\eta_p{}^2 = .92$), which showed the rating was larger for match trials (same person) than mismatch trials (different people) in both plain faces and passports conditions. We also did a by-image analysis, which shows a similar pattern as the by-subject one (see Appendix 4).

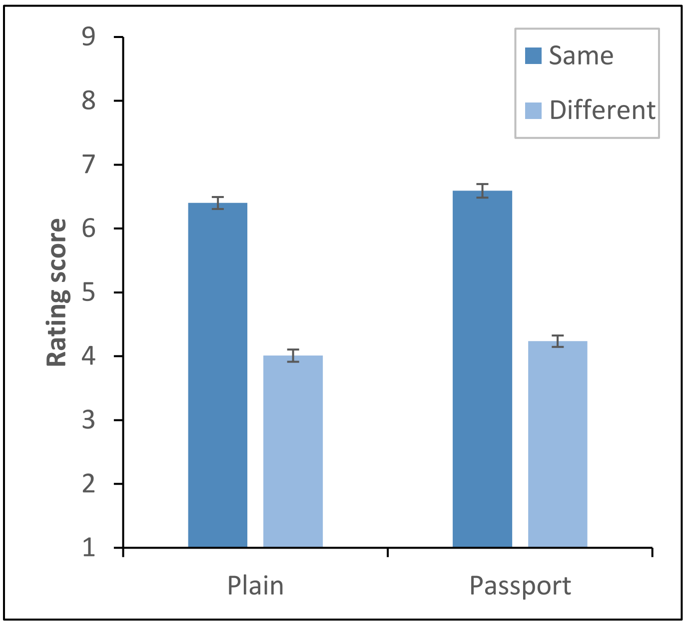

Figure 4.8 Rating scores for match types under each presentation condition. Error bars represent within-subjects standard error (Cousineau, 2005).

Better performance on more similar pairs was found in previous research (Sandford & Ritchie, 2021), suggesting an association between performance and similarity. To investigate this further, we performed a correlation between the perceived similarity and the accuracy of same/different response (data from Experiment 2) based on each face pair. Results showed significant positive correlations between similarity ratings and accuracy of face pairs showing same person for both plain faces ($r(60) = .731$, $p<.001$) and faces in passport ($r(60) = .736$, $p<.001$). Accordingly, significant negative correlations between similarity ratings and accuracy of face pairs showing different person were found in both conditions (plain faces: $r(60) = -.492$, $p<.001$; passports: $r(60) = -.687$, $p<.001$).

Discussion

From the results above, we only found significant differences in similarity ratings between match pairs (i.e., presenting same person) and mismatch pairs (i.e., presenting different people), suggesting face pairs depicting the same person were rated more similar than the face pairs depicting different people. Whether the face pair was shown in passport or not did not make a difference on perceived similarities. This suggests embedding one of the faces into a document does not change the perceived similarity between the two faces.

In particular, the correlation found in the present experiment should raise our attention. For match pairs, if the two faces are considered more similar, this leads participants to respond more ‘same person’, which increases the accuracy; for mismatch pairs, if the two faces are also considered more similar, this impairs the performance. Combined with previous studies on face matching, interestingly, viewers did not take the two faces as more similar, but they did respond more ‘same person’ to those face pairs when the document was introduced. It seems that the document bias effect only happens in the process when making decisions on face identities, even though there is a strong correlation between similarity ratings and accuracy. The more similar the face pairs are rated, the higher accuracy on responding ‘same person’ people get, and the opposite is true for face pairs showing different people.

This result shows that the bias effect is not derived from an actual change of perceived similarity but from a change of decisions, which indicates the bias may not relate to perceptual processing, but to a decision-making process.

4.6 General discussion

In this chapter, we varied the faces and the tasks to seek insights about the basis of the bias observed across previous experiments. We observed that the bias effect persists across face sets, including one which does not meet the rules of an acceptable photo when people apply for a UK passport. It seems that everyday expectations about photo-ID are not necessary to induce the bias. Likewise, it is surprising that we did not get any bias effect with familiar faces, considering the expectation of an ID card (Experiment

11). We have inferred that the bias found in unfamiliar face matching may be related to the experience of seeing familiar people's documents. However, we did not observe the bias when testing with familiar faces. As the Bruce and Young model (see Figure 4.9) suggests, there are differences between the coding of familiar and unfamiliar faces. Processing familiar faces have access to face recognition units, while processing unfamiliar faces access visual processing code. It is possible that the document context affects the processing of unfamiliar faces, but has no effect on familiar faces.

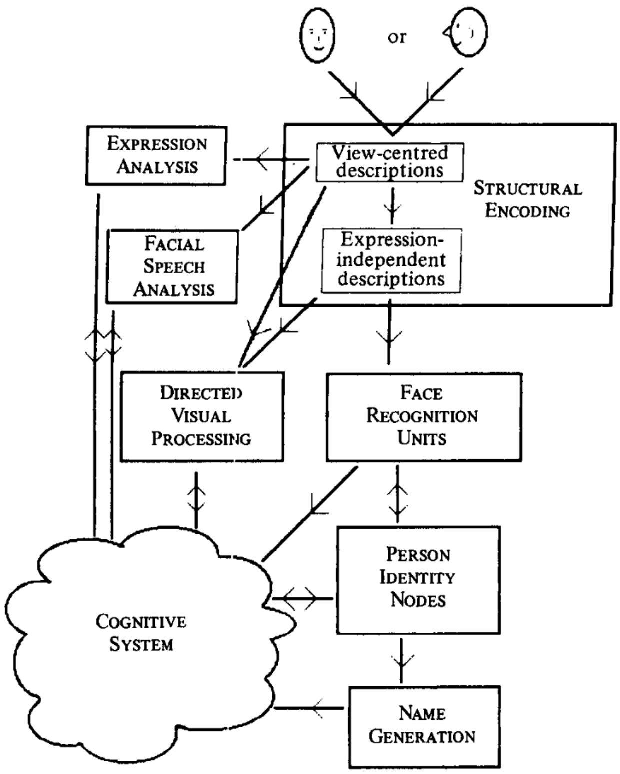

Figure 4.9 A functional model for face recognition proposed by Bruce and Young (1986).

It is also interesting to note that the direction of the bias (making more 'same person' responses) has also been observed in previous work on face matching that attempts a more realistic (rather than lab-based) context. For example, Kemp et al. (1997) asked supermarket cashiers to check the photo-credit cards of their customers. Results showed the performance was poor (67% accuracy), and the majority of the errors happened in accepting foil ID cards. The same effect is observed with professionals. White, Kemp, Jenkins, Matheson, and colleagues (2014) asked passport officers to check the photo-ID of holders and found the error rates were surprisingly high, the majority of which were made in accepting mismatches. So, it seems that experience may support the bias observed here, particularly when taken alongside an expectation that fraudulent ID use is likely to be rare. We did find even under the low-prevalence of mismatches, that the bias effect exists (Experiment 13). Although the prevalence we used was still much larger than that in real contexts, and we only tested novice participants rather than professionals, this reflects the reality in which mismatch pairs were rarely seen and provides evidence that the bias may truly exist in the officers dealing with face matching with documents.

The last important insight is from the similarity rating of faces, where we failed to detect a bias induced by an additional document frame (Experiment 14). Unlike tasks requiring identity responses, presenting a document frame did not make the two faces appear more similar. Although the similarity rating correlates with the matching performance, the document does not bias the similarity rating, but only biases the matching decisions. This indicates that the bias effect may relate to decision-making rather than the perceptual aspects of the task. Therefore, in the next chapter, we are going to run more rating tasks on perceived social traits to see if the document bias effect exists when it comes to perceptual processing.

Chapter 5 – First impressions in documents

5.1 Introduction

Previous studies we have reported are all about face matching, from which we found that matching a face image to a face embedded in a document was different from matching two isolated plain faces. The existing document context will bias viewers' decisions to give more 'same person' responses. This is possibly due to an interaction between processing faces and processing adjacent information within a card context when performing matching tasks. Face processing can be influenced by a number of different contexts (Koji & Fernandes, 2010; Rainis, 2001; Young et al., 1986), so it is reasonable to infer documents may affect other kinds of face processing. However, the presence of a document seems not to affect a rating task (Experiment 14 in Chapter 4) in which participants do not have to make a match/mismatch decision, but only rate the similarity of pairs of faces. In this chapter, we focus on the possible effects of a document on simple ratings of single faces. Instead of asking participants to make a matching decision, we simply ask them to rate faces for perceived traits, when presented either in isolation or embedded in a document.

Faces include rich information apart from identity-related information like age and gender, there are other perceived information for us to infer such as personalities (e.g., if this person is easy-going) or social traits (e.g., can I trust this person). People can form quite stable first impressions on a face (Bar et al., 2006; Todorov et al., 2010) and evaluations of faces generally show high agreement across perceivers (Willis & Todorov, 2006; Cogsdill et al., 2014).

What makes these social inferences important is that they have a great influence on social behaviour and decisions, such as 'what is beautiful is good' (Little et al., 2006), predicting voting decisions (Ballew & Todorov, 2007) and affecting court decisions (Wilson & Rule, 2016). The importance of social traits makes it particularly significant to pay attention to these factors, because even a subtle change of the social evaluations may have a large influence on behaviour.

On the other hand, emotion, stereotype, or previous experience will also affect social evaluations on faces. For example, when photos of men are labelled as being married – the ‘wedding ring effect’ (Eva & Wood, 2006) or simply paired with a female photo alongside (Waynforth, 2007), women viewers tend to rate the men as more attractive than those who are labelled as being single. Faces displaying positive emotions are evaluated as being more trustworthy (Krumhuber, Manstead, Cosker, Marshall, Rosin, & Kappas, 2007). The expectation of a person may also change the evaluation of the person independently of the face (Kelley, 1950).

Koji and Fernandes (2010) asked participants to rate faces that were shown within a scene, on a scale of -3 (negative) to 3 (positive). The faces were all neutral but the scene as background had different emotion types. They found that people were biased by the scene even when it is task-irrelevant. So, it is possible that an additional document context, which is also task-irrelevant, could affect the perceived traits of faces.

Here we ask whether the cause of the bias effect reported in earlier chapters might be viewers’ first impressions of the face. Could a document affect these first impressions? For example, the face may look more trustworthy in a document, rather than in isolation, possibly affecting a viewer’s matching decision. If a document does change first impressions, for example, the trustworthiness of the face, then it is important to establish this. According to Dangerous Decisions Theory (DDT; Porter & ten Brinke, 2009), perceptions of a defendant’s appearance may heavily bias the evaluation process about a person. It may also affect the subsequent court decisions, because of preconceived notions like ‘ugly is bad’ (Griffin & Langlois, 2006) or stereotypical beliefs may introduce a systematic bias. Thus, it is important to test if a document will affect first impressions of faces, both as a possible explanation for our observed bias, and as a more general issue of wider importance.

In this chapter, we are going to investigate if embedding a face in a document frame will affect perceived traits - first impressions of faces, including attractiveness, trustworthiness, and dominance. These are the three dimensions which have repeatedly been shown to underlie perceived face trait evaluations (Oosterhof & Todorov, 2008; Sutherland et al., 2013). We will ask participants to rate each face for these three traits with or without a passport frame, in order to establish whether the presence of a frame affects these attributions. If the presence of a document does show such effect, then the document bias found in previous matching tasks may be influenced by the perceived traits. In particular, the perceived trait of trustworthiness is important here. If a face is rated high in trustworthiness, it seems likely that this could lead to more 'same person' responses in a matching task.

5.2 Experiment 15

Introduction

In this experiment, we are going to test whether first impressions on faces are different when faces are embedded in a passport document. We show participants isolated face images or faces in passports and ask them to rate the attractiveness, trustworthiness, and dominance for each face. If a passport affects first impressions on faces, then we will find a difference between plain faces and faces in passport on attractiveness, trustworthiness, or dominance.

Method

Participants

Thirty-seven students (33 females, aged from 17 to 20, mean age = 18.6) from the University of York participated for course credit. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

Sixty face images were selected from the Kent University Face Database (KUFD). These images show students at the University of Kent, with a large variation in expression, pose, and camera lens. Because expressions affect first impressions, for example, a smile is a diagnostic of trustworthiness (Dotsch & Todorov, 2012), we only selected photos with a neutral expression with mouth closed.

For each face, we created a passport frame (see Figure 5.1), which was used in previous experiments. This was a one-factor (presentation type: plain, passport) within-subjects design. There were two blocks (plain faces, passports) in total. In each block, participants saw 30 faces. They were asked to rate each face image for trustworthiness, dominance, and attractiveness on a scale from 1 (not at all) to 9 (extremely). Each face was presented individually at the centre of the screen with the rating scale positioned below the image. Participants rated all the faces in that block for one attribute at a time. After they rated three attributes (which were presented in random order), they moved to the next block. The order of the two conditions was counterbalanced across participants.

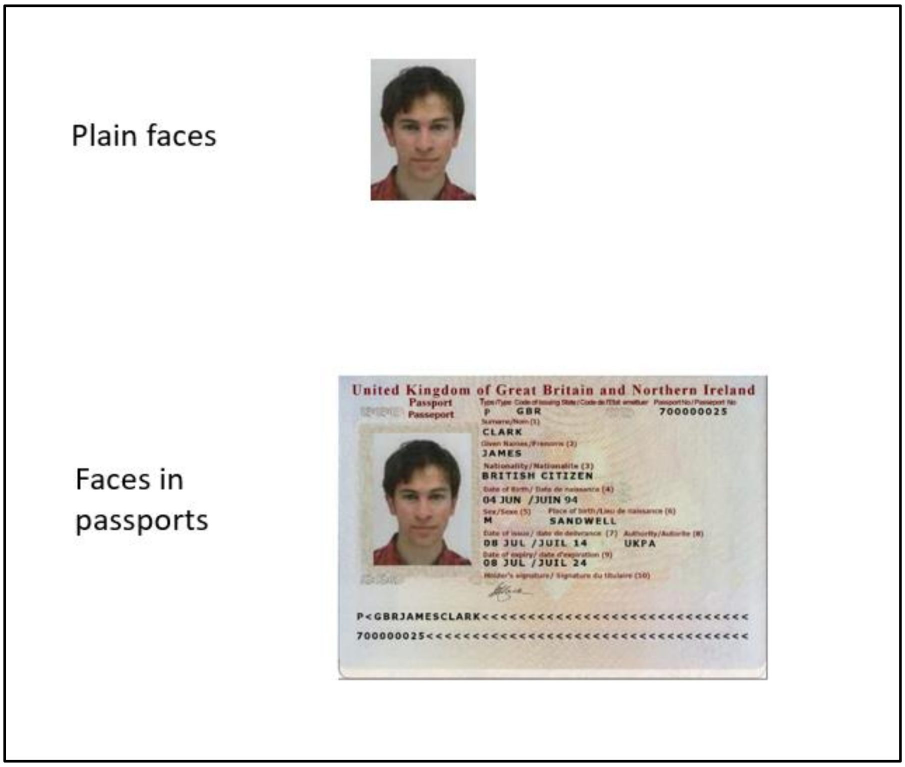

Figure 5.1 Faces from two presentation conditions in Experiment 15. Participants will only see one face identity once, either presented with a plain photo or in a passport.

Results and discussion

We analysed each trait (attractiveness, trustworthiness and dominance) separately by conducting three paired t-tests (see Figure 5.2). For attractiveness, we found a significant effect of passport ($t(36) = 3.33$, $p < .01$, Cohen's $d = .55$): participants rated faces in passport ($M = 4.79$) as more attractive than faces without passport frame ($M = 4.46$). There were no significant results between plain faces and passport for trustworthiness ($t(36) = 0.85$, $p > .05$) or dominance ($t(36) = 1.34$, $p > .05$).

We also ran a by-image analysis showing a similar pattern, which suggests a general lack of effect of additional document on first impressions. Full details are reported in Appendix 5.

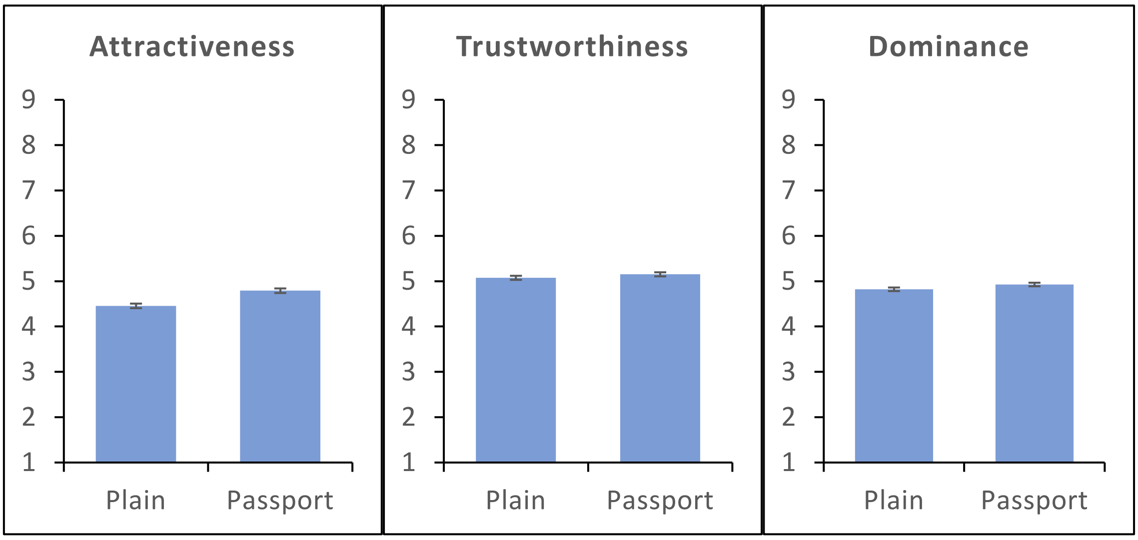

Figure 5.2 Rating scores for two conditions under each trait. Error bars represent within-subjects standard error (Cousineau, 2005).

Our results show that a document frame has no effect on trustworthiness and dominance. Although we found adding a passport frame increased the attractiveness rating for faces, the small difference between plain faces and faces in passport (only 0.33) and the effect size suggest the result is weak. The explanation will be discussed in the general discussion.

5.3 General discussion

In this chapter, we tried to investigate whether a document may affect social evaluation such as first impressions on faces. Our results showed that there were no significant differences in ratings to isolated faces and faces shown in passports, when viewers were asked to judge trustworthiness and dominance, but we found a small effect for attractiveness.

The presence of a document seems not to affect the perceived traits on faces overall, but why does an increased attractiveness rating emerge? It has been found that faces against a background that is statistically similar to natural scenes are rated as more attractive (Menzel, Hayn-Leichsenring, Langner, Wiese, & Redies, 2015). It is probable that the visual property of a document is closer to natural scenes than a white plain background, which results in an increased attractiveness rating. If this is the case, we suggest this increment related more to a visual pattern, rather than an effect of a document, or a social process of a document. In face matching tasks, the document has to be ID-like to bias viewers, indicating the concept of an ID is critical. It may be different from what we found here.

It is also possible that the increased attractiveness rating is due to some systematic variance, but not actually induced by a document, because the difference between plain faces and faces in passport is minimal. This suggests there is no general effect of embedding faces in documents. Then why did we fail to find an effect of an additional document on first impressions of the faces for the dimensions of trustworthiness and dominance?

One possible reason is that the document frame cannot influence first impressions due to the speed of processing required to perceive the document. It is well-established that first impressions of trustworthiness and dominance are formed in a very short time, within 100 milliseconds (Todorov et al., 2010; Todorov, Pakrashi, & Oosterhof, 2009). Perhaps this means that viewers form the evaluation before they process the additional document context. A document frame represents rather a complex stimulus, and it may not be as effective as other manipulations which attract attention and evoke emotional reactions (Koji & Fernandes, 2010).

Another possibility is that the document bias does not affect the rating tasks, i.e. tasks in which participants evaluate single images rather than making explicit same/different judgements. Rating a face requires a different process from making the match judgement, and it is possible that the effects of documents on face processing are largely at the decisional stage, and not at the perceptual stage of the task. There is some support for this idea from earlier experiments. For example, when the ‘same/different’ task was changed to a perceived similarity task, the document did not bias the performance (Experiment 14 in Chapter 4). In contrast, all the experiments which have shown a document bias require participants to make an explicit match/mismatch decision.

If the biasing effect of embedding a face in a document is based on decisional, rather than perceptual processes, this raises an interesting possibility. It suggests that the bias might be overcome by training people to adopt different decision criteria. Note that the effect shown in previous chapters is never to change the overall accuracy of matching performance, but only to change the response bias. It is an interesting question for future research to establish whether viewers can be trained to adjust their criterion in matching without affecting accuracy.

Chapter 6 – Face memory in documents

6.1 Introduction

In previous chapters, we found that embedding a face into a readable document context affected viewers' responses in face matching tasks, but did not affect their first impressions ratings. In addition, when we asked viewers to rate the *similarity* of two faces instead of judging the identities, there was no biasing effect of a readable document. These findings suggest the document context may primarily affect the process of identity *decisions*, for example, judging if the two images show the same person or different people. Although it is reasonable to infer that the bias affects the process of decision, it is still unclear at what stage in the process this biasing occurs.

In a face matching task, the document may affect people when they encode faces to compare; it may also take effect when they make the final decision. As the two faces are shown simultaneously in a face matching task, the stimulus that viewers encode is the same stimulus that viewers decide on, and so we cannot separate the two processes (encoding and deciding). In this chapter, we try to separate these by adopting an old/new procedure in face recognition memory. In this paradigm, participants are asked to remember several faces, then they are tested with more faces including new faces and the faces they have seen in the earlier phase. Their task is to decide whether the test face is a 'new' face or an 'old' face. Using this design, while manipulating whether faces are presented in isolation or in ID-card contexts, we should be able to isolate where the bias happens.

Will additional context affect the encoding stage in face recognition memory? A large number of studies using images or emotional pictures have demonstrated that this can occur (e.g., Van den Stock & de Gelder, 2012). For example, Rainis (2001) asked participants to remember several faces; each face was presented in a negative/positive/neutral context (e.g., concentration camps inducing negative emotion).

Results showed that a negative context impaired the recognition performance for these faces, compared with positive or neutral context.

In addition to image context, presenting words along with faces in the encoding phase affects recognition. Presenting affective (negative/ positive behavioural descriptions) or even neutral information alongside the face improved the recognition performance compared to presenting single faces (Mattarozzi et al., 2019). Even though there seems to be a puzzling effect that negative words enhance face memory while negative scenes impair it, these all demonstrate the malleable property of face memory. Words that induce social categories also show an influence on face recognition memory. Shriver et al. (2008) asked middle-class participants to remember faces that are presented in either wealthy or in impoverished contexts. They were then tested with plain faces and results showed better recognition for faces from wealthy contexts than those from impoverished contexts. It seems that participants tend to encode the faces with the context (images or plain words) altogether, which then gives rise to context affects in subsequent recognition of the faces.

There is little research on whether the context affects decision processes in face recognition. Hourihan, Fraundorf, and Benjamin (2013) found better recognition performance in the encoding phase on faces with an own-group name than faces with an other-group name. However, if the names were presented alongside the faces in the test phase (i.e., to recognize whether they had learnt the faces before), there was no difference between faces paired with own-group names and faces with other-group names. It seems that presenting a context at the decision stage in a memory task produces no effect, which is interesting because the document context seems to affect the identity decision in our matching tasks.

It should be noted that, in addition to the types of contexts described above, there is also a well-established ‘context effect’ in memory tasks. A context effect is observed when better performance of recognition is found when the original study context is preserved than when it is changed (e.g., Smith, Glenberg, & Bjork, 1978). Early research by Watkins, Ho, and Tulving (1976) demonstrated the semantic context effect on face recognition memory. Watkins et al. asked participants to remember a series of faces with short descriptive phrases, then tested them with more face-phrase pairs, in which half of the faces were paired with different phrases. Participants' task was to decide whether the face was seen before, and whether the phrase was studied with the face. Results showed an advantage of preserving context (i.e., the same phrase with the face) over changed context.

However, the bias effect induced by a document context that we observed in previous experiments is *not* the 'context effect' in the sense of encoding/recall compatibility. In this chapter, we are not preserving or changing contexts between learning and memory decisions. Instead, we are aiming to establish whether the ID-context effect reported in earlier chapters has its effect in the encoding of the face stimuli or at the decision stage. In the following experiments, we first test whether the bias effect takes place in the encoding or the decisional process of face identities by using a traditional recognition-memory task (Experiment 16A and 16B). We then adopt an immediate memory task in order to reduce memory load, and to make a comparison with a matching task (Experiment 17). Last, we test the possible effect of time passage on the document bias effect in face matching (Experiment 18).

6.2 Experiment 16A

Introduction

A traditional face memory task consists of two phases: a learning (encoding) phase and a test phase. Participants were asked to remember the faces shown in the learning phase and subsequently shown a larger set, which included those seen in the learning phase. Their task in the test phase was to decide whether or not they have seen each face during the earlier experimental phase.

In the first two experiments, we are going to test whether the document bias effect found in previous face matching experiments is related to encoding the document with the face, or related to making identity decisions. Accordingly, in Experiment 16A, we ask participants to learn plain faces or faces in documents, and test with only plain faces (see Figure 6.1). If the document context affects viewers in the encoding of faces with documents, then we expect to find a bias of responding more ‘old’ faces (i.e., a liberal criterion) when they learn faces in documents. In Experiment 16B, we ask participants to learn only plain faces, but test with plain faces or faces in documents (see Figure 6.1). If the document affects the final decisions, then we expect to see a bias when the test face is shown in a document.

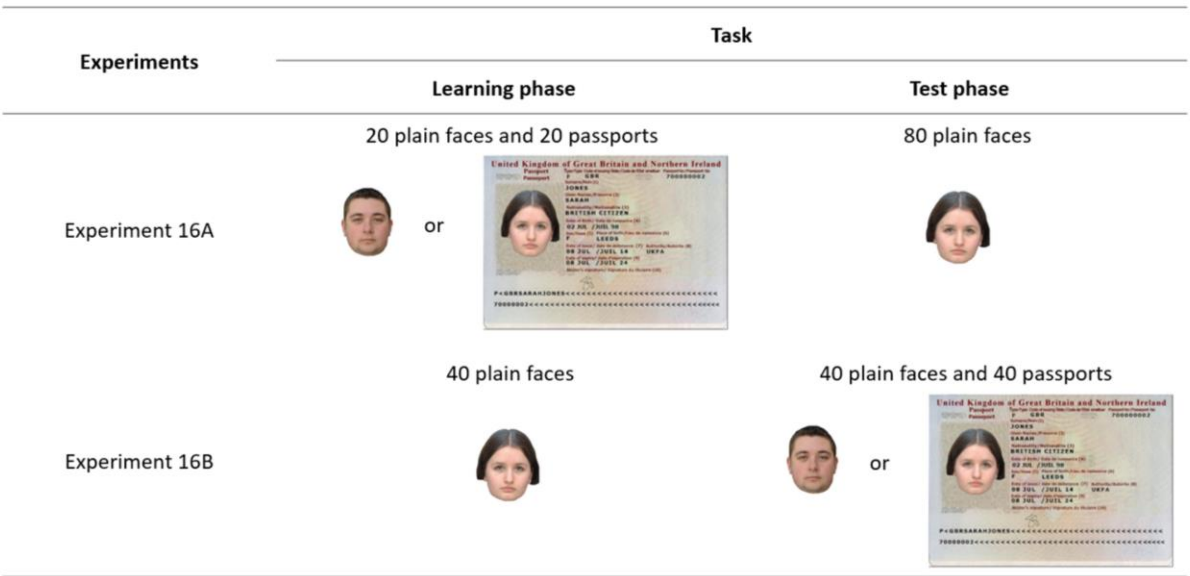

Figure 6.1 The procedures of the memory task in Experiment 16A and 16B. The passports were only shown in the learning phase of Experiment 16A and in the test phase of Experiment 16B.

Method

Participants

Thirty-five students (30 females, aged from 18 to 34, mean age = 20.9) from the University of York participated for course credit or an amount of money. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

We selected 80 face images (half male and half female) from the Glasgow University Face Database (GUFD, Burton et al., 2010). These face images were taken by the same camera with similar size and the same lighting conditions. For each face, we created a passport frame (see Figure 6.2).

This was a one-factor (presentation type: plain, passport) within-subjects design. There were two phases in this memory task (see Figure 6.1). In the learning phase, participants were instructed to remember 20 plain faces and 20 faces in passport, which were presented in random order at the centre of the screen. Each face was displayed for 5 seconds following a 500ms fixation. After the learning phase, a distractor task was used. A word search puzzle was presented to participants on a A4 paper, which lasted for about 2 minutes. Then 80 plain faces (40 old and 40 new) were presented one at a time in the test phase. Each test face was displayed until a response was made. Participants' task was to judge whether the face was an old face or a new one as correctly as possible. They were instructed to press "F" for an old face and to press "J" for a new face.

Figure 6.2 Exemplars from two presentation conditions in Experiment 16A and 16B. The faces were selected from GFUD.

Results

The overall accuracy for plain faces was 68.6%, and for faces in passport was 68.1%. We regarded participants' accurate responses of 'old face' as *hits*, and responding 'old face' to new faces as *false positives* to calculate the sensitivity (*d'*) and criterion (C) for recognition performance (see Figure 6.3). Repeated-measures ANOVA showed no significant effect of presentation type on *d'* ($F(1,34) = 0.03$, $p>.05$, $\eta_p^2 < .01$) or C ($F(1,34)=0.32$, $p>.05$, $\eta_p^2=.01$).

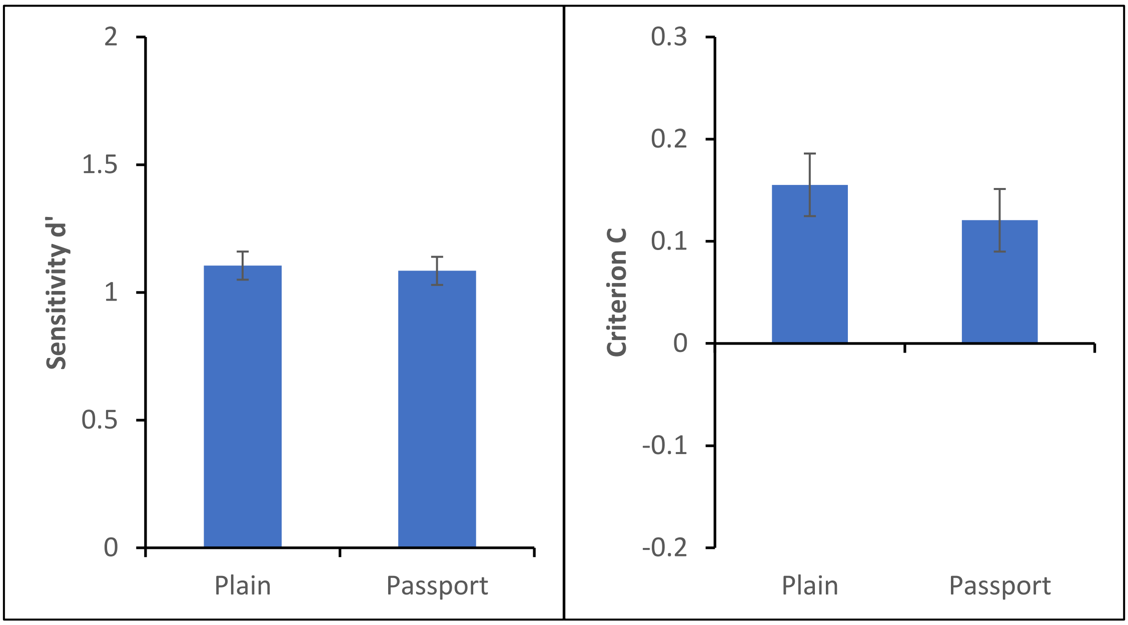

Figure 6.3 Sensitivity (*d'* on the left) and criterion (C on the right) for recognition in Experiment 16A. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

We did not find any difference between plain faces and faces in passport conditions. The presentation of a passport in the learning phase did not affect participants' later recognition performance and did not bias their responses.

The overall performance for both conditions (around 68% accuracy) was at a normal level. Participants did not treat the faces shown in passport as completely new faces to them, for the *d'* of passport is significantly above 0 ($t(34)=9.69$, $p<.001$) and has no significant difference from plain faces. This suggests that participants can perform the task well, indicating they may treat the faces in passport the same as plain faces, by ignoring the passport context.

It seems that the encoding context does not affect memory performance, which implies that the document does not affect the encoding of a face in a way that results in bias for a matching task. In the next experiment, we turn to manipulate the passport in the test phase of the task.

6.3 Experiment 16B

Introduction

In this experiment, we change to manipulate the faces in the test phase to see if the additional document will affect viewers' decisions on whether they have seen the face before. That is to say, whether the presence of a passport frame affects the identity comparison decisions.

Method

Participants

Twenty-eight students (27 females, aged from 18 to 22, mean age = 19.3) from the University of York participated for course credit. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli, Design and Procedure

The stimuli were the same as those in Experiment 16A (see Figure 6.2). The design and procedure were similar to those in Experiment 16A, except that the manipulation was in the test phase (see Figure 6.1). In the learning phase, participants were instructed to remember 40 plain faces that were presented sequentially at the centre of the screen. Each face was displayed for 5 seconds following a 500ms fixation. After the learning phase, a distractor task was used. Then 80 faces (40 old faces and 40 new faces) were presented once at a time in the test phase. Half of the faces were plain faces (including 20 old faces and 20 new faces), and the other half were faces in passport. Each test face was displayed until a response was made. Participants' task was to judge whether the face was an old face or a new one as correctly as possible.

Results and discussion

The overall accuracy for plain faces and faces in passport was 70.8% and 71.3% respectively. Figure 6.4 showed sensitivity (*d'*) and criterion (C) for recognition performance. Repeated-measures ANOVA showed no significant effect of presentation type on *d'* ($F(1,27) = 0.03$, $p>.05$, $\eta_p{}^2 < .01$) or C ($F(1,27)=0.21$, $p>.05$, $\eta_p{}^2<.01$).

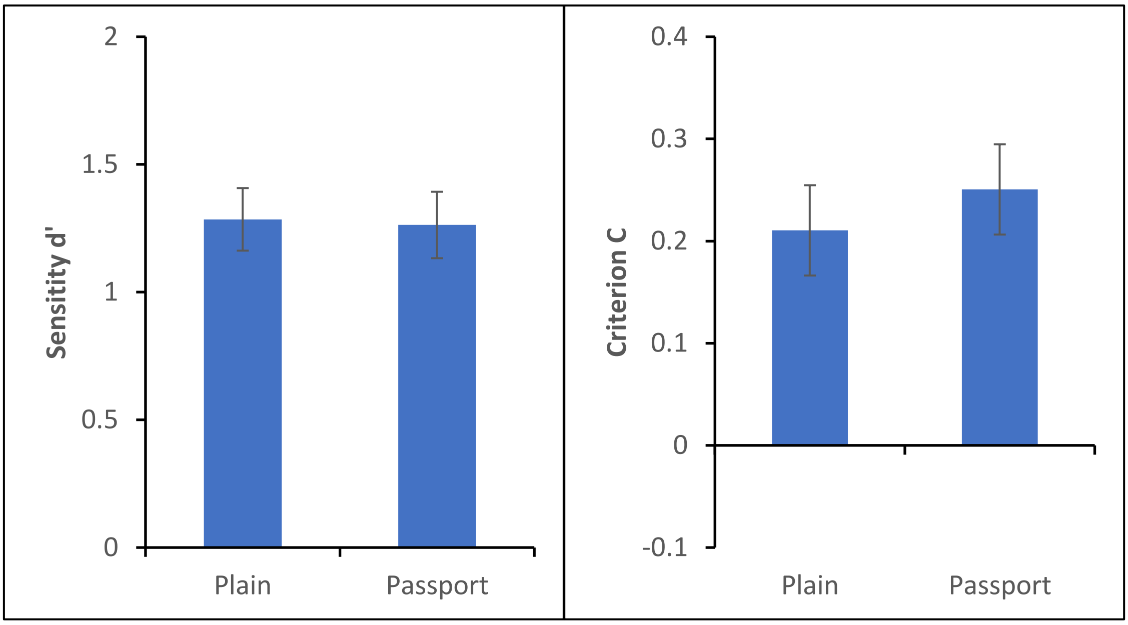

Figure 6.4 Sensitivity (*d'* on the left) and criterion (C on the right) for recognition in Experiment 16B. Error bars represent within-subjects standard error (Cousineau, 2005).

Results showed that we failed to find a difference in the memory task when asking people to remember plain faces and testing them with plain faces or faces in passport. When participants were asked to compare the remembered face to the new face, whether or not the face is in a passport, they may only focus on the face itself rather than the context.

Combined with Experiment 16A, it seems that the document bias effect found in previous experiments does not arise when the manipulation is made at the learning phase or the test phase of a recognition memory experiment for faces. This is perhaps surprising, given the highly replicable effect of document contexts in matching tasks, reported in the earlier chapters of this thesis. We considered the following possible reasons: first, it may because we use the same face image in the learning phase and the test phase, which may result in *picture* recognition rather than *identity* recognition. We did not use different images of the same face identity because we tried to avoid making the task too difficult to hinder the effect of a document. Unfamiliar face recognition is very difficult when using different images of the target people at learning and test (e.g., Bruce, 1982; Mattarozzi et al., 2019), but of course this creates an important difference with face matching – in which two different photos of each person were used in match trials. It is possible that we failed to observe a bias effect here because the 'picture memory' task was easy, by comparison to a 'person match' using different photos. The biasing effect may arise primarily in the harder 'person match' task. Having said this, note that the overall accuracy levels were nevertheless far from ceiling (around 70% accuracy in Expeirment 16A and 16B), and so our result may rely on other differences between matching and memory that are not just related to difficulty.

Second, our results may be influenced by the memory load of the task. Because we asked participants to remember 40 faces, each of which was presented for 5 seconds. In order to remember all the faces, they need to concentrate for almost 4 minutes so that they may only focus on the face and ignore other distractions or task-irrelevant things. The passport context is a complex and neutral stimulus rather than a context that evokes strong emotions, so, it may attract less attention during the encoding phase. Again, this is quite unlike a matching task, in which each test stimulus is completed in a single encounter and with unlimited time.

To address these points, in the next experiment, we are going to use different face images of one identity between the encoding and test phase. Also, we will use an immediate memory paradigm to reduce the memory load but preserve the separate presentation of two matching faces.

6.4 Experiment 17

Introduction

In order to investigate why we observe a 'document bias' in matching tasks, but not in a memory task (Experiment 16), we next investigate immediate memory. Previously, Megreya and Burton (2008) compared overall performance in a matching task to a 'delayed matching' or 'immediate memory' task in which face pairs were presented one after the other, rather than simultaneously. Megreya and Burton report similar performance levels, and so here we ask whether these two types of presentation give rise to similar biases when one of the pairs of faces is embedded in a document.

Here, we decided to only run an immediate memory task that manipulates the document context in the *test* phase, which is a similar manipulation of Experiment 16B. Since there was no effect of the document manipulation at encoding in Experiment 16A, and since previous work has found no effect of a document on first impressions (i.e. another measure of an encoding stage), we decided to manipulate test items only. Participants in the following experiment experienced both matching and immediate memory tasks. The only difference between the two tasks is the presentation of the two faces - to be presented simultaneously or not. In this manipulation, we can directly see if the bias effect is due to a document presented with a face when making identity decisions.

Method

Participants

Thirty-one students (26 females, aged from 18 to 26, mean age = 20.4) from the University of York participated for course credit. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment.

Stimuli

We selected 64 face pairs from the Kent Face Matching Test (KFMT; Fysh, & Bindemann, 2018a). For each face pair, we created a passport frame condition. These face pairs consist of a high-resolution portrait photo and a small size image, as with previous experiments (e.g., Experiment 2 in Chapter 2), so we embedded only the small size image into a passport frame (see Figure 6.5).

Figure 6.5 Exemplars of face pairs from Experiment 17. Each of these pairs shows different identities.

Design and Procedure

This was a 2 (task: immediate memory, matching) × 2 (face type: plain, passport) within-subjects design. Participants performed two tasks, an immediate memory task and a matching task. The order of the two tasks was counterbalanced across participants.

In the immediate memory task, for each trial, participants saw a face presented at the centre of the screen for 5 seconds, then the face disappeared. A second face would be displayed at the same position and at that time, participants were asked to make a judgment on if the two faces were the same person or different people (see Figure 6.6). The second face was displayed until a response was made. After participants made their judgement, they went to the next trial. There were 32 trials in total, i.e., 32 face pairs.

The matching task used the same procedure as previous experiments, in which participants were asked to respond 'same person' or 'different people' to each face pair (see Figure 6.5). There were 32 face pairs in this task. So, for each participant, there were 64 trials in total. Face pairs remained on screen until a response was made.

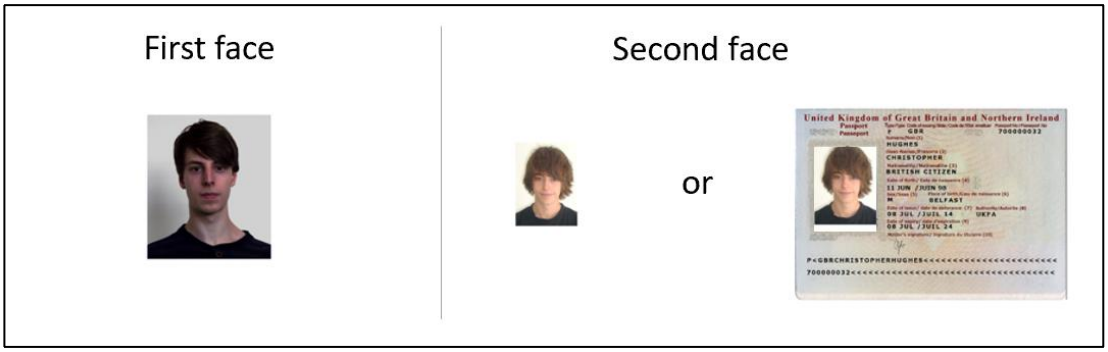

Figure 6.6 A trial of an immediate memory (or 'a delayed match') task in Experiment 17. In this task, the first face will always be plain faces, and the second face will be either plain or embedded in a passport. The first and second face shows different people in this example.

Results

Figure 6.7 shows sensitivity (*d'*) and criterion (C) for recognition performance. Repeated-measures ANOVA showed no significant main effects of face type ($F(1,30)$ = 2.44, p>.05, η_p^2 = .08) and task ($F(1,30)$ = 0.64, p>.05, η_p^2 = .02), or interactions ($F(1,30)$ = 0.23, p>.05, η_p^2 = .01) on *d'*.

But for criterion C, there was a significant main effect of face type ($F(1,30)$ = 12.20, p<.005, η_p^2 = .29), in which the criterion was much lower when faces in passports (M = -0.13) than plain faces (M = 0.10). The main effect of task ($F(1,30)$ = 0.03, p>.05, η_p^2<.01) and the interaction between face type and task ($F(1,30)$ = 0.08, p>.05, η_p^2<.01) did not reach significance.

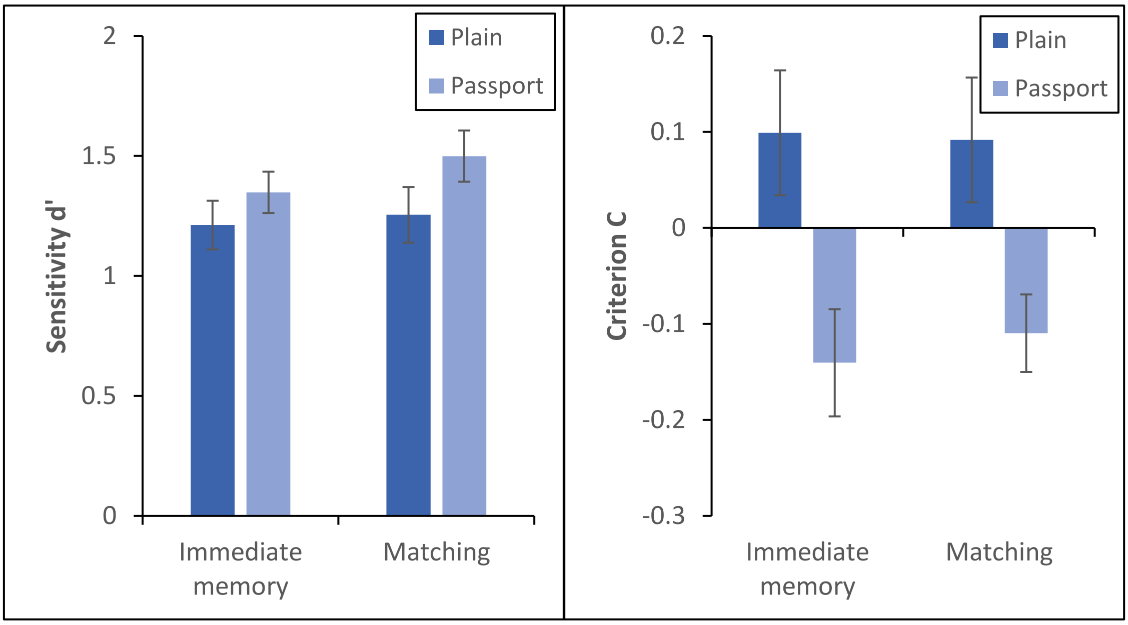

Figure 6.7 Sensitivity (*d'* on the left) and criterion (C on the right) for performance in Experiment 17. Error bars represent within-subjects standard error (Cousineau, 2005).

While immediate memory and marching tasks gave rise to very similar patterns overall, we were also interested to investigate whether the order in which participants took these tasks would influence their performance. To examine this, we performed a 2 (task: immediate memory, matching) × 2 (face type: plain, passport) × 2 (order: immediate memory first, matching first) mixed ANOVA (see Figure 6.8) on criterion C. The results showed that the main effect of face type was significant ($F(1,29) = 11.85$, $p<.005$, $\eta_p{}^2 = .29$). The three-way interaction also reached significance, $F(1,29) = 5.908$, $p<.05$, $\eta_p{}^2 = .17$. Simple main effects showed that when the matching task was presented first, participants were biased by a passport frame on the matching task ($p<.005$); when the immediate memory task was presented first, participants were biased by the passport frame on the immediate memory task ($p<.05$). However, the second task completed by participants did not produce significant biasing effects, whether this second task was matching or immediate memory.

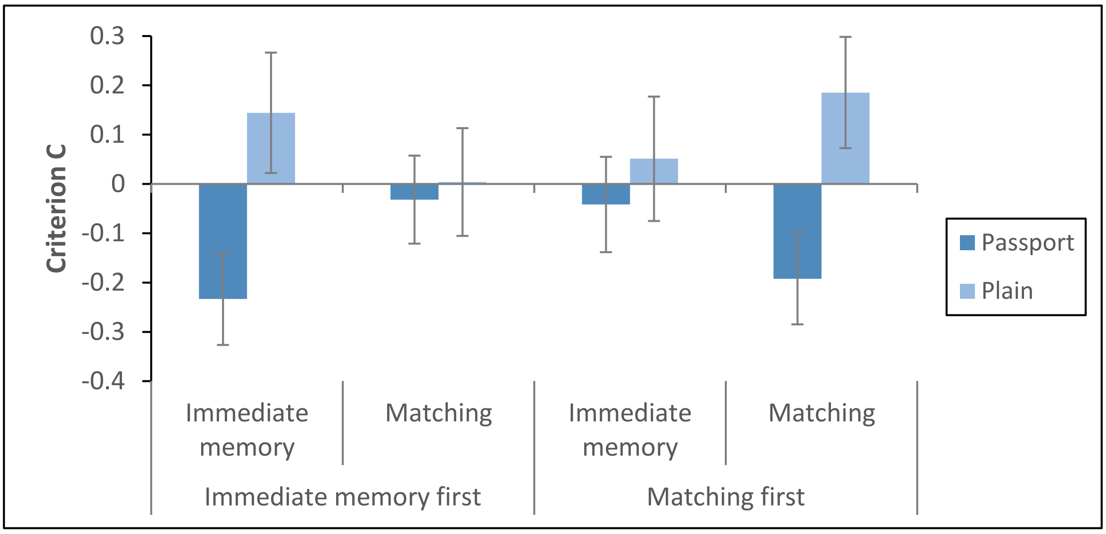

Figure 6.8 Criterion C by conditions in Experiment 17. Error bars represent the standard errors of means (SEM).

Discussion

We found that the bias effect of a passport document existed in both the immediate memory task and the face matching task, which suggests that the presence of a document frame biases viewers at the decision stage. The two matching face images need not necessarily be presented simultaneously, however, the time gap between the two faces cannot be too long, for example, compare these results to the memory task in Experiment 16A.

Our results suggest the bias may be related to the processing of identity judgment decisions. In a face matching task, viewers compare the two face images and then decide if they show the same person; while in an immediate memory task, viewers need to watch the first face carefully, remember it, and then do the comparison with the second face. The document seems to affect the criterion at the processing of making decisions, or may at the processing of comparing facial features. This result contrasts with Experiments 16A and 16B which required longer memory retention, but also was based on *picture* memory (i.e. the same photos were used at learning and test). The fact that the biasing effect re-emerges with immediate memory and using different photos of the target people, further suggests that the biasing influence of document have their effect at the decisional stage of *identification* judgements.

In this experiment, we also carried out an exploratory analysis of the counter-balancing factor: order of tasks. Although originally intended as a check, this appears to have shown something interesting. Biasing effects are clearly present in the first task completed by participants, but not in the second task. It therefore seems possible that, in either task, people are affected by the document initially, then they gradually get used to the task and show less bias. So, there may be a decreasing trend of bias over time. In the next experiment, we are going to test explicitly whether the bias effect declines over time.

6.5 Experiment 18

Introduction

Bindemann and colleagues (2016) reported an effect of time passage on the criterion for matching tasks. They asked participants to match faces under different time pressure across 5 blocks. Results showed that a response bias of accepting face pairs as matches emerges with time over the experiment. Their results seem to be in the opposite direction from our observation in Experiment 17, showing that the bias may decline with time. But the bias in Bindemann et al.'s experiment was the basic criterion observed when matching pairs of plain faces. The focus of our interest here is a different bias – that induced by embedding faces in a document context.

In this experiment, we are going to ask participants to perform a matching task as in previous experiments (i.e., the matching task in Experiment 17), but dividing the whole task into five blocks. We would like to see if there is a bias effect of document, and if the biases from each block change with time.

Method

Participants

Twenty-eight students (24 females, aged from 18 to 25, mean age = 20.1) from the University of York participated for course credit. All reported normal or corrected to normal vision. Each participant completed a consent form before the experiment. This experiment was run online.

Stimuli, Design and Procedure

We used the same 60 face images from the KFMT as those used in previous experiments (e.g., Experiment 2 in Chapter 2). For each face pair, we created a passport frame (see Figure 6.9).

This was a 2 (Face type: plain, passport) × 5 (Block: 1, 2, 3, 4, 5) within-subjects design. In each block, participants were asked to match 12 face pairs (6 identity matches and 6 mismatches), half of which were plain faces and the other half were embedded in a passport. Their task was to indicate if the faces shown in pairs were the same individual or different individuals by pressing corresponding keys. Across all five blocks of the experiment, participants completed 60 trials in total. The face pairs that appeared in each of these blocks and the order of face type were counterbalanced across participants.

Figure 6.9 Face pairs in two presentation conditions from Experiment 18. Each of these pairs shows different identities.

Results

Mean performance over blocks is shown in Figure 6.10. We ran a 2 (Face type: plain, passport) × 5 (Block: 1, 2, 3, 4, 5) repeated-measures ANOVA on both sensitivity (*d'*) and criterion (C). Results on *d'* showed a significant main effect of block ($F(4,108) = 10.77$, $p<.001$, $\eta_p{}^2 = .29$).The main effect of face type ($F(1,27) = 1.83$, $p>.05$, $\eta_p{}^2 = .06$) and the interaction ($F(4,108) = 0.38$, $p>.05$, $\eta_p{}^2 = .01$) did not reach significance. Pairwise comparisons (Tukey HSD) showed that *d'* was reliably larger for Block 3, Block 4 and Block 5 than Block 1 (*ps*<.05), and was larger for Block 3 and Block 5 than Block 2 (*ps*<.05), but no other comparisons were significant (HSD = .763; $F_{crit}(1,108) = 7.70$: *F*(Block 3 vs. Block 1) = 18.56; *F*(Block 4 vs. Block 1) = 10.88; *F*(Block 5 vs. Block 1) = 34.55; *F*(Block 3 vs. Block 2) = 8.38; *F*(Block 5 vs. Block 2) = 19.77)).

For criterion C, there was only a significant main effect of face type (F(1,27) = 12.09, p<.005, $\eta_p{}^2$ = .31), showing lower criterion in faces embedded in passport than plain faces. The main effect of block (F(4,108)=1.79, p>.05, $\eta_p{}^2$=.06) and the interaction (F(4,108)=1.32, p>.05, $\eta_p{}^2$=.05) did not reach significance.

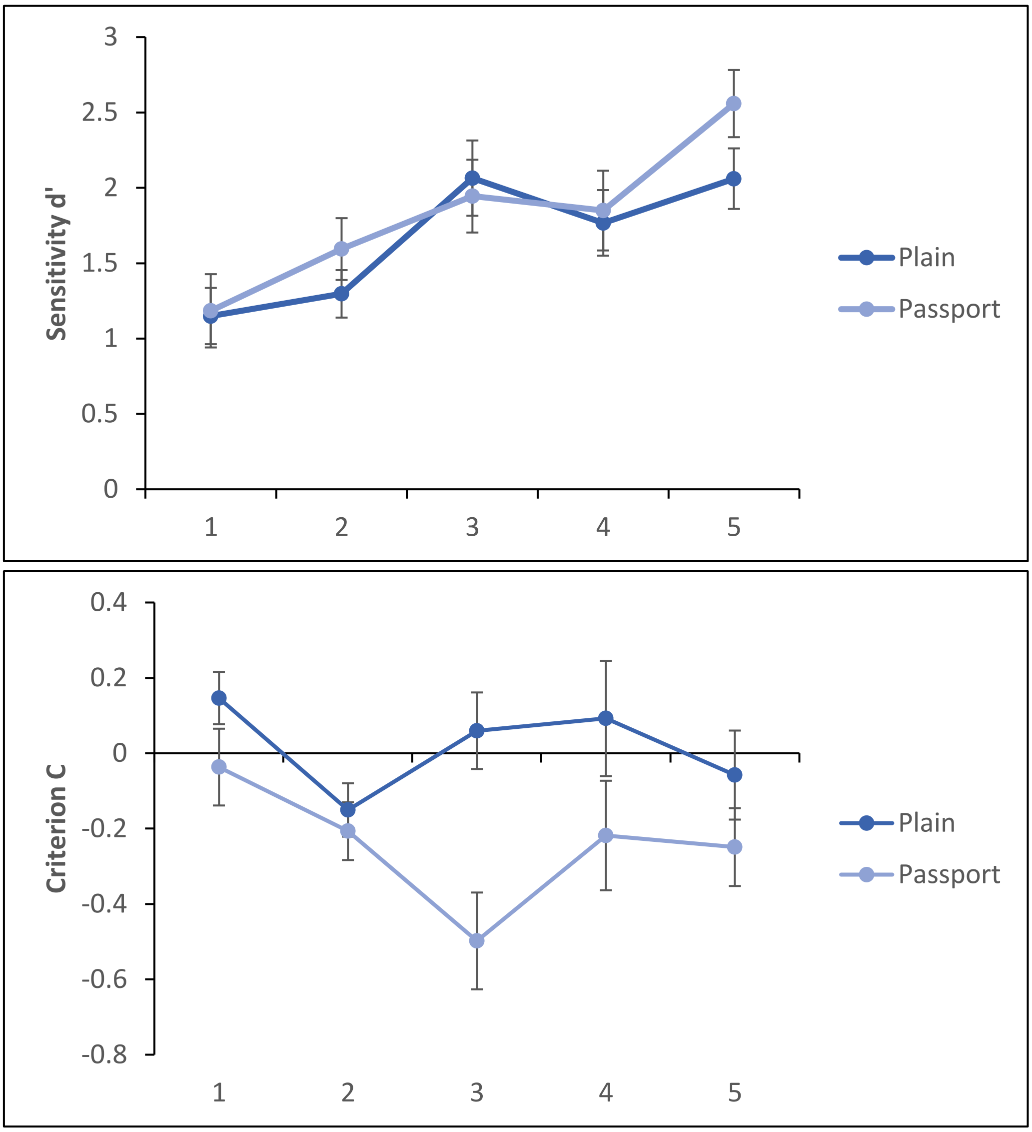

Figure 6.10 Sensitivity *d'*(top) and criterion C (bottom) for performance in Experiment 18. Error bars represent within-subjects standard error (Cousineau, 2005).

Discussion

Our result showed that the bias effect did not decline with time, but persisted during the whole task. So, the hint observed in the exploratory analysis for Experiment

17, that bias may decline over time, was not confirmed by a direct test of this hypothesis. Our result is also inconsistent with some previous research on a criterion changes for plain faces across time (Bindemann et al., 2016). From the graph (see Figure 6.10), it seems to be a tendency of decreasing criterion for both faces, though this does not reach significance.

We found better matching performance with time for both plain faces and faces embedded in passports. As time passes, participants became more accurate on matching faces, especially in the last three blocks. This suggests a beneficial effect of time on performance, even though no feedback was given. Once again, this is not consistent with all published literature. For example, some studies failed to find any difference in overall accuracy (Haggbloom & Warnick, 2003) or *d'* (Fysh & Bindemann, 2017) over time. This may reflect a lack of power in our experiments. The number of trials we use in a block was less than previous research (12 trials per block in the present experiment *vs.* 25 trials in the first block in Haggbloom and Warnick's, or 40 trials in Fysh & Bindemann's study). However, even with the power available in this experiment, there is clearly no hint the passport bias, replicated many times in this thesis, can be explained in terms of participants changing their behaviour during the course of the experiments.

6.6 General discussion

In this chapter, we tried to establish whether the bias effect happens in the encoding or the decisional processes. The presence of a document, whether in the encoding phase (Experiment 16A) or test phase (Experiment 16B), does not affect performance or criterion in a traditional recognition memory experiment. However, using an immediate memory procedure, and emphasising *person matching* rather than *picture matching* (Experiment 17), we observed the bias reported many times earlier in this thesis, i.e. the presence of a document biased viewers to a more liberal criterion (to give more 'same person' or 'old face' responses), and the effect of bias exists regardless of time (Experiment 18).

The results indicate that the bias effect we found in face matching may not exist in all identity judgment tasks. Considering why we failed to find this bias effect in the first two experiments, in addition to the reasons we have discussed before, there is another possibility that the bias is specific to matching. Maybe the bias effect is, to some extent, related to the experience or expectation from life. We normally see an ID-card in the situation that we need to decide if the card matches the person's identity, i.e., to do a face matching task. While it seems rare when we need to remember several faces in cards and recognize them with a time-gap.

What is interesting from these experiments is that we found the bias effect in the immediate memory task. This task has the same procedure of face memory: learning the identity and deciding if the test face has been learned. But after reducing the memory load and changing the image of the test face, the document effect was revealed. Because we have not manipulated the first face in the immediate memory task based on previous experiments, it is still not clear whether embedding the first face into a passport would affect performance in an immediate memory task. But, from the current results, we are sure that the document takes effect on the decisions of identities. It seems that seeing a document when viewers are making decisions, the document frame biases their choice to a more liberal criterion.

Chapter 7 - General Discussion

7.1 Summary of studies

The aim of this thesis was to investigate how additional document context affected face processing. Previous research on face processing mostly uses lab-based standard face stimuli that only include a cropped face against a white background. While in real-world situations, faces are normally encountered in contexts. This thesis focused on the photographic IDs that are needed to prove the bearer's identity, such as passports, driving licences, and workplace (e.g., student or staff) ID cards. In a situation that requires identity-checking, inspectors are dealing with unfamiliar face matching. It has been well demonstrated that unfamiliar face matching can be error-prone (e.g., Henderson et al., 2001; Kramer et al.,2019; Megreya & Burton, 2006, 2008), and professionals can be fallible (e.g., Towler et al., 2019; White, Kemp, Jenkins, Matheson, et al., 2014). In these studies, viewers see isolated plain faces, which is different from the situation of matching a face to a photo-ID. Given the importance of verifying correct identity, we need to understand whether similar results will be found when matching isolated faces and faces embedded in documents. A bias of accepting more face pairs as 'same person' emerges when one of the faces is embedded in a passport compared to isolated plain faces (McCaffery & Burton, 2016).

We proposed several possible reasons explaining this bias in the first two experimental chapters. The first hypothesis we tested was authority (Chapter 2). As the passport is issued by the government and it is hard to make a fraudulent one, viewers may tend to accept that a passport matches the holder. However, our results found the bias existed regardless of authority: with official symbols, like metropolitan police watermark in posters (Experiment 4), we did not find any bias induced by a poster; while it existed in driving licences (Experiment 3), student-ID cards (Experiment 2), and even in simple cards (Experiment 7) without any contextual information (e.g., flags, titles) suggesting authority. When we compared the documents that failed to find the bias with other documents that found the bias, we observed a difference between the layouts of these documents (see Figure 7.1). So, it is possible that the visual properties of the additional document matter.

Figure 7.1 Documents failing to induce a bias: (a) student-IDs with both sides and (b) posters showed a different layout, compared to a document eliciting a bias: (c) driving licence.

We then tested the second possible factor, the visual properties of the additional documents (Chapter 3). Considering the visual comparison between two isolated faces and one of the faces embedded in a document, the only difference is the additional document around that face. Previous experiments also suggest the layout of a document may be important, so, the bias may exist simply due to the document properties. As the document comprises a colour background and identity information, we then separated these components to see how they bias people. We found that if the background and the information were presented separately (Experiment 5 and 6), viewers were not biased. The two elements of a document had to be shown together (Experiment 7). In addition, it had to be an ID-like frame, because the information presented on a circle background did not bias people (Experiment 8). However, the information on the card need not be related to identity, as long as it was readable for viewers (Experiment 9 and 10).

This bias existed for different face sets (Experiment 12), and for more applied settings where the mismatch pairs were rare (Experiment 13). However, when it comes to familiar faces (Experiment 11), and when we changed the task of responding same/different (a decision task) to the one responding similarity between two face images (a rating task, Experiment 14), the bias disappeared.

In order to test the possibility that viewers would only be biased by documents in decision tasks rather than rating tasks, we asked participants to rate perceived traits (attractiveness, trustworthiness, and dominance) on faces either presented in isolation or in documents. Results showed a small effect of increased attractiveness when faces were within a passport, but whether the face was presented isolated or within a passport did not affect trustworthiness and dominance ratings (Experiment 15). So, it seemed that the bias primarily exists in identity-decision tasks. We then adopted the old/new paradigm in memory tasks. One reason was that the identity decisions were included in memory tasks; another reason was that this paradigm comprised an encoding phase and a test phase, from which we could separate the process where the bias may take effect. By manipulating the document either in the learning phase or the test phase in a memory task, we found no bias at all (Experiment 16A and 16B). Considering the memory load, we turned to an immediate memory task (Megreya & Burton, 2008) and compared it with the traditional matching task. We then found the bias existed in both the tasks (Experiment 17), and the bias effect did not decrease with time (Experiment 18). The results also demonstrated that the bias happened in the decision stage rather than affecting the perception of faces.

These experiments (Chapter 3, 4 and 5) explored more of the bias beyond the initial face matching tasks, and helped us form a better understanding of the generalization of document bias.

7.2 Implications and future directions

Response bias or criterion C

There are several things worth noting. The first and important one is the response bias, or the definition of bias. It may be argued that when the criterion C (i.e., response bias) score is zero, it means 'no bias' because equal numbers of 'same' and 'different' responses have been made. However, this is an ideal condition where participants show exactly equal responses of same/different in a two-alternative forced-choice (2AFC) task. In many studies that report criterion, this is not the case. There are always variabilities, even using the same face test. For example, the overall criterion C in White, Kemp, Jenkins, Matheson, and colleagues' (2014) study using GFMT (Glasgow Face Matching Test) was 0.03; while in Bobak, Dowsett, et al. (2016)'s study using the same stimlui, it was -0.11. This also happens with the KFMT (Kent Face Matching Test, whose items we use). These tasks tend to give a relatively stable mean accuracy, but a more variable C score.

The variability of C score may be due to different samples of people, different settings of stimuli, or some true factors affecting the criterion. It is important to make sure the bias we found was induced by the manipulation rather than the systematic variance. In this thesis, the same face stimuli (KFMT) were used across many of the experiments, and the C scores varied across plain face conditions which were common across many experiments (e.g., 0.22 in Experiment 2 and 0.02 in Experiment 4). This suggests that differences in mean bias were probably due to participant sampling, because the tests were the same. For these reasons, across all experiments in our study, we always compared the experimental condition with plain faces as a baseline condition. In this case, if we find a significant bias from plain faces to the faces manipulated in documents, then the bias seems to be induced by the manipulation rather than sampling. So, the bias effect we reported in each experiment was a comparison result rather than an exact value to compare with zero.

We have reported and focused on response bias in every experiment in this thesis. Why is this bias important and what does it tell us? The consistent findings are that embedding one of the faces into a readable document can lead viewers to respond more 'same person' in face matching tasks. This leads to a dangerous outcome that increases the false acceptance of wrong pairs. In a matching task with documents, there will be two types of errors: wrongly rejecting the correct person and falsely accepting the imposter. Comparing with the wrongly rejecting true pairs, if the rate of falsely accepting the wrong pairs increases, the possibility of giving imposters access to restricted things or places will rise.

We have demonstrated that the bias can generalize to different face sets and to more real-life settings like low-prevalence of mismatches, which suggests the general existence of this document bias. It has been found that even expert populations are easily affected and are vulnerable to biases (e.g., Kassin et al., 2013). So, it is reasonable to hypothesize that the document bias we found here may plausibly exist with officers dealing with matching faces with documents. However, the confirmation of this remains for future studies.

Processing of faces in documents

Although we have concluded that any readable card can lead to a bias, the underlying mechanism of this effect remains unexplained. Previous experiments on both data checking and face matching suggest that processing faces and processing the information on documents interact with each other (McCaffery & Burton, 2016; Robertson & Burton, 2021). However, in our study, the results seem to require explanation beyond this simple interaction. If the bias exists from the interaction with processing faces and processing the information, then we should get the bias using faces with adjacent information (Experiment 7), but no card context. However, it turns out that presenting only information will not influence face matching, and the card background is quite critical. It therefore appears that face matching is somehow biased by the deployment of resources diverted into task-irrelevant reading, rather than by any semantic processing of the text information.

The failure of grouping the face and information with a circle background (Experiment 8) suggests that it is not simply bringing disparate information together as a single Gestalt (a perceptual explanation) to induce the bias, but the meaning of the background matters. So, it is possible that inducing an expectation of an 'identity card' based on the social use of ID cards (a more social explanation) may help. However, the card frame did not always take effect, for example when using familiar faces in Experiment 11. Apart from these explanations, we propose further alternative explanations as follows.

The first is that faces in documents are considered as integral and processed holistically. Faces are themselves considered to be processed holistically, which means the facial features are integrated into a perceptual whole, not a collection of isolated features (Gold, Mundy, & Tjan, 2012; McKone & Yovel, 2009). So, it is possible that the faces in documents are not perceived as isolated 'face', 'information', and 'background', but as a whole 'photographic-ID'. If it is true, then the bias is easy to understand. For example, it has been found that if the facial features (e.g., a pair of eyes) are recognized within a face structure, performance will be better than recognition of the features in isolation. The face affects, or facilitates the recognition of 'its parts' (Palermo & Rhodes, 2002; Tanaka & Farah, 1993). It may be the same pattern for faces in documents. Similarly, consider the plain face as the 'eyes', and the document be the 'face'. When faces are surrounded by a document, it seems like the eyes are surrounded by a face (see Figure 7.2). This will explain why the information and the background are indispensable: the readable information and the card-like background are irreplaceable for a document just as a face comprises two eyes, a nose, and a mouth. In future studies, the investigation of whether the faces in documents are processed holistically can be conducted.

Figure 7.2 (a) An illustration of isolated eyes and the whole face (Palermo & Rhodes, 2002); (b) An example of a plain face and the face embedded in a passport.

Another suggestion concerns attention, because attentional resources have a large influence on biases like the own-group bias (Zhou, Pu, Young, & Tse, 2014). The document bias we found here may arise because more attention has been attracted by the irrelevant but readable information within a card context. Although we have never told participants to process the document, they may automatically allocate attention towards it. Eye-tracking studies may help to trace the way viewers match faces in documents. As participants have unlimited time to respond in our experiments, it is worth limiting the time to see if time pressure affects the bias. Other paradigms that manipulate attention may also help to investigate this further. For example, paradigms may be useful that manipulate perceptual load and forced attention on letters rather than faces (e.g., Jenkins, Lavie, & Driver, 2005), or paradigms that divide attention into irrelevant tasks (e.g., Palermo & Rhodes, 2002; Zhou et al., 2014).

Applied issues

In these experiments, we have manipulated the document frame in many ways, especially the information and the background, but we have not considered the face photos themselves, representing official ID photographs. All the face images we use are either from laboratory tests (GFMT, KFMT, MFMT), or from the internet (celebrities and their lookalikes), rather than official passport or driving licence photographs. These are commonly taken years before use, and are smaller or in lower quality than laboratory stimuli. Will these properties of photos affect the bias? There is research matching real document photos (see Figure 7.3, Kramer et al., 2019) with a high-resolution laboratory photo. It found no difference between matching a passport photo and a driving licence photo, with similar performance levels with standard tests (70%-90% accuracy). So, the face stimuli are probably not the key point that needed to be considered in the cause of this bias.

Figure 7.3 Example stimuli used in Kramer et al. (2019)'s study. The face images from left to right are a high-resolution photo, passport photo, and a driving licence photo of the same person.

One thing that is worth noting is the time gap between two matching photos. Photo-IDs can be valid for up to 10 years or more, which results in a large within-person variability. There are studies testing faces taken with different time gaps, for example, on the same day, a few weeks later, and a few months/years later. They found the time gap affects the performance: the larger the time gap is, the poorer the performance will be (e.g., Megreya et al., 2013). It is unclear whether different time gaps will affect bias. But based on previous research, the bias was found using GFMT whose face images were taken within one day (McCaffery & Burton, 2016), also was found using KFMT (at least three months apart) and MFMT (no specific time gap but with large variability) in our experiments. In Weatherford, Erickson, Thomas, Walker, and Schein's (2020) research, the criterion remains steady across different time-gap conditions. So, the time gap between two matching images seems not to play an important role in criterion.

It is possible that the document frame has already provided an expectation of larger variability of faces, so that the mismatch cues produced by age or quality may be neglected by participants. Another thought on why the bias is to respond more ‘same’ is that, apart from the simple student-ID card, more sophisticated ID documents contain other security features such as watermarks for officers to verify. Officers may focus more on checking whether the document is a valid and genuine one, due to the ‘halo effect’ (Thorndike, 1920), they may give little attention to actually match the faces.

The last point I would like to discuss is that we have assumed the expectation of an ID card leads people to respond more ‘same person’, but what if we tell viewers that the pair of faces are different. Will the bias still exist in this manipulation? There is one study providing ‘answers’ to each face pair, in order to mimic the real setting in border control where the computer system will return an answer suggesting whether the face pair depicting the same person or different people. Fysh and Bindemann (2018b) asked participants to do a regular face matching task using KFMT, but adding labels that showed ‘same/different/unsolved’ below those face pairs (see Figure 7.4). The labels, ostensibly the result of a match by computers, were either consistent or inconsistent with the real answer. The overall results showed that the inconsistency impaired the performance, especially when participants were informed to double-check the label (around 20% more errors) compared with when they were told to ignore the label (around 8% more errors).

Figure 7.4 Example stimuli used in Fysh and Bindemann (2018b)'s study. The face pair showed the same person with a consistent ('same'), inconsistent ('different'), and unresolved trial label.

From their results, the labels did affect performance whether viewers were told to use the labels or not. But the consistent label did not give much more improvement compared with the unresolved ones, unless feedback for each trial was given, which induced stronger compliance with the labels. These results were based on plain faces, future studies could be done on comparing plain faces with faces in documents. If the document frame does give an implicit indication to respond 'same', we may find interactions between the frame and the labels.

7.3 Conclusion

In conclusion, this thesis explored the influence of identity documents on face processing. We consistently found a bias of accepting face pairs as 'same person' more when one of the faces was embedded in any readable ID-like card context, not specifically on authoritative documents. The bias happens at the decision stage rather than the encoding process. Although this bias induced by documents appears to rely on the convergence of visual components, it resists simple explanations. For example, we still cannot rule out the influence of expectations or experience of seeing an ID, but the expectations are not strong enough to induce a bias in themselves. Both perceptual and social properties that a document carries matter. This thesis gives a better understanding of why and when the bias exists. However, it also leaves open many questions for future research.

Appendices

Appendix 1 – Calculation of *d'* and C

In these face matching tasks, we regarded viewers responding 'same' to face pair showing same identity as Hit, and viewers responding 'same' to face pair showing different identities as FA (false alarm). Sensitivity *d'* and response criterion C were calculated as:

$$d' = Z(Hit) - Z(FA)$$

$$C = -\frac{Z(Hit) + Z(FA)}{2}$$

Higher values of *d'* indicate better performance on discriminating matching pairs and mismatch pairs. Lower values of C indicate the tendency or bias of responding more 'same'.

To account for extreme performance (e.g., FA rate was zero), these extreme values are replaced by $1 - 1/2n$ for rates of 1 or $1/2n$ of 0, where *n* represents the number of targets (match pair in our tasks).

Appendix 2 – Results from Experiment 1

A by-image analysis was conducted (see Figure A1). We summarized the accuracy scores responded to each face, and calculated the *d'* and C. Repeated-measures ANOVA showed no significant effect of condition on *d'* ($F(2,118) = 0.84$, $p > .05$, $\eta_p{}^2 = .01$), or on C ($F(2,118)=0.83$, $p>.05$, $\eta_p{}^2=.01$).

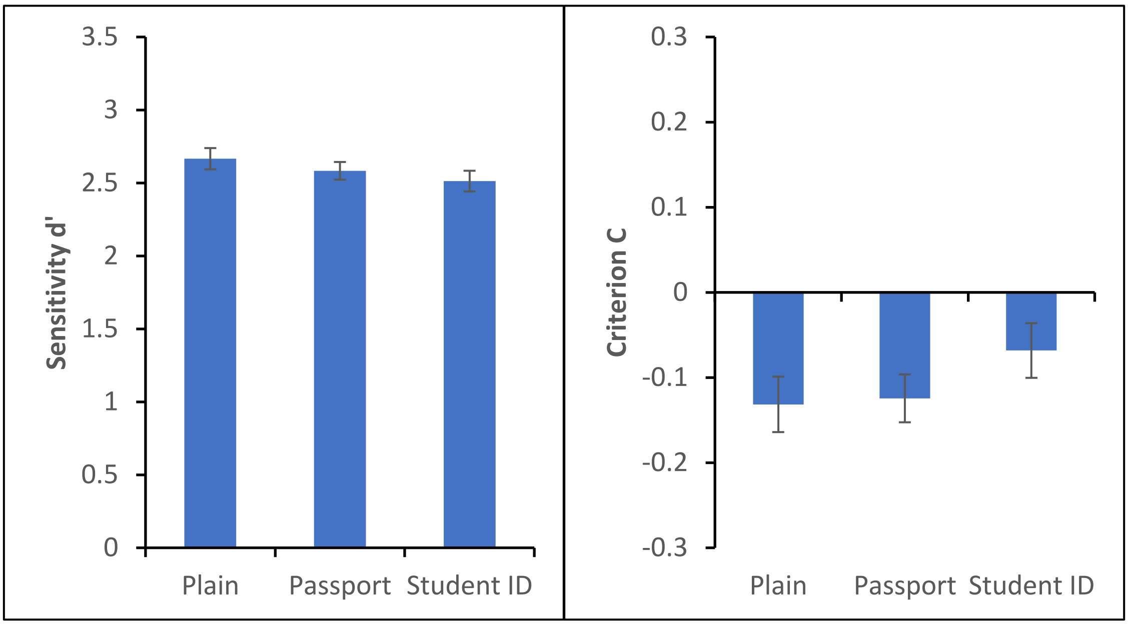

Figure A1. Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 1. Error bars represent within-subjects standard error (Cousineau, 2005).

Appendix 3 – Results from Experiment 2

A by-image analysis was conducted (see Figure A2). We summarized the accuracy scores responded to each face, and calculated the *d'* and C. Repeated-measures ANOVA showed no significant effect of condition on *d'* ($F(2,118) = 1.21$, $p > .05$, $\eta_p{}^2 = .02$), but a significant effect on C ($F(2,118)=6.95$, $p<.01$, $\eta_p{}^2=.11$). Pairwise comparisons (Tukey HSD) showed that C was reliably larger for plain photos than both passport frames ($p<.05$), but other comparisons did not differ significantly (HSD=.137; Fcrit(1,118) = 5.64: F(plain vs. passport = 13.88; F(plain vs. student-ID) = 3.90; F(student-ID vs. passport) = 3.06)).

Although the difference of ID and plain faces did not reach significance, this shows a similar pattern with the by-subject analysis.

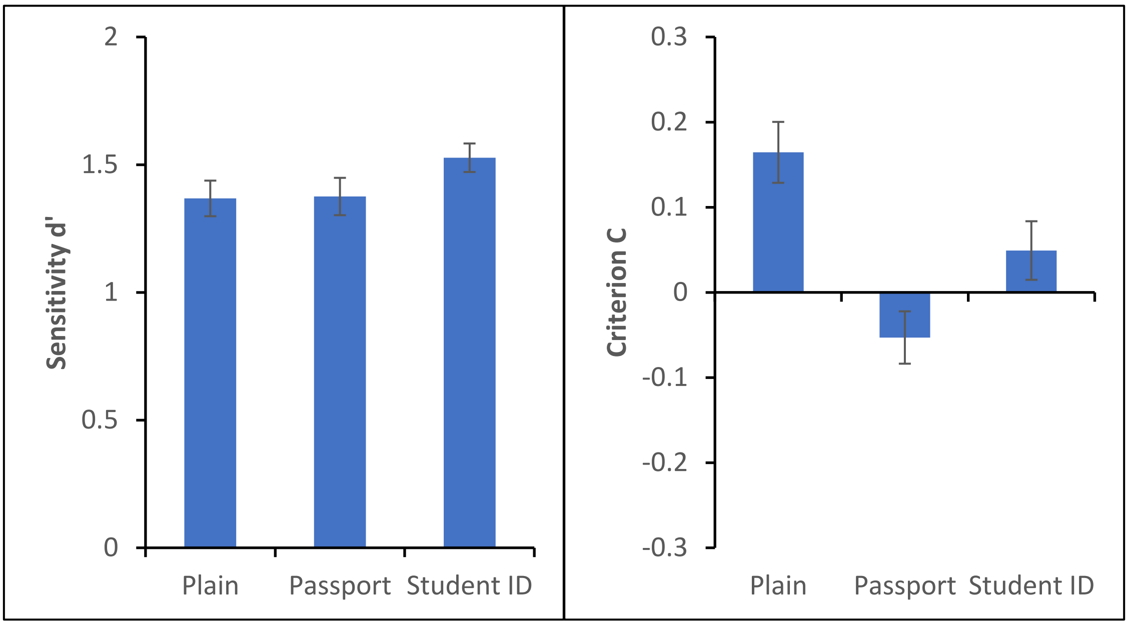

Figure A2. Sensitivity (*d'* on the left) and criterion (C on the right) for matching responses in Experiment 2. Error bars represent within-subjects standard error (Cousineau, 2005).

Appendix 4 – Results from Experiment 14

We conducted a 2 (presentation type: plain, passport) x 2 (match type: same, different) Repeated-measures ANOVA on each face pair (see Figure A3). This by-image analysis showed that only the match type reached significance ($F(1,236) = 212.59$, $p<.001$, $\eta_p^2 = .47$), that images showing the same person were rated as more similar than those showing different people. While the presentation type and the interaction did not reached significance (Fs<1.04, ps>.05).

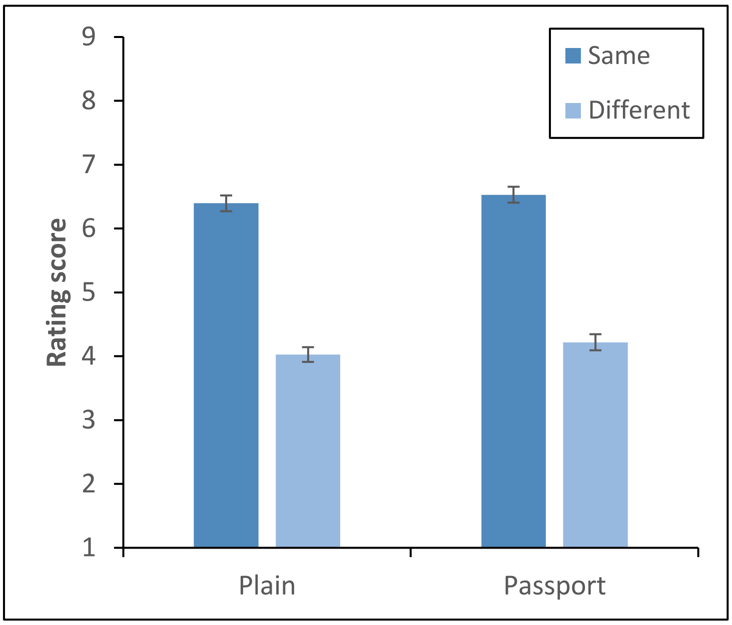

Figure A3. Rating scores for match types under each presentation condition using a by-image analysis. Error bars represent within-subjects standard error (Cousineau, 2005).

Appendix 5 – Results from Experiment 15

A by-image analysis was conducted. We summarized the scores rated for each face, and analysed each trait (attractiveness, trustworthiness, and dominance) separately. Three paired t-tests showed no significant effect between plain faces and passport for trustworthiness ($t(59) = 1.20$, $p > .05$) and dominance ($t(59) = 1.62$, $p > .05$), but a significant increased attractiveness rating on faces with passport compared to plain ones, $t(59) = 5.33$, $p<.001$, Cohen's $d = .69$ (see Figure A4).

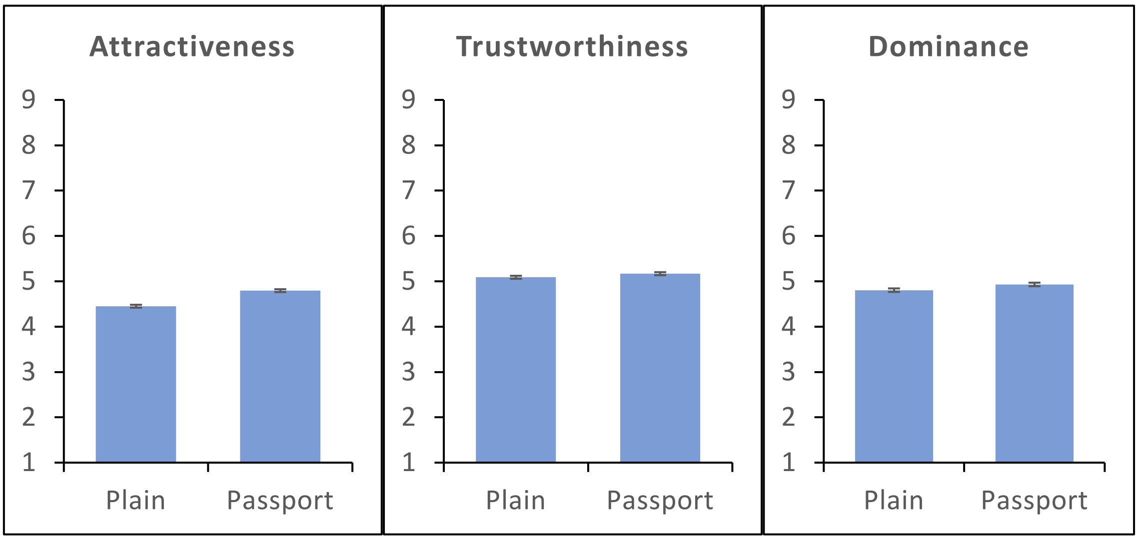

Figure A4. Rating scores for two conditions under each trait. Error bars represent within-subjects standard error (Cousineau, 2005).

In order to examine these data further, we next examined attributions to male and female faces separately. Previous research has shown gender-based differences in trait evaluation (e.g., Sutherland, Young, Mootz, & Oldmeadow, 2015), for example, male faces are generally perceived as more dominant than female faces (Boothroyd, Jones, Burt, & Perrett, 2007). So the passport effect may show different results on dominance for different gender faces. There are 36 female and 24 male face images. We did six paired t-tests (male/female faces rated on attractiveness/trustworthiness/dominance) by separating gender, but no significant results were found (see Table A1 below), except for the significant increased attractiveness found in previous analysis.

Table A1. T-tests results for each trait and separate by gender. Bold means significant results.

	Attractiveness			Trustworthiness			Dominance		
	t	*p*	*n*	*t*	*p*	*n*	*t*	*p*	*n*
Male	3.44	**.002**	24	1.98	.059	24	0.91	.374	24
Female	4.01	**.001**	36	-0.39	.697	36	1.36	.184	36

References

Alenezi, H. M., & Bindemann, M. (2013). The effect of feedback on face-matching accuracy. *Applied Cognitive Psychology, 27*(6), 735-753.

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.

Anes, M. D., & Kruer, J. L. (2004). Investigating hemispheric specialization in a novel face–word Stroop task. *Brain and Language*, *89*(1), 136-141.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617-629.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*(1), 54.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*(46), 17948-17953.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286-290.

Bauer, R. M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. *Neuropsychologia*, *22*(4), 457-469.

Benton, A. L. (1980). The neuropsychology of facial recognition. *American Psychologist*, *35*(2), 176.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science, 18*(8), 706-712.

Bhardwaj, K., & Hole, G. (2020). Effect of racial bias on composite construction. *Applied Cognitive Psychology*, *34*(3), 616-627.

Bindemann, M., Attard, J., Leach, A., & Johnston, R. A. (2013). The effect of image pixelation on unfamiliar‐face matching. *Applied Cognitive Psychology*, *27*(6), 707-717.

Bindemann, M., Avetisyan, M., & Blackwell, K. A. (2010). Finding needles in haystacks: Identity mismatch frequency and facial identity verification. *Journal of Experimental Psychology: Applied*, *16*(4), 378.

Bindemann, M., Fysh, M., Cross, K., & Watts, R. (2016). Matching faces against the clock. *i-Perception, 7*(5), 2041669516672219.

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.

Bindemann, M., & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception*, *40*(5), 625-627.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*(10), 674-679.

Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PloS one, 11*(2), e0148148.

Bobak, A. K., Hancock, P. J., & Bate, S. (2016). Super‐recognisers in action: Evidence from face‐matching and face memory tasks. *Applied Cognitive Psychology*, *30*(1), 81-91.

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.

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-movement strategies in developmental prosopagnosia and "super" face recognition. *Quarterly Journal of Experimental Psychology*, *70*(2), 201-217.

Boehm, S. G., Klostermann, E. C., Sommer, W., & Paller, K. A. (2006). Dissociating perceptual and representation-based contributions to priming of face recognition. *Consciousness and Cognition*, *15*(1), 163-174.

Bornstein, B. H., Deffenbacher, K. A., Penrod, S. D., & McGorty, E. K. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime & Law*, *18*(5), 473-490.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., ... & Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423-455.

Brierley, B., Medford, N., Shaw, P., & David, A. S. (2007). Emotional memory for words: Separating content and context. *Cognition and Emotion*, *21*(3), 495-521.

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

Bruce, V., Bindemann, M., & Lander, K. (2018). Individual differences in face perception and person recognition. *Cognitive Research: Principles and Implications*, *3*.

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.

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

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305-327.

Bruck, M., Cavanagh, P., & Ceci, S. J. (1991). Fortysomething: Recognizing faces at one's 25th reunion. *Memory & Cognition*, *19*(3), 221-228.

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.

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.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, *42*(1), 286-291.

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*.

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

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: A developmental study. *Psychological Science*, *25*(5), 1132-1139.

Davis, J. P., Forrest, C., Treml, F., & Jansari, A. (2018). Identification from CCTV: Assessing police super - recogniser ability to spot faces in a crowd and susceptibility to change blindness. *Applied Cognitive Psychology*, *32*(3), 337-353.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(4), 482-505.

Deffler, S. A., Brown, A. S., & Marsh, E. J. (2015). Judging the familiarity of strangers: does the context matter?. *Psychonomic Bulletin & Review*, *22*(4), 1041-1047.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562-571.

Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs out-perform individuals and provide a route to training. *British Journal of Psychology, 106*(3), 433-445.

Dror, I. E. (2020). Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. *Analytical Chemistry*, *92*(12), 7998-8004.

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.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109.

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

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*(1), 1-11.

Eva, K. W., & Wood, T. J. (2006). Are all the taken men good? An indirect examination of mate-choice copying in humans. *Cmaj*, *175*(12), 1573-1574.

Falvello, V., Vinson, M., Ferrari, C., & Todorov, A. (2015). The robustness of learning about the trustworthiness of other people. *Social Cognition*, *33*(5), 368-386.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77-83.

Fitzgerald, R. J., Price, H. L., & Valentine, T. (2018). Eyewitness identification: Live, photo, and video lineups. *Psychology, Public Policy, and Law*, *24*(3), 307.

Fuller, E. A., Majolo, B., Flack, T. R., & Ritchie, K. L. (2021). The importance of out-group characteristics for the own-group face memory bias. *Visual Cognition*, 1-14.

Fysh, M. C., & Bindemann, M. (2017). Effects of time pressure and time passage on face-matching accuracy. *Royal Society Open Science*, *4*(6), 170249.

Fysh, M. C., & Bindemann, M. (2018a). The Kent face matching test. *British Journal of Psychology*, *109*(2), 219-231.

Fysh, M. C., & Bindemann, M. (2018b). Human-computer interaction in face matching. *Cognitive Science*, *42*(5), 1714-1732.

Galli, G., Feurra, M., & Viggiano, M. P. (2006). "Did you see him in the newspaper?" Electrophysiological correlates of context and valence in face processing. *Brain Research*, *1119*(1), 190-202.

Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M., & Donnelly, N. (2015). The effects of increasing target prevalence on information processing during visual search. *Psychonomic Bulletin & Review, 22*(2), 469-475.

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological Science, 23*(4), 427-434.

Graham, D. L., & Ritchie, K. L. (2019). Making a spectacle of yourself: The effect of glasses and sunglasses on face perception. *Perception*, *48*(6), 461-470.

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad?. *Social Cognition*, *24*(2), 187-206.

Growns, B., & Kukucka, J. (2021). The prevalence effect in fingerprint identification: Match and non‐match base‐rates impact misses and false alarms. *Applied Cognitive Psychology*.

Haggbloom, S. J., & Warnick, J. E. (2003). Practice and incentive motivation in recognition of inverted faces. *Perceptual and Motor Skills*, *96*(2), 578-588.

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

Hassin, R., & Trope, Y. (2000). Facing faces: studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*(5), 837.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(4), 445-464.

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2013). Same faces, different labels: Generating the cross-race effect in face memory with social category information. *Memory & Cognition*, *41*(7), 1021-1031.

Jenkins, R., Lavie, N., & Driver, J. (2003). Ignoring famous faces: Category-specific dilution of distractor interference. *Perception & Psychophysics*, *65*(2), 298-309.

Jenkins, R., Lavie, N., & Driver, J. (2005). Recognition memory for distractor faces depends on attentional load at exposure. *Psychonomic Bulletin & Review*, *12*(2), 314-320.

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

Johnston, R. A., & Barry, C. (2001). Best face forward: similarity effects in repetition priming of face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *54*(2), 383-396.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42-52.

Kelley, H. H. (1950). *The warm-cold variable in first impressions of persons*.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *11*(3), 211-222.

Kerr, N. H., & Winograd, E. (1982). Effects of contextual elaboration on face recognition. *Memory & Cognition*, *10*(6), 603-609.

Kingsley, P. (2015). Why Syrian refugee passport found at Paris attack scene must be treated with caution. *The Guardian*, *15*.

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

Klatzky, R. L., Martin, G. L., & Kane, R. A. (1982). Semantic interpretation effects on memory for faces. *Memory & Cognition*, *10*(3), 195-206.

Koji, S., & Fernandes, M. (2010). Does it matter where we meet? The role of emotional context in evaluative first impressions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *64*(2), 107.

Kramer, R. S., Hardy, S. C., & Ritchie, K. L. (2020). Searching for faces in crowd chokepoint videos. *Applied Cognitive Psychology*, *34*(2), 343-356.

Kramer, R. S., Mohamed, S., & Hardy, S. C. (2019). Unfamiliar face matching with driving licence and passport photographs. *Perception*, *48*(2), 175-184.

Kramer, R. S., & Ritchie, K. L. (2016). Disguising superman: How glasses affect unfamiliar face matching. *Applied Cognitive Psychology*, *30*(6), 841-845.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*(4), 730.

Kukucka, J. (2018). Confirmation bias in the forensic sciences: Causes, consequences, and countermeasures. *The Psychology and Sociology of Wrongful Convictions: Forensic Science Reform*, 223-245.

Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. *Memory & Cognition*, *33*(6), 1017-1024.

Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological Science*, *14*(5), 510-515.

Lee, Y., Matsumiya, K., & Wilson, H. R. (2006). Size-invariant but viewpoint-dependent representation of faces. *Vision Research*, *46*(12), 1901-1910.

Little, A. C., Burt, D. M., & Perrett, D. I. (2006). What is good is beautiful: Face preference reflects desired personality. *Personality and Individual Differences*, *41*(6), 1107-1118.

Liu, H., Hu, Z., & Peng, D. (2013). Evaluating word in phrase: The modulation effect of emotional context on word comprehension. *Journal of Psycholinguistic Research*, *42*(4), 379-391.

Lloyd-Jones, T. J., & Brown, C. (2008). Verbal overshadowing of multiple face recognition: Effects on remembering and knowing over time. *European Journal of Cognitive Psychology*, *20*(3), 456-477.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, *91*(630), 2-2.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, *109*(2), 163.

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law*, *7*(1), 134.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, *13*(2), 194-196.

Malone, D. R., Morris, H. H., Kay, M. C., & Levin, H. S. (1982). Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurology, Neurosurgery & Psychiatry*, *45*(9), 820-822.

Mattarozzi, K., Colonnello, V., Russo, P. M., & Todorov, A. (2019). Person information facilitates memory for face identity. *Psychological Research*, *83*(8), 1817-1824.

Mattarozzi, K., Todorov, A., & Codispoti, M. (2015). Memory for faces: the effect of facial appearance and the context in which the face is encountered. *Psychological Research*, *79*(2), 308-317.

McCaffery, J. M., & Burton, A. M. (2016). Passport checks: Interactions between matching faces and biographical details. *Applied Cognitive Psychology*, *30*(6), 925-933.

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.

McCrackin, S. D., Lee, C. M., Itier, R. J., & Fernandes, M. A. (2021). Meaningful faces: Self-relevance of semantic context in an initial social encounter improves later face recognition. *Psychonomic Bulletin & Review*, *28*(1), 283-291.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, *16*(5), 778-797.

Megreya, A. M., Bindemann, M., Havard, C., & Burton, A. M. (2012). Identity-lineup location influences target selection: Evidence from eye movements. *Journal of Police and Criminal Psychology*, *27*(2), 167-178.

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

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

Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching face images taken on the same day or months apart: The limitations of photo ID. *Applied Cognitive Psychology*, *27*(6), 700-706.

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.

Meissner, C. A., Susa, K. J., & Ross, A. B. (2013). Can I see your passport please? Perceptual discrimination of own-and other-race faces. *Visual Cognition*, *21*(9-10), 1287-1305.

Menzel, C., Hayn-Leichsenring, G. U., Langner, O., Wiese, H., & Redies, C. (2015). Fourier power spectrum characteristics of face photographs: attractiveness perception depends on low-level image properties. *PLoS One*, *10*(4), e0122801.

Mileva, M., & Burton, A. M. (2019). Face search in CCTV surveillance. *Cognitive Research: Principles and Implications*, *4*(1), 1-21.

Mileva, M., Young, A. W., Kramer, R. S., & Burton, A. M. (2019). Understanding facial impressions between and within identities. *Cognition*, *190*, 184-198.

Nakashima, R., Kobayashi, K., Maeda, E., Yoshikawa, T., & Yokosawa, K. (2013). Visual search of experts in medical image reading: the effect of training, target prevalence, and expert knowledge. *Frontiers in Psychology*, *4*, 166.

Newell, F. N., Chiroro, P., & Valentine, T. (1999). Recognizing unfamiliar faces: The effects of distinctiveness and view. *The Quarterly Journal of Experimental Psychology: Section A*, *52*(2), 509-534.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087-11092.

Özbek, M., & Bindemann, M. (2011). Exploring the time course of face matching: Temporal constraints impair unfamiliar face identification under temporally unconstrained viewing. *Vision Research*, *51*(19), 2145-2155.

O'Brien, F., & Thorley, C. (2020). Memory of people from missing person posters: the number of posters seen, the number of times they are seen, and the passage of time matter. *Psychology, Crime & Law*, 1-17.

O'Toole, A. J., Abdi, H., Jiang, F., & Phillips, P. J. (2007). Fusing face-verification algorithms and humans. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *37*(5), 1149-1155.

O'Toole, A. J., An, X., Dunlop, J., Natu, V., & Phillips, P. J. (2012). Comparing face recognition algorithms to humans on challenging tasks. *ACM Transactions on Applied Perception (TAP)*, *9*(4), 1-13.

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *16*(4), 387.

Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive Research: Principles and Implications*, *3*(1), 1-9.

Papesh, M. H., & Goldinger, S. D. (2014). Infrequent identity mismatches are frequently undetected. *Attention, Perception, & Psychophysics*, *76*(5), 1335-1349.

Patzer, G. L. (2012). The physical attractiveness phenomena. *Springer Science & Business Media.*

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8-13.

Phillips, P. J., Beveridge, J. R., Draper, B. A., Givens, G., O'Toole, A. J., Bolme, D. S., ... & Weimer, S. (2011, March). 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.

Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., & Sharpe, M. (2009). FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(5), 831-846.

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., ... & O’Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, *115*(24), 6171-6176.

Porter, S., & ten Brinke, L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology, 14*(1), 119-134.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*(6), 477-491.

Price, H. L., Harvey, M. B., Anderson, S. F., Chadwick, L., & Fitzgerald, R. J. (2019). Evidence for the belief in live lineup superiority. *Journal of Police and Criminal Psychology*, *34*(3), 263-269.

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: a method of calibrating eyewitness identifications. *Journal of Applied Psychology*, *89*(1), 73.

Rainis, N. (2001). Semantic contexts and face recognition. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(2), 173-186.

Reynolds, J. K., & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology*, *6*(4), 279-292.

Rice, A., Phillips, P. J., Natu, V., An, X., & O'Toole, A. J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science*, *24*(11), 2235-2243.

Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision*, *8*(15), 15-15.

Ritchie, K. L., Kramer, R. S., & Burton, A. M. (2018). What makes a face photo a 'good likeness'?. *Cognition*, *170*, 1-8.

Ritchie, K. L., Mireku, M. O., & Kramer, R. S. (2020). Face averages and multiple images in a live matching task. *British Journal of Psychology*, *111*(1), 92-102.

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

Robertson, D. J., & Burton, A. M. (2021). Checking ID‐cards for the sale of restricted goods: Age decisions bias face decisions. *Applied Cognitive Psychology*, *35*(1), 71-81.

Robertson, D. J., Kramer, R. S., & Burton, A. M. (2017). Fraudulent ID using face morphs: Experiments on human and automatic recognition. *PLoS One*, *12*(3), e0173319.

Rubenstein, A. J. (2005). Variation in perceived attractiveness: Differences between dynamic and static faces. *Psychological Science*, *16*(10), 759-762.

Rubínová, E., Fitzgerald, R. J., Juncu, S., Ribbers, E., Hope, L., & Sauer, J. D. (2020). Live presentation for eyewitness identification is not superior to photo or video presentation. *Journal of Applied Research in Memory and Cognition*.

Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science*, *19*(2), 109-111.

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

Shepherd, J. W., Gibling, F., & Ellis, H. D. (1991). The effects of distinctiveness, presentation time and delay on face recognition. *European Journal of Cognitive Psychology*, *3*(1), 137-145.

Shriver, E. R., Young, S. G., Hugenberg, K., Bernstein, M. J., & Lanter, J. R. (2008). Class, race, and the face: Social context modulates the cross-race effect in face recognition. *Personality and Social Psychology Bulletin*, *34*(2), 260-274.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342-353.

Smith, A. M., Wilford, M. M., Quigley-McBride, A., & Wells, G. L. (2019). Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior*, *43*(4), 358.

Stenberg, G., Wiking, S., & Dahl, M. (1998). Judging words at face value: Interference in a word processing task reveals automatic processing of affective facial expressions. *Cognition & Emotion*, *12*(6), 755-782.

Stewart, J. E. (1980). Defendant's Attractiveness as a Factor in the Outcome of Criminal Trials: An Observational Study 1. *Journal of Applied Social Psychology*, *10*(4), 348-361.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Susa, K. J., Gause, C. A., & Dessenberger, S. J. (2019). Matching faces to ID photos: The influence of motivation on cross-race identification. *Applied Psychology in Criminal Justice*, *15*(1).

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

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46*(2), 225-245.

Thorndike, E. L. (1920). Intelligence and its uses. *Harper's magazine*.

Toet, A., & Tak, S. (2013). Look out, there is a triangle behind you! The effect of primitive geometric shapes on perceived facial dominance. *i-Perception*, *4*(1), 53-56.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*(1), 208-224.

Todorov, A., Loehr, V., & Oosterhof, N. N. (2010). The obligatory nature of holistic processing of faces in social judgments. *Perception*, *39*(4), 514-532.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813-833.

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, *25*(7), 1404-1417.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455-460.

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*, e0211037.

Towler, A., Kemp, R. I., & White, D. (2017). Unfamiliar face matching systems in applied settings. *Face Processing: Systems, Disorders and Cultural Differences. New York: Nova Science Publishing, Inc*.

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.

Van den Stock, J., & de Gelder, B. (2012). Emotional information in body and background hampers recognition memory for faces. *Neurobiology of Learning and Memory*, *97*(3), 321-325.G

Van den Stock, J., & de Gelder, B. (2014). Face identity matching is influenced by emotions conveyed by face and body. *Frontiers in Human Neuroscience*, *8*, 53.

Van den Stock, J., Vandenbulcke, M., Sinke, C. B., Goebel, R., & de Gelder, B. (2014). How affective information from faces and scenes interacts in the brain. *Social Cognitive and Affective Neuroscience*, *9*(10), 1481-1488.

Vredeveldt, A., Charman, S. D., den Blanken, A., & Hooydonk, M. (2018). Effects of cannabis on eyewitness memory: A field study. *Applied Cognitive Psychology*, *32*(4), 420-428.

Warrington, E. K., & James, M. (1967). An experimental investigation of facial recognition in patients with unilateral cerebral lesions. *Cortex*, *3*(3), 317-326.

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 505-517.

Watson, D. G., Blagrove, E., Evans, C., & Moore, L. (2012). Negative triangles: simple geometric shapes convey emotional valence. *Emotion*, *12*(1), 18.

Waynforth, D. (2007). Mate choice copying in humans. *Human Nature*, *18*(3), 264-271.

Weatherford, D. R., Erickson, W. B., Thomas, J., Walker, M. E., & Schein, B. (2020). You shall not pass: How facial variability and feedback affect the detection of low-prevalence fake IDs. *Cognitive Research: Principles and Implications*, *5*(1), 1-15.

Weatherford, D. R., Roberson, D., & Erickson, W. B. (2021). When experience does not promote expertise: security professionals fail to detect low prevalence fake IDs. *Cognitive Research: Principles and Implications*, *6*(1), 1-27.

Wells, G. L. (1993). What do we know about eyewitness identification?. *American Psychologist*, *48*(5), 553.

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied*, *20*(2), 166.

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error rates in users of automatic face recognition software. *PloS One*, *10*(10), e0139827.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PloS One*, *9*(8), e103510.

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, *21*(1), 100-106.

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.

White, D., Sutherland, C. A., & Burton, A. L. (2017). Choosing face: The curse of self in profile image selection. *Cognitive Research: Principles and Implications*, *2*(1), 1-9.

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological and Personality Science*, *7*(4), 331-338.

Wilson, J. P., See, P. E., Bernstein, M. J., Hugenberg, K., & Chartier, C. (2014). Differences in anticipated interaction drive own group biases in face memory. *PloS One*, *9*(3), e90668.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592-598.

Wirth, B. E., & Carbon, C. C. (2017). An easy game for frauds? Effects of professional experience and time pressure on passport-matching performance. *Journal of Experimental Psychology: Applied*, *23*(2), 138.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision*, *13*(3), 33-33.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*(7041), 439-440.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*(2), 121-124.

Young, A. W., & Burton, A. M. (2018). Are we face experts?. *Trends in Cognitive Sciences*, *22*(2), 100-110.

Young, A. W., Ellis, A. W., Flude, B. M., McWeeny, K. H., & Hay, D. C. (1986). Face–name interference. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 466.

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

Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to "bad genes" and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of Nonverbal Behavior*, *28*(3), 167-185.

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research*, *105*, 61-69.

Zhou, G., Pu, X., Young, S. G., & Tse, C. S. (2014). Effects of divided attention and social categorization on the own-race bias in face recognition. *Visual Cognition*, *22*(9-10), 1296-1310.

Zhou, X., & Jenkins, R. (2020). Dunning–Kruger effects in face perception. *Cognition*, *203*, 104345.