Effects of Lineup Members' Facial Characteristics and Demeanour on Eyewitness Identification Performance

Tomoko Nishizawa

B.Psych (Hons)

School of Psychology

Faculty of Social and Behavioural Sciences

Flinders University

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Clinical Psychology)

February 2014

Summary

The composition of a police lineup can affect the accuracy of identification decisions made by witnesses. This has been demonstrated, for example, by experimental manipulations of variables such as lineup functional size, suspect-filler similarity and the closeness of the match between a suspect and a witness' description of the culprit. This thesis investigated whether other characteristics of lineup members, such as how familiar, distinctive and memorable they appeared, affected witnesses' perceptions of the likelihood that a particular member was the culprit, perceptions that might affect the likelihood of false identifications of innocent suspects or incorrect rejections of culprits.

Experiments 1 and 2 explored the possibility that a smile displayed by an innocent suspect in a photospread may arouse a sense of familiarity, biasing the photospread against the suspect and increasing the risk of false identifications. Although a smile consistently aroused a sense of familiarity, Experiments 1 and 2 provided mixed results regarding its effect on witnesses' perceptions of the degree of resemblance between the suspect and the culprit. Using an odd-looking smile, Experiment 3 unexpectedly showed the reverse effect, with the smile making the innocent suspect appear unfamiliar, leading to the perception that the suspect was unlikely to be the culprit. Experiment 4 demonstrated that, in addition to inducing unfamiliarity, an odd smile on the face of the culprit in a photospread made the culprit appear distinctive and memorable, leading to witnesses perceiving the culprit presented in the lineup as an unlikely match to be the culprit.

Experiments 5 and 6 manipulated the perceived distinctiveness, memorability and unfamiliarity of the culprit using various other cues to explore further how such perceptions might contribute to the culprit presented in a photospread being falsely

perceived as an unlikely match to be the culprit. That is, encoding conditions were such that participants could not see all of the culprit's face. Later they viewed a photospread containing the culprit, who had a prominent physical feature (e.g., beard, tattoo) on the previously concealed part of his or her face. The presentation of the physical feature increased the likelihood of an inaccurate perception that the culprit presented in the lineup was not the actual culprit, particularly when participants felt certain that the culprit did not have the feature.

Taken together, these studies indicate that the facial characteristics and the demeanour of the suspect or the culprit in a lineup have the potential to affect the accuracy of eyewitness identifications. For example, the presence of some cue that makes an innocent suspect appear familiar may increase the risk of false identifications. Conversely, if for some reason, a culprit appears unexpectedly more memorable than the witnesses' memory of that person, the risk of incorrect rejections may increase.

Declaration

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

……

Tomoko Nishizawa

Acknowledgement

Neil, simply put, thank you for being the awesome person that you are. I am so grateful that you said yes when I asked you to be my supervisor (despite my choice to cross over to the dark side). I feel extremely lucky to have had the opportunity to learn from you. Thank you for sharing your wisdom, generosity and sense of humour with me.

Matt, thank you for being my co-supervisor and providing me moral support. If it wasn't for your enthusiasm and kindness, I would not have had the courage to attempt this.

Staff at Flinders University – special thank you to Tracey Q and W for going out of their way to ensure that I could complete the program without any financial stress.

My friends,

Thank you all for your support over the last 4 years – the best thing about attempting PhD was meeting you all wonderful people.

Special thank you to Nicole for being an awesome lab manager and importantly, for keeping me sane throughout this process with our daily breakfast, lunch and sometimes, afternoon tea break.

Ammo, thank you for your inappropriate jokes – you made me laugh so much over the 4 years.

My Aikido family, thank you for allowing me a place where I could refresh my energy every week. You guys make me smile.

My family,

Dad and mum, thank you for your support and encouragement. You always have wise advice to give (although I never actually listen to you!). You have taught me never to quit and to keep on smiling - no matter how tough things get. I hope I make you proud.

Heath, Nao, Aki and Nobu, thanks for looking after your lil sis and making me laugh when I needed to.

And last but not least,

Paul, thank you for being there for me (even when I started to turn into Ms Grumpy Grumble). I could not have done this without your patience, support and encouragement - I am extremely lucky to have you in my life.

Table of Contents

CHAPTER 1

Introduction

In current Western legal systems, eyewitness testimony often provides crucial evidence. In some cases, convictions are made solely based on eyewitness evidence (Wells et al., 1998). However, as has often been noted by eyewitness memory researchers, human memories do not function like a video-recorder and eyewitnesses are prone to making errors. Although fallible, eyewitness testimonies likely continue to exert a powerful influence within criminal justice systems, as is suggested by mock-juror studies demonstrating the powerful impact of eyewitness testimony on jurors' decisions (Penrod & Cutler, 1995). For example, Loftus (1974) found that guilty verdicts against a suspect increased from 18% of the sample to 72% when incriminating evidence was accompanied by an eyewitness' testimony. Alarmingly, even when the witness was discredited, the proportion of guilty verdicts remained relatively unchanged at 68%. Considering its highly influential effect regardless of its accuracy, it is important to investigate factors that could cause eyewitnesses to make erroneous identification decisions. This may lead to development of procedures that might eliminate such errors. In line with this, the research program described in the following chapters investigated how the demeanour of the members of a photospread (or photographic lineup)[1] may affect the possibility of inaccurate eyewitness identification decisions. The primary focus of the research was to explore how the demeanour of the members of a photospread, such as how familiar, distinctive or memorable they appear, may influence witnesses' perceptions regarding the likelihood of a lineup member being the culprit.

[1] In this thesis the use of the terms photospreads, photoarray, photographic lineups and lineups reflects their usage in the eyewitness literature where they are typically used interchangeably – except where a live identification parade is used.

The procedure employed to conduct a police lineup can differ significantly from one law enforcement agency to another. For example, some agencies primarily use photographic lineups, whilst other agencies tend to use live lineups. Within the United States, a national survey completed by 619 law enforcement agencies (Police Executive Research Forum, 2013) reported that the use of a photo lineup (94.1%) was the most commonly used method within their agencies. Alarmingly, of these agencies, only 68% reported providing training regarding how to conduct a lineup procedure. Furthermore, only 45.7% reported that they had a written policy as to how to conduct a photo lineup. Such data suggest that the law enforcement agencies would benefit from further research, which could be used to further shape the current practice employed by the agencies.

Commonly, a lineup consists of one (sometimes more) suspect who is placed amongst fillers (i.e., those who are known to the police to be innocent). The suspect is someone who is suspected by the police of being the culprit. Therefore, in reality, a suspect can be either guilty or innocent. In experimental conditions however, the status (i.e., innocence vs. guilt) of the suspect is pre-determined and, thus, when composing a lineup, researchers can refer to the lineup as being either target-present or –absent. A target-present lineup refers to a lineup, which consists of the guilty suspect (i.e., the actual culprit). Therefore, when an eyewitness identifies the suspect from a target-present lineup, the identification decision would be considered the correct decision. On the other hand, if the eyewitness chooses to reject the lineup (i.e., identifies no one) or identifies one of the fillers instead, such decisions would be considered incorrect decisions, leading to a *miss* of the culprit. In contrast, a target-absent lineup refers to a lineup, which includes an innocent suspect. For example, the police might accidently suspects an innocent man to have committed a particular

crime, leading to the man being placed in a lineup. In this instance, the rejection of the lineup would be considered a correct decision, whilst the identifications of either the suspect or the fillers would be considered incorrect decisions.

Broadly speaking, erroneous eyewitness identifications can be grouped into two categories: false identifications of innocent suspects or known-to-be-innocent lineup fillers, and incorrect lineup rejections. False eyewitness identifications lead to less severe consequences for the individual identified than identifications of innocent suspects as, in an appropriately constructed lineup, fillers are known to the police to be innocent. On the other hand, false suspect identifications can lead to major detrimental consequences, potentially causing false convictions. Laboratory findings using staged crimes indicate that the proportion of false identifications varies widely depending on encoding and test conditions, under some conditions being as low as 12% and reaching as high as 70% under other conditions (Wells, 1984). Not surprisingly, false eyewitness identifications have been identified as a major cause of wrongful convictions (Huff, Rattner, Sagarin, & MacNamara, 1986 & MacNamara 1986; Wells et al., 1998). Indeed, to date, 321 cases of wrongful convictions have been overturned via use of forensic DNA testings in the United States, with approximately 72% constituted to false eyewitness identifications (Innocence Project, 2014).

As suggested earlier, it is also possible for a witness to incorrectly reject a target-present lineup, or pick a known-innocent filler instead of the culprit, resulting in a *miss* of the culprit. Although the literature has perhaps focused more on false identifications than misses, the latter type of identification error can also lead to detrimental consequences. For example, an incorrect lineup rejection might lead to the release of the actual culprit and erroneous redirections in police investigations.

Research shows that incorrect lineup rejections can also occur quite frequently, perhaps approaching up to 50% of lineups (Steblay, Dysart, Fulero, & Lindsay, 2001). Together, these major types of incorrect identification decisions can lead to injustices as well as wasted investigative time and resources.

Eyewitness research has established a list of factors that can increase the risk of identification errors (Brewer & Wells, 2011). These factors can be categorised into two classes: *estimator* and *system* variables. (Wells, 1978). Estimator variables are factors that are beyond the control of the legal system, mostly linked to the conditions of the criminal events such as characteristics of the encoding conditions (e.g., exposure duration, retention interval), the witness, the defendant and the testimony. For example, poor viewing conditions of the culprit and a long viewing distance can increase the likelihood of a witness making an identification error. System variables are factors over which the legal system may exert control, such as how lineups are constructed and what procedures are used to administer the lineups. For example, research indicates that a use of a biased lineup instruction and a biased lineup presentation increase the risk of identification errors being made. Wells (1978) suggests that research should perhaps focus on exploring system variables, for the findings may lead to more fruitful outcomes in determining ways to reduce identification errors.

Lineup Composition

Lineup composition has been well established as one of the system variables that can influence witnesses' choosing behaviour, increasing the risk of identification errors when the lineup is poorly composed. Generally, a lineup is suggested to be poorly composed if the lineup contains few plausible fillers or if it is composed in a biased manner. A lineup is considered biased if a witness does not require any

memory of the culprit or the criminal event to know who the suspect is (Malpass, Tredoux, & McQuiston, 2007). For example, a lineup may contain fillers who all look relatively similar to the suspect, yet the suspect may stand out from the lineup as the only member who matches the witness description of the culprit. In this instance, the lineup would be biased against[2] the suspect, increasing the likelihood of suspect identifications by suggesting to the witness that the suspect is the only plausible choice. Although less researched, a lineup can also become biased in a way that favours[3] the suspect, if the suspect deviates from the witness' description of the culprit. For example, a lineup may consist of a suspect and fillers who look alike, yet compared to the fillers, the suspect may deviate significantly from the witness' description of the culprit. In this instance, the witness may consider the suspect to be an unlikely choice by virtue of having fillers who are more plausible picks. This would likely lead to increased likelihood of lineup rejections (i.e., missing the suspect).

Eyewitness research has consistently shown that a biased lineup increases the likelihood of identification errors (Brigham, Ready, & Spier, 1990; R. C. L. Lindsay & Wells, 1980; Smith, Lindsay, & Pryke, 2000). For example, R. C. L. Lindsay and Wells (1980) presented participants with either a lineup containing match-to-description fillers or a biased lineup containing fillers who were dissimilar to the culprit. Specifically, for a Caucasian, blond haired culprit, the biased lineup consisted of black haired Asian fillers and brown haired Caucasian fillers. For a target-present lineup, the proportions of correct suspect identifications remained at reasonable levels for both match-to-description and biased lineups, with only a slight increase being observed (58% vs. 71%). However, in a target-absent lineup, when the culprit was

[2] A lineup that is biased against the suspect is composed in a way that increases the likelihood of the suspect identifications, when compared to a fair lineup.
[3] The term "favours" does not infer that this is the favourable way of composing a lineup.

replaced with a similar looking suspect (i.e., blond haired Caucasian), the proportion of false suspect identifications more than doubled from the match-to-description lineup to the biased lineup (31% vs. 70%). That is, by making the innocent suspect the only plausible pick, the composition of a biased lineup significantly increased the risk of false suspect identifications.

The results obtained from the use of such biased lineups provide some critical information regarding eyewitnesses' decision-making processes. In particular, the increase in the proportions of the suspect identifications observed for the biased lineups demonstrates that eyewitnesses do not rely just on their memory for the culprit to make identification decisions. Indeed, Wells (1984) suggested that eyewitnesses tend to come into view a lineup with an assumption that the culprit must be in the lineup and, thus, their job is to pick the lineup member who is the best match to their memory for the culprit, relative to the other lineup members. Wells (1984) termed such decision-making process a *relative judgement process*. Under the use of the relative judgement process, the presentation of a biased lineup would likely encourage an eyewitness to choose the lineup member against whom the lineup is biased. Therefore, when a suspect stands out from a lineup, suggesting to the eyewitness that the suspect is the culprit, regardless of whether the suspect is guilty or innocent, the likelihood of suspect identifications increases.

A recent meta-analysis of 17 eyewitness studies (Fitzgerald, Price, Oriet, & Charman, 2013) provides further support regarding the effect of lineup compositions on identification errors. The meta-analysis found that for target-absent lineups, false identifications of innocent suspects were more common from lineups composed of dissimilar fillers than from lineups composed of moderately similar fillers (.40 vs. .24) or highly similar fillers (.37 vs. .19). In other words, in line with R. C. L. Lindsay

and Wells (1980), the likelihood of false suspect identifications increased when the innocent suspects stood out from the lineups as resembling the culprits the most. For target-present lineups, the use of similar rather than dissimilar fillers led to a lower proportion of correct suspect identifications (.44 vs. .65) and therefore increased the likelihood of miss of the culprits. Together, the existing research shows that a poorly composed lineup increases the risk of identification errors. In particular, a lineup that is biased against the suspect would likely increase the risk of false identifications, if the suspect happens to be innocent. On the other hand, a lineup that is biased in a way that favours the suspect would likely increase the risk of miss of the culprit, if the suspect happens to be the actual culprit.

To minimise identification errors, eyewitness researchers have recommended that anything that can attract unwanted attention to the suspect should be eliminated from a lineup, avoiding the possible composition of a biased lineup (Malpass, 1981; Wells & Seelau, 1995). Therefore, the definition of an unbiased lineup should extend to how well the suspect and fillers match in terms of their demeanour, not just how well they fit the witness description of the culprit. However, to my knowledge, currently no detailed recommendations have been provided regarding the importance of ensuring that the lineup members match one another in their demeanour. Therefore, there is a potential risk that in some instances, a lineup that is judged to be unbiased due to it consisting of match-to-description suspect and fillers may in fact be biased due to the suspect conveying a distinctively different demeanour to the fillers.

Demeanour Influences on Legal Decision Making

The notion that the demeanour of lineup members may lead to a biased lineup may sound absurd. However, it is not so long ago that some experts believed that criminals could be identified by aspects of their appearance such as distinguishing

physical features. For example, in the 1900s, criminologist Lombroso (1911) believed that murderers could be distinguished by their thin lips and robbers by their beak like noses. In the 1940s, some psychologists believed that criminals were born rather than made and that they could be distinguished by their solid and well-built muscular bodies (Sheldon, Stevens, & Tucker, 1940). Whilst the notion of *born criminals* and the link between criminality and body types has been dismissed (Rafter, 2007), people continue to stereotype criminals as possessing particular traits and physiques. For instance, people often presume that criminals possess a distinctive and memorable appearance (MacLin & Herrera, 2006). Similarly, facial characteristics such as attractiveness have been linked to criminality, where highly unattractive as opposed to attractive individuals are more likely to be judged as being criminals (Saladin, Saper, & Breen, 1988). In sum, people continue to hold stereotypical beliefs regarding the type of demeanour that is associated with criminals. Furthermore, social psychology research indicates that such impressions of strangers can be formed rather rapidly and based on very little information (Ambady & Rosenthal, 1993). For example, after merely 100ms of observing a photographed image of a person, people can form impressions of others on traits such as attractiveness, likability, trustworthiness and aggressiveness (Willis & Todorov, 2006).

Alarmingly, mock-juror studies provide ample evidence that impressions regarding criminal stereotypes can influence legal decisions (Heath, Grannemann, & Peacock, 2004; Porter & ten Brinke, 2009; Wessel, Drevland, Eilertsen, & Magnussen, 2006). In particular, mock-juror studies have repeatedly shown that, in comparison to a positively perceived defendant, the jurors are likely to judge a negatively perceived defendant more harshly. In a typical mock-juror experiment, participants are presented with a mock-crime case and they are asked to provide a

verdict (i.e., guilty/ not guilty) and suggest a sentence. Using this method, Downs and Lyons (1991) found that in comparison to unattractive defendants, attractive defendants were more likely to receive not guilty verdicts, shorter sentences and were considered less dangerous. Similarly, Kassin (1983) found that a plaintiff received a more favourable verdict when his courtroom behaviour was positive (e.g., polite, attentive and confident) as opposed to negative (e.g., impolite, cautious and annoyed). Furthermore, Porter, ten Brinke, and Gustaw (2010) demonstrated that when a severe crime was thought to be committed by defendants whose photographs either cued trustworthiness or untrustworthiness, mock-jurors needed less evidence to arrive at a guilty verdict for defendants judged to be untrustworthy than for defendants judged to be trustworthy. These results indicate that information conveyed by individuals' demeanour may affect a range of decisions made in legal contexts.

In a lineup context, there is some, albeit limited, evidence to suggest that lineup members' demeanour can influence witnesses' identification decisions. For example, according to the *criminal face bias* effect (Flowe & Humphries, 2011; Flowe, Klatt, & Colloff, 2014), under some circumstances, eyewitnesses might rely on the facial characteristics of lineup members to make identification decisions, such as how criminal-like a lineup member's face appears. In particular, Flowe and Humphries (2011) found that when mock-witnesses were presented with a lineup without a description of the culprit and asked to choose who the suspect might be, the mock-witnesses often reported choosing a lineup member who they felt were most criminal-like.

Buckhout, Figueroa, and Hoff (1975) further demonstrated that a behavioural presentation of a suspect can bias the lineup against the suspect, increasing the likelihood of suspect identifications. In their study, participants were presented with

either a biased or unbiased photospread. In the biased photospread, the photo of the suspect was displayed at an off angle (i.e., crooked) from the other photos presented on the photospread and also had an emotional facial expression[4]. The unbiased photospread displayed the suspect in a neutral manner (i.e., photo presented at a straight angle and with a neutral expression). The photos of the fillers were presented with neutral expressions and looking straight ahead. These photospreads were presented with either biased (i.e., no warning of the possibility of the target absence) or unbiased lineup instructions. The results showed that in comparison to the unbiased photospread + unbiased instruction condition, the biased photospread + biased instruction condition produced a higher proportion of suspect identifications (61.3% vs. 40%). In his follow-up study, Buckhout (1975) replicated the results when the biasing cues presented by the suspect were altered to a grinning expression and a tilted head. In particular, the study found that in comparison to the unbiased photospread + unbiased instruction condition, the biased photospread + biased instruction condition produced a higher proportion of suspect identifications (61% vs. 40%). The results therefore provide very tentative evidence for the claim that the demeanour of lineup members may influence witnesses' identification decisions.

However, the evidence for the claim is somewhat limited, for several obvious reasons. The article only reports the combined effects of the photospread and the instruction types. Considering that the use of a biased instruction alone can increase the proportion of positive identifications (Malpass & Devine, 1981), it is unclear as to what the contribution of the biased photospread was. Furthermore, even if the biased photospread made an independent contribution, no explanations are provided as to why a biased demeanour increased the likelihood of suspect identifications. Did the

[4] No explicit description is given as to what kind of facial expression was used. However, an example of the biased photospread displayed in the article shows the suspect displaying a negative expression (perhaps anger or disgust).

suspect appear more familiar and therefore resemble the witnesses' memory for the culprit better when presented with the biased demeanour rather than in a neutral manner? Or did the biased demeanour make the suspect fit the stereotypical image of a criminal better? In short, the studies conducted by Buckhout and colleagues are, at best, suggestive and do not explain how a witness' identification decision may be influenced by the demeanour of photospread members. In the following chapters, in addition to exploring how the demeanour of photospread members may contribute to a biased or poorly composed lineup, I further explored how such demeanours may influence witnesses' metacognitions and, in turn, contribute to an increased risk of identification errors.

Metacognition in Memory Processing

According to the metacognition literature (Koriat, 2006; Koriat and Levy-Sadot, 2000), memory processing involves two types of metacognitive judgments: judgements based on subjective feelings such as a sense of familiarity and judgments based on subjective beliefs regarding one's own memory processing, such as how easily one believes that a stimulus can be memorised. For example, in a recognition task, a subjective feeling of familiarity may play a key role in determining whether a stimulus is judged to be old (i.e., seen before) or new (i.e., never seen before). A strong feeling of familiarity would likely lead to a stimulus being judged old, whilst a sense of unfamiliarity (or a lack of familiarity) would likely lead to the stimulus being judged new.

People's beliefs regarding their own memory capacities and limitations also play a key role in shaping recognition decisions (Koriat & Levy-Sadot, 2000). For example, in a recognition task, individuals may believe that their ability to encode and recall a stimulus would depend on how memorable the stimulus is. A highly

memorable stimulus may cue a belief that a previous encounter with that stimulus should cue a clear recollection. Conversely, a much less memorable stimulus may cue a belief that, because the stimulus was rather difficult to encode, it is likely to be difficult to recall. Based on these beliefs, a lack of recollection experienced for a highly memorable stimulus would likely lead to the stimulus judged as being new. Conversely, a lack of recollection experienced for an unmemorable stimulus may lead to the stimulus judged as being old due to an assumption that, perhaps, the previous encounter with the stimulus has been forgotten.

As memory research suggests, such metacognitive judgements do not always lead to accurate recognition decisions (Strack & Bless, 1994; Whittlesea, Jacoby, & Girard, 1990). For example, a novel stimulus may cue a false sense of familiarity, leading to the stimulus judged as being old. Conversely, an old stimulus may cue a false sense of unfamiliarity, leading to the stimulus judged as being new. Similarly, an old stimulus may fail to cue a recollective experience despite its memorable appearance, leading to the stimulus judged as being new. In a lineup context, a falsely cued sense of familiarity may lead to a false identification decision, whilst a falsely cued sense of unfamiliarity, as well as failure to cue a recollection, may lead to a false rejection of a lineup. In short, adjusting the demeanour of photospread members, in particular, their apparent familiarity and memorability may increase the risk of witnesses making erroneous identification decisions.

Familiarity. Several facial recognition studies have demonstrated that a smile can arouse a false feeling of familiarity, increasing the likelihood of novel faces being misidentified as belonging to known people (Baudouin, Gilibert, Sansone, & Tiberghien, 2000; Dobel et al., 2008; Lander & Metcalfe, 2007). For example, Baudouin et al. (2000) asked participants to look at 20 famous and 20 unknown faces

and to categorise each face as being *familiar* or *unfamiliar* (Experiment 1) or to rate each face using a scale of 1 (unknown) to 7 (famous) (Experiment 2). They found that the unknown faces were more often falsely categorised as being familiar as well as rated higher on the 7-point scale when displaying a smile as opposed to a neutral expression. In line with this, Garcia-Marques, Mackie, Claypool, and Garcia-Marques (2004) conducted a facial recognition study in which participants studied some unknown faces and, later, were asked to make recognition decisions from a set of old and novel faces. Half of the old and new faces were presented smiling, whilst the other half were presented with a neutral expression. The results showed that the proportion of false recognitions made for the novel faces increased when they were presented smiling rather than with a neutral expression. Together, the two studies provide evidence that an expression of a smile can cue a feeling of familiarity, producing a false sense of a previous encounter.

Two theories further suggest that an expression of a smile may produce a feeling of familiarity. First, according to a perceptual fluency theory, a feeling of familiarity arises when a stimulus is processed rapidly and with ease (Jacoby & Dallas, 1981; Johnston, Dark, & Jacoby, 1985; Johnston, Hawley, & Elliott, 1991). In line with this, numerous studies suggest that faces are processed more rapidly when displaying a smile rather than neutral or negative expressions (Becker, Anderson, Mortensen, Neufeld, & Neel, 2011; Juth, Lundqvist, Karlsson, & Öhman, 2005; Leppänen & Hietanen, 2004; Williams, Moss, Bradshaw, & Mattingley, 2005). Thus, it is plausible that an expression of a smile could allow a face to be processed with ease, producing a feeling of familiarity for that face.

Second, a *warm glow* heuristic (Monin, 2003) suggests that a feeling of familiarity is linked to liking. That is, when a stimulus is familiar, it also arouses a

feeling of positivity. Accordingly, since a familiar stimulus arouses positive affect, the reverse effect has been suggested whereby positive affect can arouse a feeling of familiarity (Monin, 2003). Assuming that a positive stimulus arouses positive affect, for example, a smile may make a face appear more likable, Garcia-Marques et al. (2004) suggest that positive facial expressions such as a smile may also arouse a feeling of familiarity. Therefore, the two theories provide additional support for the proposition that an expression of a smile may cue a sense of familiarity.

The purpose of the current research was not to compare the two theories directly but rather, to explore how in a lineup-context a smile on the face of one lineup member might lead to the face being perceived as more familiar. Presuming that a lineup member who is perceived as more familiar than the others is probably more likely to be perceived as the culprit, a smile may increase the risk of false identifications when the smile is displayed by a lineup member who is not the culprit. Accordingly, further research was considered necessary to explore the possibility that a smile on one lineup member's face may bias the lineup against the smiling member. The possible effect of a smile functioning as a familiarity cue in a target-absent lineup is discussed, and tested, in Chapters 2 and 3 (Experiments 1 and 2).

Assuming that a feeling of familiarity would increase the likelihood of a lineup member being perceived as the culprit, a feeling of unfamiliarity might have the reverse effect. That is, a lineup member who appears more unfamiliar than the other lineup members would be considered an unlikely pick from the lineup, perhaps, leading to a lineup becoming biased in a way that favours the unfamiliar-looking member. The effect of unfamiliarity is discussed, and tested, in Chapter 3 (Experiment 3).

Memorability. In addition to a subjective feeling of unfamiliarity, subjective beliefs regarding stimulus memorability can also cause people to falsely assume that an old stimulus is new. For example, if one face in a lineup is judged to be particularly memorable, yet fails to cue a clear recollection, the lineup member may be rejected as being the culprit. If the memorable face happens to belong to the culprit, the lack of a clear recollection would likely increase the risk of incorrect lineup rejections.

So how can the culprit presented in a lineup become surprisingly more memorable whilst cuing no clear recollection? One possibility is that the culprit's appearance may change between encoding (e.g., crime scene) and test (e.g., lineup), due to a deliberate (e.g., use of disguise) or a natural change over time (e.g., age, facial expression, posing), leading to the culprit appearing unexpectedly more memorable when presented in a lineup. Alternatively, a witness may fail to encode a particular physical feature belonging to the culprit, due to the feature being concealed at the time of the crime. For example, due to a poor lighting condition or a limited viewing condition of the culprit, a witness may fail to encode that the culprit has a distinctive feature such as a scar or a tattoo on his or her face. Later, when the culprit is presented in a lineup with the originally concealed feature being visible, the feature may make the culprit appear unexpectedly more memorable than the witness' memory for the culprit. This unexpectedness in perceived memorability of the culprit may lead to the witness falsely rejecting the culprit, with the witness thinking to herself, "If this is the guy who committed the crime, surely, I should have a more vivid memory of the culprit".

In support of the view presented so far, a theory of memory for nonoccurrences (Förster & Strack, 1998; Strack & Bless, 1994; Strack, Förster, &

Werth, 2005) suggests that in order for a stimulus to be judged new, the stimulus should not only fail to cue recollection but it should also evoke a subjective belief that, if it had been seen previously, the stimulus is memorable enough that it would have been remembered. To demonstrate the role of expected memorability, Strack et al. (2005) conducted a word recognition task in which participants were led to believe in superior memorability of either low frequency or high frequency word types. The results showed that those led to believe in the memorability of the low frequency words judged a greater proportion of the low frequency words as being new than they did for the high frequency words. In contrast, those led to believe in the memorability of the high frequency words judged a greater proportion of the high frequency words as being new than they did for the low frequency words. The patterns were observed regardless of the actual status of the words. These results provide evidence that when a stimulus is considered highly memorable, the decision criterion for judging the stimulus as old may be raised, where the absence of a clear recollection would likely lead to the stimulus judged as being new.

In line with the theory of memory for nonoccurrences, facial recognition studies have repeatedly shown that distinctive and therefore presumably more memorable faces are better discriminated than typical faces (Courtois & Mueller, 1981; Going & Read, 1974; Knapp, Nosofsky, & Busey, 2006; Light, Kayra-Stuart, & Hollander, 1979; Semmler & Brewer, 2006). That is, when serving as targets, distinctive faces are better recognised than typical faces, whilst when serving as distractors distinctive faces are more readily rejected than typical faces (Bartlett, Hurry, & Thorley, 1984; Courtois & Mueller, 1981). In line with the theory of memory for nonoccurrences, Brewer, Weber, and Semmler (2005) suggest that a highly distinctive face may evoke a feeling of memorability, leading to face

distinctiveness being used as a basis for estimating one's memory accuracy. For example, when one face in a lineup appears highly distinctive, the face would likely cue an expectation that, if seen previously, the face should arouse a clear recollection. Due to the expectation regarding the memorability of the face, if the face fails to cue a clear recollection, it would likely be rejected as belonging to the culprit.

In short, as the theory of memory for nonoccurrences suggests, judged memorability of a face would likely play a key role in determining whether a face has been encountered before or not. Furthermore, existing facial recognition studies suggest that a perceived memorability of a face can be affected by the distinctiveness of the face: the more distinctive the face appears more memorable it would be judged. It is further proposed that an appearance change made to a face from encoding to test, which causes the face to appear more distinctive and, thus, memorable would likely increase the likelihood of the face falsely judged as being new. The role of face distinctiveness and memorability in a lineup context is further discussed, and tested, in Chapters 4 and 5 (Experiments 4, 5 and 6).

Summary

This research examined whether the demeanour of the members of a photospread could bias witnesses' perceptions regarding the likelihood of a particular photospread member being the culprit. In particular, the research explored the roles of perceived familiarity, memorability and distinctiveness in a lineup context. Experiments 1 and 2 examined whether having one smiling face in a lineup would lead to an increased feeling of familiarity about the face, increasing the risk of false identifications being made from a target-absent lineup. Using target-absent lineups, Experiment 3 tested whether a sense of unfamiliarity (as opposed to familiarity) would increase the likelihood of a lineup member perceived as being unlikely to be

the culprit. In addition to the role of unfamiliarity, Experiments 4, 5 and 6 examined the roles of distinctiveness and, thus, memorability in shaping witnesses' tendency to perceive a lineup member as being unlikely to be the culprit, increasing the risk of incorrect lineup rejections being made form target-present lineups. Together, the research attempted to provide additional knowledge regarding how the demeanour of photospread members may shape witnesses' metacognition regarding their memory processing and, in turn, witnesses' perceptions regarding the likelihood that a particular individual is the culprit.

CHAPTER 2

Experiment 1

As outlined in Chapter 1, a lineup is considered biased if a lineup member stands out as the only member who matches the witness description of the culprit. Such a lineup would likely lead to the witness choosing the lineup member against whom the lineup is biased. For example, if a target-absent lineup is biased against an innocent suspect, the risk of false suspect identifications would likely increase. To avoid this, Wells and Seelau (1995) recommend that "the suspect should not stand out in the lineup or photospread as being different from the distractors[5] on the basis of the eyewitness's previous description of the culprit or other factors that would draw extra attention to the suspect" (p. 779). Although eyewitness research has focused on ensuring that all members of a lineup match the witness' description of the culprit, little attention has been given to exploring whether other factors such as facial expressions of the members of a lineup could introduce bias into the lineup. Experiment 1 aimed to address this overlooked form of a biased lineup by investigating the possibility that a smile on one lineup member could arouse a feeling of familiarity, thereby increasing the likelihood of the member being perceived as resembling the culprit better than the other lineup members. Experiment 1 only employed target-absent lineups, as the primary interest was to determine whether the aroused sense of familiarity caused by a smile could contribute towards the increased risk of false identifications.

Memory Strength and Exposure Duration

Considering multiple factors can influence a witness' identification decision (Brewer et al., 2005), a mere presentation of a smiling face in a lineup may not always

[5] The term distractor can be used interchangeably with the term filler.

influence the witness' identification decision. For example, if a witness has a very strong memory of the culprit, regardless of how familiar a smiling face appear, the witness is unlikely to identify the face as the culprit's if it is not. In line with this, probabilistic mental models (Chase, Hertwig, & Gigerenzer, 1998; Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Kleinbölting, 1991) provide a theoretical framework within which to consider when a potentially biasing cue presented in a photospread, such as a smile presented by a lineup member might influence a witness' judgement regarding the likelihood of the lineup member being the culprit. The models suggest that two paths are available to making a binary yes/no decision, for example, when trying to decide whether a face has been seen before or not. The first path involves a process where an answer to the question is available in one's memory and, therefore, the answer is directly retrieved from the memory (Gigerenzer et al., 1991). The models suggest that people first attempt to make a decision by taking this path, thereby relying on their memory. When the first path fails to provide an answer, people take the second path, where they search for information that is related to the question to make an educated guess as to what the answer might be (Gigerenzer & Goldstein, 1996; Gigerenzer et al., 1991).

For example, imagine undergoing a facial recognition task where you are asked to study a face and, later, identify whether a face presented at test is old (i.e., seen at study phase) or new (i.e., not seen). If you have a strong memory (i.e., recollection) of the studied face, then a recognition decision would be made based on your memory (i.e., after taking the first path). A match between the memory of the studied face and the face presented at test would likely lead to a decision that the test face is old, whilst a mismatch would likely lead to a decision that the test face is new. On the other hand, if you have a poor memory of the studied face, you would likely

struggle to make a recognition decision based on your memory. In this instance, you might look for information other than the memory of the studied face to make a decision (i.e., take the second path). For example, you may rely on subjective feelings such as how familiar the test face appears. If the face presented at test arouses a feeling of familiarity, you may decide that the face is old even though you do not recollect that face in the context of the prior presentation. Conversely, if the face presented at test arouses no sense of familiarity, you would likely decide that the face is new. Now imagine that a novel face is presented at test with a biasing cue of a smile, which makes the face appear familiar. If you have a clear memory of the studied face, regardless of how familiar the novel face appears, based on your memory, you would likely make a correct decision that the test face is new. Conversely, if you have a poor memory of the studied face and, therefore, are relying on alternative information to make a recognition decision, the biasing cue may sway you to incorrectly answer that the test face is old. Therefore, the probabilistic mental models would suggest that a biasing cue such as a smile would influence people's recognition decisions only when, due to a poor memory, the decisions are made by primarily relying on alternative information (i.e., the second path) rather than on their own memory (i.e., the first path).

In a lineup context also, depending on how well a witness recollects the culprit, a sense of familiarity produced by a potentially biasing cue presented in a photospread may or may not influence the witness' identification decision. For example, a witness who clearly recollects the culprit in the context of the crime would likely make an identification decision by primarily relying on that recollection. In this instance, even if the witness is presented with a biased photospread, the biasing cue would not influence the witness' identification decision. A witness with a poor recollection,

however, may try to make an identification decision by relying on alternative information that is available from the lineup, looking for clues that might suggest one of the lineup members being the culprit. In this instance, the witness might try to make an identification decision by primarily relying on feelings of familiarity. Therefore, if the witness is presented with a biased photospread, a false sense of familiarity aroused by the biasing cue might be used to guide the identification decision.

In line with this, a number of memory models suggest that recognition judgements can be based on two distinct forms of memory: recollection of details about previous events (i.e., recollection) or assessment of stimulus familiarity (i.e., familiarity) (e.g., Yonelinas, 2002). Familiarity is thought to reflect a continuous index of memory strength, whilst recollection is thought to reflect the retrieval of specific information. Thus, recollection is believed to be more sensitive than familiarity to the quality of one's memory for a studied stimulus.

In line with my predictions based on the probabilistic mental models, the Yonelinas' dual-processing model (Yonelinas, 2002) assumes that when people are engaged in a memory based tasks, they would first attempt to make a decision by retrieving aspects of a studied event (i.e., recollection). The model assumes that it is when people are unable to retrieve any accurate qualitative information about the studied event (i.e., failure to gain adequate information from recollection), they would rely on their assessment of familiarity. Thus, the Yonelinas dual-processing model provides further support that a biasing cue presented in a photospread is more likely to exert influence over a witness' identification decision when the conditions are such that the witness' memory for the culprit is rather poor.

One way to manipulate the strength of a witness' memory for the culprit is to vary exposure duration to the culprit at encoding. For example, Memon, Hope, and Bull (2003) found that the proportion of false identifications made from a target-absent lineup almost doubled when witnesses were exposed to the close-up of the culprit's face for 12 seconds instead of 45 seconds (45.5% vs. 85%). Similarly, D. S. Lindsay, Read, and Sharma (1998) compared witnesses' performance between four memory conditions: best (i.e., three minutes exposure to the culprit, full attention to the culprit, no interval), good (i.e., one minute exposure, full attention, no interval), medium (i.e., 10 seconds exposure, full attention, 15 minutes interval), and worst (i.e., 10 seconds exposure, attention to peripheral information, 15 minutes interval). They found that the proportion of accurate identifications declined significantly as the memory condition worsened from the best to the worst conditions (86%, 78%, 44%, and 11%). Thus, exposure duration can have a strong effect on recognition performance, with a shorter exposure generally linked to poorer recognition accuracy than with a long exposure (Horry, Halford, Brewer, Milne, & Bull, 2014; Shapiro & Penrod, 1986). Accordingly, Experiment 1 employed two exposure duration conditions, where participants observed the culprit's face in a close-up view for either two seconds or one minute. In line with the probabilistic mental models, it was predicted that the relatively long exposure of one minute would lead to the biasing cue of a smile having no influence on witnesses' perception regarding the likelihood of the smiling lineup member being the culprit. Conversely, the shorter exposure of two seconds was predicted to lead to a sense of familiarity aroused by the smile cue being used as supporting evidence that the smiling lineup member was the culprit.

In sum, Experiment 1 explored whether an expression of a smile could function as a biasing cue when presented in a photospread. To test this, participants

were exposed to the culprit's face in close-up for either two seconds (short exposure condition) or one minute (long exposure condition). After a delay, participants saw a target-absent photospread, which contained an innocent suspect, who was presented either smiling (smile cue/biased condition) or with a neutral expression (neutral condition). Other members (i.e., fillers) were presented with neutral expressions. It was predicted that the smiling innocent suspect would appear more familiar, potentially increasing the risk of false suspect identifications. Given that the short exposure condition was predicted to lead to a weaker memory of the culprit than the long exposure condition, this effect was considered to be more likely to occur in the short than the long exposure condition.

Dependent Measures

Perceived familiarity. To measure the perceived familiarity of the innocent suspect relative to the fillers, participants were asked to rate how familiar each member of the photospread appeared, using a 7 point-Likert scale (1 = *not at all*; 7 = *very much so*). To determine whether a smile made the suspect appear more familiar than the fillers, a difference score represented by the suspect rating minus the maximum filler rating ($S_{uspect}–F_{iller\ Max}$) was calculated. For example, imagine that the suspect obtained a familiarity rating of 7 and the highest familiarity rating given to any fillers was 5. The difference score in this instance would become 7 – 5 = 2. Alternatively, if the suspect obtained a rating of 3 and the highest value given to the fillers was 6, the difference score would become 3 – 6 = -3. Therefore, a positive difference score would indicate that the suspect stood out from the photospread as the most familiar member, whilst a negative score would indicate otherwise. It was predicted that the familiarity difference score given to the smile cue condition would

be more positive than the neutral condition, indicating that the smile cue made the suspect stand out from the photospread as appearing highly familiar.

Main dependent measure. As discussed earlier, considering that a witness' identification decision can be influenced by various factors, a smile cue presented in a photospread may affect a witness' judgments about the stimuli without actually changing the witness' identification decision. For example, when a suspect is presented smiling in a photospread, the suspect may be judged as more familiar and, thus, stand out as most likely to be the culprit, yet this may not lead to a positive identification of the suspect if the witness feels that there is not enough evidence to do so. Thus, the effect of a smile cue may influence a witness' judgement about the smiling lineup member presented in the lineup, yet this effect may not be evident when a standard lineup procedure is used and a witness is asked to either make a positive identification (i.e., choose a member) or reject the lineup. The effect of a smile may be more likely to be detected with a more sensitive measure that is analogous to a continuous rather than a dichotomous measure of a match between each lineup member and a witness' memory for the culprit. One such measure is to ask witnesses to provide an indication as to how confident they are that each lineup member is the culprit. Developed by Sauer, Brewer, and Weber (2008), the measure enables witnesses to compare each lineup member to their memory of the culprit. This multiple confidence procedure has been shown to provide valuable information about the likelihood that a face has or has not been seen before (Brewer, Weber, Wootton, & Lindsay, 2012; Sauer et al., 2008; Sauer, Brewer, & Weber, 2012). Compared to the standard lineup procedures, the multiple confidence procedure has been shown to improve the accuracy of eyewitness performance when classification algorithms are

applied to the multiple confidence data to determine a positive response (for further detail, see Sauer et al., 2008).

The multiple confidence procedure is based on an accumulator model (Vickers, 1979), which proposes that when making a decision (e.g., old or new face), information is accumulated through a sequential sampling process over time. A decision is reached only when the accumulated information exceeds the decision criterion for one of the response alternatives. For example, when presented with a lineup member, a witness would come to either reject or accept the lineup member as being the culprit. In this instance, the model would suggest that a decision would be made by collecting evidence in two counters: a counter in support of accepting the lineup member as being the culprit and a counter in support of rejecting the lineup member. It is thought that only when the witness accumulates enough evidence towards one of the counters, an identification decision may be made in favour of the counter with the accepted level of accumulated evidence. According to a balance of evidence hypothesis, response confidence (e.g., obtained through the use of the multiple confidence procedure) reflects the evidential discrepancy between the two decision counters (Van Zandt, 2000; Vickers, 1979). Therefore, a high confidence value given to a lineup member would suggest that the eyewitness accumulated more information in favour of the lineup member being the culprit, as opposed to the lineup member not being the culprit. In a given lineup, the lineup member who receives the highest confidence rating would be considered as the member who resembled the culprit the most. Thus, the more the lineup member stands out as the likely match to be the culprit (i.e., likelihood of being identified from the lineup), the higher rating given to that lineup member would be compared to the ratings given to the other lineup members.

In line with this, instead of a standard lineup procedure, Experiment 1 used a multiple confidence procedure in which witnesses were presented with a simultaneous (i.e., all members presented at once) photospread and asked to provide a confidence rating (0-100%) for each photospread member. Therefore, from a six-persons photospread, six separate confidence ratings were obtained, one for each member. To index the difference between suspect and filler confidence ratings, a difference score of the suspect rating minus the maximum filler rating ($S_{uspect}–F_{iller\ Max}$) was calculated. For example, if a witness gave a confidence rating of 10% to the suspect and the highest confidence rating of 90% to one of the fillers, a difference score would be 10 – 90 = -80. Similarly, if a witness gave a suspect confidence rating of 50% and the highest filler rating of 20%, the difference score would be 50 – 20 = 30. As indicated by the two examples, a positive difference score would indicate that the suspect resembled the witness' memory for the culprit better than the fillers and, thus, when translated to the standard lineup procedure, the suspect had the highest chance of being identified from the photospread. On the other hand, a negative score would indicate that one of the fillers resembled the culprit better than the suspect. Therefore, I operationalized likelihood of a suspect being chosen from a photospread in terms of a confidence difference score. It was predicted that the confidence difference score would be more positive (or less negative) when the suspect was presented in a photospread smiling than with a neutral expression. Furthermore, as suggested earlier, I predicted that the effect of the smile cue would be more marked in the short than the long exposure condition.

Method

Each participant viewed two mock-crime videos (one after another) and one distractor video, then completed two multiple confidence procedures using

photospreads designed to correspond to each mock-crime. Finally, each participant was shown each photospread again and they were asked to rate how familiar each photospread member appeared. That is, two sets of data points were obtained from each participant. Eyewitness studies often employ a single encoding stimulus and lineup. However, considering multiple factors can affect witnesses' identification performance, a finding from a typical single stimulus design can be somewhat limited (Brewer et al., 2005). That is, significant findings arising from a single stimulus test cannot be confidently generalised to other situations (Wells & Windschitl, 1999). To assess generalizability, each participant was asked to undergo two lineup procedures.

Participants and Design

Two-hundred and fifty-eight (164 female, 94 male) paid community participants and first year psychology students for course credit participated in Experiment 1. Ages ranged from 16 to 42 years (M = 21.21, SD = 5.64). The experiment consisted of a 2 (cue: smile, neutral) × 2 (exposure duration: short, long) × 2 (mock-crime: male, female) design. Each participant saw one female and one male mock-crime video: one short and one long in duration. They then viewed one female and one male photospread: one had the suspect presented with a neutral expression (i.e., neutral cue condition) and another had the suspect presented smiling (i.e., smile cue condition).

Materials

Materials included two stimulus videos, a distractor video, and a set of photos used to create photospreads.

Stimulus videos. Two mock-crime videos were used, one displaying a Caucasian female culprit and another displaying a Caucasian male culprit. The female mock-crime video displayed a woman walking into a clothing store, shoplifting a

product and walking out of the store. The male mock-crime video showed a man walking into a food-court and taking a wallet that was sitting on a woman's table. For each video, short and long versions were created where the culprit's face was shown in close-up for either two seconds or one minute. Aside from the difference in the duration of the close-up, the two versions of each video were identical in content. The short versions of the mock-crime videos lasted approximately 20-30 seconds in total, whilst the long versions lasted approximately 1 minute and 20-30 seconds in total. The distractor video used was a 20 minutes scene taken from an animated movie.

Photospreads. For each culprit, two target-absent photospreads were created: one containing a smiling innocent suspect (smile cue) and another containing the same suspect with a neutral expression (neutral). All fillers were presented with a neutral expression. The neutral photospread and the photospread with a smile cue were identical to one another, with the exception of the suspect's photo. Each photospread consisted of six photos: five match-to-description fillers and one match-to-description suspect. Two actors were employed to act as innocent suspects, one for the male photospread and one for the female photospread. The actors were selected based on their match to the general descriptions of the culprits (e.g., age, gender, race, hair colour, face shape ,etc.). Photos of the fillers were collected from various face databases available online[6]. All photos were presented with white backgrounds. To ensure the appropriateness of the materials used for the photospreads, several pilot-tests were conducted.

Match-to-description and similarity tests. To select appropriate fillers for the photospreads, a group of participants (N = 5) viewed the two mock-crime videos and provided a written description for each culprit. These descriptions were

[6] Most of the images were obtained from the Florida Department of Corrections website, http://www.dc.state.fl.us/.

summarised into one modal description for each culprit. The male culprit was described as "Caucasian male; aged in early 20s; short, dark hair; medium to heavy build; roundish face". The female culprit was described as "Caucasian female; aged late teens to early 20s; medium height and build; wavy, shoulder length, light-brown to blond hair". In a classroom setting, another group of participants (N = 15) was presented with each of these descriptions along with the appropriate photos of one innocent suspect and eight potential fillers. All photos showed the faces in neutral expressions. Participants were asked to rate the extent to which each photo matched the given description, using a 7-point Likert scale (1 = *not match the description at all*; 7 = *absolutely matches the description*). Each photo was presented one at a time for approximately five seconds. To ensure that without a smile, the suspects did not stand out from the photospreads as resembling the culprits the most, participants were then shown the same set of photos, but this time, each photo was presented in a pair with a photo of the appropriate culprit (e.g., each male photo with the male culprit). Each pair was shown one at a time and participants were asked to rate the similarity between the two photos, using a 7-point Likert scale (1 = *not similar at all*; 7 = *very similar*). For each photospread, five photos that obtained the closest match-to-description and similarity ratings to the innocent suspect were chosen as the fillers.

As shown in Table 2.1, the female and the male photospreads showed that the average match-to-description and similarity ratings given to the photospreads were close to the ratings given to the suspects. Importantly, as shown by the ranges of ratings given to the photospreads, the suspects did not stand out as obtaining the highest match-to-description or similarity ratings. Therefore, when presented without a smile, the photospreads were considered not biased against the suspects.

Smiling manipulation check test. To obtain the smiling photos of the innocent suspects, each actor posing as the innocent suspect was asked to smile at the camera[7]. The appropriateness of the photos were then tested using the same group of participants (N = 15) who did the match-to-description and the similarity pilot-tests. They were presented with all of the photos used to create the photospreads (i.e., fillers, suspects smiling and in neutral expression), one at a time in a random order. For each photo, they were asked to rate the extent to which each person was smiling using a 7-point Likert scale (1 = *not smiling at all*; 7 = *definitely smiling*).

Table 2.1

Means (Standard Deviations) for the Match-to-Description and Similarity Ratings Given to the Female and the Male Photospreads: Suspect Ratings, Average Photospread Ratings and Ranges of Ratings Given to the Photospreads

	Female			Male		
	Suspect	Average	Ranges	Suspect	Average	Ranges
Match-to-description	4.60 (1.06)	4.09 (0.77)	3.57 – 4.73	3.40 (1.50)	3.89 (1.19)	3.00 – 4.53
Similarity	2.47 (1.55)	2.79 (0.85)	2.00 – 3.27	2.33 (1.23)	2.60 (1.16)	2.20 – 3.13

Paired samples t-tests showed that the smiling female suspect (M = 6.33, SD = 1.59) was perceived as smiling significantly more than when she was presented with a neutral expression (M = 1.13, SD = 0.35), $t(14) = 11.06$, $p < .001$, $f = 2.68$, or when compared to the average rating given to the fillers, (M = 1.46, SD = 0.55), $t(14) = 9.90$, $p < .001$, $f = 2.28$. Similarly, the smiling male suspect (M = 4.67, SD = 1.72)

[7] Photos of the suspects with neutral expressions were also taken at the same session, ensuring that there were minimal differences between the smiling and the neutral photos of each suspect.

was perceived as smiling more in comparison to when he was presented with a neutral expression, (M = 1.63, SD = 0.51), $t(14) = 8.41$, $p < .001$, $f = 1.36$, or when compared to the average rating given to the fillers (M = 2.8, SD = 1.21), $t(14) = 5.14$, $p < .001$, $f = .64$. Therefore, the smiling suspects were seen as smiling more than the fillers and the neutral versions of the suspects.

Effective size and defendant bias measure. To further ensure that the photospreads were fair without a smile cue, in a classroom setting, another group of participants were presented with the neutral female (N = 198) and male (N = 214) photospreads along with the written descriptions of the culprits (used in the match-to-description test). From each photospread, they were asked to choose one member who best matched the written description of the culprit. From the data obtained, the photospreads' effective sizes (i.e., number of plausible members) were calculated (Malpass, 1981; Malpass & R. C. L. Lindsay, 1999). The effective sizes were then used to calculate the chance expectations of the suspects being chosen (i.e., 1/ effective size). Tredoux (1999) suggests that one way of obtaining a defendant bias measure (i.e., ensuring the fairness of a lineup against the suspect) is to see if the 95% confidence interval of the proportion of the choosing of the suspect and that expected by the chance (i.e., calculated from the effective size) overlap with one another.

The effective size of the female photospread was 4.85 (i.e., there were around five plausible members). The effective size indicated that the chance expectation of the suspect being chosen was .21 (95% CI = .15, .27), whilst the actual choosing was .18 (95% CI = .13, .23). Therefore, the female photospread was considered not biased against or in favour of the suspect, when she was presented without a smile. For the male photospread, the effective size was 4.6 (i.e., there were around four to five plausible members). Using the effective size, the chance expectation of the suspect

being chosen was calculated to be .22 (95% CI = .16, .27), whilst the actual choosing was .07 (95% CI = .04, .10). Therefore, the male photospread tended to favour the suspect when he was presented without a smile. Considering that the main concern was to ensure that the photospreads were not biased against the suspects when presented without a smile, the photospreads were considered acceptable.

Procedure

Upon arrival, participants were randomly allocated to one of the combinations of the conditions, and were then led to an individual cubicle and seated at a computer. The computer provided all of the instructions and, thus, participants' progress throughout the experiment was self-paced. Participants began the experiment by watching either of the following combinations of the two mock-crime videos: (1) the male short video and the female long video, (2) the female short video and the male long video. The videos were shown one after the other, with the orders of the videos counterbalanced between participants. Prior to each video, participants were asked to pay attention to the video as they would be asked some questions about it later. Participants then viewed the distractor video, which lasted approximately 20 minutes.

The computer screen then instructed participants to think back to the two mock-crime videos they saw earlier. They were informed that they were about to view two photospreads, one for each mock-crime video. Each participant saw a combination of either (1) the female photospread with a smile cue and the male neutral photospread, or (2) the female neutral photospread and the male photospread with a smile cue. The presentation order of the photospreads was counterbalanced between participants and, thus, some participants initially saw the female photospread, whilst others initially saw the male photospread. The positioning of the photos were also randomised by the computer. Prior to viewing each photospread,

participants were instructed that the photospread may or may not contain the actual culprit. For each photospread, participants were presented with six photos (i.e., one innocent suspect and five fillers) with a text box beneath each photo. They were instructed to fill each and every text box with percentages between 0 and 100%, indicating how confident they felt that the each person was the culprit, using the number keys to enter the confidence rating. Once all of the text boxes for the first photospread were filled, the screen moved on to the second photospread, where the instruction and the procedure were repeated.

Finally, participants were presented with each photospread five more times. Each time, they were asked to rate all of the photos on each of the following five perceived personal characteristics: familiarity, trustworthiness, nervousness, likableness and attractiveness, using a 7-point Likert scale (1 = *not at all*; 7 = *very much so*). The primary interest was the familiarity rating, with the other items acting as distractors.

Results

An alpha level of .05 was adopted for all inferential analyses. The effect size estimate reported throughout was Cohen's *f*, with approximate cut-off guidelines for small, medium and large effects being .10, .25 and .40. Where comparison of two means was made, *f* was also reported for ease of interpretation, (*f* is equal to half *d*).

Whilst each subject completed two sets of mock-crimes, these mock-crimes were treated separately where each analysis was conducted separately for the female and the male stimuli[8]. Of the 258 subjects' data obtained for the female stimuli, 18

[8] The analyses were split between the male and the female stimuli for it was considered that a smile cue could impact the two stimuli differently and, thus, when analysed together, the effect found for one stimuli might skew the effect found for the other stimuli. For example, a smile could increase the confidence difference score for the female stimuli whilst have no impact on the male stimuli's confidence difference score and, thus, when analysed together, the effect of the smile cue might not be evident, leading to the true effects of the smile cue being missed.

subjects' data were excluded due to the misuse of the multiple confidence procedure. In particular, some participants used a 1–6 ranking system to rate the photospread members, whilst one participant indicated that he failed to read the instruction and therefore randomly entered 100 to all of the members. Of the 258 subjects' data obtained for the male stimuli, 15 subjects' data were excluded for the same reason. Therefore, the analyses of the female stimuli were carried out using 240 subjects' data, whilst the analyses of the male stimuli were carried out using 243 subjects' data.

As explained earlier, for each of the familiarity ratings and the confidence ratings, a difference score of the suspect rating minus the filler maximum rating ($S_{uspect}–F_{iller\ Max}$) was calculated. The possible range for the familiarity difference score was -6 to 6, -6 indicating that the suspect obtained a familiarity rating of 1 and a filler obtained the maximum rating of 7 (i.e., 1 – 7 = -6), whilst 6 indicating that the suspect obtained a familiarity rating of 7 and a filler obtained the maximum rating of 1 (i.e., 7 – 1 = 6). The possible range for the confidence difference score was -100 to 100, -100 indicating that the suspect obtained 0% and a filler obtained the maximum value of 100% (i.e., 0 – 100 = -100), whilst 100 indicating that the suspect obtained 100% and a filler obtained the maximum value of 0% (i.e., 100 – 0 = 100).

Perceived Familiarity Rating

For the female and the male stimuli, independent samples t-tests were conducted, treating the cue variable as the independent variable and the familiarity difference ($S_{uspect}–F_{iller\ Max}$) score as the dependent variable. It was predicted that the smile cue condition would produce a more positive familiarity difference score than the neutral condition. As shown in Table 2.2, for both female and male stimuli, the smile cue condition produced less negative familiarity difference scores than the neutral condition. Therefore, relative to the fillers, the suspects obtained higher

familiarity ratings when presented smiling than with a neutral expression. However, as evident by the negative values obtained in the smile cue condition, the cue did not make the suspects stand out as appearing most familiar.

Table 2.2

Summary of Descriptive and Inferential Statistics for the Effect of the Cue on the Familiarity Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Smile	-0.69	1.90	-1.03, -0.35	-0.59	1.97	-0.95, -0.23
Neutral	-1.67	2.03	-2.04, -1.30	-1.15	2.13	-1.53, -0.77
	$t(237) = 3.84$, $p < .001$, $f = .25$			$t(240) = 2.12$, $p = .04$, $f = .14$		

Multiple Confidence Rating

As explained earlier, a maximum confidence value given to the suspect would indicate that the suspect was perceived as resembling the culprit the most. Overall, the proportions of the suspects receiving the maximum confidence values were rather low. Of the 240 subjects who viewed the female stimuli, 34 (14.2%) gave a unique maximum confidence value to the suspect, 152 (63.3%) gave it to a filler, and 54 (22.5%) gave no single unique maximum confidence value to the photospread (e.g., all members received 0 or the suspect and at least one of the fillers received the same values). Similarly, of the 243 subjects who viewed the male stimuli, 28 (11.5%) gave a unique maximum confidence value to the suspect, 164 (67.5%) gave it to a filler, and 51 (21%) gave no single unique maximum confidence value to the photospread[9].

[9] Throughout this thesis, further analyses were conducted using the unique confidence values as the dependent measure instead of the confidence difference scores. In particular, I further explored if the

To determine whether the photospread with a smile cue led to a more positive (or less negative) confidence difference (S_{uspect}–$F_{iller\ Max}$) score than the neutral photospread, a 2 (cue: smile, neutral) × 2 (exposure: short, long) factorial ANOVA analysis was conducted separately for the female and the male stimuli[10]. The outcomes for the male and the female stimuli differed. For the female, no main effects of cue, $F(1, 236) = .26$, $p = .61$, $f = .03$, nor the exposure condition, $F(1, 236) = 1.57$, $p = .21$, $f = .08$, were observed. However, there was a significant interaction between the cue and the exposure condition, $F(1, 236) = 3.79$, $p = .05$, $f = .13$. Simple effects analyses showed that, in the short exposure condition, the smile cue condition produced a non-significant weak increase in the confidence difference score than the neutral condition, $t(104.89) = 1.81$, $p = .07$, $f = .17$ (see Table 2.3). In the long exposure condition, the confidence difference score given to the two cue conditions did not differ, $t(119) = -0.98$, $p = .33$, $f = .09$. For the male stimuli, the interaction between the cue and the exposure condition was non-significant, $F(1, 239) = .2.68$, $p = .10$, $f = .11$. Furthermore, no main effects of cue, $F(1, 239) = .02$, $p = .90$, $f = .01$, nor the exposure condition, $F(1, 239) = 1.21$, $p = .27$, $f = .07$, were observed. The descriptive statistics are provided in Table 2.3.

Discussion

Experiment 1 produced three key findings. First, relative to the fillers, both female and male suspects were perceived as appearing more familiar when presented in the photospreads smiling rather than with a neutral expression. Second, for the

---

IVs such as the cue condition would have any effects on the unique confidence values provided to the suspect/culprit, fillers and to neither. No significant findings emerged and, thus, the results are not reported.

[10] The variables of the presentation orders of the mock-crime videos and the photospreads were omitted from the analyses for they were peripheral to the main interests of Experiment 1 and also, they showed no significant effects.

Table 2.3

Summary of Means, Standard Deviations and 95% CIs for the Effects of the Cue and the Exposure Condition on the Confidence Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Short						
Smile	-22.41	43.71	-33.90, -10.92	-25.21	38.27	-35.36, -15.06
Neutral	-35.16	32.39	-43.46, -26.87	-32.56	39.51	-42.68, -22.44
Overall	-28.95	38.70	-35.98, -21.93	-29.01	38.92	-36.11, -21.91
Long						
Smile	-26.02	44.37	-37.38, -14.66	-38.57	35.28	-47.46, -29.69
Neutral	-18.52	39.66	-28.77, -8.28	-29.94	38.94	-39.83, -20.05
Overall	-22.30	42.09	-29.88, -14.72	-34.29	37.25	-40.09, -27.70
Overall						
Smile	-24.26	43.90	-32.23, -16.29	-32.23	37.18	-38.95, -25.51
Neutral	-26.90	36.98	-33.56, -20.24	-31.24	39.09	-38.22, -24.26
Overall	-25.60	40.50	-30.75, -20.45	-31.72	38.08	-36.53, -26.91

female stimuli, as predicted, the effect of the smile cue on the confidence difference score was observed in the short exposure condition but not in the long exposure condition. Third, in contrast, the male stimuli failed to show any effects of the smile cue on the confidence difference score.

The results of the female stimuli provide some evidence that an expression of a smile on one lineup member's face might function as a familiarity cue under some circumstances, potentially biasing a lineup. It is important to note, however, the

moderating effects of the cue and the exposure duration conditions on the confidence difference score were rather weak, indicating that replications are necessary before clear conclusions can be made. The results are perhaps caused by the fact that the short exposure condition did not sufficiently undermine the participants' memory for the culprit. Considering that participants did not make identification decisions, the proportions of accurate decisions made between the two exposure conditions could not be compared. However, the maximum confidence value given to a photospread can be considered a proxy for memory strength. That is, if the participants' memory for the culprit was significantly impaired in the short exposure condition, the strongest match to their memory for the culprit might have been rated lower in confidence in the short than the long exposure conditions. In contrast to this prediction, for the female stimuli, the maximum confidence values given to the short and the long exposure duration conditions did not differ (M = 47.98, SD = 33.51 vs. M = 43.93, SD = 34.89), $t(238) = 0.92$, $p = .36$, $f = .06$[11]. Therefore, it is feasible that the exposure duration of two seconds led to the participants having somewhat weakened memory of the culprit, yet not weak enough to rely heavily on feelings of familiarity, thus, leading to the smile cue only having a weak effect. Thus, it could be argued that, perhaps, further undermining participants' memory for the culprit could lead to a more robust effect of the smile cue being observed. This notion is further discussed, and tested, in Experiment 2 (Chapter 3).

Alternatively, given that many factors can influence a witness' perceptions of lineup members (Wells & Olson, 2003), perhaps the weak effect of the smile cue observed for the female stimuli in Experiment 1 is not surprising. For example, it might be that there were fillers who resembled the culprit much better than the

[11] The male photospread showed the same results. That is, the maximum confidence values given to the short (M = 42.73, SD = 27.83) and the long (M = 38.24, SD = 31.33) exposure duration conditions did not differ, $t(240.13) = 1.18$, $p = .24$, $f = .08$.

innocent suspect, leading the smile cue to have little impact. In this instance, regardless of the presence or absence of the smile cue, the participants might have primarily relied on alternative information such as the close resemblance of the fillers to the culprit to make their confidence assessment. Indeed, collapsed across the cue conditions, the female photospread contained one filler who consistently received higher confidence rating than the female suspect (M = 20.71, SD = 29.02 vs. M = 15.50, SD = 25.26), $t(239) = -2.30$, $p = .02$, $f = .10$. This likely indicates that even when the smile cue aroused a sense of familiarity for the female suspect, the cue did not lead to the suspect stand out as resembling the culprit the most. Therefore, the effect of the smile cue might be more prevalent when it is displayed by a lineup member who resembles the culprit relatively better than the fillers. Experiment 3 (Chapter 3) examined this possibility.

Two reasons are apparent as to why the male stimuli failed to replicate the results of the female stimuli. First, as shown by the defendant bias measure obtained from the pilot-test, even when presented with a neutral expression, the male photospread was biased in a way that favoured the innocent suspect. Thus, regardless of how familiar the smile cue made the suspect appear, he was likely to be perceived as resembling the culprit poorly relative to the fillers. Therefore, for the smile cue to have an impact, the cue may need to be displayed by a member of a photospread who resembles the culprit enough to be considered at least a plausible match.

Second, the smile cue led to a much smaller increase in the familiarity difference score for the male than the female stimuli. Thus, a smile displayed by the male suspect was perhaps, overlooked by the participants as being potentially valuable evidence to be incorporated into their confidence assessments. Whilst it is unclear as to why this difference occurred, such results further reflect that as the

probabilistic mental models would suggest, additional information such as a sense of familiarity would not always bias witnesses' confidence assessments. Adjustments were made to the male photospread in an attempt to eliminate these potential methodological limitations in Experiments 2 and 3 (Chapter 3).

CHAPTER 3

Experiment 2

As predicted, Experiment 1 showed that a smile displayed by an innocent suspect presented in a photospread could arouse a feeling of familiarity. However, the expression did not necessarily result in the smiling face being assessed as resembling the culprit the most. In Experiment 2, I aimed to replicate the results obtained in Experiment 1 with the female stimulus materials and also examined if several methodological adjustments allowed a similar effect to emerge with the male stimuli. In Chapter 2, I suggested two possible reasons for the smile cue failing to produce the same effects for the male stimuli as were observed with the female stimuli. First, the effect of the smile cue might have been negated for, as suggested by the defendant bias measure calculated from the pilot-test, the male neutral photospread used in Experiment 1 was biased in a way that favoured the suspect. Second, for unknown reason, the male suspect's smile only showed a weak effect of familiarity, likely leading to the smile cue having no impact. To address these issues, in Experiment 2 a new male photospread was constructed. In particular, when the new innocent suspect was presented with a neutral expression instead of a smile, the photospread was considered not biased against or in favour of the suspect. As in Experiment 1, at the end of Experiment 2, participants were asked to rate the familiarity of the suspects and the fillers to see if the innocent suspects' smiles aroused a sense of familiarity.

Divided Attention and Memory Quality

Another objective in Experiment 2 was to strengthen the manipulation of memory strength in an attempt to provide a more decisive test of the predictions based on the probabilistic mental models (Gigerenzer et al., 1991). This was done by introducing a manipulation designed to further weaken the quality of participants'

memory for the culprit so that they might be more likely to rely on the sense of familiarity aroused by a smile cue.

Wells and Olson (2003) suggest that the accuracy of an identification decision is not simply predicted by the length of the exposure to the culprit that a witness has but also by the level of attention the witness pays to the culprit. Thus, in an attempt to further undermine participants' memory for the culprits, Experiment 2 manipulated both the exposure duration at encoding and the level of attention participants paid to culprits, whilst watching the mock-crime videos. Modelling a procedure used by Palmer, Brewer, McKinnon, and Weber (2010), the mock-crime videos were accompanied by a soundtrack consisting of series of tones randomised for pitch (high versus low) and intervening interval (1s vs. 2s). In the long exposure duration condition, participants were told to ignore the tone, attending to the videos under a full attention condition (i.e., better encoding condition). In the short exposure duration condition, participants were told to identify whether each tone was low or high pitched as quickly as they could (i.e., poor encoding condition). Using a standard eyewitness identification procedure, Palmer et al. (2010) found that this attention manipulation significantly reduced the proportion of correct identifications made in the divided attention condition compared to the full attention condition (49.7% vs. 60.4%). Accordingly, I predicted that in Experiment 2, by further undermining participants' memory quality for the culprits, the poorer encoding condition of the divided attention and the short exposure duration manipulations would cause their confidence assessments to be influenced by the smile cue.

Method

Participants and Design

Experiment 2 consisted of 267 (165 female, 98 male) paid community participants and first year psychology students who participated for course credit. Ages ranged from 16 to 60 years (M = 21.68, SD = 6.19). As discussed earlier, the main interest of Experiment 2 was to use the combined effects of the attention manipulation and the exposure duration at encoding to gain a stronger manipulation of the quality of participants' memory for the culprit. Thus, the aim was to create and compare two encoding conditions: the poor (i.e., short exposure duration/ divided attention) and better (i.e., long exposure duration/ full attention) conditions. However, due to an error made to the experimental software, the actual design consisted of a 2 (cue: smile, neutral) × 2 (exposure duration: short, long) × 2 (attention: divided, full) × 2 (mock-crime: male, female) design. That is, instead of having the two proposed cells (i.e., short/divided and long/full), Experiment 2 collected data for four cells, some with the combined effects of strengthening and weakening encoding manipulations (i.e., short/full and long/divided), which were of no interest. The issue regarding the design error will be further discussed in the Results section. As with Experiment 1, participants were randomly assigned to some combination of variables, with the exception of the attention variable, which was manipulated between-subjects. Presentation orders of the mock-crime videos and the photospreads were counterbalanced across subjects.

Materials

Stimulus videos. Four stimulus videos (short and long versions of the male and the female mock-crimes) from Experiment 1 were used. However, for the purpose of the attention manipulation, a soundtrack was added to each video. The soundtrack

consisted of series of tones randomised for pitch (high versus low) and intervening interval (1s versus 2s).

Photospreads. The female photospread (i.e., smile cue and neutral) from Experiment 1 was used. To create the new male photospreads (i.e., smile and neutral), photos of the individuals who matched the general description of the male culprit were collected. Photos (i.e., smiling and in a neutral expression) of an innocent suspect were obtained by employing a new male actor. Photos of potential fillers were obtained from various face databases available online[12]. Using these photos, the pilot-tests described in Experiment 1 (Chapter 2, Method section) were conducted to create the photospreads.

Match-to-description, similarity and smile manipulation check tests. A group of participants (N = 15) completed the tests. When the male suspect was presented with a neutral expression, on the match-to-description scale, he was rated similarly to the average rating given to the photospread (scale of 1-7, M = 5.60, SD = 1.12 vs. M = 5.28, SD = 0.89; range = 4.20 to 5.87). Importantly, the suspect did not receive the highest match-to-description rating. On the similarity scale, the suspect rating and the average photospread rating (scale of 1-7, M = 3.40, SD = 1.12 vs. M = 3.82, SD = 1.07; range = 2.47 to 4.40) also showed that, when presented with a neutral expression, the suspect did not stand out from the photospread as resembling the culprit the most. Therefore, the results indicated that the male photospread consisted of a suspect who matched the description of the culprit just as well as the fillers, yet did not stand out as resembling the culprit the most.

For the smile manipulation check test, a series of paired samples t-tests showed that the male suspect was rated higher on the smiling scale (i.e., more

[12] The majority of the photos were obtained from the Florida Department of Corrections website.

smiling) when he was presented smiling (M = 6.80, SD = 0.41) than with a neutral expression (M = 1.2, SD = 0.41), $t(14) = -29.44$, $p < .001$, $f = 6.83$, or when compared to the average rating given to the fillers (M = 1.59, SD = 0.49), $t(14) = -30.35$, $p < .001$, $f = 5.79$. Therefore, the smiling photo of the male suspect was considered an appropriate replacement for the suspect's neutral photo to compose a photospread with a smile cue.

Effective size and defendant bias measure. Another group of participants (N = 219) completed the effective size/ defendant bias measure pilot-test, which was conducted in a classroom setting. The general description of the male culprit, which was obtained for the match-to-description test (Experiment 1) was used. When the suspect was presented in a neutral manner, the effective size of the male photospread was estimated to be 5.6 (i.e., around five effective members). The chance expectation of the suspect being chosen based on the effective size was .18 (95% CI = .13 - .23), overlapping with the proportion of the actual choosing of the suspect, which was also .18 (95% CI = .13 - .23). Therefore, the male photospread was considered not biased against or in favour of the suspect when he was presented without a smile.

Procedure

The procedure was identical to Experiment 1 except for the addition of the attention manipulation at encoding. Each participant saw one male mock-crime video and one female mock-crime video with the added sound track. Those in the full attention condition were told that the sound track was irrelevant to the experiment and to ignore the tones. Those in the divided attention condition were instructed to acknowledge the occurrence of the tones by pressing keys marked "low" for low pitched tones and "high" for high pitched tones. Participants were reminded that it was very important to be as fast as they could with their responses without making

any mistakes. In order to ensure that participants understood the instruction regarding the tones, they experienced some practice tones prior to the presentation of the mock-crime videos. During the practice phase, those in the full attention condition were asked to practice ignoring the tones, whilst those in the divided attention condition were asked to respond to each tone.

Results

Of the 267 subjects' data obtained for the female stimuli, 29 were omitted: following Palmer et al. (2010), 23 were due to the subjects making more than 1/3 errors on the divided attention task and, therefore, failing to comply with the divided attention manipulation, whilst six were due to the misuse of the confidence scale (as described in Chapter 2). For the male stimuli, of the 267 subjects' data, 26 were omitted: 21 were due to the subjects making more than 1/3 errors on the divided attention task and five were due to the misuse of the confidence scale. This led to 238 subjects' data available for the female stimuli and 240 subjects' data available for the male stimuli.

As discussed earlier, due to a technical error made with the software, the two memory manipulations of attention and exposure duration were crossed with each other, leading to four combinations of encoding conditions, some with combined effects of strengthening and weakening encoding manipulations (i.e., short/full vs. long/divided) instead of the two conditions of the interest (i.e., short/divided vs. long/full). The main interest of Experiment 2 was to use the combined effects of the manipulations of the exposure duration and attention at encoding to gain a stronger manipulation of the quality of the witnesses' memory for the culprit. To reflect this, of the available data, analyses focused on comparing the poor (i.e., short encoding duration/ divided attention) and the better (i.e., long encoding duration/ full attention)

encoding conditions, eliminating the other half of the data[13]. This led to the analyses being conducted using 116 subjects' data for the female stimuli and 117 subjects' data for the male stimuli.

As in Experiment 1, analyses were conducted separately for the female and the male stimuli. Prior to running the analyses, difference (S_{uspect}–$F_{iller\ Max}$) scores were calculated for the multiple confidence ratings and the familiarity ratings.

Perceived Familiarity Rating

Independent samples t-tests revealed that, consistent with Experiment 1, the female stimuli showed that the smile cue condition produced a less negative familiarity difference score than the neutral condition. However, as displayed in Table 3.1, the mean difference score in the smile cue condition was negative, indicating that the smiling suspect did not stand out from the photospread as appearing most familiar. In contrast, the male stimuli showed no difference between the two cue conditions.

Table 3.1

Summary of Descriptive and Inferential Statistics for the Effect of the Cue on the Familiarity Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Smile	-0.77	2.16	-1.32, -0.22	-1.60	2.03	-2.12, -1.08
Neutral	-1.56	1.98	-2.10, -1.03	-1.68	2.44	-2.34, -1.02
	$t(114) = 2.06$, $p = .04$, $f = .19$			$t(115) = 0.19$ $p = .85$, $f = .02$		

[13] Analyses conducted with the full data set showed no significant effects, with the exception of the results of the familiarity difference score for the male stimuli, which showed a non-significant trend for the smile cue condition (M = -1.34, SD = 2.05) to have a less negative familiarity difference score than the neutral condition (M = -1.85, SD = 2.11 $t(238) = 1.90$, $p = .06$, $f = .12$.

Multiple Confidence Rating

Of the 116 subjects who viewed the female photospread, 13 (11.2%) gave a unique maximum confidence value to the suspect, 79 (68.1%) gave it to one of the fillers and 24 (20.7%) gave no single unique maximum confidence value (i.e., the suspect and at least one of the fillers obtained equal values). Similarly, of the 117 subjects who viewed the male photospread, 18 (15.4%) gave a unique maximum confidence value to the suspect, 79 (67.5%) gave it to one of the fillers and 20 (17.1%) gave no single unique maximum confidence value.

To test the effect of a smile cue, using the dependent measure of the confidence difference (S_{uspect}–$F_{iller\ Max}$) score, a 2 (cue: smile, neutral) × 2 (encoding: short/divided, long/full) factorial ANOVA analysis was conducted separately for the female and the male stimuli[14]. The female stimuli showed no main effects of cue, F (1, 112) = 1.22, p = .27, f = .11, encoding condition, F (1, 112) = 1.12, p = .29, f = .10, nor an interaction between the cue and the encoding condition, F (1, 112) = 2.28, p = .13, f = .15. The male stimuli also showed no main effects of cue, F (1, 113) = 2.74, p = .10, f = .16, encoding condition, F (1, 113) = 0.62, p = .43, f = .08, nor an interaction between the cue and the encoding condition, F (1, 113) = 0.01, p = .93, f = .16. The descriptive statistics are shown in Table 3.2. To note, analyses conducted with the other half of the data set (i.e., short exposure duration/ full attention and the long exposure duration/ divided attention conditions) showed no significant effects (the descriptive statistics are provided in Appendix A).

[14] The variables of the presentation orders of the mock-crime videos and the photospreads were excluded from the analyses for the analyses showed no presentation order effects, with the exception of the female stimuli, which showed that presenting the photospread first (M = -41.05, SD = 35.72) led to a more negative confidence difference score than presenting it second (M = -22.00, SD = 42.25), t(114) = -2.63, p = .01, f = .24.

Table 3.2

Summary of Means (Standard Deviations) for the Effects of the Cue and the Encoding Condition on the Confidence Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Short/ Divided						
Smile	-34.81	35.94	-47.99, -21.63	-36.15	35.67	-50.56, -21.74
Neutral	-37.80	30.52	-50.40, -25.20	-23.28	38.35	-37.11, -9.45
Overall	-36.14	33.36	-45.07, -27.21	-29.05	37.41	-38.89, -19.21
Long/ Full						
Smile	-38.13	38.23	-52.41, -23.86	-43.28	49.05	-61.94, -24.62
Neutral	-18.83	49.84	-37.44, -0.22	-29.03	51.23	-48.16, -9.90
Overall	-28.48	45.10	-41.12, -17.83	-36.03	50.25	-49.13, -22.94
Overall						
Smile	-36.44	36.81	-45.87, -27.01	-39.91	43.01	-51.54, -28.28
Neutral	-27.45	42.88	-39.04, -15.86	-26.06	44.76	-37.43, -14.69
Overall	-32.18	39.88	-39.51, -24.85	-32.57	44.30	-40.68, -24.46

Discussion

Consistent with Experiment 1, relative to the fillers, the female suspect was perceived as more familiar, when she was presented in the photospread smiling than with a neutral expression. Despite this, the smile cue did not increase the perceived resemblance of the female suspect to the culprit, as measured by the confidence difference score. For the male stimuli, the new male suspect's smile did not make the suspect appear more familiar nor did it increase the perceived resemblance between

the suspect and the culprit. Thus, Experiment 2 failed to replicate the findings of the female stimuli from Experiment 1.

Three findings from Experiment 2 merit comment. First, the female stimuli failed to replicate the results of Experiment 1, despite the fact that the same stimulus materials were used in Experiment 2. This may have been due to the fact that the female suspect was perceived as resembling the culprit rather poorly in Experiment 2 compared to Experiment 1. Indeed, the proportion of witnesses who gave a unique confidence value to the female suspect was somewhat less in Experiment 2 compared to Experiment 1, indicating that, for some reason, the same female suspect was perceived as resembling the culprit relatively less than the fillers in Experiment 2. Therefore, in Experiment 2, even when the smile cue aroused a moderate level of familiarity for the suspect, participants might have rejected the possibility that the suspect could be the culprit.

Second, despite the adjustments made to the male photospread, there was no effect of the smile cue on the familiarity difference score for the male stimuli. Perhaps, this suggests that unlike the smile expressed by the female suspect, something about the male suspect's smile was odd, failing to cue a sense of familiarity. Or perhaps, the results suggest that particular facial characteristics of the female and the male suspects used in Experiments 1 and 2 produced the difference in effects. For example, a smile expressed by feminine versus masculine faces, or roundish versus slim faces might lead to a smile influencing the perceived characteristics of faces differently. The existing facial recognition research has not thus far reported any associations between particular facial characteristics, a smile and a sense of familiarity (Baudouin et al., 2000; Garcia-Marques et al., 2004). Whilst this is an interesting point to consider, it is beyond the scope of the current research.

Nevertheless, the results of Experiment 2 suggest that, an expression of a smile might not always lead to a face appearing familiar. In Experiment 3, I attempted to further standardise an expression of a smile expressed by the female and the male suspects to see if the difference would disappear.

Third, Experiment 2 showed no apparent effect of the encoding manipulation. This may indicate that the addition of the divided attention manipulation did not affect the quality of participants' memory for the culprit as intended. Indeed, as with Experiment 1, further analyses showed that on average, participants in the poor encoding condition (i.e., short exposure duration/ divided attention) provided just as high maximum confidence values to the photospreads as those in the better encoding condition (i.e., long exposure duration/ full attention) (M = 49.20 vs. 46.50, for the female; M = 57.91 vs. 59.42, for the male). Thus, the combined effects of the exposure duration and attention manipulations at encoding had no impact on the maximum confidence values given to the photospreads.

Alternatively, it may be argued that, regardless of the effect of the divided attention manipulation on participants' memory for the culprit, any increased sense of familiarity produced by the smile was negated because the photospreads contained highly plausible fillers, who resembled the witnesses' memory for the culprits better than the suspects. Indeed, an additional analysis showed that regardless of the cue condition, the female photospread contained one filler who consistently received a higher confidence rating than the female suspect (M = 27.97, SD = 33.69 vs. M = 12.84, SD = 22.09), $t(237) = -6.14$, $p < .001$, $f = .27$. Similarly, the male photospread contained one filler who consistently received a higher confidence rating than the male suspect (M = 25.08, SD = 31.14 vs. M = 20.06, SD = 28.13), $t(239) = -2.06$, $p = .04$, f = 08. Therefore, both female and male photospreads contained fillers who

tended to stand out as resembling the culprits better than the suspects. In saying so, such speculation needs to be considered with caution for in Experiment 1, whilst the female stimuli showed the same patterns of results regarding the memory manipulation and the presence of a better matched filler, some limited evidence regarding the effect of the smile cue on the confidence difference score was observed. Therefore, as suggested earlier, it is likely that in addition to these factors, the non-significant results obtained in Experiment 2 likely point towards the notion that the female suspect was perceived to be even a less of a match to be the culprit in Experiments 2 than1, leading to the smile cue to have no effects on the confidence assessments.

Together, the results further point towards the notion that a smile presented in a photospread would not always function as a biasing cue[15]. However, when displayed by a photospread member who closely resembles the culprit, the cue may exert its influence, leading to witnesses perceiving that the smiling photospread member could be the culprit due to a sense of familiarity aroused by the smile. Experiment 3 attempted to test whether the smile cue would exert its predicted influence on witnesses' confidence assessments when the cue was displayed by innocent suspects who were already vulnerable of being perceived as resembling the culprits better than the fillers. That is, Experiment 3 used new innocent suspects, who were rated higher on the culprit-similarity scale (i.e., pilot-test) than the fillers, yet did not stand out from the photospreads as matching the general descriptions of the culprits better than the fillers.

[15] Of course, the halving of the sample size due to the errors made to the experimental software would likely have reduced the power and, thus, the non-significant results observed in Experiment 2 needs to be interpreted with caution. Indeed, post-hoc power analysis (using GPower software) estimated low power (i.e., < 0.5) for the results of both the female and the male stimuli. Therefore, further replications of the study with larger sample sizes could perhaps lead to the effect of the smile cue emerging in the predicted direction. To note, no significant effects (including the main effect of the smile cue) were observed when the full data were used in the analyses.

Experiment 3

In Experiment 3, two major adjustments were made to the stimuli used thus far. First, adjusted sets of match-to-description photospreads were composed for the female and the male stimuli. To do so, the data collected from the match-to-description and similarity pilot-tests conducted for the preceding experiments were reanalysed. For each photospread, the photo, which was rated highly on the similarity scale (to the culprit) was chosen as the new innocent suspect. Fillers consisted of the photos, which were rated similarly on the match-to-description scale, yet significantly lower on the similarity scale to the new innocent suspects. The aim was to ensure that the innocent suspects would be perceived as resembling the culprits more than the fillers, thereby allowing for the smile cue to have an impact.

Second, attempts were made to standardise the smiles expressed by the female and the male suspects. To do so, each suspect's smile was created by taking two steps. Step one involved a use of face morphing software called *FaceGen Modeller 3.5*. Step two involved a mobile phone application called *Face Switch* (available for Apple iPhone). As outlined in Figure 3.1, the *FaceGen Modeller 3.5* software was used to upload the suspect's photo so that the facial expression could be manipulated. For the current research, an open mouth smile was chosen. As shown in step 1 of Figure 3.1, the software only replicates the internal facial features of a face. Considering people often rely on external facial features to make facial recognition decisions (Bruce et al., 1999; Ellis, Shepherd, & Davies, 1979), eliminating a crucial feature such as hair can cause recognition of a face difficult. To overcome this, in step 2, the *Face Switch* application was used to merge the smiling image of the suspect and the original photo of the suspect.

Original photo **Step1** **Step 2**

Morphed via use of *FaceGen Modeller*

Two images morphed together via *Face Switch*

Figure 3.1 An example of the process taken to create a smiling innocent suspect.

Method

Participants, Design and Procedure

Experiment 3 used 120 (77 female, 43 male) paid community participants and first year psychology students who participated for course credit. Ages ranged from 17 to 38 years (M = 23.19, SD = 4.77). Experiment 3 employed the same design and procedure as Experiment 1, involving a 2 (cue: smile, neutral) × 2 (exposure duration: short, long) × 2 (mock-crime: male, female) design.

Materials

Stimulus videos. The four mock-crime videos (i.e., the short and the long versions of the female and the male mock-crimes) used in Experiment 1 were used (i.e., there was no divided attention manipulation in this experiment).

Photospreads. To create new sets of female and male photospreads, the pilot-test data obtained (i.e., match-to-description and similarity) for Experiments 1 and 2 were reviewed. Then, to test the appropriateness of the photospreads and the smile cue, additional pilot-tests were conducted.

Match-to-description and similarity tests. To create new target-absent photospreads, match-to-description and similarity data collected from 15 participants in Experiments 1 (female stimulus materials) and 2 (male stimulus materials) were reanalysed with the sample size being increased by 10 subjects (i.e., N = 25), who completed the same pilot-tests. For each photospread, a photo that was rated the highest on the similarity rating (i.e., resembled the culprit the most) was chosen as the new suspect[16]. The photos that were rated similarly to the suspects on the match-to-description scale were chosen as fillers.

Paired samples t-tests showed that, when presented with a neutral expression, the female photospread consisted of a suspect who was rated similarly to the fillers on the match-to-description scale (all *ps* > .05), yet was rated significantly higher on the similarity scale than the fillers (all *ps* < .05). The male photospread also consisted of a suspect who was rated similarly on the match-to-description scale to all (*ps* > .05) but one filler, who was rated slightly higher than the suspect (M = 5.80, SD = 1.26), $t(24) = -2.35$, $p = .03$, $f = .29$. Importantly though, with the exception of one filler who failed to differ from the suspect, (M = 4.44, SD = 1.83), $t(24) = 1.83$, $p = .08$, $f = .10$, the suspect was rated significantly higher on the similarity scale than all of the fillers, including the filler who was rated higher on the match-to-description scale (*ps* < .05). The descriptive statistics are provided in Table 3.3. Overall, considering that these fillers did not receive significantly higher similarity ratings than the suspects, they were considered acceptable fillers.

Smile manipulation check test. As described earlier, for each suspect, an image of an open mouth smile was constructed. The image replaced the neutral photo of the suspect to create a photospread with a smile cue. To ensure that the

[16] As explained in Chapter 2, both match-to-description and similarity scales were on 7 point-Likert scales, the higher the rating, more that the photo matched the description of the culprit or that the photo appeared similar to the culprit.

Table 3.3

Summary of Match-to-Description and Similarity Ratings given to the Female and Male Photospreads: Means (Standard Deviations) for the Suspect Rating, Average Filler Rating and Range of a Ratings given to each Photospread

	Female			Male		
	Suspect	Average	Range	Suspect	Average	Range
Match-to-description	4.80 (1.26)	4.67 (1.03)	4.12 – 5.12	5.24 (1.20)	5.14 (1.09)	4.48 – 5.80
Similarity	4.84 (1.65)	3.43 (1.28)	3.12 – 3.88	5.08 (1.58)	3.97 (1.12)	3.28 – 4.44

smiling expressions displayed by the suspects did appear smiling, a group of participants (N = 10) completed the smile manipulation check test. The results showed that the female suspect was rated higher (i.e., more smiling) on the smile scale when presented smiling (M = 5.80, SD = 1.62) than with a neutral expression (M = 1.90, SD = 1.10), $t(9) = 5.65$, $p < .001$, f = 1.43, and when compared to the average rating given to the fillers (M = 1.58, SD = 0.48), $t(9) = 7.90$, $p < .001$, f = 2.01. Similarly, the male suspect was rated higher on the smile scale when presented smiling (M = 6.10, SD = 1.45) than with a neutral expression (M = 1.30, SD = 0.67), $t(9) = 9.80$, $p < .001$, f = 2.26, and when compared to the average rating given to the fillers (M = 2.26, SD = 0.75), $t(9) = 6.97$, $p < .001$, f = 1.75. Therefore, both female and male suspects' smiles made the suspects stand out as smiling.

Effective size and defendant bias measure. Another group of participants (N = 105) completed the effective size/defendant bias measure pilot-test. The female neutral photospread's effective size was 5.19 (i.e., around five effective members). The chance expectation of the female suspect being chosen based on the effective size

was .21 (95% CI = .13 - .29), overlapping with the proportion of the actual choosing of the suspect of .13 (95% CI = .07 - .19). Similarly, the effective size of the male neutral photospread was 4.74 (i.e., approximately five effective members). The chance expectation of the male suspect being chosen based on the effective size was .19 (95% CI = .12 - .27), overlapping with the actual choosing of the suspect of .26 (95% CI = .18 - .34). Thus, the photospreads were considered not biased against or in favour of the suspects when presented without a smile cue.

Results

Of the 120 subjects' data obtained for the female stimuli, five subjects' data were omitted due to the misuse of the multiple confidence procedure (as per preceding experiments). For the same reason, of the 120 subjects' data obtained for the male stimuli, six subjects' data were omitted. Therefore, in total, the analyses of the female stimuli consisted of 115 subjects' data, whilst the male stimuli consisted of 114 subjects' data. As with the preceding experiments, difference (S_{uspect}–$F_{iller\ Max}$) scores were calculated for the familiarity and the multiple confidence ratings. The following analyses were conducted separately for the female and the male stimuli.

Perceived Familiarity Rating

As shown in Table 3.4, and contrary to the previous experiments' findings, the smile cue condition led to more negative familiarity difference scores than the neutral conditions for both stimuli. That is, relative to the fillers, both female and male suspects were perceived as appearing less familiar (rather than more familiar), when presented smiling than with a neutral expression. As shown by the Cohen's *f* displayed in Table 3.4, the effects were moderate in size for both female and male stimuli. Thus, whilst the difference in the effect sizes observed between the two stimuli in the preceding experiments disappeared, the effect of the smile on perceived

familiarity of the suspects versus the fillers was reversed. This finding is discussed in detail later.

Table 3.4

Descriptive and Inferential Statistics for the Effect of the Cue on the Familiarity Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Smile	-2.63	1.96	-3.19, -2.13	-2.54	2.09	-3.10, -1.98
Neutral	-1.26	2.24	-1.85, -0.67	-1.16	1.94	-1.77, -0.71
	$t(113) = -3.50$, $p = .001$, $f = .33$			$t(112) = -3.65$, $p < .001$, $f = .32$		

Multiple Confidence Rating

Of the 115 subjects who viewed the female photospread, 12 (10.4%) gave a unique maximum confidence value to the suspect, 70 (60.9%) gave it to one of the fillers and 33 (28.7%) gave no single unique maximum confidence value. Of the 114 subjects who viewed the male photospread, 14 (12.2%) gave a unique maximum confidence value to the suspect, 80 (70.4%) gave it to one of the fillers and 20 (17.4%) gave no single unique maximum confidence value.

To further investigate the effect of the smile cue, treating the confidence difference (S_{uspect}–$F_{iller\ Max}$) score as the dependent variable, a 2 (cue: smile, neutral) × 2 (exposure: short, long) factorial ANOVA analysis was conducted separately for the two stimuli[17]. In line with the results of the familiarity difference score, both the

[17] Again, the variables of the presentation orders of the mock-crime videos and the photospreads were excluded from the analyses for they provided no significant effects, with the exception of the female stimulus, which showed that presenting the photospread first ($M = -35.73$, $SD = 41.85$) led to a more

female and the male stimuli showed the opposite effect of the smile cue to that predicted. For the female stimuli, the smile cue condition produced a more negative confidence difference score than the neutral condition, $F(1,111) = 13.46$, $p < .001$, $f = .35$(see Table 3.5). The main effect of exposure condition, $F(1,111) = 0.06$, $p = .80$, $f = .02$, and the interaction between the cue and the exposure condition, $F(1,111) = 0.001$, $p = .97$, $f = .01$, were non-significant.

For the male stimuli, the smile cue condition also produced a more negative confidence difference score than the neutral condition, $F(1,110) = 7.33$, $p = .008$, $f = .26$ (see Table 3.5). Again, neither the main effect of exposure condition, $F(1,110) = .16$, $p = .69$, $f = .05$, nor the interaction between the cue and the exposure condition were observed, $F(1,110) = 0.16$, $p = .69$, $f = .04$. Therefore, relative to the fillers, the confidence ratings given to the suspects declined when the suspects were presented smiling rather than with a neutral expression.

Mediation Analyses

The results presented thus far indicate that, instead of the smile cue causing the suspects to appear familiar and to resemble the culprits more than the fillers, exactly the opposite occurred. Given that the smile cue produced a significantly more negative confidence difference score for both stimuli, I tested whether the relationship between the cue condition and the confidence difference score was mediated by the effect of the familiarity difference score. To test this, for each stimulus, an analysis was conducted using an INDIRECT Statistical Package for the Social Sciences (SPSS) macro, treating the cue condition as the independent variable, the familiarity difference score as the mediator and the confidence difference score as the dependent variable. The cue condition was coded as smile cue = 1 and neutral = 2 and, thus, a

---

negative confidence difference score than presenting it second ($M = -20.65$, $SD = 36.96$), $F(1,111) = 4.59$, $p = .04$.

Table 3.5

Summary of Means, Standard Deviations and 95% CIs for the Effects of the Cue and the Exposure Condition on the Confidence Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Short						
Smile	-42.07	37.07	-56.17, -27.97	-38.90	35.79	-52.26, -25.54
Neutral	-15.66	41.68	-31.51, 0.19	-23.21	34.84	-36.46, -9.96
Overall	-28.86	41.30	-31.72, -18.00	-31.19	35.90	-40.55, -21.83
Long						
Smile	-40.00	32.46	-52.59, -27.41	-38.85	30.60	-51.21, -26.49
Neutral	-14.14	40.69	-29.62, 1.34	-17.76	41.14	-32.70, -1.70
Overall	-26.84	38.81	-37.14, -16.54	-27.73	38.31	-37.46, -17.04
Overall						
Smile	-41.05	34.59	-50.23, -31.87	-38.88	33.18	-47.77, -29.99
Neutral	-14.90	40.83	-25.64, -4.16	-20.48	38.42	-30.10, -10.20
Overall	-27.86	39.92	-35.23, -20.49	-29.52	36.96	-36.09, -22.46

positive relationship between the independent variable and the mediator or the dependent variable indicated that the neutral condition scored higher than the smile cue condition.

The analysis calculated total, direct and indirect effects. For the direct effects (i.e., IV to DV), tests of significance were based on ordinary least squares regression analyses. The indirect effect (i.e., mediation) was based on a bootstrapping procedure with a resampling of 5000 bootstrap samples. The bootstrapping procedure does not

presume a normal distribution or require a large sample size and, thus, is considered preferable over procedures such as the Sobel test (Hayes, 2009; MacKinnon, Lockwood, & Williams, 2004; Preacher & Hayes, 2008). Bias-corrected 95% confidence intervals were also computed, with an indirect effect considered significant when zero was not contained within the interval.

Table 3.6

Summary of Mediation Analyses: Unstandardized Regression Coefficients (Standard Errors) for the Relationships between IV (Cue), Mediator (Familiarity Difference Score) and DV (Confidence Difference Score)

	IV - Mediator	Mediator - DV	Indirect effect	Bootstrap 95% CIs	Direct effect (IV - DV)
Female	1.44*** (0.39)	4.78** (1.60)	6.88 (3.61)	(1.63 - 16.82)	17.26* (6.98)
Male	1.26** (0.39)	4.42** (1.47)	5.55 (2.64)	(1.37 – 12.02)	9.31 (6.24)

*$p \leq .05$, ** $p \leq .01$, ***$p \leq .001$, IV coded as smile cue = 1, neutral condition = 2

As shown in Table 3.6, consistent results were observed across the two stimuli. That is, the effect of the smile cue on the confidence difference score was observed through the indirect effect of the familiarity difference score. That is, the smile cue led to more negative familiarity difference scores, which in turn led to more negative confidence difference scores, producing positive indirect effects. The female stimuli also showed a direct effect of the cue on the confidence difference score, whilst the male stimuli did not. Together, the results indicate that, when presented smiling, the suspects were perceived as appearing unfamiliar relative to the fillers,

which in turn contributed to the suspects being perceived as resembling the culprits less than the fillers.

Discussion

In contrast with Experiments 1 and 2, Experiment 3 showed that the smiling innocent suspects were perceived as relatively more unfamiliar than the fillers, reducing the likelihood of the suspects being perceived as resembling the culprits the most. These unexpected results are likely caused by the modification made to create the smile cue in Experiment 3. That is, perhaps, the use of morphing software led to the smiling innocent suspects appear odd and distorted, leading to a sense of unfamiliarity rather than familiarity. Indeed, after completing Experiment 3, several participants made comments regarding the smiling suspects such as "the face looked different" and "he/she looked odd". These comments suggest that perhaps the smile cue used in Experiment 3 made the suspects appear somewhat distinctive and memorable. Such perceptions might have led participants to conclude that such memorable faces without clear recollections could not be the culprits, also making them judge the suspects as appearing unfamiliar.

Indeed, the theory of memory for nonoccurrences (Strack & Bless, 1994; Strack & Förster, 1998) suggests that judged memorability of a stimulus aroused by factors such as the stimulus distinctiveness plays a crucial role in determining the newness of a stimulus. The theory suggests that in order for a stimulus to be judged new (i.e., never seen before), two criteria needs to be met: first, the stimulus should cue no recollective experience and, second, the stimulus should be judged to be memorable, arousing an expectation that if the stimulus was seen previously, it should have been remembered. For example, when presented with a distinctive face, a person would likely form an expectation that if it had been seen previously, the face would

have been memorable enough to cue a recollection. Based on this expectation, a lack of recollection experienced for the face would likely lead to an assumption that the face had never been seen before. Conversely, when presented with a non-distinctive face, the face would likely be judged as appearing rather unmemorable. Therefore, even if the face cues no recollection, this may be interpreted as a sign of forgetting, where the face had been seen before but was not encoded well enough to be remembered. In line with this, by making the suspects appear more distinctive, the smile cue used in Experiment 3 might have made participants think that, if the suspects were the culprits, they should have cued clear recollections. Therefore, when the smiling suspects failed to cue recollections, participants might have interpreted the experiences as indications that the suspects were not the culprits.

Given the informal comments made by participants exiting the experiment, to examine the possibility that the smile cue also made the suspects appear distinctive, the last 40 participants in Experiment 3 were also asked to rate how distinctive each member of the male[18] photospread appeared, using a 7 point-Likert scale (1 = *very typical*; 7 = *very distinctive*). As with other dependent measures, a difference (S_{uspect}–$F_{iller\ Max}$) score was calculated for the distinctiveness ratings. The results showed that the smile cue condition led to a less negative distinctiveness difference score than the neutral condition (M = -0.95, SD = 2.89 vs. M = -2.77, SD = 2.33), $t(42) = 2.30$, $p = .03$, $f = .35$. That is, relative to the fillers, participants perceived the male suspect as appearing more distinctive when he was presented smiling than with a neutral expression. Therefore, it may be that the effect of the smile cue found for the confidence difference score was, in part, due to the cue making the suspects appear

[18] An attempt was made to collect the same data from the female photospread. However, due to technical errors made to the software, the data were not collected.

rather distinctive and, thus, memorable, perhaps fuelling the cognition that if the suspects were the culprits, then they should have cued clear recollections.

Additionally, beyond the effect of the facial expression of a smile, the results might point towards the fact that the use of the morphing software itself had impacted the results. That is, irrespective of the type of facial expression used, by making the suspects appear rather unrealistic in appearance, the morphing technique could have caused the suspects to be perceived as an unlikely match to be the culprits. Therefore, the results obtained in Experiment 3 may not be specific to an expression of a smile. The results of Experiment 3 therefore are open to criticism. Nevertheless, some theoretical speculation is warranted. That is, the results tap into how an eyewitness may come to reject a particular lineup member and, more importantly, the results suggest that a biasing cue that could make a lineup member appear unfamiliar and/or distinctive could potentially make the lineup biased in a way that favours the lineup member. Based on this theoretical rationale, Experiments 4 (Chapter 4), 5 and 6 (Chapter 5) further explored how potentially biasing cues that could arouse a sense of unfamiliarity as well as a face appear distinctive could make a lineup become biased.

In sum, the results of Experiment 3 suggest that, judged distinctiveness and, thus, memorability of a lineup member may play an important role in reducing witnesses' judgement of the degree of resemblance between the lineup member and the culprit. The potential roles of perceived distinctiveness and memorability was further explored in Chapters 4 and 5 (Experiments 4, 5 and 6).

CHAPTER 4

Experiment 4

Experiment 3 unexpectedly showed that an expression of a smile made participants perceive the innocent suspects as appearing unfamiliar and, thus, resembling the culprits less than the fillers. Moreover, there was a suggestion from data gathered from a subset of participants that this unfamiliarity may have been related to the suspects appearing more distinctive and memorable than the fillers, with these perceptions perhaps fuelling the cognition that the suspects could not have been seen before because they would have been remembered if they had been. Experiment 4 explored this possibility using the same stimulus materials as in Experiment 3, but with the addition of a target-present photospread condition. That is, either the actual culprit or an innocent suspect was presented in a photospread with an odd smile or with a neutral expression. The fillers were presented with neutral expressions. It was predicted that, regardless of who was displaying a smile (i.e., culprit or innocent suspect), compared to the fillers the smiling face would be perceived as an unlikely match to the culprit. Therefore, even when the smiling face belonged to the culprit, it was predicted that the smile would cause one of the fillers to be perceived as resembling the culprit better than the actual culprit.

As suggested in Experiment 3, according to the theory of memory for nonoccurrences (e.g., Strack & Bless, 1994), there is one clear distinction between one's interpretation of a lack of recollection experienced for a stimulus as an indication that the stimulus is new (i.e., never been seen before) as opposed to its encoding being forgotten. According to the theory, when a lack of recollection is experienced for a seemingly non-distinctive and, thus, unmemorable stimulus, the experience is likely to be interpreted as a case of the stimulus being encoded but

forgotten. Conversely, it is suggested that when a lack of recollection is experienced for a seemingly distinctive and, thus, memorable stimulus, the stimulus is likely to be interpreted as being new. Accordingly, the theory would suggest that when a face is made to appear unexpectedly distinctive and, thus, memorable, for example, due to its usual facial expression, a lack of recollection experienced for that face would likely be interpreted as the face being new rather than its encoding experience being forgotten.

If, as the theory of memory for nonoccurrences (Strack & Bless, 1994) suggests, expected memorability of a face plays a crucial role in determining the newness of the face, it is quite possible that an old face will be falsely judged as being new when the face, which happens to appear highly memorable, fails to cue a clear recollection. For example, imagine that a witness encodes a culprit who appears non-distinctive and/or unmemorable committing a crime. Then, at a later date, the witness is presented with a lineup which contains the culprit who has an odd facial expression that leads to the culprit being perceived as rather distinctive and probably much more memorable than the image of the culprit encoded by the witness. A likely reaction from the witness would be "If this person in the lineup is the culprit, surely, I would have remembered him clearly". In turn, the witness may reject the possibility that the culprit presented in the lineup is in fact the culprit, possibly leading to the rejection of the lineup. Similarly, if the witness is presented with an innocent suspect who happens to appear highly distinctive due to a display of a rather odd facial expression, a lack of recollection experienced for such a distinctive face would likely lead the witness to conclude that the face does not belong to the culprit.

Manipulation Check Measures: Distinctiveness and Memorability

To ensure that the expression of a smile made the culprits and the innocent suspects appear distinctive and memorable, at the end of Experiment 4, participants were asked to rate the distinctiveness and memorability of all of the photos used to create the photospreads: the fillers, the smiling and neutral photos of all of the innocent suspects and the culprits. I was particularly interested in seeing if the expression of a smile made the innocent suspects and the culprits stand out from the photospreads as appearing most distinctive and memorable. To examine if the cue did make them stand out, difference (S_{uspect}–$F_{iller\ Max}$) scores were calculated for each of the ratings. That is, from the each of the ratings given to the smiling and the neutral photos of the innocent suspects and the culprits, the maximum value given to a filler was subtracted. Therefore, a positive score would have indicated that the innocent suspects or the culprits stood out from the photospread as appearing most distinctive or memorable, whilst a negative score would have indicated otherwise. For each innocent suspect and culprit, the difference scores given to the smiling and neutral photos were compared to see if the smile cue made them appear more distinctive and memorable.

As discussed in Chapter 4, I was also interested in whether the perceived unfamiliarity aroused by the smile cue observed in Experiment 3 was, in part, associated with the cue leading to the perception that the suspects were too distinctive and memorable to not have been remembered. Thus, as in Experiment 3, in Experiment 4, the smile cue was predicted to make the innocent suspects and the culprits appear unfamiliar. It was further predicted that, the perception of unfamiliarity would be related to the perceptions of distinctiveness and memorability:

the more distinctive and memorable the innocent suspects and the culprits were perceived to be, the more unfamiliar they would be perceived to be.

Method

Participants and Design

Experiment 4 consisted of 120 (70 female, 50 male) paid community participants and first year psychology students who participated for course credit. Ages ranged from 17 to 63 years (M = 23.53, SD = 9.01). The general design and procedure of Experiment 4 were similar to Experiment 3, with the exceptions of the omission of the exposure duration manipulation[19] and the addition of a target-present photospread condition. At the end of Experiment 4, manipulation check ratings of distinctiveness and memorability were also added. Thus, Experiment 4 consisted of a 2 (cue: smile, neutral) × 2 (target presence: target-present, target-absent) × 2 (mock-crime: male, female) design. Each participant saw one male and one female mock-crime and later, saw one male and one female photospread, which consisted of combinations of one target-present (TP), one target-absent (TA), one smile cue and one neutral photospread. Thus, each participant saw one of the following combinations of the photospreads: (1) male TP smile cue photospread and female TA neutral photospread, (2) male TP neutral photospread and female TA smile cue photospread (3) male TA smile cue photospread and female TP neutral photospread, (4) male TA neutral photospread and female TP smile cue photospread.

Materials

Photospreads. The TA photospreads from Experiment 3 were used. TP photospreads were created by replacing the photos of the innocent suspects with the photos of the culprits. For the TP neutral photospreads, the photos of the innocent

[19] The exposure duration manipulation was omitted as Experiments 1-3 showed no significant effects of this manipulation.

suspects were replaced with the neutral photos of the culprits. For the TP smile cue photospreads, the photos of the innocent suspects were replaced with the smiling photos of the culprits. The photos of the smiling culprits were created using the morphing software and the *FaceSwitch* application, as described in Experiment 3 (Chapter 4).

Smile manipulation check test. To ensure the appropriateness of the smiling photos of the culprits, a group of participants (N = 10) were shown the smiling and neutral photos of the culprits and asked to rate each photo on the smile scale (refer to Chapter 2 for details). Paired samples t-tests showed that the female culprit was rated higher on the smile scale (i.e., smiling more) when she was presented smiling than with a neutral expression (scale of 1-7; M = 5.10, SD = 1.73 vs. M = 2.50, SD = 1.58), $t(9) = -3.62$, $p = .006$. Similarly, the male culprit was perceived as smiling more when he was presented smiling than with a neutral expression (M = 6.00, SD = 0.94, vs. M = 2.50, SD = 1.51), $t(9) = -6.22$, $p < .001$. Furthermore, independent samples t-tests showed that compared to the average filler ratings (obtained for Experiment 3, refer to Chapter 3 for descriptive statistics), the female and the male culprits were rated higher on the smile scale when presented smiling, $t(18) = 4.76$, $p < .001$, f = 1.91, and, $t(13.30) = 13.24$, $p < .001$, f = 2.12, respectively. Therefore, the smiling culprits were perceived as smiling more than the fillers as well as the neutral versions of the culprits.

Effective size and defendant bias measure. To ensure that the TP neutral photospreads were not biased against or in favour of the culprits, another group of participants (N = 194 for the female; N = 195 for the male) completed the effective size and defendant bias measure test in a classroom setting. The effective size of the female photospread was calculated to be 4.26 (i.e., approximately four effective

members). Based on the effective size, the chance expectation of the female culprit being chosen was .23 (95% CI = .17 - .29), whilst the proportion of the actual choosing of the culprit was .30 (95% CI = .24 - .36). Therefore, the 95% CIs of the two proportions overlapped with one another, suggesting that the photospread was not biased against or in favour of the female culprit. The male photospread's effective size was calculated to be 4.23 (i.e., approximately four effective members). Based on the effective size, the chance expectation of the male culprit being chosen was .24 (95% CI = .18 - .30), whilst the proportion of the actual choosing of the culprit was .34 (95% CI = .27 - .41). Again, the overlapping of the 95% CIs of the two proportions suggested that the photospread was not biased against or in favour of the male culprit.

Procedure

Experiment 4 mirrored the general procedure of Experiment 3 with the exception of the added ratings of distinctiveness and memorability at the end. That is at the end of Experiment 4, participants were presented twice with the photos of the fillers as well as the smiling and neutral photos of the innocent suspects and the culprits. In the first presentation, they rated the perceived distinctiveness; in the second, they rated the perceived memorability of each photo. The photos were presented one at a time in the centre of a computer screen. The order of the presentation of the photos was randomised. In the first round, participants were asked to rate the perceived distinctiveness of each photo, using a 7 point Likert-scale (1 = *very typical*; 4 = *neither*; 7 = *very distinctive*) and the following definition: "A TYPICAL looking face is one that would easily blend into a crowd and that you would not remember very well. If you think that a face is typical, rate it low on the scale (i.e., give 1 or 2). A DISTINCTIVE face is one that would stand out from a

crowd and make you take notice of it. A face might be distinctive because it is particularly attractive or particularly ugly. Or there may just be something about it that would make you remember it, for example, it may have a particularly prominent feature or a set of features that do not match well. If you think a face is distinctive, rate it high on the scale (i.e., give it a 6 or 7). Conversely, if you think a face is typical, rate it low on the scale (i.e., give a 1 or 2)"[20]. Underneath each photo, buttons labelled 1 to 7 were presented on the screen. Participants were asked to click on the appropriate button using their mouse. Once all of the photos were rated, the procedure was repeated to rate the perceived memorability of each photo using a 7 point-Likert scale (1 = *not at all memorable*; 7 = *extremely memorable*) displayed as buttons on the screen. The following definition of the term memorability was provided: "A MEMORABLE face is defined as a face that would be relatively easy to remember". The experiment ended once the participants rated all of the photos.

Results

Of the 120 subjects' data obtained for the female stimuli, one subject's data were omitted because the subject knew the actor who played the female culprit. Of the 120 subjects' data obtained for the male stimuli, one subject's data were omitted due to a misuse of the multiple confidence procedure (as described in Chapter 2). Therefore, for both female and male stimuli, analyses were conducted using 119 subjects' data. As with the preceding experiments, analyses were conducted separately for the female and the male stimuli. Difference (S_{uspect} - F_{iller} M_{ax}) scores were calculated from the multiple confidence, familiarity, distinctiveness and memorability ratings.

[20] The definition was obtained from a study conducted by Semmler and Brewer (2006).

Manipulation Check Ratings: Distinctiveness & Memorability

One participant indicated that she failed to read the instruction when rating the distinctiveness of the female photos. Therefore, one additional subject's data were omitted, leading to 118 subjects' data for the female stimuli being use for the manipulation check analysis.

As discussed earlier, distinctiveness and memorability difference scores were calculated for the neutral and smiling photos of each innocent suspect and culprit. For example, imagine that the following ratings were given: one of the fillers obtained the highest distinctiveness rating of 3 (out of 7), the culprit's neutral photo obtained a rating of 2 and the smiling photo a rating of 7, the innocent suspect's neutral photo obtained a rating of 3 and the smiling photo a rating of 5. These ratings would have led to the culprit's neutral photo obtaining a score of -1 (i.e., 2 – 3) and the smiling photo a score of 4 (i.e., 7 – 3), the innocent suspect's neutral photo obtaining a score of 0 (i.e., 3 – 3) and the neutral photo a score of 2 (i.e., 5 – 3). The distinctiveness and memorability difference scores could range from -6 to 6.

For each photo, the scores of distinctiveness and memorability were positively correlated (r = .24 - .62) and, thus, the two scores were collapsed together as a single variable called memorability difference score (i.e., the total of the two scores was divided by two)[21]. The new variable suggested that, the more positive the score was, the more that the innocent suspects and the culprits stood out as appearing distinctive and memorable. For each innocent suspect and culprit, a repeated measures ANOVA analysis was conducted, comparing the scores calculated for the smiling and the neutral photos.

[21] Similar results were obtained when the distinctiveness and memorability difference scores were treated separately.

Table 4.1

Means, Standard Deviations and 95% CIs for the Memorability Difference (S_{uspect}– $F_{iller\ Max}$) Scores Given to the Innocent Suspects' and the Culprits' Smiling and Neutral Photos

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Culprit						
Smile	-0.12	1.86	-0.46, 0.22	-0.70	1.61	-0.99, -0.41
Neutral	-1.57	1.46	-1.84, -1.30	-1.98	1.28	-2.12, -1.75
	$t(117) = -7.99$, $p < .001$, $f = .44$			$t(118) = -8.34$, $p < .001$, $f = .43$		
Innocent Suspect						
Smile	0.01	1.76	-0.31, 0.33	-1.18	1.56	-1.46, -0.90
Neutral	-1.59	1.32	-1.83, -1.35	-2.54	1.37	-2.79, -2.29
	$t(117) = -10.04$, $p < .001$, $f = .52$			$t(118) = -8.91$, $p < .001$, $f = .46$		

As shown in Table 4.1, across the stimuli, the smiling photos were given less negative memorability difference scores than the neutral photos. That is, relative to the fillers, the innocent suspects and the culprits were perceived as appearing more distinctive and memorable when presented smiling than with a neutral expression. However, with the exception of the female innocent suspect, as indicated by the negative scores observed in the smile cue condition, the cue did not make them stand out as appearing most memorable[22].

[22] Prior to rating the distinctiveness and memorability of the photospread members, participants were always asked to complete the multiple confidence procedure. Therefore, it could be argued that, the mental processing associated with completing the multiple confidence procedure influenced the participants' perceptions regarding how distinctive and memorable the photospread members appeared, rather than such perceptions affected the participants' multiple confidence assessment. To ensure that this was not the case, a separate group of participants ($N = 67$) watched the two mock-crime videos and after 20 minutes delay, rated the perceived distinctiveness and memorability of the photospread members instead of rating the multiple confidence. In line with Experiment 4, the results showed that

Perceived Familiarity Rating

To test if the smile cue also made the innocent suspects and the culprits appear less familiar, a 2 (cue: smile, neutral) × 2 (target presence: TP, TA) factorial ANOVA analysis was conducted, treating the familiarity difference score as the dependent variable. As the descriptive statistics shown in Table 4.2 indicate, the female stimuli showed a main effect of cue, with the smile cue condition producing a more negative familiarity difference score than the neutral condition, $F(1, 115) = 25.35$, $p < .001$, $f = .46$. The TA photospread produced a more negative familiarity difference score than the TP photospread, $F(1, 115) = 4.66$, $p = .03$, $f = .18$. No interaction between the cue and the target presence condition was observed, $F(1, 115) = 0.15$, $p = .70$, $f = .04$.

The male stimuli also showed a main effect of cue, with the smile cue condition producing a more negative familiarity difference score than the neutral condition, $F(1, 115) = 6.07$, $p = .02$, $f = .23$. No main effect of target presence condition was detected, $F(1, 115) = 0.05$, $p = .82$, $f = .02$, nor was there an interaction between the cue and the target presence condition, $F(1, 115) = 0.004$, $p = .95$, $f = .01$. Thus, in line with Experiment 3, relative to the fillers, the suspects and the culprits were given lower familiarity ratings when presented smiling than with a neutral expression[23].

---

across the stimuli, the smile cue condition led to less negative distinctiveness and memorability difference scores than the neutral condition (all $t > 3.20$, $p \leq .002$, $f > .25$). Therefore, relative to the fillers, the innocent suspects and the culprits were perceived as appearing more distinctive and memorable when presented smiling than with a neutral expression.

[23] As with the scores of distinctiveness and memorability, considering that the rating familiarity was always preceded by the multiple confidence procedure, it could be argued that, completing the multiple confidence procedure influenced the familiarity ratings, rather than the perceived familiarity of the photospread members affected the participants' multiple confidence assessment. To ensure this was not the case, a separate group of participants ($N = 128$) completed Experiment 4 without the multiple confidence procedure. The photospread members were presented sequentially and, thus, the effect of the smile cue on the familiarity rating was examined within-subjects. In line with Experiment 4, paired samples t-tests showed that the smile cue condition led to more negative familiarity difference scores than the neutral condition for all (all $t > 3.56$, $p \leq .001$, $f \geq .15$), except the male culprit, which showed no difference between the smile and neutral conditions ($M = -1.27$, $SD = 2.35$ vs. $M = -1.37$, $SD = 2.56$), $t(127) = 0.49$, $p = .63$, $f = .02$. Thus, the elimination of the multiple confidence procedure as

Table 4.2

Means, Standard Deviations and 95% CIs for the Effects of the Cue and the Target Presence on the Familiarity Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Target-Present						
Smile	-2.33	2.99	-3.45, -1.21	-2.59	2.01	-3.36, -1.83
Neutral	-0.45	1.72	-1.10, 0.20	-1.57	2.06	-2.34, -0.80
Overall	-1.41	2.61	-2.09, 0.51	-2.07	2.08	-2.81, -1.73
Target-Absent						
Smile	-3.37	1.96	-4.10, -2.64	-2.47	2.01	-3.22, -1.72
Neutral	-1.17	1.95	-1,91, -0.43	-1.50	2.64	-2.49, -0.51
Overall	-2.27	2.23	-2.85, -1.69	-1.98	2.38	-2.60, -1.37
Overall						
Smile	-2.85	2.56	-3.51, -2.19	-2.53	1.99	-3.11, -2.07
Neutral	-0.81	1.86	-1.30, -0.33	-1.53	2.35	-2.14, -0.92
Overall	-1.84	2.45	-2.29, -1.40	-2.03	2.23	-2.44, -1.63

To see if the familiarity and memorability difference scores were related to one another, for each innocent suspect and culprit, a correlational analysis was conducted using the two scores. Prior to conducting the analysis, the memorability difference scores were reorganised such that, for each participant, the scores that reflected his or her allocated cue and target presence conditions were selected, one for the male stimulus and one for the female stimulus. For example, if a participant was allocated to the combined conditions of the smile cue/TP conditions for the male

well as the use of the sequential presentation led to the smile cue failing to make the male culprit appear relatively less familiar than the fillers.

stimuli and the neutral/TA conditions for the female stimuli, the memorability difference scores calculated for the male culprit's smiling photo and for the female innocent suspect's neutral photo were selected. The correlational analyses were conducted separately for each innocent suspect and culprit.

The analyses showed mixed results: for the female innocent suspect, the familiarity and memorability difference scores were negatively correlated (r = -.40, N = 59, p = .002), indicating that, as predicted, the more the female innocent suspect appeared memorable, the less familiar she appeared. No relationship was observed between the two difference scores for the female culprit (r = .16, N = 59, p = .23), the male culprit (r = -.19, N = 59, p = .16) or the male innocent suspect (r = -.10, N = 60, p = .45).

Multiple Confidence Rating

For the female stimuli, of the 59 subjects who viewed the female TP photospread, 28 (47.5%) gave a unique maximum confidence value to the culprit, 27 (45.8%) gave it to the fillers and 4 (6.8%) gave no single unique maximum confidence value. Of the 60 subjects who viewed the female TA photospread, 7 (11.7%) gave a unique maximum confidence value to the suspect, 40 (66.7%) gave it to the fillers and 13 (21.7%) gave no single unique maximum confidence value. For the male stimuli, of the 59 subjects who viewed the male TP photospread, 17 (28.8%) gave a unique maximum confidence value to the culprit, 34 (57.6%) gave it to the fillers and 8 (13.6%) gave no single unique maximum confidence value. Of the 60 subjects who viewed the male TA photospread, 5 (8.3%) gave a unique maximum confidence value to the suspect, 51 (85%) gave it to the fillers and 4 (6.7%) gave no single unique maximum confidence value.

To test the effect of the cue, treating the confidence difference ($S_{uspect} - F_{iller\ Max}$) score as the dependent variable, a 2 (cue: smile, neutral) × 2 (target presence: TP, TA) factorial ANOVA analysis was conducted separately for the female and the male stimuli[24]. The female stimuli showed a main effect of cue, with the smile cue condition producing a more negative confidence difference score than the neutral condition, $F(1,115) = 24.35$, $p < .001$, $f = .43$. The TA photospread produced a more negative confidence difference score than the TP photospread, $F(1,115) = 14.36$, $p < .001$, $f = .32$. No interaction between the cue and the target presence condition was observed, $F(1,115) = 0.06$, $p = .81$, $f = .02$. The descriptive statistics are provided in Table 4.3.

The male stimuli also showed a main effect of cue, with the smile cue condition producing a more negative confidence difference score than the neutral condition, $F(1,115) = 28.97$, $p < .001$, $f = .46$. The TA photospread produced a more negative confidence difference score than the TP photospread, $F(1,115) = 14.04$, $p < .001$, $f = .31$ (see Table 4.3). A significant two-way interaction between the cue and the target presence condition was also observed, $F(1,115) = 13.20$, $p < .001$, $f = .34$. Simple effects analyses showed that the interaction was due to the effect of the cue being significant for the male TP photospread, $t(57) = -5.79$, $p < .001$, $f = .76$, but not for the male TA photospread, $t(58) = -1.39$, $p = .17$, $f = .17$. The descriptive statistics are provided in Table 4.3. Thus, with the exception of the male innocent suspect, relative to the fillers, the confidence ratings given to the innocent suspect and the culprits declined when they were presented smiling rather than with a neutral expression.

[24] No presentation order effects of the mock-crime videos or the photospreads were observed and, thus, the variables are omitted from the analyses.

Table 4.3

Means, Standard Deviations and 95% CIs for the Effects of the Cue and the Target Presence Condition on the Confidence Difference ($S_{uspect} - F_{iller\ Max}$) Score for the Female and Male Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Target-Present						
Smile	-21.83	57.65	-43.36, -0.30	-50.34	35.73	-63.93, -36.75
Neutral	18.45	53.00	-1.34, 38.24	12.17	46.36	-5.14, 29.48
Overall	-2.03	58.57	-17.16, 13.10	-18.56	51.81	-32.06, -5.06
Target-Absent						
Smile	-56.50	31.19	-68.15, -44.85	-51.13	30.12	-62.38, -39.88
Neutral	-12.00	41.20	-27.67, 3.67	-39.00	37.17	-52.88, -25.12
Overall	-34.25	42.61	-45.35, -23.15	-45.07	34.10	-53.88, -36.26
Overall						
Smile	-39.17	49.16	-51.87, -26.47	-50.75	32.71	-59.27, -42.23
Neutral	2.97	49.40	-9.90, 15.84	-13.42	49.00	-26.08, -0.76
Overall	-18.28	53.44	-27.98, -8.58	-31.92	45.58	-40.19, -23.65

Mediation Analyses

It was predicted that the smile cue would make the innocent suspects and the culprits appear memorable, leading to the smile cue condition producing a more negative confidence difference score than the neutral condition. Thus, the effect of the smile cue on the confidence difference score was predicted to be observed through the indirect effect of the memorability difference score with, perhaps, the associated effect of the familiarity difference score. In line with this, the results presented thus

far indicate that, with the exception of the male innocent suspect, the smile cue produced a more negative confidence difference score than the neutral condition. Furthermore, across the stimuli, the smile cue produced more positive memorability difference scores and more negative familiarity difference scores than the neutral condition. However, only the results of the female innocent suspect showed a moderately strong relationship between the memorability and familiarity difference scores. Considering these results, mediation analyses were conducted separately for each innocent suspect and culprit[25].

For the female innocent suspect, the potential mediator consisted of the memorability and familiarity difference scores being collapsed together as a single variable called total memorability. That is, the familiarity difference score was reverse coded (i.e., higher rating suggested less familiar) and it was combined with the memorability difference score (i.e., adding the two scores and dividing the total by two). Therefore, a high total memorability difference score suggested that the female innocent suspect appeared low on familiarity, whilst highly distinctive and memorable. For the male innocent suspect as well as the female and the male culprits, the original memorability difference score (i.e., without the combined effect of the familiarity difference score) was treated as a potential mediator. For each mediator, an analysis was conducted using an INDIRECT Statistical Package for the Social Sciences (SPSS) macro where the cue condition was treated as the independent variable and the confidence difference score was treated as the dependent variable. The cue condition was coded as the smile cue condition = 1 and the neutral condition = 2. Thus, a negative relationship between the independent variable and either the

[25] Whilst the male innocent suspect showed no relationship between the cue and the confidence difference score, Hayes (2009) suggests that the absence of a relationship between IV and DV does not necessarily mean that the IV and DV cannot be related to one another through the effect of a mediator. Therefore, the male innocent suspect was included in the mediation analysis.

mediator or the dependent variable suggested that the neutral condition produced a higher difference score than the smile cue condition.

As shown in Table 4.4, the results varied between the stimuli. The female innocent suspect showed that, as predicted, the effect of the cue was found on the confidence difference score through the indirect effect of the total memorability difference score. In particular, the smile cue condition produced a more positive total memorability difference score, which in turn led to a more negative confidence difference score, producing a positive indirect effect of the total memorability difference score. The direct effect of the cue on the confidence difference score was also observed. The female and male culprits, as well as the male innocent suspect, showed no indirect effect of the memorability difference score on the relationship between the cue and the confidence difference score[26].

Additional analyses showed that, in line with Experiment 3, the smile cue for the female culprit affected the confidence difference score through the indirect effect of the familiarity difference score (B = 17.02, SE_B = 7.81, 95% CI, 5.26; 37.48). The indirect effect of the familiarity difference score was not observed for the male culprit (B = 3.59, SE_B = 3.32, 95% CI, -0.50; 13.40) or the male innocent suspect (B = 5.63, SE_B = 4.83, 95% CI, -0.20; 19.56).

Discussion

Experiment 4 showed several consistent results across the stimuli. As predicted, relative to the fillers, the innocent suspects and the culprits were perceived to be more distinctive and memorable when they were presented in the photospreads smiling than with a neutral expression. Furthermore, in line with Experiment 3, the smile cue made them appear unfamiliar. Importantly, with the exception of the male

[26] Treating the distinctiveness and memorability difference scores separately instead of combining them as a single memorability scale produced no indirect effects.

Table 4.4

Summary of Mediation Analyses: Unstandardized Regression Coefficients (Standard Errors) for the Relationships between IV (Cue), Mediator (Total Memorability or Memorability) and DV (Confidence Difference Score)

	IV - Mediator	Mediator - DV	Indirect effect	Bootstrap 95% CIs	Direct effect
		Total Memorability			
Female Innocent Suspect	-1.76*** (0.35)	-6.98 (3.57)	12.31 (5.45)	(2.54 – 24.42)	31.77** (22.29)
		Memorability			
Female Culprit	-0.56 (0.47)	5.58 (4.00)	-3.14 (3.85)	(-15.75 – 1.25)	43.42** (14.49)
Male Innocent Suspect	-1.18** (0.36)	1.28 (3.18)	-1.51 (4.80)	(-13.43 – 6.42)	13.65 (9.57)
Male Culprit	-1.60** (0.38)	-1.04 (3.80)	1.67 (7.85)	(-10.60 – 20.09)	60.84*** (12.47)

*$p \leq .05$, ** $p \leq .01$, ***$p \leq .001$, IV coded as smile cue = 1, neutral = 2

innocent suspect, the smile cue led to the smiling individuals being perceived as less likely to be the culprits, even when the individual was the culprit. Thus, as predicted, the results provide evidence that, when a culprit is presented in a lineup with an odd smile, the expression can be used as a piece of evidence in favour of incorrectly rejecting the culprit. However, Experiment 4 also showed some inconsistent results across the stimuli. For the female innocent suspect, the judged memorability and familiarity of the suspect were related to one another: the more memorable the suspect appeared, the more unfamiliar she appeared. Together, such perceptions influenced participants' confidence assessments, whereby the smile cue made the innocent suspect appear more memorable and unfamiliar, leading to the reduction in

participants' judgement of the degree of the resemblance between the innocent suspect and the culprit. On the other hand, for the female culprit, the smile cue made the culprit appear relatively unfamiliar when compared to the fillers, leading to participants perceiving the culprit as poorly resembling their memory for the culprit. The male culprit and the innocent suspect showed no evidence of the mediating effect of the judged memorability or familiarity for the effect of the smile cue on the confidence difference score.

Together, the results provide some evidence that, in some circumstances, a face that appears distinctive and, thus, memorable might also arouse a feeling that the face is unfamiliar, encouraging the perception that the face has never been seen before. However, the results also suggest that, a judgement of face memorability would not always be associated with a perception of face unfamiliarity and moreover, such judgements would not always influence witnesses' perception regarding the resemblance between a face presented in a lineup and the witnesses' memory for the culprit.

Face Distinctiveness, Memorability and Familiarity

The focus of this thesis is not to explore the nature of the relationships between face distinctiveness, memorability and familiarity. However, considering the mixed results observed in Experiment 4, the topic merits some comment. In particular, two possible explanations are considered regarding the inconsistent results observed across the stimuli. First, the results may perhaps reflect the fact that face distinctiveness, memorability and familiarity are extremely complex variables that the single item measures used in Experiment 4 are unable to capture adequately. For example, the single scale of face distinctiveness used in Experiment 4 was defined in several ways. In particular, the definition suggested that a face could be distinctive

due to it being particularly attractive or ugly, or due to it having a prominent feature or having several features that did not match well. Guided by such a definition, perhaps the particular facial characteristics of the female innocent suspect led the participants to focus on a particular aspect of the distinctiveness definition, which happened to correlate highly with their familiarity judgement. Conversely, perhaps something about the facial characteristics of the other members (i.e., the culprits and the male innocent suspect) led the participants to focus on the different aspects of the distinctiveness definition, which happened to correlate poorly with their familiarity judgement. Therefore, depending on how individuals conceptualise terms such as face distinctiveness, memorability and familiarity, as well as on the aspects of such terms on which they choose to focus, the relationships between such perceptions may or may not be evident when single item measures are used.

In line with this view, several studies have demonstrated that, depending on how the term face distinctiveness is defined, different conclusions can be drawn about the relationships between judgements of face distinctiveness, memorability, and familiarity. In these studies, participants are generally asked to look at a series of faces and rate them on various characteristics such as distinctiveness, memorability, familiarity and positivity (i.e., attractiveness, pleasantness, etc.). The data are then analysed using principal component analyses. Defining a typical (as opposed to distinctive) face as a face that is average in appearance, Vokey and Read (1992) found that the rating of face typicality loaded negatively with the memorability component, which consisted of the memorability rating, whilst positively with the familiarity component, which consisted of the familiarity and positivity ratings. O'Toole, Deffenbacher, Valentin, and Abdi (1994) attempted to replicate the results using a slightly different definition of face distinctiveness: a distinctive face as a face that

would stand out from a crowd of other faces. Participants were asked to rate either Caucasian (same-race) or Asian (different-race) faces. In contrast to Vokey and Read (1992), their distinctiveness definition led to the distinctiveness rating loading weakly (for the Caucasian faces) or failing to load (for the Asian faces) with the component of familiarity. On the other hand, regardless of the race of the faces, the distinctiveness rating loaded strongly with the memorability component. Morris and Wickham (2001) further tested the relationships between face distinctiveness, memorability and familiarity, using four measures of face distinctiveness: (1) whether a face would stand out in a crowd of more typical faces, (2) degree of deviation from an average/typical face, (3) comparison of faces as to which face was more distinctive, (4) comparison of faces as to how similar they appeared to one another. They found that, whilst the ratings obtained using all four measures of face distinctiveness loaded strongly with the memorability component, only the distinctiveness rating using the measure 2 (deviation from an average/typical face) loaded moderately with the familiarity component. Together, the studies demonstrate that, perhaps, there are several aspects to the definition of face distinctiveness, with some aspects relating to the judgement of face familiarity, whilst others are not.

Second, on a related note, assuming that there are several aspects to face distinctiveness, memorability and familiarity, to assume that the facial character judgements of distinctiveness and memorability have linear relationships with a judgment of familiarity may be misguided. That is, in some circumstances, a face that is distinctive and, thus, memorable might be associated with low familiarity whilst, under other circumstances, a face that is distinctive and memorable might be associated with high familiarity. For example, if a person comes across a face, which has a distinctive witch-like nose, the face would most likely be judged as appearing

memorable. If, the person does not know of anyone who has a witch-like nose, he or she might perceive the face as appearing unfamiliar. However, if the person happens to have a friend who has a similarly distinctive witch-like nose, the face might remind the person of the friend, leading to the face arousing a feeling of familiarity. Thus, what one person in one situation perceives as distinctive, memorable and familiar would likely differ from what another person considers as distinctive, memorable and familiar under alternative circumstances. Together, the inconsistent results observed in Experiment 4 point towards the need for further investigations as to how people generally conceptualise terms such as face distinctiveness, memorability and familiarity. Although this is an important issue, it is beyond the scope of this thesis. However, further discussions as well as recommendations regarding future directions are provided in Chapter 6 (General Discussion).

Summary

In sum, consistent with Experiment 3, Experiment 4 provided evidence that, when a lineup member stands out from a photospread as the only member with an odd smile, the member may be judged as appearing relatively more memorable and unfamiliar than the other lineup members. Due to such judgements, the expression can lead to witnesses perceiving the smiling lineup member to be an unlikely match to the culprit, even if the member is the culprit. Following from the results of Experiment 4, Experiments 5 and 6 (Chapter 5) further explored the effects of face distinctiveness, memorability and unfamiliarity on witnesses' confidence assessments by manipulating such perceptions of lineup members using cues other than an odd smile.

CHAPTER 5

Experiment 5

Experiment 4 demonstrated that an expression of a smile made the innocent suspects and the culprits appear distinctive and memorable. Moreover, some evidence suggested that the expression might have cued participants to think that the previous encounter with such a distinctive face should have cued a clear recollection, leading to the smiling innocent suspects and culprits being perceived as unlikely to be the culprits. The evidence also showed that, perhaps due to the smile causing the innocent suspects and the culprits to appear distinctive and memorable, or perhaps due to other reasons, the cue made the innocent suspects and culprits appear unfamiliar, also contributing to the smiling individuals being perceived as unlikely to be the culprits. Experiment 5 investigated other ways in which a face presented in a photospread might appear unexpectedly distinctive and, thus, (a) shape the view that the face is too memorable not to cue a clear recollection or (b) make the face appear unfamiliar, motivating witnesses to consider that the face does not belong to the culprit. In Experiment 5, participants received a limited view of the culprit with the encoding conditions such that, even though participants had a prolonged exposure to the culprit's face, they did not get a view of the whole face. Later, they were presented with a photospread, which consisted of an innocent suspect or the culprit, who was presented either with or without a prominent physical feature such as a scar or a beard on the part of the face that was not visible at encoding.

It was predicted that the presence of an unexpected physical feature would lead to a face presented in a lineup perceived as being new (i.e., not belonging to the culprit), which would be a different experience from assuming that the face could have been encoded but it was simply forgotten. That is, as the theory of memory for

nonoccurrences would suggest, it was predicted that due to the unexpected presence of the physical feature making the face appear distinctive and memorable (as opposed to appearing non-distinctive and unmemorable), when a lack of a clear recollection was experienced for that face, the face would be perceived as being new as opposed to it being forgotten. Thus, in Experiment 5, the presence of the unexpected physical features of a scar and a beard were thought to make the innocent suspects and the culprits be perceived as though they had never been seen before.

During a crime, due to the condition of the viewing (i.e., lighting, angle of the viewing), only a limited view of the culprit might be available to the witness. In line with this, in Experiment 5, throughout the female mock-crime video the angle of the camera shot resulted in participants only seeing the left side of her face and, thus, the right side of her face was concealed. Later, those in the cue present condition were presented with a photospread, which showed either an innocent suspect or the culprit with a scar on their right cheeks. Similarly, throughout the male mock-crime video, the male culprit's chin was concealed due to the culprit holding a newspaper in front of his chin and, thus, participants were provided with a limited view of the culprit. Later, those in the cue present condition were presented with a photospread, which showed either an innocent suspect or the culprit with a dark beard on their chins. It was predicted that participants would perceive the members displaying the features as surprisingly distinctive, memorable, and perhaps unfamiliar. Whether together or independently, such perceptions were expected to lead to the photospread members with the distinguishing features being perceived as unlikely to be the culprits, even when the members were in fact the culprits.

Third Cue Condition

In addition to the two cue conditions (feature present vs. absent), Experiment 5 incorporated a third cue condition where the feature of the scar or the beard was replicated across the photospread. Thus, Experiment 5 consisted of three cue conditions: the absence of the feature from the photospread (i.e., cue absent), the feature being presented by the innocent suspect/culprit alone (i.e., solo cue), and the feature being presented by all of the members of the photospread (i.e., all cue). The third cue condition was added, for at times, the police could come across a situation where they have a suspect who has a physical feature that was not mentioned by the witness. In such instance, Wells, Seelau, Rydell, and Luus (1994) suggest that the police should select fillers who have a similar feature to the suspect to ensure that the suspect does not stand out from a lineup. However, there are several plausible reasons why the description of the culprit provided by an eyewitness might not match the description of the suspect obtained by the police. First, the eyewitness might have encoded the feature, yet simply forgotten to mention the feature to the police. Second, the eyewitness might not have mentioned the feature due to the feature not being available in his or her memory (e.g., due to forgetting or due to the feature never being encoded). Therefore, depending on the rationale behind the discrepancy found between the eyewitness' description of the culprit and the suspect obtained by the police, such recommendation could potentially increase the risk of incorrect rejections being made from target-present lineups. That is, when a particular feature of the culprit is not mentioned by the eyewitness due to the feature being absent from the eyewitness' memory, the unmentioned feature might force the eyewitness to conclude that none of the photospread members could be the culprit, even if the actual culprit was in the photospread. Therefore, the recommendation provided by Wells et al.

(1994) might not always be favourable. In line with this, in Experiment 5 I explored what the impact of this recommended practice would be when a prominent feature, which the eyewitness did not see on the culprit at the time of encoding, was replicated across a photospread.

It was predicted that, compared to the feature being presented only by the innocent suspect or the culprit, replicating the feature across the photospread would stop the innocent suspect and the culprit from appearing more distinctive and memorable than the fillers, thereby possibly reducing the likelihood of them standing out as resembling the culprit rather poorly. Thus, on average, the confidence difference score given to the all cue condition was predicted to be less negative than the score given to the solo cue condition.

When comparing the all cue and the cue absent conditions, replicating as opposed to eliminating the cue across the photospread was predicted to harm the confidence assessment, whereby the cues (i.e., scar, beard) might confuse participants and, thus, make them less certain of their memory for the culprit. This effect was predicted to be particularly robust under the condition in which there would be a relatively large familiarity difference between the suspect and the best filler (i.e., target-present photospread). Therefore, the TP photospreads were predicted to lead to the all cue condition producing a more negative confidence difference score than the neutral cue condition. On the other hand, for the TA photospreads, considering that the suspects were not the actual culprits, there would be a relatively small familiarity difference between the innocent suspect and the best filler. Due to this limited discrepancy in familiarity, regardless of whether the cue was replicated or eliminated across the photospreads, it was predicted that the perceived similarity of the innocent suspects to the culprits in relation to the fillers would not be strongly impacted by the

presence of the cue across the photospreads. Therefore, it was predicted that the difference observed between the all cue and the cue absent conditions would be much smaller in the TA photospreads than the TP photospreads. Overall, the third cue condition was added as an exploratory condition.

Viewing Condition

Experiment 5 was conducted with the assumption that, whilst the mock-crime videos only showed limited views of the culprits' faces, participants would falsely recall seeing the whole of the culprits' faces. Thus, it was conducted with an assumption that people are generally poor at recalling the viewing conditions of their past events. In support of this, Intraub and colleagues (Intraub, Bender, & Mangels, 1992; Intraub & Richardson, 1989) found that when participants were presented with a series of scenic photographs and later asked to draw what they saw, participants consistently extended the boundaries of the scenes by adding more background details than what they saw. Similarly, Kraft (1987) presented participants with photographs of people conducting everyday activities and later, tested their recognition ability by presenting them with the same photographs or photographs of the same events shot from different angles. They found that participants generally struggled to identify the changes in the viewing angles, often falsely indicating that the photographs shot from different angles were the exact replica of the photographs they saw at encoding. Together, the studies suggest that people are rather poor at recalling the exact details of past events, such as the viewing conditions to which they were previously exposed.

Schema theory suggests that people are not only poor at recalling the details of past events but often elaborate and modify their memories to make them fit with their schema, that is, expectations based on their past experiences (Alba & Hasher, 1983). To test this, Tuckey and Brewer (2003) presented witnesses with a bank robbery,

which contained several pieces of ambiguous information. For example, the female robber's gender was made ambiguous due to the robber wearing a balaclava and the absence of a weapon was made ambiguous due to the robber placing a hand in her pocket throughout the robbery as though she was holding a gun inside it. When the witnesses were asked to recall the details of the bank robbery, they tended to replace the ambiguous information with schema consistent information, for example, reporting that the bank robbery involved a male robber and a gun. Thus, evidence suggests that, when witnesses are presented with ambiguous information, they may elaborate their memory to make it fit with their expectations, leading to false information being recalled. In line with the schema theory, I predicted that when participants encoded a limited view of the culprit's face, they would falsely recall viewing the whole face of the culprit. Therefore, when the participants were later presented with the photospreads with the scar or the beard being visible on the members of the photospreads, it was predicted that they would falsely conclude that the features could not belong to the culprit.

Additional Manipulation Check Measure: Physical Feature Questionnaire

To check whether participants falsely believed that they got a view of the whole face of the culprit, at the end of Experiment 5, questions were provided asking if the female culprit had a scar on her right cheek and if the male culprit had a beard on his chin. Participants were required to provide one of the following response options: "yes" if they clearly recalled that the culprit had the feature, "no" if they clearly recalled the feature was absent, "unsure" if they could not recall as to whether the feature was present or absent and "couldn't see" if for some reason, the feature could not be seen at the time of encoding. The "no" response was thought to reflect that participants falsely believed seeing the concealed parts of the culprits' faces in

the mock-crime videos. Thus, it was predicted that regardless of the cue conditions, the majority of the participants would falsely choose the "no" option rather than the "unsure" option or the correct "couldn't see" option, indicating that they were certain that the culprits did not have such features. The response option of "yes" was added as a distractor.

Viewing Condition Pilot-Test

Prior to Experiment 5, a pilot-test was conducted to address two issues: (1) whether the culprits would be perceived as appearing more distinctive or unfamiliar when they were presented with unexpected physical features of a scar or a beard than without, and (2) whether a retention interval of 20 minutes between the viewing of the mock-crime videos and the presentations of the photospreads was long enough for witnesses to demonstrate their inability to remember the details of the viewing conditions of the mock-crime videos.

Method

Stimulus videos. Prior to conducting the pilot-test, the mock-crime videos used in Experiment 4 were edited to limit the visibility of the culprits' faces. In particular, the female mock-crime video was edited in such a way that only the left side of the culprit's face was visible throughout the video. The male mock-crime video showed the male culprit with his chin being concealed due to the culprit holding a newspaper or by his face being outside of the camera shot. Each video lasted around 20 seconds, with the culprit's face appearing in a close-up for approximately 5 seconds.

Procedure. A group of participants (N = 30) took the pilot-test, which consisted of three phases: encoding, retention interval and test. Placed in an individual cubicle, participants first completed the encoding phase by watching the two mock-

crime videos one after the other on a computer screen. During the retention interval, participants completed a maze booklet for 20 minutes. At the beginning of the test phase, the computer instructed participants that they were going to be asked some questions about the mock-crimes videos. Participants were provided with a booklet (see Appendix B) and asked to write down all of their answers in it.

The test phase consisted of two parts. In the first part, participants saw six images of each culprit, one neutral and five with various physical features being added to the face. The female culprit was presented with an addition of one of the followings: a scar on her right cheek, a scar on her forehead, a mole on the left side of her mouth, a tattoo on her forehead, and a nose piercing. The male culprit was presented with an addition of one of the followings: a beard on his chin, a scar on his forehead, a mole next to his nose, a moustache and a piercing on the bridge of his nose. Each image was also presented with a written description of the feature. For example, the image of the female culprit with a mole on her nose was presented with a description, "mole on the nose". For each culprit, after the presentations of the five images with the various physical features being added, the neutral image of the culprit was presented, along with a description, "she/he had none of the physical features described". The right cheek scar (for the female) and the beard (for the male) served as target images, the neutral images served as controls, and the rest served as distractors. For each image, participants were asked to rate: (i) how confident (0-100%) they were that the culprit had the physical feature, (ii) how atypical/distinctive and (iii) familiar each face appeared. The latter two ratings were conducted using a 7-point Likert scale (1 = *not at all*; 7 = *very much so*). It was predicted that the target images would be rated lower on the confidence scale than the neutral images, suggesting that participants were certain that the culprits did not have the physical

features of scar or beard. It was further predicted that the target images would be rated much higher on the atypical/distinctiveness scale and much lower on the familiarity scale than the neutral images.

In the second part of the test phase, participants saw 18 images related to the mock-crime videos: nine for each mock-crime. The nine images consisted of three types of images: same, similar and dissimilar. The same images consisted of three still images that were extracted from the actual mock-crime video. The similar images consisted of three still images of the mock-crime event but showed the full face of the culprit. The dissimilar images consisted of three still images of the culprit at a different event. Participants saw each image one at a time in a random order for 15 seconds. For each image, participants indicated either "Yes – The image matches what I saw in the crime scene" or "No – It is different to the crime scene I saw". Participants were warned that some of the images may belong to the mock-crime videos they saw earlier but something about the viewing condition (e.g., camera angle, use of disguise etc.) would be different. They were instructed to say "yes" only when they felt that the image explicitly replicated the mock-crime videos they saw. It was predicted that participants would most often say "yes" to the similar images, indicating that they generally believed that the mock-crime videos showed the whole faces of the culprits. The dissimilar images acted as distractors.

Results

Test phase part one. As shown in Table 5.1, the results of the test phase part one showed that, on average, participants were more confident that the culprits had none of the proposed physical features (i.e., neutral) than that the culprits had the features of the right cheek scar or the beard. Further, the male culprit was rated as appearing more distinctive and less familiar when presented with the beard than

without. For the female culprit, the distinctiveness rating failed to reach significance. However, the scar made the female culprit appear less familiar and as such, the scar was considered an acceptable cue to be used in Experiment 5.

Test phase part two. The results of the test phase part two showed that, as predicted, collapsed across the two mock-crimes, the majority of the responses given to the similar images were "yes" (83%). Furthermore, paired samples t-tests showed that compared to the proportions of correct "yes" responses given to the same images,

Table 5.1

Descriptive and Inferential Statistics for the Confidence, Distinctiveness and Familiarity Ratings given to the Female and the Male Culprits with the Cue (Right Cheek Scar/ Beard) and Without (Neutral)

	Female Culprit		Male Culprit	
	M (*SD*)	95% CIs	*M* (*SD*)	95% CIs
Confidence				
Cheek Scar/ Beard	21.73 (28.94)	10.92, 32.54	20.50 (32.23)	8.47, 32.54
Neutral	77.83 (26.32)	68.00, 87.66	71.00 (30.33)	59.68, 82.33
	$t(29) = -5.92$, $p < .001$, $f = 1.02$		$t(29) = -5.29$, $p < .001$, $f = .81$	
Distinctiveness				
Cheek Scar/ Beard	3.86 (2.00)	3.11, 4.61	4.17 (2.07)	3.40, 4.94
Neutral	3.10 (2.34)	2.23, 3.97	2.93 (2.08)	2.15, 3.71
	$t(29) = 1.44$, $p = .16$, $f = .18$		$t(29) = 2.73$, $p = .01$, $f = .30$	
Familiarity				
Cheek Scar/ Beard	3.62 (2.04)	2.86, 4.38	3.07 (1.98)	2.33, 3.81
Neutral	5.83 (1.72)	5.19, 6.47	5.47 (2.08)	4.69, 6.25
	$t(29) = -4.72$, $p < .001$, $f = .59$		$t(29) = -4.43$, $p < .001$, $f = .59$	

the similar images produced much lower proportions of correct "no" responses, for both female and male mock-crimes (M = .94, SD = .13; 95% CI = .89, .99 vs. M = .18, SD = .34; 95% CI = .04, .30), and (M = .94, SD = .20; 95% CI = .87, 1.02 vs. M = .18, SD = .31; 95% CI = .05, .29), respectively, all $ts > 8.8$, $ps < .001$, $fs > 1.0$. Thus, the pilot-test confirmed that the use of the newly edited mock-crime videos, followed by a 20 minutes retention interval would likely lead to participants falsely believing that they received views of the whole faces of the culprits. Therefore, Experiment 5 was conducted using the methodology described in the following section.

Method

Experiment 5 consisted of 180 (130 female, 50 male) paid community participants and first year psychology students who participated for course credit. Ages ranged from 17 to 48 (M = 22.08, SD = 6.31). The general procedure and the design of Experiment 5 were similar to Experiment 4 with the exception of the addition of the third cue condition (i.e., all cue), leading to a 3 (cue: solo, all, absent) × 2 (target presence: TP, TA) × 2 (mock-crime: male, female) mixed design. In addition, at the end of Experiment 5, a manipulation check questionnaire regarding the culprits' appearances was added.

Materials

Stimulus videos. The mock-crime videos from the viewing condition pilot-test were used.

Photospreads. The neutral TP and TA photospreads used in Experiment 4 were used. For each member of the photospreads, two types of photos were produced: one neutral and one with the cue (i.e., scar/beard) added. That is, using the *Face Switch* application (Apple iPhone), the image of the scar (from the viewing condition

pilot-test) was added to the right cheek of each female face (i.e., innocent suspect, culprit and fillers) and the image of the beard (from the viewing condition pilot-test) was added to the chin of each male face. See Figure 5.1 for examples.

Using the two types of photos (i.e., cue present and absent), for each TP and TA photospread, three cue conditions were created: all members presented without the cue (i.e., cue absent condition), the cue presented only by either the innocent suspect or the culprit (i.e., solo cue condition) and all members presented with the cue (i.e., all cue condition). Therefore, in total, 12 photospreads were produced: female and male TP and TA photospreads, each with three cue conditions (i.e., cue absent, solo cue, all cue).

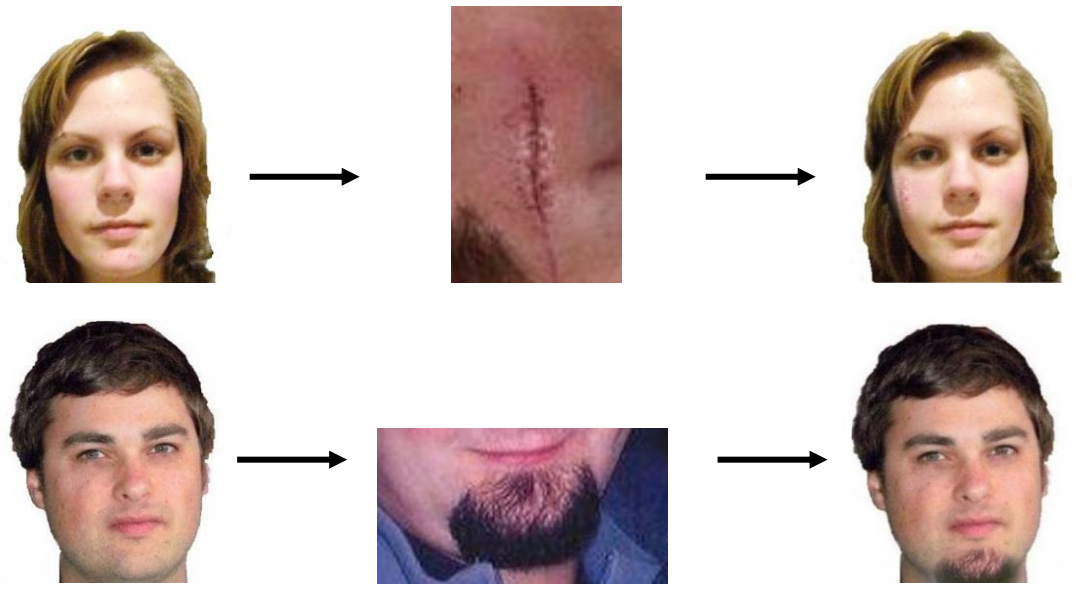

Figure 5.1. Female and male culprits' transitions displayed from neutral (left) to the cues being added (right)

Procedure

As suggested earlier, the general procedure of Experiment 5 mirrored Experiment 4, with the exception of the addiiton of the third cue condition. That is, during the multiple confidence procedure, participants saw one female and one male photospread, which consisted of one TP and one TA photospread with one of the following combinations of the cue conditions: (1) one cue absent and one all cue, (2) one cue absent and one solo cue, (3) one solo cue and one all cue. For example, if participants saw the TP solo cue photospread for the male mock-crime, for the female mock-crime, they saw either the TA cue absent photospread or the TA all cue photospread. Simiarly, if participants saw the TP all cue photospread for the female mock-crime, for the male mock-crime, they saw either the TA solo cue photospread or the TA cue absent photospread.

In line with Experiment 4, at the end of Experiment 5 each participant rated the perceived distinctiveness and memroability of all of the photos that were used to create the photospreads. That is, for each scale (i.e., distictiveness and memorability), the photos of all of the photospread members were presented to each participant twice, once with the physical feature of the scar or the beard being added and once without. Therefore, each participant provided the ratings for the inncoent suspects and the culprits with and without the physical features, as well as all of the fillers with and without the physical features, leading to the ratings being obtained within-subjects.

After the manipulation check ratings of distinctiveness and memorability, as an additional manipulation check measure a questionnaire was provided to participants to see whether they correctly recalled not being able to see the scar or the beard on the culprits' faces. In particular, a paper questionnaire (example in Appendix C) was provided, which asked participants to think back to the culprits' appearances

during the mock-crime videos. For the male culprit, participants were asked if he had physical features such as a beard on his chin, a tattoo on his forehead, a moustache, glasses, an eyebrow piercing or a lip piercing. For the female culprit, participants were asked if she had physical features such as a scar on her right cheek, a tattoo on her forehead, a mole on her nose, glasses, a nose piercing or a lip piercing. For each feature, participants were asked to choose one of the answer options, "yes", "no", "unsure" or "couldn't see". For each answer, participants were asked to provide a confidence rating (0-100%). At the end of the questionniare, participants were also provided with a blank space for additional comments. The primary interests were the questions of right cheek scar (for the female culprit) and the beard (for the male culprit), with the others acting as distractors.

Results

As with the preceding experiments, difference (S_{uspect}–$F_{iller\ Max}$) scores were calculated for the familiarity, distinctiveness, memorability and multiple confidence ratings. The effect size estimate reported for the comparisons of proportions was Cohen's *w*, with the approximate cut-off guidelines for small, medium and large effects being 0.10, 0.30 and 0.50.

Manipulation Check Measures

Physical feature questionnaire. One subject failed to complete the questionnaire and, thus, the analyses of the questionnaire were conducted using 179 subjects' data. From the questionnaire, it was predicted that across the cue conditions, participants would most often choose the "no" option, more so than the "couldn't see" or the "unsure" options.

For the female stimuli, a chi-square test showed no significant association between the cue condition and the response options, χ^2 (6, n = 179) = 7.52, p = .28, w

= .20. That is, as shown in Figure 5.2, regardless of the cue conditions, participants were just as likely to choose the options of "couldn't see" or "unsure" as they were to choose the "no" option. Therefore, the prediction was not supported.

For the male stimuli, as shown in Figure 5.3, in the solo cue and the cue absent conditions, more participants chose the "no" rather than the "unsure" or "couldn't see" response options. Furthermore, a significant association was observed between the cue condition and the response options, χ^2 (6, n = 179) = 12.60, p = .05, w = .21. Using the adjusted standardised residuals, in the solo cue condition, the proportion observed for the "couldn't see" response option was significantly lower than the expected count. In the all cue condition, the proportion observed for the "no" response option was significantly lower than the expected count. In the cue absent condition, the proportion observed for the "yes" response option was significantly lower than the expected count. Thus, when presented with the photospread consisting of either the innocent suspect or the culprit as the only photospread member with the beard, participants were unlikely to consider the possibility that the beard could not be seen on the culprit at encoding. When the beard was replicated across the photospread instead, less participants were certain that the culprit did not have the beard. In contrast, when the beard was eliminated from the photospread, perhaps, this information was used as a validation that the male culprit did not have the beard, leading to less participants choosing the "yes" response option. The confidence ratings obtained for the response options showed no significant effects of cue. The descriptive statistics for the confidence data are reported in Appendix D.

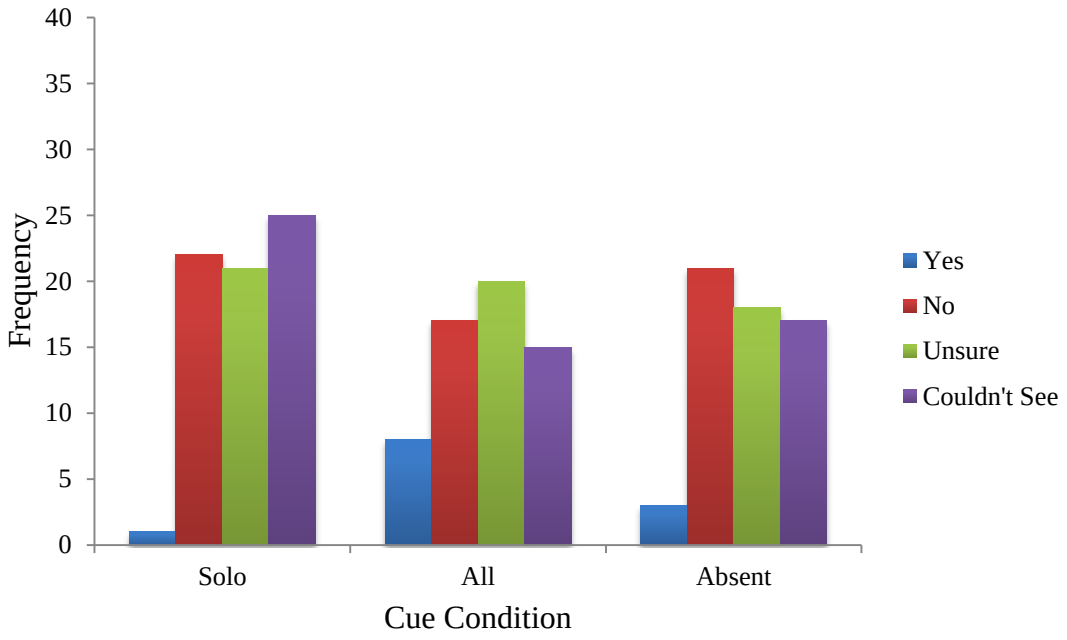

Figure 5.2. Frequencies with which "yes", "no", "unsure" and "couldn't see" response options were given to the female culprit's right cheek scar question in the three cue conditions

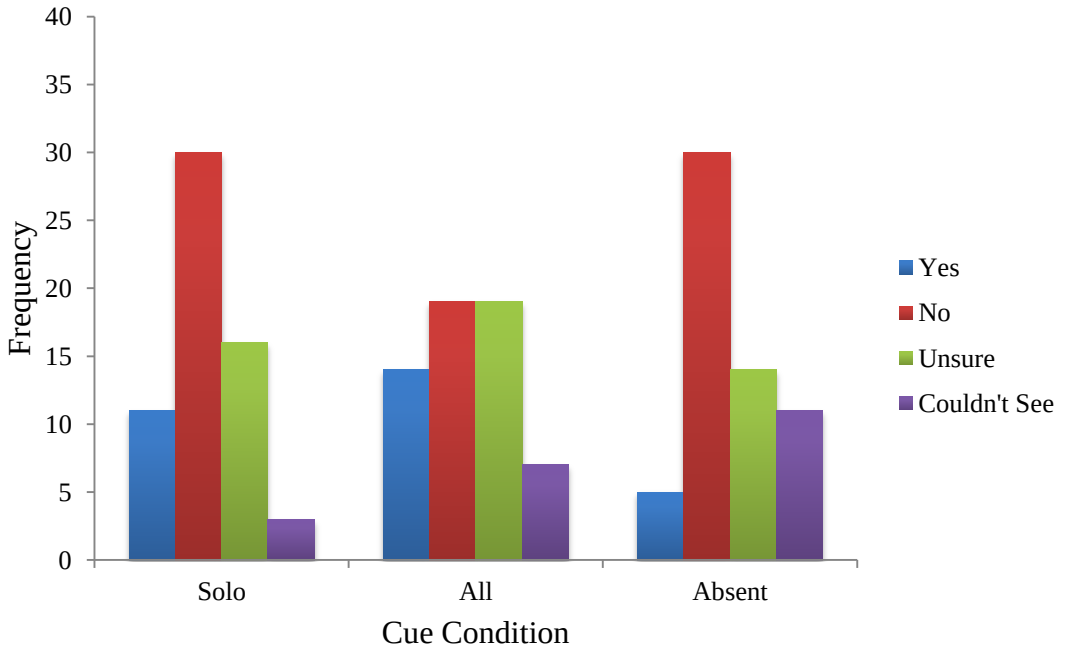

Figure 5.3. Frequencies with which "yes", "no", "unsure' and 'couldn't see" response options were given to the male culprit's beard question in the three cue conditions

Memorability. For each innocent suspect and culprit, three difference scores were calculated to reflect the three cue conditions. That is, the scores indexed the differences between: (1) the value given to the neutral innocent suspect/ culprit and the maximum value given to a filler without the cue (i.e., cue absent condition), (2) the value given to the innocent suspect/ culprit presented with the cue and the maximum value given to a filler without the cue (i.e., solo cue condition), (3) the value given to the innocent suspect/ culprit presented with the cue and the maximum value given to a filler presented with the cue (i.e., all cue condition).

As in Experiment 4, the distinctiveness and memorability difference scores given to each innocent suspect and culprit were positively correlated (r = .20 - .49). Thus, the difference scores of distinctiveness and memorability were collapsed into a single variable called memorability difference score (i.e., adding the two scores and dividing the total by two)[27].

To determine whether the cue presented solely by the innocent suspects and the culprits made them appear more distinctive and memorable than when the cue was replicated or eliminated across the photospreads, for each innocent suspect and culprit, the memorability difference scores reflecting the three cue conditions (i.e., solo, all, absent) were compared using a repeated measures ANOVA analysis. As shown in Table 5.2, consistent results were observed across stimuli. That is, the memorability difference scores reflecting the solo cue condition were less negative than the scores reflecting the all cue (ps < .01, fs > .13) or cue absent conditions (ps > .01, fs > .23). The scores reflecting the all cue condition were also less negative than the scores reflecting the neutral condition (ps > .01, fs > .12). Thus, when the innocent suspects and the culprits were the only photospread members presented with the scar

[27] Analyses were repeated treating the distinctiveness and memorability difference scores separately, with these showing similar patterns of results.

or the beard, they were perceived as appearing more distinctive and memorable than when the features were replicated or eliminated across the photospreads.

Table 5.2

Means, Standard Deviations and 95% CIs for the Memorability Difference (S_{uspect}–$F_{iller\ Max}$) Score Calculated for the Photos of the Innocent Suspects and the Culprits, Reflecting the Three Cue Conditions (Solo, All and Cue Absent)

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Culprit						
Solo	-0.39	1.69	-0.64, -0.14	-0.91	1.48	-1.13, -0.69
All	-1.21	1.08	-1.40. -1.05	-1.27	1.24	-1.45, -1.09
Absent	-1.72	1.39	-1.92, -1.52	-1.70	1.37	-1.90, -1.49
	Wilks' Lambda = .52, $F(2, 178)$ = 82.10, $p < .001$, $f = .40$			Wilks' Lambda = .76, $F(2, 178)$ = 28.89, $p < .001$, $f = .34$		
Innocent Suspect						
Solo	0.19	1.56	-0.04, 0.42	-1.17	1.62	-1.41, -0.93
All	-0.62	1.04	-0.77, -0.47	-1.53	1.34	-1.73, -1.33
Absent	-1.08	1.29	-1.27, -0.89	-1.85	1.36	-2.05, -1.65
	Wilks' Lambda = .56, $F(2, 178)$ = 69.88, $p < .001$, $f = .40$			Wilks' Lambda = .80, $F(2, 178)$ = 22.16, $p < .001$, $f = .19$		

Perceived Familiarity Rating

Using the familiarity difference score, a 3 (cue: solo, all, absent) × 2 (target presence: TP, TA) factorial ANOVA analysis was conducted separately for the female and the male stimuli. The female stimuli showed a main effect of cue, $F(2, 174)$ = 5.70, $p = .004$, $f = .25$. Post-hoc comparisons using the Tukey HSD tests showed that

the solo cue condition led to a more negative familiarity difference score than for the all cue condition, $p = .003$, $f = .31$(see Table 5.3). No differences were observed between any other combinations of the cue conditions ($ps > .05$, $fs = .22$). The main effect of target presence was not significant, $F(1, 174) = 2.79$, $p = .10$, $f = .12$, nor was the interaction between the cue and the target presence condition, $F(2, 174) = 2.52$, $p = .08$, $f = .18$. The male stimuli show a main effect of cue, $F(2, 174) = 3.21$, $p = .04$, $f = .19$. Post-hoc comparisons using the Tukey HSD tests showed that the solo cue condition led to a more negative familiarity difference score than for the cue absent condition, $p = .05$, $f = .41$ (see Table 5.3). No differences were observed between any other combinations of the cue conditions ($ps > .05$, $fs \leq .19$). The TP photospread produced a less negative familiarity difference score than the TA photospread, $F(1, 174) = 7.33$, $p = .007$, $f = .12$. No significant interaction between the cue and the target presence condition was observed, $F(2, 174) = 0.39$, $p = .68$, $f = .07$.

To see if the memorability and the familiarity difference scores were associated, for each innocent suspect and culprit, a correlational analysis was conducted using the two scores. The results showed no relationship between the two scores for the male innocent suspect ($r = .08$), the female innocent suspect ($r = -.004$) or the culprit ($r = .08$). The male culprit showed a positive but weak relationship ($r = .23$). That is, opposite to the prediction, there was a tendency for the male culprit to be rated higher on the familiarity scale when he was rated higher on the memorability scale.

Multiple Confidence Rating

Of the 90 subjects who viewed the female TP photospread, 45 (50%) gave a unique maximum confidence value to the culprit, 32 (35.6%) gave it to one of the fillers and 13 (14.4%) gave no single unique maximum confidence value. Of the 90

Table 5.3

Means, Standard Deviations and 95% CIs for the Effect of the Cue and the Target Presence on the Familiarity Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Female and the Male Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Target-Present						
Solo	-1.17	2.95	-2.77, -0.57	-1.17	2.32	-2.04, -0.30
All	-0.47	2.27	-1.32, 0.38	-0.50	2.99	-1.62, 0.62
Absent	-0.10	2.19	-0.92, 0.72	-0.50	2.50	-1.43, 0.43
Overall	-0.58	2.50	-1.10, -0.06	-0.72	2.61	-1.27, -0.17
Target-Absent						
Solo	-2.10	2.31	-2.96, -1.24	-2.50	2.06	-3.27, -1.73
All	0.03	1.59	-0.56, 0.62	-1.50	2.03	-2.26, -0.74
Absent	-1.40	2.42	-2.30, -0.50	-1.07	2.33	-1.94, -0.20
Overall	-1.16	2.29	-1.64, -0.68	-1.69	2.21	-2.15, -1.23
Overall						
Solo	-1.63	2.67	-2.32, -0.94	-1.83	2.28	-2.42, -1.24
All	-0.22	1.96	-0.73, 0.29	-1.00	2.58	-1.67, -0.33
Absent	-0.75	2.38	-1.37, -0.14	-0.78	2.42	-1.41, -0.16
Overall	-0.87	2.41	-1.22, -0.52	-1.21	2.46	-1.57, -0.85

subjects who viewed the female TA photospread, 18 (20%) gave a unique maximum confidence value to the innocent suspect, 48 (53.3%) gave it to one of the fillers and 24 (26.7%) gave no single unique maximum confidence value. Of the 90 subjects who viewed the male TP photospread, 42 (46.7%) gave a unique maximum confidence

value to the culprit, 37 (41.1%) gave it to one of the fillers and 11 (12.2%) gave no single unique maximum confidence value. Of the 90 subjects who viewed the male TA photospread, 8 (8.9%) gave a unique maximum confidence value to the suspect, 73 (81.1%) gave it to one of the fillers and 9 (10%) gave no single unique maximum confidence value.

Treating the confidence difference score as the dependent variable, a 3 (cue: solo, all, absent) × 2 (target presence: TP, TA) factorial ANOVA analysis was conducted separately for the female and the male stimuli[28]. As shown by the descriptive statistics provided in Table 5.4, for the female stimuli, the TP and the TA photospreads showed different patterns regarding the effect of the cue. For the TP photospread, the three cue conditions showed similar confidence difference scores. For the TA photospread, the all cue condition led to a less negative confidence difference score than for the solo cue and the cue absent conditions. Despite this, no interaction between the cue and the target presence condition was observed, $F(2,174) = 1.36$, $p = .26$, $f = .13$. Furthermore, no main effect of cue was observed, $F(2,174) = 1.60$, $p = .21$, $f = .13$. There was a main effect of target presence, with the TP photospread producing a more positive confidence difference score than the TA photospread, $F(1,174) = 6.05$, $p < .001$, $f = .30$.

The male stimuli showed a main effect of cue, $F(2,174) = 5.59$, $p = .004$, $f = .23$. Post-hoc comparisons using the Tukey HSD test showed that the solo cue condition led to a significantly more negative confidence difference score than for the cue absent condition, $p < .05$, $f = .28$ (see Table 5.4). No differences were observed between any other combinations of the cue conditions ($ps > .05$, $fs \leq .15$). The TP

[28] No presentation order effects of the mock-crime videos nor the photospreads were observed.

Table 5.4

Means, Standard Deviations and 95% CIs for the Memorability Difference (S_{uspect}–$F_{iller\ Max}$) Score Calculated for the Photos of the Innocent Suspects and the Culprits, Reflecting the Three Cue Conditions (Solo, All and Cue Absent)

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Target-Present						
Solo	1.17	60.94	-21.59, 23.93	-4.70	42.97	-20.75, 11.35
All	3.20	51.22	-15.93, 22.33	-1.67	50.62	-20.57, 17.23
Absent	8.53	44.35	-8.03, 25.09	18.50	55.90	-2.37, 39.37
Overall	4.30	52.11	-6.61, 15.21	4.04	50.61	-6.56, 14.64
Target-Absent						
Solo	-35.83	44.24	-52.35, -19.31	-50.17	35.17	-63.30, -37.04
All	-8.33	26.76	-18.32, 1.66	-28.23	42.16	-43.97, -12.49
Absent	-24.83	39.51	-39.58, -10.08	-18.17	42.03	-33.86, -2.48
Overall	-23.00	38.84	-37.50, -8.50	-32.19	41.69	-40.92, -23.46
Overall						
Solo	-17.33	55.99	-31.79, -2.87	-27.43	45.18	-39.10, -15.76
All	-2.57	40.93	-13.14, 8.00	-14.95	48.09	-27.37, -2.53
Absent	-8.15	44.91	-19.75, 3.45	0.17	52.40	-13.37, 13.71
Overall	-9.35	47.83	-16.39, -2.32	-14.07	49.68	-21.38, -6.76

photospread produced a more positive confidence difference score than the TA photospread, $F(1,174) = 28.79$, $p < .001$, $f = .39$. No significant interaction between the cue and the target presence condition was observed, $F(2,174) = 0.65$, $p = .52$, $f = .09$. Thus, for the male stimuli, compared to when the beard was eliminated from the

photospread, presenting the innocent suspect or the culprit as the only photospread members with the beard led them being perceived as resembling the culprit relatively less than the fillers[29].

Discussion

In Experiment 5, the physical features of a right cheek scar and a beard led to both female and male faces being perceived as appearing distinctive and memorable when they were the only faces with the features, rather than when the features were replicated or eliminated across the photospreads. Whilst some evidence suggested that the features also made the innocent suspects and the culprits appear unfamiliar, the familiarity difference score was not related to the memorability difference score, with the exception of the male culprit which showed a weak relationship in the opposite direction to that predicted. Importantly, despite some consistent results observed between the female and the male stimuli, only the male stimuli's beard affected the witnesses' confidence assessment. That is, compared to when the beard was eliminated from the male photospread, making the innocent suspects or the culprits stand out as the only photospread members with the beard led to them being perceived as resembling the culprit relatively less than the fillers. Importantly, for the male culprit, the beard reduced the tendency for participants to perceive the culprit as the best match to be the culprit. Replicating the beard across the photospread did not increase or decrease participants' perceptions of the degree of the resemblance

[29] Attempts were made to test if the memorability or the familiarity difference score would mediate the relationship between the cue and the confidence difference score. Hayes and Preacher (2013) suggest that when an IV has more than two levels, the analyses should be conducted by constructing *k*-1 dummy variables (i.e., *k* = levels of IV) and then running INDIRECT *k*-1 times, with each run, treating one dummy variable as an IV and the other as a covariate. Using this method, a mediation analysis was conducted separately for each stimulus, treating the cue condition as the IV, the memorability or the familiarity difference score as the DV and the confidence difference score as the DV. No indirect effects of the memorability or the familiarity difference score on the relationship between the cue and the confidence difference score were observed.

between the innocent suspect and the culprit or between the culprit presented in the photospread and the image of the culprit held by the participants.

The results of the female stimuli likely reflect that the effect of the right cheek scar was considered a relatively non-salient cue, as indicated by the results of the viewing condition pilot-test where the scar failed to make the female culprit appear distinctive. This might have led to one of two things. First, participants might have been more willing to consider the possibility that they either had forgotten seeing the female culprit's scar or had failed to encode the scar. Indeed, whilst "no' was the most common option chosen for the question of the male culprit having the beard, when presented with the question regarding the female culprit's scar, similar proportions of participants chose the options of "unsure", "couldn't see" and "no". Thus, as the theory for the memory of nonoccurrences (Förster & Strack, 1998) would suggest, in order for a previously concealed feature to influence witnesses' confidence assessment, perhaps the feature needs to produce a sense of certainty that the culprit did not have the feature and, if seen previously, it would not have been forgotten. To test this possibility, the female stimuli's confidence difference score data were reanalysed only using the subjects' data who responded "no" to the female culprit's right cheek scar question. However, the results showed no effect of the cue on the confidence difference score, $F(2, 54) = 0.91$, $p = .41$, $f = .12$[30].

Second, it may be that regardless of whether participants were certain or uncertain of the female culprit having the right cheek scar, perhaps the scar was considered relatively minor information compared to other information available from the photospread, leading to the presence of the scar being dismissed. For example, when the female culprit was presented with the right cheek scar, even when

[30] To note, analysing the data separately for those who said "yes", "unsure" or "couldn't see" did not provide any significant results.

participants felt that the culprit was unlikely to have had the scar, perhaps additional information about her (e.g., her hair colour and facial structure) might have convinced them to think that, despite the scar, she was the best match to the culprit. Thus, in order for a face with a previously concealed feature to be perceived as unlikely to be the culprit, perhaps the feature needs to be more salient than the scar used in Experiment 5, leading witnesses to become more reliant on the feature to make their confidence assessment. To test this assumption, a follow-up experiment was conducted, replacing the female stimuli's physical feature of the right cheek scar with a facial tattoo.

Experiment 6

Experiment 6 aimed to replicate the results of Experiment 5 by adjusting the female stimuli by replacing the right cheek scar with a tattoo of a black cross. A facial tattoo was chosen because the reanalysis of the viewing condition pilot-test showed that the female culprit was rated higher on the distinctiveness scale when she was presented with a facial tattoo (scale of 1-7; $M = 5.66$, $SD = 2.39$) than with the right cheek scar or without any features, $ts > 3.80$, $ps \leq .001$[31]. In addition, the outcomes of the physical feature questionnaire used in Experiment 5 showed that when participants were asked if the female culprit had a facial tattoo (distractor item), the majority responded "no" (89.4%). Thus, it was predicted that the facial tattoo would produce a more salient effect than the scar used in Experiment 5, causing participants to think that the face with the tattoo is unlikely to be the culprit for such a salient feature on the culprit would not have been forgotten. Thus, the use of a more salient cue was thought to cause a degree of discrepancy between expected and actual

[31] The descriptive statistics for the latter two variables are reported in the viewing condition pilot-test section.

experience of recollection felt towards the female stimuli, whereby the experience would not be attributed to a case of forgetting.

In addition to the change made to the female stimuli, in Experiment 6, I used a new set of male stimuli to test the generalizability of the results of the male stimuli obtained in Experiment 5. That is, a new male actor was used as a culprit in a new mock-crime video and new photospreads were composed. The male stimuli's cue remained the same. That is, in the solo cue and the all cue conditions, the image of the beard used in Experiment 5 was added to the photos of the male photospread members.

Method

Experiment 6 consisted of 180 (99 female, 81 male) paid community participants and first year psychology students who participated for course credit. Ages ranged from 17 to 61 years (M = 21.84, SD = 5.77). Experiment 6 was similar to Experiment 5 with the exception of using a facial tattoo of a black cross as the cue for the female stimuli and the use of new male stimulus materials.

Materials

Stimulus video. The new male mock-crime video depicted a young Caucasian male wearing a blue scarf walking into a female bathroom, stealing a woman's yellow handbag and fleeing the scene. Prior to entering the bathroom, the culprit's face was shown in a close-up for five seconds. Throughout the video, the blue scarf covered the culprit's chin. The video lasted approximately 28 seconds.

Photospreads. For the female photospreads, the right cheek scar cue was replaced by a right cheek tattoo of a black cross. For the male, new TP and TA photospreads were composed using new sets of photographs. The photo of the culprit was taken by the experimenter. The photos of the innocent suspect and the fillers were

collected from various face databases available online. To create the solo cue and the all cue conditions, the image of the beard used in Experiment 5 was added to each photo. A pilot-test was conducted to ensure that in the absence of the beard, the photospreads were not biased against or in favour of the innocent suspect or the culprit.

Effective size and defendant bias measure. Placed in an individual cubicle with a computer, a group of participants (N = 12) watched the new male mock-crime video and provided written descriptions of the male culprit. The descriptions were then summarised into one modal description, describing the culprit as "Caucasian male, 20s, short brown/sandy hair, long face". Then, on a computer screen placed in an individual cubicle, another group of participants (N = 160 for TP and N = 107 for TA) were presented with the description along with eight photos of male individuals: one culprit and seven others. Participants were asked to pick one photo that matched the written description of the culprit the best. Out of the seven other photos, five were selected to serve as fillers and one was selected to serve as an innocent suspect, replacing the culprit in the TA photospread. The effective sizes and the defendant bias measures were then calculated for the TP and the TA photospreads. Both TP (effective size = 4.96) and TA (effective size = 4.73) photospreads were estimated to consist of approximately five effective members. Based on the effective size, the chance expectation of the culprit being chosen from the TP photospread was .20 (95% CI = .12 - .28), overlapping with the actual choosing of .12 (95% CI = .06 - .18). Thus, the TP photospread was considered not biased against or in favour of the male culprit. From the TA photospread, the chance expectation of the innocent suspect being chosen was .21 (95% CI = .13 - .29), whilst the proportion of the actual choosing of the suspect was .36 (95% CI = .27 - .45). Thus, the 95% CIs of the two proportions

overlapped with one another, indicating that the photospread was not biased against or in favour of the innocent suspect when the beard was absent from the photospread.

Results

For both female and male stimuli, of the 180 subjects' data obtained, two subjects' data were excluded due to the failure to follow the multiple confidence procedure (as described in Chapter 2). Therefore, both female and male stimuli's analyses were conducted using 178 subjects' data. As in Experiment 5, difference (S_{uspect}–$F_{iller\ Max}$) scores were calculated for the distinctiveness, memorability, familiarity and multiple confidence ratings. In addition, in line with Experiment 5, the distinctiveness and the memorability difference scores calculated for each innocent suspect and culprit showed positive correlations (r = .24 - .51). Therefore, the two scores were collapsed together as a single variable called memorability difference score.

Manipulation Check Measures

Physical feature questionnaire. For the female stimuli, as shown in Figure 5.4, in the solo cue and the cue absent conditions, more participants chose the "no" response option than the "unsure" or "couldn't see" response options. Furthermore, a Chi-square test showed a significant association between the cue condition and the response option, χ^2 (6, n = 178) = 13.25, p = .04, w = .24. Using the adjusted standardised residuals, some significant effects were observed. In the all cue condition, the proportion observed for the "yes" response option was significantly more than the expected count, whilst the proportion observed for the "no" response option was significantly less than the expected count. In the cue absent condition, the proportion observed for the "no" response option was significantly more than the expected count. The results likely reflect that, when the tattoo was replicated across

the photospread, more participants were convinced that the culprit could have had the tattoo. In contrast, when the tattoo was eliminated from the photospread, this information may have been taken as a validation that the female culprit did not have the proposed tattoo.

For the male stimuli, as shown in Figure 5.5, across the three cue conditions, most of the participants chose the "no" or "couldn't see" options, with the proportion of the "couldn't see" response option being particularly high in the solo cue condition. A Chi-square test showed a significant association between the cue and the response option, χ^2 (6, n = 178) = 17.35, p = .008, w = .28. Using the adjusted standardised residuals, some significant effects were observed. In particular, in the solo cue condition, the proportion observed for the "couldn't see" response option was significantly greater than the expected count. In the all cue condition, the proportion observed for the "unsure" response option was significantly greater than the expected count. In the cue absent condition, the proportion observed for the "no" response option was significantly greater than the expected count. Thus, importantly, after viewing the photospreads with the beard displayed only by the innocent suspect or the culprit, a large proportion of participants correctly indicated that they could not see the male culprit's beard at encoding. As in Experiment 5, the confidence ratings given to these response options were unaffected by the cue condition. Therefore, the results are only reported in Appendix E.

Memorability. For each innocent suspect and culprit, the memorability difference scores reflecting the three cue conditions (i.e., solo, all, absent) were compared using a repeated measures ANOVA analysis. As shown in Table 5.5, consistent results were observed across stimuli. For both female and male stimuli, the memorability difference scores reflecting the solo cue condition were less negative

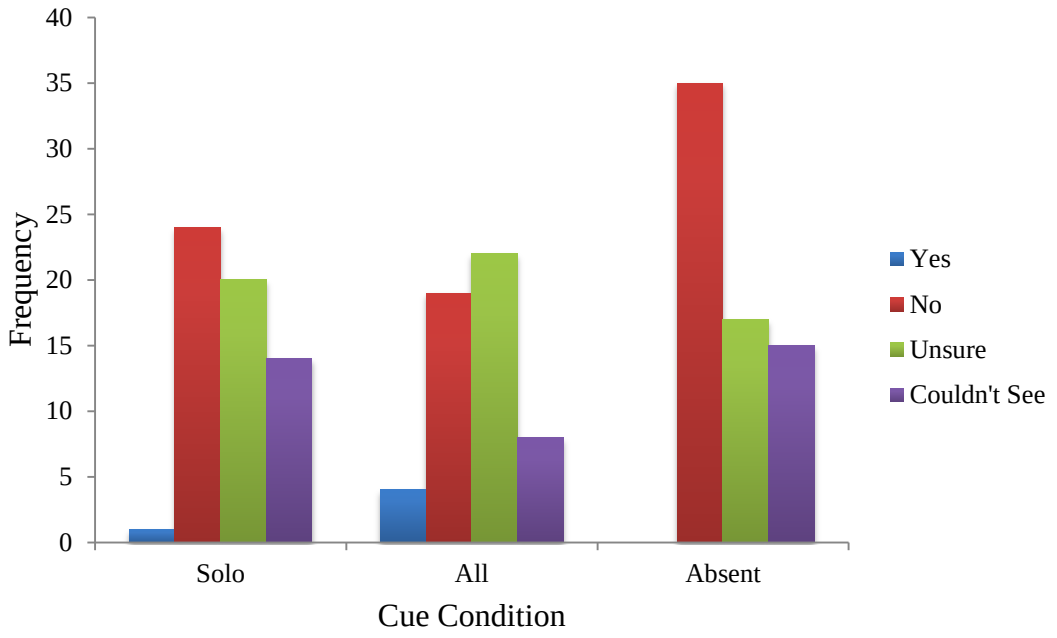

Figure 5.4. Frequencies with which "yes", "no", "unsure' and 'couldn't see" response options were given to the female culprit's tattoo question in the three cue conditions

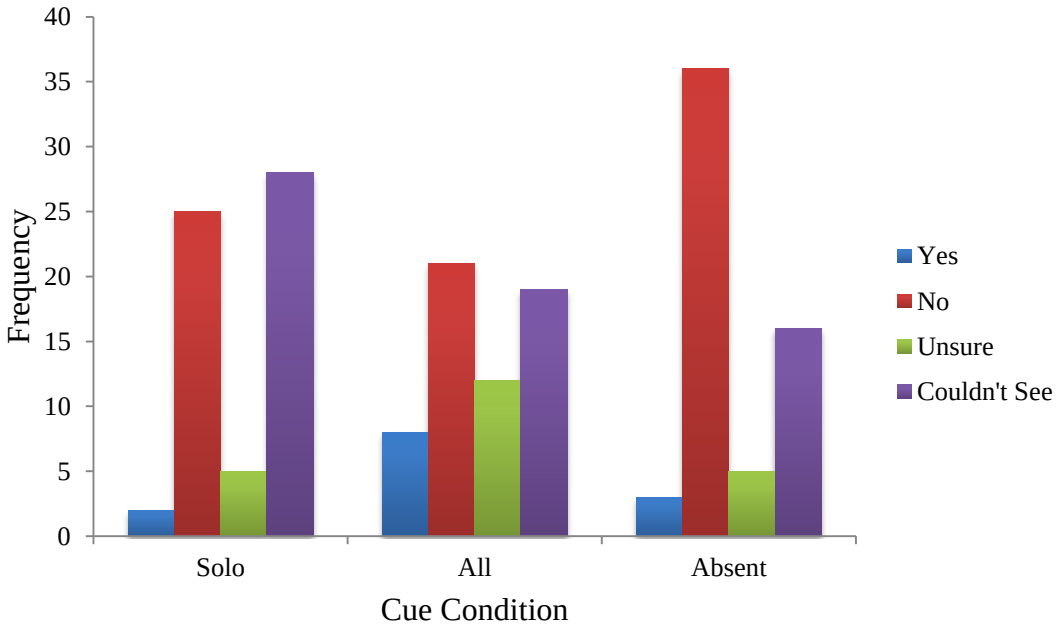

Figure 5.5. Frequencies with which "yes", "no", "unsure' and 'couldn't see" response options were given to the male culprit's beard question in the three cue conditions

than the scores reflecting the all cue (*ps* < .01, *fs* > .11) or the neutral conditions (*ps* >.01, *fs* > .18). The scores reflecting the all cue condition were also less negative than the scores reflecting the neutral condition (*ps* > .01, *fs* > .08). Thus, as in Experiment 5, relative to the fillers, the innocent suspects and the culprits were perceived as appearing more memorable when they stood out as the only members with the features of the tattoo or the beard, as opposed to when the features were replicated or eliminated across the photospreads.

Table 5.5

Means, Standard Deviations and 95% CIs for the Memorability Difference (S_{uspect}– $F_{iller\ Max}$) Score Calculated for the Photos of the Innocent Suspects and the Culprits, Reflecting the Three Cue Conditions (Solo, All and Cue Absent)

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Culprit						
Solo	-0.01	1.68	-0.26, 0.24	-0.78	1.44	-0.99, -0.57
All	-0.82	1.15	-0.99, -0.65	-1.07	1.28	-1.26, -0.88
Absent	-1.59	1.33	-1.79, -1.39	-1.30	1.52	-1.53, -1.08
	Wilks' Lambda = .51, $F(2, 175)$ = 85.46, $p < .001$, $f = .47$			Wilks' Lambda = .78, $F(2, 175)$ = 25.11, $p < .001$, $f = .28$		
Innocent Suspect						
Solo	0.01	1.66	-0.24, 0.26	-1.46	1.24	-1.64, -1.28
All	-0.81	1.08	-0.97, -0.65	-1.74	1.13	-1.91, -1.57
Absent	-1.19	1.37	-1.39, -0.99	-2.14	1.22	-2.32, -1.96
	Wilks' Lambda = .56, $F(2, 175)$ = 69.49, $p < .001$, $f = .37$			Wilks' Lambda = .73, $F(2, 175)$ = 33.17, $p < .001$, $f = .23$		

Perceived Familiarity Rating

Using the familiarity difference score, a 3 (cue: solo, all, absent) × 2 (target presence: TP, TA) factorial ANOVA analysis was conducted separately for the female and the male stimuli. The female stimuli showed a main effect of cue, $F(2, 172) = 8.14$, $p < .001$, $f = .29$. Post-hoc comparisons using the Tukey HSD tests showed that the solo cue condition produced a more negative familiarity difference score than the cue absent and the all cue conditions ($ps < .02$, $fs > .27 > .35$) (see Table 5.6). No differences were observed between the latter two cue conditions, $p > .05$, $f = .23$. The TP photospread produced a less negative familiarity difference score than the TA photospread, $F(1, 172) = 4.26$, $p = .04$, $f = .16$. No significant interaction between the cue and the target presence was observed, $F(2, 172) = 0.44$, $p = .64$, $f = .09$. The male stimuli showed no main effects of cue, $F(2, 171) = 1.56$, $p = .21$, $f = .13$, target presence, $F(1, 171) = 1.89$, $p = .16$, $f = .07$, nor interaction between the two, $F(2, 171) = 2.15$, $p = .12$, $f = .17$[32].

Further correlational analyses showed no associations between the familiarity and the memorability difference scores for the female culprit ($r = .08$), female innocent suspect ($r = .03$), male culprit ($r = .01$) and male innocent suspect ($r = .12$). Thus, similar to Experiment 5, participants' perceptions of familiarity and memorability as measured by the difference scores were unrelated.

Multiple Confidence Rating

Of the 90 subjects who viewed the female TP photospread, 36 (40%) gave a unique maximum confidence value to the female culprit, 34 (37.8%) gave it to the fillers and 20 (22.2%) gave no single unique maximum confidence value. Of the 88 subjects who viewed the female TA photospread, 12 (13.6%) gave a unique maximum

[32] For the male stimuli, for an unknown reason, the software failed to encode one subject's familiarity rating. Therefore the analysis was conducted using 177 subjects' data.

Table 5.6

Means, Standard Deviations and 95% CIs for the Effect of the Cue and the Target Presence on the Familiarity Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Female and the Male Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Target-Present						
Solo	-1.43	2.64	-2.42, -0.44	-1.63	2.44	-2.54, -0.72
All	-0.27	2.13	-1.07, 0.53	-1.47	2.10	-2.25, -0.69
Absent	0.53	2.26	-0.31, 1.37	-1.07	2.96	-2.18, 0.04
Overall	-0.39	2.46	-0.91, 0.13	-1.39	2.51	-1.92, -0.86
Target-Absent						
Solo	-1.83	2.17	-2.64, -1.02	-2.27	2.02	-3.02, -1.52
All	-0.83	2.44	-1.74, 0.08	-0.87	2.54	-1.82, 0.08
Absent	-0.73	1.36	-1.24, -0.22	-2.10	2.19	-2.92, -1.28
Overall	-1.12	2.08	-1.56, -0.68	-1.74	2.32	-2.23, -1.25
Overall						
Solo	-1.63	2.41	-2.25, -1.01	-1.95	2.24	-2.53, -1.37
All	-0.55	2.29	-1.14, 0.04	-1.17	2.33	-1.77, -0.57
Absent	-0.10	1.95	-0.60, 0.40	-1.58	2.64	-2.26, -0.90
Overall	-0.75	2.30	-1.09, -0.41	-1.56	2.42	-1.92, -1.20

confidence value to the innocent suspect, 59 (67%) gave it to the fillers and 17 (19.3%) gave no single unique maximum confidence value. Of the 90 subjects who viewed the male TP photospread, 22 (24.4%) gave a unique maximum confidence value to the male culprit, 55 (61.1%) gave it to the fillers and 13 (14.4%) gave no

single unique maximum confidence value. Of the 88 subjects who viewed the male TA photospread, 4 (4.5%) gave a unique maximum confidence value to the innocent suspect, 67 (76.1%) gave it to the fillers and 17 (19.3%) gave no single unique maximum confidence value.

Treating the confidence difference score as the dependent variable, a 3 (cue: solo, all, absent) × 2 (target presence: TP, TA) factorial ANOVA analysis was conducted separately for the female and the male stimuli[33]. The female stimuli showed a main effect of cue, $F(2,172) = 4.05$, $p = .02$, $f = .21$. Post-hoc comparisons using the Tukey HSD test showed that as predicted, the solo cue condition produced a more negative confidence difference score than the cue absent condition, $p = .02$, $f = .25$ (see Table 5.7). No differences were observed between any other combinations of the cue conditions, $ps > .05$, $f < .21$. The TP photospread produced a less negative confidence difference score than the TA photospread, $F(1,172) = 13.98$, $p < .001$, $f = .28$. No interaction between the cue and the target presence condition was observed, $F(2,172) = 0.06$, $p = .94$, $f = .03$.

In contrast, the male stimuli showed no main effect of cue, $F(1,172) = 1.10$, $p = .35$, $f = .11$. Although the TP and the TA photospreads showed different pattern regarding the effect of the cue (see Table 5.7), the interaction between the cue and the target presence condition was non-significant, $F(2,172) = 1.03$, $p = .36$, $f = .12$. A main effect of target presence was observed with the TP photospread producing a less negative confidence difference score than the TA photospread, $F(1,172) = 8.28$, $p = .005$, $f = .24$. Thus, in contrast to Experiment 5, only the results of the female stimuli supported the prediction[34].

[33] Again no presentation order effects of the mock-crime videos nor the photospreads were observed.
[34] As in Experiment 5, the INDIRECT analyses showed no evidence of the indirect effects of the memorability or the familiarity different score on the relationship between the cue and the confidence difference score.

Table 5.7

Means, Standard Deviations and 95% CIs for the Effect of the Cue and the Target Presence on the Confidence Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Female and the Male Stimuli

	Female			Male		
	M	*SD*	95% CIs	*M*	*SD*	95% CIs
Target-Present						
Solo	-13.43	52.72	-33.12, 6.26	-23.97	54.94	-44.49, -3.46
All	3.30	47.17	-14.31, 20.91	-20.50	47.59	-36.84, -1.10
Absent	7.00	52.00	-12.42, 26.42	-12.93	55.59	-33.69, 7.83
Overall	-1.04	50.91	-11.70, 9.62	-19.13	52.44	-29.64, -7.60
Target-Absent						
Solo	-42.33	40.17	-57.33, -27.33	-47.47	36.38	-60.69, -32.75
All	-23.37	49.63	-41.90, -4.84	-26.83	38.07	-41.40, -12.86
Absent	-16.07	35.52	-29.84, -2.30	-42.67	41.41	-59.36, -35.46
Overall	-27.51	43.29	-36.68, -18.34	-38.99	39.25	-48.04, -32.52
Overall						
Solo	-27.88	48.90	-40.51, -15.25	-35.72	47.69	-47.63, -22.68
All	-10.03	49.85	-22.91, 2.85	-23.67	42.84	-34.24, -12.00
Absent	-4.14	45.92	-16.21, 7.93	-27.80	50.86	-42.38, -17.38
Overall	-14.13	49.00	-21.38, -6.88	-29.06	47.25	-36.24, -22.52

Discussion

In Experiment 6, changing the cue of the right cheek scar to the facial tattoo led to the female stimuli showing the predicted effects. That is, compared to when the

tattoo was replicated or eliminated across the photospreads, making the female innocent suspect or the culprit stand out as the only photospread members with the tattoo led to them being perceived as more distinctive and memorable, whilst less familiar. Moreover, in contrast to Experiment 5, participates were most likely to say "no" to the question of the female culprit having the tattoo. Not surprisingly, compared to when the tattoo was eliminated across the photospread, when the female innocent suspect and the culprit were the only photospread members presented with the tattoo, relative to the fillers, they were perceived as resembling the culprit less. Importantly, from the TP photospread, participants correctly perceived the female culprit as the most plausible match to be the culprit when she was presented without the tattoo. However, when the culprit stood out as the only photospread member with the tattoo, the culprit was no longer perceived as the most plausible match. On the other hand, replicating the tattoo across the photospread did not reduce nor improve participants' judgements of the degree of the resemblance between the innocent suspect and the culprit, or between the culprit presented in the photospread and the participants' memory for the culprit.

As in Experiment 5, the male stimuli showed that making the innocent suspect and the culprit stand out as the only photospread members with the beard led to them being perceived as more distinctive and memorable than when the beard was replicated or eliminated across the photospreads. However, the beard did not make them appear less familiar. Moreover, in contrast to Experiment 5, a rather large proportion of participants correctly indicated not being able to see the male culprit's beard during encoding, especially when they were in the solo cue condition. Not surprisingly, the results of the confidence difference score failed to replicate the effect of the beard observed in Experiment 5.

There are at least two explanations as to why the male stimuli failed to replicate the results of Experiment 6. First, as evident by the results of the physical feature questionnaire, perhaps something about the male mock-crime video used in Experiment 6 made the participants recall that the male culprit's chin was concealed at encoding. Indeed, the proportion of participants who correctly indicated that they could not see the male culprit's beard more than doubled in Experiment 6 compared to Experiment 5. Furthermore, in the open comment section of the questionnaire, several participants correctly wrote down that the male culprit's chin was covered by a blue scarf. Therefore, perhaps, having the chin covered by the blue scarf (Experiment 6) as opposed to a newspaper (Experiment 5) led to more attention being drawn to the male culprit's chin, making the concealment overly memorable. To test whether participants' ability to recall the concealment of the culprit's chin underpinned the non-significant results in Experiment 6, the male stimuli's confidence difference score data were reanalysed only using the subjects' data who responded "no" to the beard question. Using 82 subjects' data, the results showed no main effect of cue on the confidence difference score, $F(2, 76) = 2.80$, $p = .07$, $f = .29$[35].

The explanation provided for the non-significant results obtained for the male stimuli (Experiment 6) might sound rather contradictory to the earlier explanations provided for the non-significant results obtained for the female stimuli (Experiment 5). However, there is one clear difference between the two accounts put forward. For the female stimuli (Experiment 5), I predicted that perhaps the non-significant results were caused by the fact that the cue was not salient enough cue, possibly leading to the presence of the cue being dismissed. In contrast, I propose that for the male

[35] As with Experiment 5, analysing the data separately for those who said "yes", "unsure" or "couldn't see" did not provide any significant results.

stimuli (Experiment 6), the non-significant results could have arisen due to the concealment of the cue (not the cue itself) being rather salient, possibly leading to the presence of the cue being accepted as a feature that could have belonged to the culprit. Therefore, the former proposes the need for the cue itself to be salient, whilst the latter proposes the need for the concealment of the cue to be non-salient.

Second, the non-significant results found for the male stimuli might have been caused by the fact that for some reason, regardless of the presence or the absence of the beard, relative to the fillers, participants perceived the male culprit and the innocent suspect as resembling the culprit rather poorly. Indeed, even in the absence of the beard, both the male culprit and the innocent suspect tended to produced more negative confidence difference scores in Experiments 6 than 5. Thus, the beard might have failed to influence participants' perceptions of the innocent suspect and the culprit where even when they were presented without the beard, they were perceived as unlikely match to be the culprit. This might indicate that, perhaps, an unexpected presence of a physical feature is most likely to affect witnesses' confidence assessment when the feature is displayed by a photospread member, who would not be readily rejected as being the culprit.

Overall Discussion

Together, the results of Experiments 5 and 6 suggest that, when a member of a photospread stands out as the only member with a physical feature, which was unseen on the culprit at the time of the encoding, the feature could reduce the witness' judgement of the resemblance between the member and the culprit. However, the results of Experiments 5 and 6 also suggest that, unexpected presence of a previously concealed feature on one photospread member's face would not always lead to a witness judging the member with the feature as an unlikely match to be the culprit.

Two aspects of these results merit comment. First, some evidence suggest that a witness may be unaffected by the presentation of an unexpected feature on a member of a photospread when the witness is able to consider the possibility that the feature could not have been seen on the culprit at the time of encoding. For example, as indicated by the results of the male stimuli in Experiment 6, when the concealment of a culprit's face is rather clear, a witness might be able to recall the concealment, allowing the witness to consider the possibility that the culprit could have had the feature. Alternatively, the results of the female stimuli in Experiment 5 suggest that when a witness is able to consider the possibility that an unexpected feature presented in a photospread could have been displayed by the culprit at encoding, but the feature was forgotten, the feature is unlikely to influence the witness' confidence assessment. Thus, in order for a witness to perceive a face with an unexpected physical feature as an unlikely match for the culprit, perhaps the feature needs to cause the witness to feel that the previous encounter with the feature could not have been forgotten. This is consistent with what is suggested by the theory of memory for nonoccurrences (Förster & Strack, 1998; Strack & Bless, 1994).

Second, unless a feature is highly salient, the presence of an unexpected physical feature may not affect witnesses' confidence assessment. For example, as shown by the contrasting results of the female stimuli in Experiments 5 and 6, when presented with features that are relatively low in salience such as a scar or a mole, witnesses might consider the features as minor cues, leading to the features being overlooked. In contrast, when presented with features that are relatively salient such as a facial tattoo or a birthmark, witnesses might consider the features as valuable cues to rely on, thereby leading to the features affecting the witnesses' confidence assessment.

Replication of a Cue Across the Photospread

The addition of the third cue condition, where the feature of the scar, beard or tattoo was replicated across the photospreads did not influence participants' confidence assessments. That is, the TP photospreads showed that replicating the features across the photospreads instead of only presenting the culprits with the features, or eliminating the features entirely, did not improve or reduce participants' perceptions of the resemblance between the culprits presented in the photospreads and their memory for the culprits. In contrast, although non-significant, the TA photospreads showed some mixed patterns regarding the effect of the all cue condition. In particular, in some instances, the same patterns of results as the TP photospreads were observed. However, in instances when the confidence difference scores given to the innocent suspects without the features were rather negative, there was a tendency for participants to perceive the innocent suspects as resembling the culprits relatively better than the fillers, when the features were replicated rather than eliminated across the photospreads. Therefore, the results might point towards the possibility that, under some circumstances, adjusting one physical feature of photospread members could change the composition of the photospread, leading to a photospread member who would have appeared as an unlikely match to be the culprit become a more plausible match. Of course, without further exploration into this matter, such claim remains speculation.

Summary

In sum, Experiments 5 and 6 provide evidence that, when one photospread member is presented with an unexpected physical feature, the photospread can become biased in a way that leads to that member not being rated as a likely culprit. Therefore, from the applied perspective, the results of Experiments 5 and 6 suggest

that, perhaps, as recommended by Wells et al. (1994), when the police have a suspect who has a distinctive feature that has not been mentioned by the witness, the suspect should not be made to stand out as the only lineup member with the feature. If the witness did not mention the feature due to forgetfulness, the feature may result in a lineup that is biased against the suspect. Alternatively, as demonstrated by Experiments 5 and 6, if the witness did not mention the feature due to a failure to encode the feature, the lineup might become biased in a way that favours the suspect, which could potentially lead to a detrimental outcome if the suspect is the culprit.

CHAPTER 6

General Discussion

Eyewitness research shows that a poorly composed lineup increases the risk of mistaken eyewitness identifications (Brigham, Ready, & Spier, 1990; Lindsay & Wells, 1980), potentially leading to detrimental consequences. This thesis reports six experiments designed to advance our knowledge in this area. In particular, I explored whether the characteristics of lineup members, such as how familiar, distinctive and memorable they appear, could bias witnesses' perceptions regarding the likelihood of a particular member being the culprit. A sense of familiarity associated with a lineup member was predicted to increase the judged resemblance of the lineup member to the culprit. Conversely, it was predicted that perceived unfamiliarity, as well as unexpected degree of distinctiveness and memorability would decrease the judged resemblance of the lineup member to the culprit.

Experiments 1 and 2 examined if a sense of familiarity aroused by an expression of a smile could lead to an innocent suspect being perceived as resembling the culprit better than the fillers. Overall, there was a tendency for participants to perceive the innocent suspects as appearing more familiar when they were presented smiling than with a neutral expression. However, this effect was not always consistent (i.e., the male stimuli in Experiment 2 failed to show this effect) and, thus, the results need to be interpreted with caution. In particular, it may be argued that it was not the smile per se but other confounding factors that attributed to the arousal of familiarity. For example, as the *warm glow* heuristic (e.g., Monin, 2003) would suggest, a smile might have caused the suspect to appear rather positive, and this positivity (as opposed to the expression per se) might have led to the suspect being perceived as

appearing rather familiar. Such consideration points towards the need for further exploration in this area of research.

Nevertheless, there was some (admittedly limited) evidence that, when the exposure duration was very brief (i.e., 2 seconds), presenting the innocent suspect with a smile rather than with a neutral expression led to an increase in participants' judgments of the resemblance between the innocent suspect and the culprit. However, the effect of the smile was not robust enough to make the innocent suspect stand out from the photospread as resembling the culprit the most. Together, the results indicate that an expression of a smile displayed by a photospread member will not always influence a witness's judgment regarding the degree of resemblance between a photospread member and a culprit. This is consistent with the notion that multiple factors contribute to a witness' identification decision and, thus, one particular factor will not always influence eyewitness identification decisions the same way (Brewer, Weber, & Semmler, 2005).

Using morphing software to create the expression of a smile, Experiment 3 unexpectedly showed that the smiling innocent suspects appeared unfamiliar rather than familiar, leading to participants perceiving the innocent suspects as resembling the culprits less than the fillers. In Experiment 4, the TP photospreads also showed similar results. That is, when the culprits were presented in the photospreads smiling rather than with a neutral expression, participants perceived the culprits as less likely to be the culprits than the fillers. Furthermore, Experiment 4 demonstrated that, in addition to making the innocent suspects and the culprits appear unfamiliar, the smile made them appear distinctive and memorable. Thus, consistent with the theory of memory for nonoccurrences (Förster & Strack, 1998; Strack & Bless, 1994), perhaps the smile made the innocent suspects and the culprits appear unexpectedly distinctive

and, thus memorable, leading to an assumption that, if seen previously, they should have cued clear recollections. It was thought that these perceptions of (unexpected) distinctiveness and memorability could be associated with participants perceiving the smiling innocent suspects and the culprits as appearing unfamiliar. Interestingly though, when exploring the relationships between participants' judgments of face distinctiveness, memorability and familiarity, inconsistent results were observed between the stimuli. That is, whilst the scores of face distinctiveness and memorability were consistently shown to correlate positively, only the female innocent suspect's familiarity score correlated negatively with the combined score of the distinctiveness and memorability. The mixed results likely reflect the complex nature of the relationships between face distinctiveness, memorability and familiarity, and their effects on recognition performance (Hosie & Milne, 1995). This issue will be discussed later in further detail.

Experiments 5 and 6 further explored the roles of face distinctiveness, memorability and familiarity by presenting photospread members with physical features such as facial scar, tattoo and beard, which were not visible on the culprits at encoding. As predicted, participants perceived the innocent suspects and the culprits as appearing more distinctive and memorable and, in most cases, less familiar when they were presented with the features than without. Moreover, compared to when the features of beard or tattoo were eliminated from the photospreads, when the innocent suspects and the culprits stood out as the only photospread members with the features, relative to the fillers, both the innocent suspects and the culprits were perceived as less likely to be the culprits. However, in instances where participants felt that the culprits could have had such features due to the features being relatively minor or due to the concealment of such features at encoding being rather clear, the participants'

judgments of the degree of resemblance between the individuals (i.e., the innocent suspects and the culprits) displaying the features and the culprits were unaffected by the features. The results likely reflect that a mere increase in perceived distinctiveness of a face between encoding and test would not always increase the likelihood of the face being perceived as never been seen before. In particular, when an increase in perceived distinctiveness and, thus, memorability of a previously seen face can be justified (e.g., "I failed to see the scar on the face"), people are able to consider the possibility that the distinctive face could have been seen before. Rather, the likelihood of a face being perceived as not being seen before increases only when the face appears unexpectedly distinctive and, thus, memorable (e.g., "the face I saw definitely did not have a scar"). Thus, as suggested by the theory of memory for nonoccurrences (Förster & Strack, 1998; Strack & Bless, 1994), a sense of unexpectedness likely play a key role in determining the newness of a face.

In sum, Experiments 1-6 provide a rather complex picture regarding how an eyewitness' memory processing can be influenced by a potentially biasing cue that is linked to the demeanour of the members of a lineup. In particular, depending on the circumstances, a smile presented by a suspect within a lineup can cue a sense of familiarity (Experiments 1 and 2), unless the smile is seen to be particularly distinctive and memorable (Experiments 3 and 4). The increased sense of familiarity may not always be associated with an increased confidence assessment (Experiments 1 and 2). However, a decreased sense of familiarity is likely to be associated with a decreased confidence assessment (i.e., perceived as unlikely match to be the culprit) (Experiments 3, 4, 5 and 6). In addition, the presence of a distinctive and, thus, memorable physical feature presented by a lineup member is likely to decrease the confidence assessment given to that member (i.e., unlikely match to be the culprit)

only when (a) the lineup member is the only member with the feature and (b) witnesses believe that the feature could not have been present on the culprit at the time of the crime (Experiments 5 and 6).

Metacognition and Memory Processing

Together, Experiments 1-6 support the notion that memory processing can be influenced by metacognitive judgments that are based on subjective feelings (e.g., familiarity) as well as subjective beliefs regarding one's own memory processing (e.g., unexpected memorability) (Koriat & Levy-Sadot, 2000). Importantly, such metacognitive knowledge can, under some conditions, lead witnesses to make false judgments as to whether a face has been seen before or not. In particular, arousal of a sense of familiarity can lead witnesses to falsely perceive an innocent suspect as resembling the culprit relatively better than the fillers. Conversely, when a culprit presented in a photospread appears unfamiliar or unexpectedly distinctive and, thus, memorable, such perceptions might lead witnesses to perceive the culprit as an unlikely match to be the culprit. Together, these results add to the literature investigating the effects of metacognitive factors on memory performance (Förster & Strack, 1998; Koriat, 2006; Koriat & Levy-Sadot, 2000; Strack & Bless, 1994; Whittlesea, 1993, 2002), and how such factors might, under some circumstances, contribute towards erroneous identification decisions being made by eyewitnesses.

In saying so, the use of the multiple confidence procedure rather than the standard identification procedure means that the results presented thus far need to be interpreted with some caution. As discussed in Chapter 2, the multiple confidence procedure was chosen for it was thought to provide a more sensitive measure of the effects of the potentially biasing cues, which in any given context may or may not affect a witness' identification decision. For example, a smile was thought to make a

lineup member be perceived as resembling the culprit more than the fillers. However, it was predicted that the smile would not necessarily lead to a positive identification of the member, unless the witness felt that there was enough evidence to do so. Similarly, it was predicted that, cues such as an odd smile and various physical features would make a lineup member perceived as less likely to be the culprit than their fillers, yet such potentially basing cues would not affect witnesses' identification decisions, unless the witnesses felt that there was enough evidence to reject the lineup member. Thus, the effects of the potentially basing cues were tested by measuring the likelihood of a particular innocent suspect or a culprit being chosen from a photospread, instead of measuring participants' actual choosing behaviours. Therefore, without further investigations, the current findings cannot provide any definitive conclusions as to under what conditions such biasing cues would influence a witness' identification decision and under what conditions they would not.

However, as stated in Chapter 2, the advantage of the multiple confidence procedure was that it could detect potentially important effects that may sometimes, even if occasionally, be important in a lineup decision and yet be undetected if the standard identification procedure was used. Thus, by using the multiple confidence procedure, I provided evidence that factors that make a lineup member appear distinctive, memorable or unfamiliar (or familiar) could potentially bias a witness' perception regarding the likelihood of the lineup member being the culprit, which could under some circumstances function as key evidence for a witness to reject or identify the lineup member as being the culprit. Some practical implications regarding the current findings will be discussed later.

Facial Characteristics: Distinctiveness, Memorability & Familiarity

The focus of this set of studies was not the nature of the relationships between the facial characteristics of distinctiveness, memorability and familiarity. However, given the complex nature of the relationships described in Chapter 4, the topic deserves attention. In the facial recognition literature, face typicality and distinctiveness are thought to lie on a continuum, with typicality at the one end of the continuum and distinctiveness at the other end. Vokey and Read (1992) suggest that face distinctiveness/typicality consists of two orthogonal components: context-free familiarity and memorability. Context-free familiarity is defined as a sense of familiarity without indexing of the source of the memory (Busey, 2001). According to this view, at the one end of the distinctiveness/typicality continuum, distinctive faces are perceived as appearing highly memorable, whilst generating a low level of familiarity, perhaps leading to high hit and low false alarm rates. At the other end of the continuum, typical faces are perceived as appearing highly familiar and perhaps unmemorable, leading to high false alarm rates (Hosie & Milne, 1995). However, as outlined in the following sections, the relationships between face distinctiveness, memorability and familiarity are likely to be far more complex than the view portrayed by Vokey and Read (1992).

When measuring the effects of face distinctiveness, memorability and familiarity in facial recognition studies, participants are generally provided with single item scales of distinctiveness, memorability and familiarity, accompanied by definitions derived from theoretical frameworks, and they are asked to rate a series of faces on such characteristics. However, this methodology may give rise to several issues.

First, a gap might exist between how a term such as face distinctiveness is conceptualized in existing theory and how individual participants generally conceptualize the terms. For example, Valentine's (1999) face-space model conceptualizes distinctive faces as those that fall away from the central tendency of the general population. That is, the model suggests that faces can be represented on a multi-dimensional space, where faces are normally distributed along each dimension. Typical faces are thought to be located around the central tendency of the face-space, where it is densely populated (i.e., many faces look similar to one another and, thus, they are spaced closely together). On the other hand, distinctive faces are thought to be located away from the central tendency and, thus, away from other faces (i.e., dissimilar to other faces). Thus, researchers often consider a typical face to be a statistically average face. In contrast to this view, Wickham, Morris, and Fritz (2000) found that, when participants were asked to rate 88 faces regarding their perceived distinctiveness using a typicality-distinctiveness continuous scale and a commonly used definition that a distinctive face is a face that would stand out in a crowd of typical faces, their ratings were normally distributed. That is, instead of perceiving the typical end of the scale as the average or the norm (which the face-space model would suggest), participants tended to pick the mid-point of the typicality-distinctiveness scale as the average point. On the other hand, when participants were asked to rate the same 88 faces using a less commonly used definition of a distinctive face as a face that would deviate from appearing average or typical, the ratings were skewed with most of the faces being rated close to the typical end of the scale. Thus, whilst face distinctiveness is theoretically conceptualized as a face that would deviate from the typical/ average pool of faces, depending on how the term distinctiveness is defined, participants' rating of face distinctiveness might not always share the same view.

Indeed, research shows that depending on which definition of face distinctiveness is used, participants' distinctiveness ratings given to the same face could change, leading to different conclusions being drawn about the relationships between face distinctiveness, memorability and familiarity (Morris & Wickham, 2001). For example, as discussed in Chapter 4, Morris and Wickham (2001) demonstrated that the rating of face distinctiveness loaded negatively with the rating of familiarity when face distinctiveness was defined as a face that would deviate from the norm, but failed to do so when it was defined as a face that would stand out in a crowd. In my thesis, considering that I emphasized the "stand out in a crowd" concept when defining the term face distinctiveness, it is not surprising that, in general, the familiarity variable failed to relate to the distinctiveness and memorability variables in Experiments 4-6. Therefore, the results of Experiments 4-6 might reflect the choice of the distinctiveness definition that I used, not necessarily the nature of the relationships between face familiarity, distinctiveness and memorability.

Second, as Burton and Vokey (1998) suggest, the definition of face distinctiveness might consist of some composite of the multiple dimensions on which a faces varies and, thus, the use of a single item scale to measure characteristics such as face distinctiveness might not provide an accurate view of the relationships between face distinctiveness, memorability and familiarity. In line with this view, Bruce, Burton, and Dench (1994) suggest that, whilst face distinctiveness can be defined as a deviation from the norm, a face can deviate from the norm in various ways and, thus, a degree of distance from the norm might not always lead to a face also appearing unfamiliar and memorable. That is, a face that is deviant in some regard might be seen as appearing unfamiliar but not distinctive or memorable, whilst a face that is deviant in other regard might be seen as appearing distinctive and

memorable but not unfamiliar. Thus, the three perceptions might not always be related to one another. However, if a face deviates in a particular way that taps into the perceptions of distinctiveness, memorability and unfamiliarity, the relationships might be observed between the three characteristics. Bruce et al. (1994) attempted to capture the different aspects of the three characteristics by asking participants to provide distinctiveness ratings and memory performance to 175 photographed faces, then later comparing the results to the 175 faces' actual measures of various facial features (e.g., forehead height, pointiness of chin). However, the results were inconclusive, perhaps, further reflecting the complexity of the nature of characteristics of face distinctiveness, memorability and familiarity.

Third, as Hosie and Milne (1995) suggest, when making face distinctiveness judgments, participants could engage in different types of judgment strategies, which could affect how faces are rated. In particular, it is thought that participants may make either local or global judgments (Hosie & Milne, 1995). For example, when asked to rate the distinctiveness of a face, a person might judge a face to be highly distinctive in relation to the other faces presented in the stimulus set (i.e., local judgment) or in relation to their experiences of faces in general (i.e., global judgment). Consequently, depending on the judgment strategies used, the relationships between judgments of face distinctiveness, memorability and familiarity might differ. For example, when using the local judgment strategy, a participant might rate a face as appearing highly distinctive due to its distinctiveness when compared to the other faces presented in the experiment. Yet, the participant might not perceive the face as appearing unfamiliar due to it reminding the participant of someone he knows. In support of this view, Wickham et al. (2000) have indicated that, in their studies, some participants reported rating a face as appearing highly distinctive, despite the fact that they felt that the face

closely resembled someone they knew. Conversely, when using the global judgment strategy, a face would be judged as appearing highly distinctive only if the face looks dissimilar in comparison to the other faces seen in the past and, thus, the face would likely be perceived as appearing unfamiliar rather than familiar. Therefore, unless explicitly told to make local or global judgments about a face, the use of subjective ratings of face distinctiveness, memorability and familiarity might not provide the whole picture regarding the relationships between the three characteristics.

Future Research Directions

Together, the existing research and the findings presented in this thesis suggest that there is need for further investigation regarding the basis upon which participants make judgments such as face distinctiveness, memorability and familiarity. For example, although face distinctiveness has been theoretically defined as a face that deviates from the norm, as suggested by the existing research, the ways in which people generally conceptualize the term face distinctiveness might differ, leading to inconsistent results being obtained across different studies. Thorough investigations into how people conceptualize the terms face distinctiveness, memorability and familiarity could potentially lead to developments of psychological measures that could more accurately capture such definitions. In particular, instead of consisting of single item scales as often used in the existing research, such measures could consist of multiple items to capture the multi-dimensionality and, thus, complexity of such constructs. The measures could then, perhaps, provide us with in-depth understanding of people's metacognitions and, in particular, the mechanisms underlying the effects of such characteristics on people's memory performance. The development of such measures could then be used to further advance our knowledge as to how the combined effects of such facial characteristics of the culprit and other lineup members

might influence a witness' identification decision. By doing so, a deeper insight could be gained into how eyewitnesses' metacognitions impact the accuracy of their identification decisions.

Practical Implications

As suggested earlier, some inconsistent results were obtained throughout the thesis. Therefore, without further investigation, implications regarding these findings remain speculative. Nevertheless, the results of this thesis have several important practical implications. In line with existing recommendations regarding the composition of a lineup (Wells & Seelau, 1995), the results clearly demonstrate that any potential biasing cues, should be eliminated from a lineup. As the results of my research demonstrate, this would include ensuring that a suspect does not stand out from a lineup due to his behavioural presentation, such as due to having a distinctively different facial expression from the other lineup members. In contrast with those recommendations, a survey of 220 US police jurisdictions (Wogalter, Malpass, & McQuiston, 2004) showed that an alarming majority (94%) of the respondents reported that when composing a lineup they relied on their own subjective judgments to determine the fairness of the lineup. Moreover, none of the respondents mentioned the importance of standardizing the facial expressions of the members of a lineup. Thus, assuming future investigations produce findings consistent with those reported here, future recommendations could emphasize the importance of standardizing the general appearances of the members of a lineup, but also in their behavioural presentations, such as their facial expressions as well as other presentations that could potentially make the lineup members stand out as appearing familiar, distinctive or memorable.

In regards to when the police have a suspect who has a distinctive feature that has not been mentioned by the witness, the results suggest that it is best not to present the suspect as the only lineup member with the feature. Yet, the survey showed that when composing a lineup, 30% of the US police officers reported doing nothing when the suspect had a distinctive facial marking (Wogalter et al., 2004). Furthermore, currently, there is no standard procedure used in such situations amongst UK or US police forces (Zarkadi, Wade, & Stewart, 2009). As a potential solution, the US Department of Justice recommends warning eyewitnesses that the culprit's appearance might have changed since the time of the crime. However, the use of the warning instruction has been shown to inflate false identifications, whilst having no impact on correct identifications of the culprits (Charman & Wells, 2007; Molinaro, Arndorfer, & Charman, 2013). Thus, further investigation in this area is vital for it could lead to a development of a standardized procedure, which enables eyewitnesses to consider the possibility that the culprit's appearance might have changed over time without increasing the risk of erroneous identifications.

In sum, by deepening our understanding about potentially biasing influences infect the composition of a lineup, this thesis provides a useful starting point for understanding the role of such biases in the production of identification errors. In saying so, considering the inconsistent results obtained throughout the thesis, it is vital that further research is conducted in this area before any solid conclusions are drawn.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*, 203-231.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431-441. doi: 10.1037/0022-3514.64.3.431

Bartlett, J., Hurry, S., & Thorley, W. (1984). Typicality and familiarity of faces. *Memory & Cognition, 12*, 219-228. doi: 10.3758/BF03197669

Baudouin, J.-Y., Gilibert, D., Sansone, S., & Tiberghien, G. (2000). When the smile is a cue to familiarity. *Memory, 8*, 285-292. doi: 10.1080/09658210050117717

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S. L., & Neel, R. (2011). The face in the crowd effect unconfounded: Happy faces, not angry faces, are more efficiently detected in single- and multiple-target visual search tasks. *Journal of Experimental Psychology: General, 140*, 637-659. doi: 10.1037/a0024060

Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness identification. . In N. Brewer & K. Williams, D (Eds.), *Psychology and law: An empirical perspective*. New York: The Guilford Press.

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science, 23*, 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science, 20*, 24-27. doi: 10.1177/0963721410389169

Brigham, J. C., Ready, D. J., & Spier, S. A. (1990). Standards for evaluating the fairness of photograph lineups. *Basic and Applied Social Psychology, 11*, 149-163. doi: 10.1207/s15324834basp1102_3

Bruce, V., Burton, M. A., & Dench, N. (1994). What's distinctive about a distinctive face? *The Quarterly Journal of Experimental Psychology Section A, 47*, 119-141. doi: 10.1080/14640749408401146

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*, 339-360. doi: 10.1037/1076-898X.5.4.339

Buckhout, R. (1975). Eyewitness testimony. *Scientific American, 231*, 23-31.

Buckhout, R., Figueroa, D., & Hoff, E. (1975). Eyewitness identification: Effects of suggestion and bias in identification from photographs. *Bulletin of the Psychonomic Society, 6*, 71-74.

Burton, A. M., & Vokey, J. R. (1998). The Face-Space Typicality Paradox: Understanding the Face-Space Metaphor. *The Quarterly Journal of Experimental Psychology Section A, 51*, 475-483. doi: 10.1080/713755768

Busey, T. A. (2001). Formal models of familiarity and memorability in face recognition. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Context and challenges. Scientific psychology series* (pp. 147 - 191). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Charman, S. D., & Wells, G. L. (2007). Eyewitness lineups: Appearance-change instruction a good idea? *Law and Human Behavior, 31*, 3-22.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences, 2*, 206-214. doi: http://dx.doi.org/10.1016/S1364-6613(98)01179-6

Courtois, M. R., & Mueller, J. H. (1981). Target and distractor typicality in facial recognition? *Journal of Applied Psychology, 66*, 639-645. doi: 10.1037/0021-9010.66.5.639

Dobel, C., Geiger, L., Bruchmann, M., Putsche, C., Schweinberger, S., & Junghöfer, M. (2008). On the interplay between familiarity and emotional expression in face perception. *Psychological Research, 72*, 580-586. doi: 10.1007/s00426-007-0132-4

Downs, A. C., & Lyons, P. M. (1991). Natural observations of the links between attractiveness and initial legal judgments. *Personality and Social Psychology Bulletin, 17*, 541-547. doi: 10.1177/0146167291175009

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

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19*, 151-164. doi: 10.1037/a0030618

Flowe, H. D., & Humphries, J. E. (2011). An examination of ncriminal faces in a random sample of police lineups. *Applied Cognitive Psychology, 25*, 265 - 273. doi: 10.1002/acp.1673

Flowe, H, D,m Klatt, T., & Colloff, M. F. (2014). Selecting fillers on emotional appearance improves lineup identification accuracy. *Law and Human Behavior, 38*, 509-519. doi: 0147-7307/14/$12.00

Förster, J., & Strack, F. (1998). Subjective theories about encoding may influence recognition: Judgmental regulation in human memory. *Social Cognition, 16*, 78-92. doi: 10.1521/soco.1998.16.1.78

Garcia-Marques, T., Mackie, D. M., Claypool, H. M., & Garcia-Marques, L. (2004). Positivity can cue familiarity. *Personality and Social Psychology Bulletin, 30*, 585-593. doi: 10.1177/0146167203262856

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669. doi: 10.1037/0033-295X.103.4.650

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506-528. doi: 10.1037/0033-295X.98.4.506

Going, M., & Read, J., Don. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills, 39*, 109-110.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*, 408-420. doi: 10.1080/03637750903310360

Hayes, A. F., & Preacher, K. J. (2013). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67, 451-470. doi: 10.1111/bmsp.12028

Heath, W. P., Grannemann, B. D., & Peacock, M. A. (2004). How the defendant's emotion level affects mock jurors' decisions when presentation mode and

evidence strength are varied. *Journal of Applied Social Psychology, 34*, 624-664. doi: 10.1111/j.1559-1816.2004.tb02563.x

Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law & Human Behavior, 38*, 94-108. doi: 10.1037/lhb0000060

Hosie, J. A., & Milne, A. B. (1995). Distinctiveness and memory for unfamiliar faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition: Exploration in face space* (pp. 95 - 112). London: Routledge.

Huff, C. R., Rattner, A., Sagarin, E., & MacNamara, D. E. J. (1986). Guilty until proved innocent: Wrongful conviction and public policy. *Crime & Delinquency, 32*, 518-544. doi: 10.1177/0011128786032004007

Innocence Project. (2014). 2014, from http://www.innocenceproject.org/

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 180-191.

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 179-187.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306-340. doi: 10.1037/0096-3445.110.3.306

Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 3-11.

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 210-223. doi: 10.1037/0278-7393.17.2.210

Juth, P., Lundqvist, D., Karlsson, A., & Öhman, A. (2005). Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. *Emotion, 5*, 379-395. doi: 10.1037/1528-3542.5.4.379

Kassin, S. M. (1983). Deposition testimony and the surrogate witness: Evidence for a "messenger effect" in persuasion. *Personality and Social Psychology Bulletin, 9*, 281-288. doi: 10.1177/0146167283092013

Knapp, B., Nosofsky, R., & Busey, T. (2006). Recognizing distinctive faces: A hybrid-similarity exemplar model account. *Memory & Cognition, 34*, 877-889. doi: 10.3758/BF03193434

Koriat, A. (2006). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289 - 326). New York: Cambridge University Press.

Koriat, A., & Levy-Sadot, R. (2000). Conscious and unconscious metacognition: A rejoinder. *Consciousness and Cognition, 9*, 193-202. doi: http://dx.doi.org/10.1006/ccog.2000.0436

Kraft, R. (1987). The influence of camera angle on comprehension and retention of pictorial events. *Memory & Cognition, 15*, 291-307. doi: 10.3758/BF03197032

Lander, K., & Metcalfe, S. (2007). The influence of positive and negative facial expressions on face familiarity. *Memory, 15*, 63-69. doi: 10.1080/09658210601108732

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research, 69*, 22-29. doi: http://dx.doi.org/10.1007/s00426-003-0157-2

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 212-228. doi: 10.1037/0278-7393.5.3.212

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science, 9*, 215-218. doi: 10.1111/1467-9280.00041

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? *Law and Human Behavior, 4*, 303-313. doi: 10.1007/BF01040622

Loftus, E. (1974). Reconstructing memory: The incredible eyewitness. *Psychology Today, 8*, 116-119.

Lombroso, C. (1911). *Criminal man: According to the classification of Cesare Lombroso*. New York: The Knickerbocker Press.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99-128. doi: 10.1207/s15327906mbr3901_4

MacLin, M. K., & Herrera, V. (2006). The criminal stereotype. *North American Journal of Psychology, 8*, 197-207.

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law & Human Behavior, 5*, 299-309. doi: 10.1007/BF01044945

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482-489. doi: 10.1037/0021-9010.66.4.482

Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring line-up fairness. *Applied Cognitive Psychology, 13*, S1-S7.

Malpass, R. S., Tredoux, C. G., & McQuiston, D. (2007). Lineup construction and lineup fairness. In R. C. L. Lindsay, D. F. Ross, R. J. D & T. M. P (Eds.), *Handbook of eyewitness psychology: Memory for people* (Vol. 2, pp. 155-178). United States of America: Lawrence Erlbaum Associates, Inc.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339-354. doi: 10.1348/000712603767876262

Molinaro, P. F., Arndorfer, A., & Charman, S. D. (2013). Appearance-change instruction effects on eyewitness lineup identification accuracy are not moderated by amount of appearance change. *Law and Human Behavior, 37,* 432-440. doi: 10.1037/lhb0000049

Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality and Social Psychology, 85*, 1035-1048. doi: 10.1037/0022-3514.85.6.1035

Morris, P. E., & Wickham, L. H. V. (2001). Typicality and face recognition: A critical re-evaluation of the two factor theory. *The Quarterly Journal of Experimental Psychology Section A, 54,* 863-877. doi: 10.1080/713755992

O'Toole, A., Deffenbacher, K., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22*, 208-224. doi: 10.3758/BF03208892

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica, 133*, 137-145. doi: http://dx.doi.org/10.1016/j.actpsy.2009.11.002

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, & Law Special Theme: Witness Memory and Law, 1*, 817-845.

Police Executive Research Forum, (2013). *A national survey of eyewitness identification procedures in law enforcement agencies* (Research Report). Retrieved from https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=264692

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*, 119-134. doi: 10.1348/135532508X281520

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*, 477-491. doi: 10.1080/10683160902926141

Preacher, K., & Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891. doi: 10.3758/BRM.40.3.879

Rafter, N. (2007). Somatotyping, antimodernism, and the production of criminological knowledge. *Criminology, 45*, 805-833. doi: 10.1111/j.1745-9125.2007.00092.x

Saladin, M., Saper, Z., & Breen, L. (1988). Perceived attractiveness and attributions of criminality: What is beautiful is not criminal. *Canadian Journal of Criminology, 30*, 251-259.

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General, 137*, 528-547. doi: 10.1037/a0012712

Sauer, J. D., Brewer, N., & Weber, N. (2012). Using confidence ratings to identify a target among foils. *Journal of Applied Research in Memory and Cognition, 1*, 80-88. doi: http://dx.doi.org/10.1016/j.jarmac.2012.03.003

Semmler, C., & Brewer, N. (2006). Postidentification feedback effects on face recognition confidence: evidence for metacognitive influences. *Applied Cognitive Psychology, 20*, 895-916. doi: 10.1002/acp.1238

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139-156. doi: 10.1037/0033-2909.100.2.139

Sheldon, W. H., Stevens, S. S., & Tucker, W. B. (1940). *The varieties of human physique*. Oxford, England: Harper.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology, 85*, 542-550. doi: 10.1037/0021-9010.85.4.542

Steblay, N. K., Dysart, J. E., Fulero, S. M., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law & Human Behavior, 25*, 459-473. doi: 10.1023/A:1012888715007

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33,* 203-217. doi: 10.1006/jmla.1994.1010

Strack, F., & Förster, J. (1998). Self-Reflection and recognition: The role of metacognitive knowledge in the attribution of recollective experience. *Personality and Social Psychology Review, 2,* 111-123. doi: 10.1207/s15327957pspr0202_4

Strack, F., Förster, J., & Werth, L. (2005). "Know thyself!" The role of idiosyncratic self-knowledge in recognition memory. *Journal of Memory and Language, 52,* 628-638. doi: http://dx.doi.org/10.1016/j.jml.2005.02.005

Tredoux, C. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology, 13,* S9-S26.

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied, 9,* 101-118.

Valentine, T. (1999). face-space models of face recognition. In M. Wenger, J & J. Townsend, T (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Vokey, J., R, & Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition, 20,* 291-302. doi: 10.3758/BF03199666

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36,* 1546-1557. doi: 10.1037/0022-3514.36.12.1546

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295. doi:10.1146/annurev.psych.54.101601.145028

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, and Law, 1*, 765-791. doi: 10.1037/1076-8971.1.4.765

Wells, G. L., Seelau, E. P., Rydell, S. M., & Luus, C. A. E. (1994). Recommendations for properly conducted lineup identification tasks. In D. Ross, F, J. Read, D. & M. Toglia, P (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 223-244). Cambridge: Cambridge University Press.

Wells, G. L., Small, M., Penrod, S. D., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads (Vol. 22, pp. 603-647). Germany: Springer.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus Sampling and Social Psychological Experimentation. *Personality and Social Psychology Bulletin, 25*, 1115-1125. doi: 10.1177/01461672992512005

Wessel, E., Drevland, G. C. B., Eilertsen, D. E., & Magnussen, S. (2006). Credibility of the emotional witness. *Law & Human Behavior, 30*, 221-230. doi: 10.1007/s10979-006-9024-1

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235-1253. doi: 10.1037/0278-7393.19.6.1235

Whittlesea, B. W. A. (2002). False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General, 131*, 96-115. doi: 10.1037/0096-3445.131.1.96

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29*, 716-732. doi: http://dx.doi.org/10.1016/0749-596X(90)90045-2

Wickham, L. H. V., Morris, P. E., & Fritz, C. O. (2000). Facial distinctiveness: Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology, 91*, 99-123. doi: 10.1348/000712600161709

Williams, M., Moss, S., Bradshaw, J., & Mattingley, J. (2005). Look at me, I'm smiling: Visual search for threatening and nonthreatening facial expressions. *Visual Cognition, 12*, 29-50. doi: 10.1080/13506280444000193

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592-598. doi: 10.1111/j.1467-9280.2006.01750.x

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime & Law, 10*, 69-82. doi: 10.1080/10683160410001641873

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517. doi: http://dx.doi.org/10.1006/jmla.2002.2864

Zarkadi, T., Wade, K. A., & Stewart, N. (2009). Creating fair lineups for suspects with distinctive features. *Psychological Science, 20,* 1448-1453. doi: 10.1111/j.1467-9280.2009.02463.x

Appendix A

Table A.1

Experiment 2: Summary of Means, Standard Deviations and 95% CIs for the Effects of the Cue and the Encoding Condition (Short/ Full vs. Long/ Divided) on the Confidence Difference (S_{uspect}–$F_{iller\ Max}$) Score for the Two Stimuli

	Female			Male		
	M	*SD*	95% CI	*M*	*SD*	95% CI
Short/ Full						
Smile	-30.10	36.74	-44.08, -16.13	-21.89	44.64	-39.20, -4.58
Neutral	-30.65	41.31	-45.06, -16.24	-26.75	46.88	-44.93, -8.57
Overall	-30.40	38.96	-40.21, -20.59	-24.32	45.42	-36.48, -12.16
Long/ Divided						
Smile	-34.50	38.80	-48.99, -20.01	-37.21	35.76	-49.69, -24.73
Neutral	-39.62	32.69	-42.06, -17.19	-20.33	42.69	-35.47, -5.19
Overall	-37.02	35.71	-46.33, -27.71	-28.90	39.94	-38.64, -19.16
Overall						
Smile	-32.34	37.54	-42.12, -22.56	-30.29	40.41	-40.55, -20.03
Neutral	-34.78	37.57	-44.24, -25.32	-23.28	44.40	-34.65, -11.91
Overall	-33.60	37.42	-40.31, -26.89	-26.81	42.41	-34.38, -19.24

Appendix B

Example of the questionnaire for the Viewing Condition Pilot Test (Chapter 5)

Task 1: Culprits

a) How confident (0-100%) you are that the culprit in the video had the distinctive feature presented in the image.

b) How atypical/distinctive the face appears in general.
(Scale: 1 = very typical, 7 = very atypical/distinctive)

c) How familiar the face appears.
(Scale: 1 = not at all familiar, 7 = very familiar)

Female Culprit

(i) Right cheek scar

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(ii) Mole on left-side of mouth

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(iii) Tattoo on face

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(iv) Scar on forehead

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(v) Nose piercing

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(vi) Nothing

Confidence_____% Atypical/ Distinctive_____ Familiar_____

Male Culprit

(i) Scar on forehead

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(ii) Nose-bridge piercing

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(iii) Mole next to nose

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(iv) Beard

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(v) Moustache

Confidence_____% Atypical/ Distinctive_____ Familiar_____

(vi) Nothing

Confidence_____% Atypical/ Distinctive_____ Familiar_____

Task 2: Crime Scenes

Female mock-crime video		Male mock-crime video	
Image 1:	Yes/No	Image 1:	Yes/No
Image 2:	Yes/No	Image 2:	Yes/No
Image 3:	Yes/No	Image 3:	Yes/No
Image 4:	Yes/No	Image 4:	Yes/No
Image 5:	Yes/No	Image 5:	Yes/No
Image 6:	Yes/No	Image 6:	Yes/No
Image 7:	Yes/No	Image 7:	Yes/No
Image 8:	Yes/No	Image 8:	Yes/No
Image 9:	Yes/No	Image 9:	Yes/No

Appendix C

Example of the manipulation check questionnaire for Experiments 5 and 6

Please think back to the culprits you saw in the mock-crime videos. Whilst thinking back to the culprits you have witnessed in the videos, please complete this questionnaires by **circling one of the four response options** provided (i.e., Yes, No, Unsure, Couldn't see) for each question provided on following pages. You can only circle **one** option for each question. Then for each option, please indicate how confident you are about your answer by providing a confidence rating (between 0-100%) in the allocated space below each answer option.

Example:

1. The male culprit had shaved head. Yes • No • Unsure • Couldn't see

   How confident are you about your answer: _____%

In above statement:
- If I clearly remember seeing the culprit having a shaved head, I would circle "Yes".
- If I clearly remember that he did not have a shaved head, I would circle "No".
- If I cannot recall what type of hair-style the culprit had, I would circle "Unsure".
- If for some reason, I could not see the type of hair the culprit had, I would circle "Couldn't see".

After choosing the appropriate response option, I would indicate how confident I am about the response.

If you have any questions about the questionnaire, please open your door and ask for your experimenter.

The **male culprit** in the crime video:

1. Had a beard on his chin Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

2. Had a tattoo on his forehead Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

3. Had a moustache Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

4. Had glasses Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

5. Had an eyebrows piercing Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

6. Had a lip piercing Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

Are there any other specific features, details etc. that you remember about the male culprit?

_____

_____

_____

_____

_____

_____

The **female culprit** in the crime video:

1. Had a tattoo on her forehead* Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

2. Had a mole on her nose Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

3. Had glasses Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

4. Had an nose piercing Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

5. Had a scar on her right cheek** Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

6. Had a lip piercing Yes • No • Unsure • Couldn't see
   How confident are you about your answer: _____%

Are there any other specific features, details etc. that you remember about the female culprit?

_____

_____

_____

_____

_____

_____

* In Experiment 6, the question was rephrased to "tattoo on the right cheek".

** Similarly, in Experiment 6, the question was rephrased to "scar on her forehead"

Appendix D

Table D.1

Experiment 5: Means (Standard Deviations) for the Confidence Ratings given to the Female Culprit's Right Cheek Scar Question Broken Down by the Cue Condition and the Response Options

	Cue			
	Solo	Neutral	All	Overall
Yes	60.00 (-)*	76.67 (25.17)	76.88 (18.31)	75.42 (18.76)
No	78.18 (21.47)	80.95 (22.11)	83.82 (13.41)	80.75 (19.59)
Unsure	66.05 (26.80)	74.71 (25.77)	58.06 (31.30)	66.11 (28.40)
Couldn't See	86.92 (24.28)	80.00 (19.76)	70.77 (31.48)	79.30 (25.32)
Overall	75.73 (24.90)	78.62 (22.30)	71.52 (26.95)	75.33 (24.79)

*Only one data was available in this cell.

Table D.2

Experiment 5: Means (Standard Deviations) for the Confidence Ratings given to the Male Culprit's Beard Question Broken Down by the Cue Condition and the Response Options

	Cue			
	Solo	Neutral	All	Overall
Yes	74.09 (17.72)	86.00 (11.40)	70.71 (19.79)	74.50 (18.21)
No	83.63 (14.70)	83.10 (20.70)	85.21 (13.93)	83.81 (16.89)
Unsure	68.00 (32.78)	60.42 (35.19)	68.68 (30.36)	66.30 (31.91)
Couldn't See	93.33 (5.77)	91.82 (11.89)	87.14 (12.20)	90.48 (11.17)
Overall	78.37 (21.90)	80.31 (24.63)	76.68 (22.72)	78.44 (23.02)

Appendix E

Table E.1

Experiment 6: Means (Standard Deviations) for the Confidence Ratings given to the Female Culprit's Right Cheek Tattoo Question Broken Down by the Cue Condition and the Response Options

	Cue			
	Solo	Neutral	All	Overall
Yes	60.00 (-)*	-	72.50 (41.93)	70.00 (36.74)
No	87.71 (17.57)	88.48 (14.66)	77.63 (18.51)	85.53 (17.02)
Unsure	64.44 (28.95)	77.14 (25.55)	69.47 (26.56)	69.80 (27.10)
Couldn't See	83.46 (31.05)	76.43 (24.28)	80.83 (22.75)	80.94 (26.04)
Overall	78.75 (26.68)	83.98 (19.72)	75.09 (24.17)	79.27 (23.87)

*Only one data was available in this cell.

Table E.2

Experiment 6: Means (Standard Deviations) for the Confidence Ratings given to the Male Culprit's Beard Question Broken Down by the Cue Condition and the Response Options

	Cue			
	Solo	Neutral	All	Overall
Yes	85.00 (21.21)	90.00 (17.32)	71.25 (26.42)	77.69 (23.86)
No	81.20 (15.70)	77.91 (18.30)	81.90 (15.93)	79.96 (16.82)
Unsure	64.00 (25.10)	54.00 (33.62)	66.00 (32.04)	62.50 (29.71)
Couldn't See	83.80 (18.56)	92.56 (16.02)	88.42 (17.08)	87.60 (17.53)
Overall	80.96 (18.33)	80.47 (21.46)	79.83 (22.34)	80.42 (20.68)