Diagnosing Eyewitness Accuracy

Andrew Russ

School of Psychology

University of Kent

A thesis submitted for the Degree of Ph.D. in the Faculty of Social Sciences at the University of Kent

March 2015

Abstract

Eyewitnesses frequently mistake innocent people for the perpetrator of an observed crime. Such misidentifications have led to the wrongful convictions of many people. Despite this, no reliable method yet exists to determine eyewitness accuracy. This thesis explored two new experimental methods for this purpose. Chapter 2 investigated whether repetition priming can measure prior exposure to a target and compared this with observers' explicit eyewitness accuracy. Across three experiments slower responses to target faces were consistently observed irrespective of eyewitness accuracy in a lineup task. This indicates that repetition priming can provide a covert index of eyewitness accuracy. However this method could not reliably assess the accuracy of *individual* eyewitnesses. Chapter 3 therefore explored an alternative test of eyewitness accuracy which was based on a multiple lineup procedure for faces. The characteristics of this method were assessed over five experiments which showed that only some eyewitnesses can *actually* identify a perpetrator repeatedly. Chapter 4 then showed that such repeat-identifications can provide a direct index of eyewitness accuracy in a field study. Over two experiments, the success of this method was such that eyewitnesses who consistently acted on the same identity over six lineups were always accurate eyewitnesses. These results demonstrate that multiple lineups of faces could provide a useful method for assessing eyewitness accuracy. The implications of these findings, both for further study and for forensic application, are discussed.

Acknowledgements

I would like to thank Dr. Markus Bindemann for the best supervision a student could wish for. Also, Dr. Melanie Sauerland for her considerable help and supervision during my time in Maastricht. I would also like to express my gratitude to all my housemates and teammates for keeping me sane over the course of my PhD.

Finally thanks to my parents and my brother for their encouragement and support, and most of all to Claire.

Declaration

I declare that this thesis is my own work carried out under the normal terms of supervision.

Andrew Russ

Publications

Chapter 4

Russ, A. J., Sauerland, M., & Bindemann, M. (submitted). The assessment of eyewitness accuracy with multiple lineups of faces.

Chapter 1

General Introduction

1.1. Introduction

In criminal proceedings, eyewitnesses are routinely required to identify a previously seen perpetrator from a police lineup. These lineups or, identity parades, refer to a collection of people who resemble the suspect of a crime presented in an array which includes this suspected perpetrator (Police and Criminal Evidence Act Code D, 1984, henceforth referred to as PACE; Memon, Havard, Clifford, Gabbert & Watt, 2011; Lindsay & Wells, 1985; Sporer, 1993; Steblay, Dysart, Fulero & Lindsay, 2001). Lineups can be 'live' or made up of computerised presentations of the identities comprising either photographs or videos (Memon et al., 2011). In the UK alone, tens of thousands of identity lineups are administered every year (e.g., VIPER, n.d.).

These lineups are important to the criminal justice system because in many instances eyewitness testimony is the only available evidence in the identification of a perpetrator (Wells & Olson, 2003). Although there is a long history of using eyewitnesses to identify criminals, the formal study of the effectiveness of this practice has been restricted to psychology. Since the initial concerns about the accuracy of eyewitness testimony (Münsterberg, 1908; Loftus, 1979; Huff, 1987) countless psychological studies have demonstrated it is an extremely error-prone process (for reviews see Cutler & Penrod, 1995; Wells, 1993).

The exact scale of this problem in the UK remains difficult to assess but archival studies indicate that more than one in five eyewitness identifications from police lineups might be misidentifications (Slater, 1994; Wright & McDaid, 1996). A more recent field investigation suggests an even higher number, whereby at least half of all eyewitness identifications might reflect such errors (Memon et al., 2011). This problem also translates to other criminal justice systems. In the USA, where the death penalty is still enforced in several states, the Innocence Project was introduced to identify and exonerate wrongfully convicted people using advances in DNA evidence. This organisation’s own statistics state that eyewitness misidentification is the single-most prevalent cause of wrongful convictions, playing a part in 72% of those later overturned (Innocence Project, 1992).

Despite well-documented problems with eyewitness testimony there are many people who are capable of making perfectly correct identifications. Furthermore, due to the criminal justice system’s heavy dependence on eyewitnesses, their removal from police proceedings on the grounds of general unreliability is not realistic (Wells & Olson, 2003). These factors indicate a method must be developed for differentiating between eyewitnesses who make correct identifications and those who make errors. Recent advances in face identification research have shown it is possible to distinguish accurate from inaccurate eyewitnesses (e.g., Bindemann, Brown, Koyas, & Russ, 2012). It is the aim of this thesis to build on these ideas and to take the next steps in diagnosing eyewitness accuracy.

This thesis explores methods of dissociating between eyewitness responses across three themes. The first is concerned with priming as an implicit measure of recognition in person identifications. The second theme examines multiple identity lineups as an improved tool in distinguishing accurate from inaccurate eyewitnesses. The third then develops the multiple lineup procedure and introduces a field setting. The thesis begins by outlining the state of the field of eyewitness accuracy research. Then the variables that can affect this accuracy are discussed. This is followed by a short review of what is known about the diagnosis of eyewitness accuracy and how recent advances could help improve it.

1.2. The accuracy of eyewitness identification

1.2.1 Archival studies.

Archival studies refer to the analysis of actual post-crime information (Wells, Memon & Penrod, 2006). Examining real eyewitness' responses is the most ecologically valid way to account for factors such as emotion and the sense of importance of the identification, factors which are not easily replicable in a laboratory (Wells et al., 2006; Fisher, Geiselman & Amador, 1989; Trollestrup, Turtle & Yuille, 1994; Yuille & Cutshall, 1986). As a result, these studies demonstrate the scope of the problem of eyewitness identification accuracy. For example, in a large study examining 843 witnesses viewing 302 suspects, Slater (1994) found that fillers, otherwise known as foils, which are innocent people included in the lineup to disguise the suspect, were identified on 23% of occasions. This finding has been replicated in seven other studies (Wright & McDaid, 1996; Behrman & Davey, 2001; Valentine, Pickering & Darling, 2003; Behrman & Richards, 2005; Wright & Skagerberg, 2007; Horry, Memon, Wright & Milne, 2012; Horry, Halford, Brewer, Milne & Bull, 2014). Across all available data the filler identification rate equates to roughly a third of eyewitnesses who identified the wrong person in a lineup. Clearly this level of inaccuracy constitutes an area of serious concern for the police, who rely on eyewitnesses to inform their actions.

Memon et al. (2011) examined one of the electronic lineup systems currently employed by the UK police. The Video Identification Parade Electronic Recording, or VIPER system is an alternative to 'live' lineups where an array of head and shoulders videos of lineup members is shown to an eyewitness. In this study, identification rates were measured and although the occasions of a suspect identification were comparable to that found in other archival studies at 44%, the filler identification rate (i.e. occurrence of mistaken identifications) was higher, at 42%. The authors explanation for this finding is that due to the large size of the VIPER video library, fillers who bear closer resemblance to the suspect than in most eyewitness studies were available, making this a harder test. In any event, these misidentifications are problematic for the UK authorities who now use this system as standard in real investigations (PACE, 1984).

There is obvious environmental validity of archival research, however, there are problems in collecting data in this way. Police records often do not differentiate between witnesses who made a filler identification and those who made no identification (Wells & Olson, 2003). Also, there are cases where it is unclear whether a suspect identification is in fact a perpetrator identification. The police do not conduct a lineup unless there is a suspect (Wells et al., 2006) and this suspect could be the perpetrator or an innocent person who has been mistakenly identified. It is impossible for the authorities to be certain whether this person was the perpetrator or not based only on an eyewitness identification. It then follows that the misidentification rate stated in each study is in fact an underestimate since presumably some innocent people were identified and considered correct identifications of the suspect (Wells, 2014). This would mean that these results may obscure some wrongly prosecuted suspects who were in fact misidentified as the perpetrator.

1.2.2 Field studies.

An alternative to studying actual crime data is to recreate the conditions of criminal proceedings within field experiments. By conducting experiments under ecologically valid circumstances researchers are able to know with certainty whether the target was present or not in the lineup whilst still maintaining a realistic context. The ecological validity is maintained through the limited control over the stimuli (Read, 1995) and because it is possible (and common) not to warn participants that they are part of an experiment until after the perpetrator has been seen. These factors allow a more natural observation of an event than is possible with a crime video in a laboratory study (Wells, 1993). In fact some field studies have even maintained the appearance of an actual crime throughout the identification stage (see Malpass & Devine, 1980; Murray & Wells, 1982).

One such experiment, conducted by Hosch, Leippe, Marchioni and Cooper (1984) used the natural settings to investigate the effects of victimisation on eyewitness accuracy. Participants sat in a classroom and were told to begin work on a preliminary experimental task. After a short while a confederate (posing as another participant) entered the room and stole either a calculator belonging to the department, or the participants' watches that had been collected earlier. Witnesses of the calculator theft were able to identify the perpetrator on 25% of occasions while witnesses of their own watch being stolen were able to identify him on only 17% of occasions. This study was able to use natural conditions to realistically examine a phenomenon not easily replicated in a laboratory. However, due to the limitations of field study methodology, it is not possible to determine if all participants saw the suspect equally or were similarly affected by the situation, undermining the experimental assumptions that only the variable of interest was different between conditions.

1.2.3 Laboratory studies.

The final form of eyewitness identification experiments take place in the laboratory. Lab studies are necessarily less realistic than actual or staged events due to the participant's knowledge that they are taking part in an experiment from the outset. However, the increased control over the conditions of lab experiments make this method of examining variables valuable. Outside of the lab there may always be additional factors which are difficult to control for or even identify, and many of these factors can be eliminated in a laboratory setting.

Early lab studies reported high identification accuracy but this was determined to be a product of using the same *images* rather than the same identities (Hancock, Bruce & Burton, 2000). Image matching is a much easier task than *identity* matching from different images. In a study conducted in 1982, Bruce found that recognition rates fell from 90% to 60% when same image matching was replaced by same identity matching between study and test. More recent laboratory evidence shows that identification rates similar to those in archival studies are found both for suspect identifications (Bruce et al., 1999) and filler identifications (Wells et al., 2006) supporting the notion that this work can be used to examine eyewitness accuracy.

Studies using different images of a perpetrator have shown that difficulties in eyewitness identification are not simply a product of memory failures (Henderson, Bruce, & Burton, 2001; Megreya & Burton, 2006a). Matching tasks require participants to decide whether two simultaneously presented images feature the same or different people. They represent an optimum situation where memory and choice are minimised and even under these conditions errors persist. Bruce et al. (1999) introduced the 1-in-10 task, a test of identification accuracy that requires the participant to select a target face from an array of distractors (see Figure 1.1). The target identities were exposed using footage taken from CCTV cameras and were tested with a static image lineup. When asked to match a target face to frontal images with a neutral expression, participants identified the target on only 70% of occasions and were correct in rejecting target-absent arrays to approximately the same level of accuracy. When given only target-present lineups and asked to choose the target with no option of rejection,

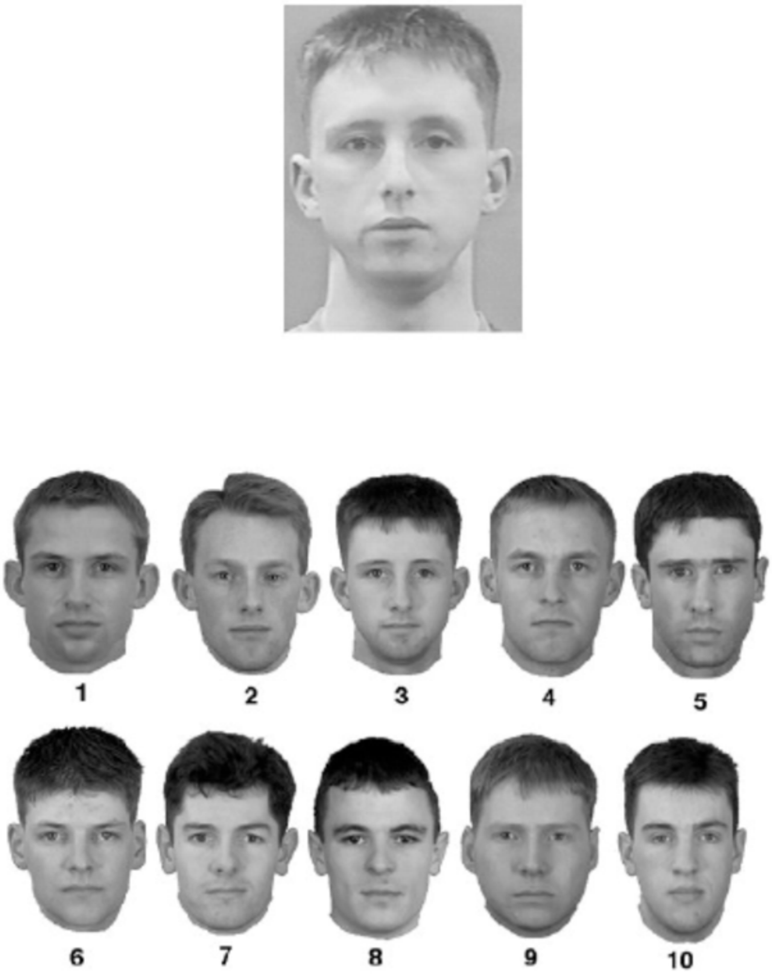

Figure 1.1 An example of a typical target exposure and test array used in lineups research. Participants must choose a face from the ten options that matches the identity of the face above, or must declare the target to be absent from the lineup. Taken from Bruce et al. (1999).

accuracy still only rose to 79%. Accuracy on arrays where the target and lineup faces were of different orientations was consistently lower at 61% and 68% respectively.

Megreya and Burton (2008) tested this poor identification accuracy rate with 'live' exposure to the targets. The targets were asked to stand in front of the participants for 30 seconds maintaining a neutral expression. Participants were then presented with a typical photo-lineup. Accuracy at this task was again found to be approximately 70%.

It is important to note that these experiments represent an optimal test of identification. In many of the experiments the memory requirements of the eyewitness task were entirely removed (by simultaneously presenting the targets and lineups) demonstrating the difficulty in identifying unfamiliar people from different presentations of their face. It is clear from this evidence that eyewitness errors are not simply a case of an initially perfect memory, degrading over time. In a landmark review, Wells (1978) identified many other factors that can affect eyewitness accuracy in his descriptions of system and estimator variables.

1.3 System Variables

A system variable is defined as a factor that could be under the control of the criminal justice system (Wells, 1978). In other words, the impact of a system variable on eyewitness accuracy can be changed by altering police procedures. One of the first examples of a system variable was offered by Loftus and Palmer (1974) in their now famous study of suggestive interrogation. The authors conducted interviews about a previously shown video of a car crash. In the interview, participants were either asked how fast the cars were travelling when they 'hit each other' or when they 'smashed into each other'. The latter condition left the participants with a greater impression of speed so that a week later they more often reported broken glass at the scene than those who had heard the word 'hit'.

This research suggested the importance of neutral wording in criminal proceedings. Since this study, neutral wording has been investigated in association with eyewitnesses several times. In a meta-analysis containing 12 studies, Clark (2005) argued that biased lineup instructions such as telling the witness that the perpetrator *is* present in the lineup increases the likelihood of an identification. This was true both in target-present and target-absent lineups meaning that although perpetrators were identified more often, so were filler identities who may have been innocent suspects. Police procedures have changed to take this evidence into account and now require officers to give unbiased instructions to eyewitnesses (PACE, 1984).

Highlighting the importance of the construction of the lineup, Doob and Kirshenbaum (1973), and Wells, Leippe and Ostrom (1979) drew attention to problems of lineup bias by identifying some real-world identity parades where the suspect could be identified just by a description. Such circumstances do not require the observer to recognise the target in order to identify them. It is possible, for example, for both guilty and innocent suspects to stand out from a lineup if the lineup foils resemble this person poorly (Lindsay & Wells, 1980). This can also be the case if this suspect is used as the basis for selecting the other lineup members (Clark & Tunnicliffe, 2001; Navon, 1992; Wogalter, Marwitz, & Leonard, 1992).

Malpass (1981; Malpass & Devine, 1983) identified effective size of a lineup as another potential problem, where some foil identities are so unlike the target they are effectively not options. Lindsay and Wells (1980) conducted the first study into the effective size of a lineup, where they manipulated the resemblance of the fillers to the suspect. In the high resemblance condition, there were fewer identifications of the perpetrator and of an innocent suspect than in the low resemblance condition. This is because in a low-resemblance condition the suspect 'stands out' and is more easily identifiable than the other faces. Importantly, when the resemblance of the foils was increased, the rate of eyewitness identifications of the perpetrator was less affected than that of the innocent suspect. It would seem, from this evidence, that a larger effective lineup is not only a more rigorous investigation of an eyewitness' memory, it is also superior to an unfair lineup in eliciting accurate identifications.

Fairness of lineups has also been scrutinised as a system variable. Lineup foils are intended to look like the target, but selecting them is a subjective decision made by the researchers or the police during the construction of arrays. Doob and Kirshenbaum (1973) improved fairness, introducing the practice of giving mock-eyewitnesses a written description of a suspect which could equally apply to all members of a lineup. This description typically includes information such as gender, skin, hair and eye colour, face shape and approximate age. A fair lineup stipulates that resulting identifications should be evenly distributed across all identities in the array since no suspect has actually been seen. If a lineup is fairly constructed, any preference for a face in an experiment should then be explained by previous exposure rather than any superficial factors. If this is not the case, this constitutes evidence that a bias may be present. Fairness of a lineup can now be examined with a binomial test which detects deviances away from the expected distribution of identifications by mock-witnesses (Tredoux, 1998; 1999). This practice has begun to be adopted by researchers within the field of eyewitness accuracy. However, criminal procedures still do not call for such measures in the assembly of lineups. Particularly in the case of 'live' lineups, where arrays are made up of available police officers (PACE, 1984), and are therefore limited by their appearances.

These variables represent procedural elements of a case, which can be changed should the evidence recommend such an action. System variables are factors of high importance because they can help prevent inaccuracies from occurring, rather than estimating the impact of inaccuracies later (as estimator variables do), and are easier to apply to police procedures directly (Wells & Olson, 2003).

1.4 Estimator Variables

Factors beyond the control of the authorities are called estimator variables. These factors are typically characteristics of the crime itself and due to their uncontrollable nature, their effects must be estimated after the event. Better eyewitness procedures could eliminate the need for some estimator variables, for example, the period of time between exposure to the perpetrator and test (response latency) is controllable to an extent once the suspect has been apprehended. However, if a suspect is not found for some time this cannot be controlled and the effect on accuracy must then be estimated (see Shapiro & Penrod, 1986).

Cross-race identification problems, for example, are well researched (for reviews see Meissner & Brigham, 2001; Sporer, 2001). They refer to the added difficulty in identifying a perpetrator of a different race compared to the eyewitness' own. In an environment such as an airport where passport control officers must match faces to identification documents many different-race officials can be employed making this a system variable. However, when dealing with a limited number of eyewitnesses to any particular crime, it will not always be possible to match race and therefore the impact on recognition accuracy must be investigated. The relative difficulty in identifying other-race faces is not surprising. People generally have more experience with faces from their own racial group compared to those from outside of it. In fact, there is evidence that supports this experiential explanation that has shown a reduction in the deficit when people have had increased contact with another race (Hancock & Rhodes, 2008).

Exposure duration, which refers to the amount of time the eyewitness was able to see the perpetrator for during the criminal activity, is another estimator variable that is important to understand. A meta-analysis conducted by Shapiro and Penrod (1986) found a positive correlation between exposure time and identification accuracy. This finding was supported by Memon, Hope and Bull (2003) who presented participants with a video of a crime that lasted 12 seconds in the short exposure condition or 45 seconds in the long exposure condition. Accuracy was higher in both target-present and target-absent lineups after the longer exposure. In a more recent meta-study, Bornstein, Deffenbacher, Penrod, and McGorty (2012) analysed exposure times recorded in 25 studies. The range of times was 0.7s to 3570s (median difference was 4.7s). Although longer exposure times were associated with higher performance at test, this correlation was non-linear. Despite some positive results, there are other factors which affect the exposure quality. Weapon focus is a well-known effector which reduces accuracy considerably because of the presence of a dangerous item which draws attention away from the perpetrator's face (Loftus et al., 1987; Steblay, 1992). Disguises, such as face masks, are also an obvious detriment to later identifications (Mansour et al., 2012; Shapiro & Penrod, 1986) and introduce new problems to account for in the initial exposure.

Due to the complexity of the associations between factors and the difficulty that arises in applying the findings of estimator variable research, Wells (1978) suggested it may not be highly fruitful to investigate this area. It is true that much estimator research is not easy to apply. For example, outside of the lab it is difficult to predict how long an eyewitness saw a perpetrator for, although there are recent studies that have addressed this problem (e.g., Attard & Bindemann, 2014) and there may be many other undetected variables that also affected the exposure conditions. However, if eyewitness accuracy is to be diagnosed to any degree of certainty estimator variables cannot be ignored. Even if police procedures are excellent, errors will still occur (Smith, Lindsay & Pryke, 2000), and research into quantifying these errors after an identification is important. The term '*post*diction' was introduced because a sub-set of estimator variables can be used to analyse eyewitness identification after the event and these postdictor variables could be used to diagnose eyewitness accuracy after a response has been made.

1.5 Diagnosing Eyewitness Accuracy

Postdictor variables can be used to diagnose accuracy because, due to their utilisation after the event, the measured effect incorporates all preceding factors, and there is no assumption of a causal effect on the eyewitness (Wells et al., 2006). These variables are measured separately for every circumstance, meaning that an observer's performance on one day may be different to another. Due to this specific focus, postdictor variables might be able to give a good estimate of accuracy for any particular instance that takes multiple variables into account. For example, if an eyewitness had seen a perpetrator for a full three minutes but they had been wearing a hockey mask for this period, exposure duration will be a limited tool in estimating their ability to recognise the target later. In contrast, a variable measured after the event for this particular identification should give a better indication of accuracy.

1.5.1 Confidence.

One of the most researched postdictor variables is post-decision confidence. This refers to the self-reported belief that the decision the eyewitness made was the correct one. In order to test confidence, a lineup is shown to the eyewitness requiring them to choose an identity or declare the perpetrator absent. Following this they are asked how sure they were that the decision they made was the correct one (for an early meta-analysis see Wells & Murray, 1984). It is an intuitive measure that is widely accepted by the general public to be a strong indicator of accuracy (for a summary see Sporer, Penrod, Read & Cutler, 1995). However, despite its intuitive appeal, the utility of post-decision confidence as a postdictor has been debated and has produced much conflicting research (Charman & Cahill, 2012).

Initial research found that confidence was only moderately correlated with accuracy (Deffenbacher, 1980; Sporer, 1993; Luus & Wells, 1994) but more recent developments have suggested a stronger correlation exists between these variables when additional constraints are put on the estimates. The introduction of a longer target exposure (Bothwell, Deffenbacher, & Brigham, 1987) and the removal of eyewitnesses who stated the lineup to be target-absent (Sporer et al., 1995) both increase the confidence-accuracy correlation.

Sauerland and Sporer (2007) conducted an eyewitness study using confidence as a postdicting variable. They showed a filmed theft to participants and then presented them with a lineup one week later. As well as asking for decisions as to the identity of the person in the film they also asked for an estimate of confidence in the decision. They found that eyewitnesses who rated their post-decision confidence as 50% or higher *and* made a choice (i.e. correctly identified the perpetrator or incorrectly chose a lineup member) were correct on 43% of occasions. Those choosers who rated their confidence as lower than 50% were correct on only 11% of occasions. This study, along with much other modern confidence research, used another modification that has improved the correlation between confidence and accuracy, the introduction of calibration into the confidence estimate (Brewer, Keast & Rishworth, 2002; Juslin, Olsson & Winman, 1996; Sauerland & Sporer, 2009). Calibration is calculated by examining the correlation between a participant's confidence estimate and their proportion of actual correct identifications. For example, a perfectly calibrated eyewitness who is 70% confident of a decision will be accurate in 70% of such decisions (for a full explanation see Juslin et al., 1996).

Although beneficial to the understanding of the relationship between confidence and accuracy, calibration is not a viable method for the practical testing of eyewitness accuracy. A stable estimate of calibration requires at least 200 participants (Weber & Brewer, 2003), meaning research using it aims to use calibration to develop confidence as a postdictor rather than to employ calibration directly.

1.5.2 Decision time.

A measure often combined with confidence is response latency or decision time. This is the length of time between an eyewitness being shown the lineup and making a decision (Sporer, 1992; Sauerland & Sporer, 2007, 2009). Studies have reported that eyewitnesses who make correct identifications do so faster than those who make incorrect identifications (Smith et al., 2000; Sporer, 1992, 1994). This is held to occur because of the involuntary attentional draw of the known identity from the distracter images, a common occurrence in the recognition of familiar faces (Charman & Cahill, 2012). However, if this variable is to be a useful postdictor it must be shown to differentiate between accurate and inaccurate eyewitnesses over varying circumstances. This has led researchers to attempt to find an optimum time boundary where correct decisions are most likely to be made.

Considerable research has investigated the possibility of an optimal time boundary, or window, for identification decisions (see Brewer, Caon, Todd, & Weber, 2006; Sauer, Brewer, & Wells, 2008; Weber, Brewer, Wells, Semmler, & Keast, 2004). However, substantial differences exist between initial exposures to perpetrators, both between contexts of different experiments and concerning the eyewitnesses within each experiment. Due to these inter- and intra- experiment differences an optimal time window encompassing all remains elusive.

1.5.3 Filler identifications.

Based upon the observation that within-study response latency was consistently lower for accurate decisions due to the 'pop-out' effect for targets, Charman and Cahill (2012) added a new postdictor with an innovative study which measured memory for *filler* identities. They argued that the postdictors described to this point represent decisions that are made quickly and confidently because the witness experiences an automatic recognition and is able to make their decision without needing to attend to the other faces in the lineup. Those participants who do not experience such recognition must spend more time comparing photographs with one another to find the best match and then decide whether the match is good enough to be a positive identification. It has been well-documented that this relative matching is a source of error (e.g., Lindsay & Wells, 1985; Steblay et al., 2001; Sporer, 1993; Lindsay & Bellinger, 1999). It stands to reason, therefore, that eyewitnesses who do not need to go through this process should achieve greater accuracy.

Charman and Cahill (2012) presented participants with a crime video and lineup presentation followed by a surprise test of recognition of the lineup faces. They found a negative correlation between accuracy on the lineup and recognition of the filler faces, suggesting that participants who had looked at the fillers for longer had done so because they were not able to identify the target from the video. This approach avoids many of the problems that arise with confidence and decision time studies. It does not rely on self-reports which have been questioned previously (Nisbett & Wilson, 1977), and does not rely on a time boundary which has varied across studies using both sequential (Sauer et al., 2008) and simultaneous lineups (Brewer et al., 2006).

Despite the elegance of this approach there are still some problems. Differences in decision-making styles and in conscientiousness may affect the results, since some eyewitnesses will likely check the other faces despite experiencing automatic recognition. Also, Charman and Cahill (2012) themselves acknowledge that police officers and jurors may not appreciate the importance of a good eyewitness *not* remembering the filler faces, and may in fact question the testimony of someone who does not remember these details. These issues can be addressed with another, more intuitive, postdictor in multiple lineups.

1.5.4 Multiple lineups.

Multiple lineups provide a more direct measure of eyewitness accuracy than many other postdictors. In this method, eyewitnesses are required to identify the same perpetrator repeatedly, but from different person aspects that might have been observed at a crime scene (Lindsay, Wallbridge, & Drennan, 1987; Pryke, Lindsay, Dysart, & Dupuis, 2004; Sauerland & Sporer, 2008; Sauerland, Stockmar, Sporer, & Broers, 2013). For example, in a mock-directions task Sauerland and Sporer (2008) found correct identifications of 61% in portrait face lineups, 19% in body lineups, 11% in bags lineups and 29% in profile face lineups. Identification of a suspect's body from a lineup indicates a 0.6 probability that the identified person is, in fact, guilty. However, this number rises to 0.9 when the separate identification of body and face cues, from two different lineups, are considered together.

These results have been replicated, over two staged live encounters, Pryke et al. (2004) found average correct identifications of 72% lineups of faces, 38% for bodies, 27% for voices and

50% for clothing. Combinations of these lineups led to a probability of 0.9 that the guilty person had been identified. For a staged theft Sauerland et al. (2013) found portrait faces were correctly identified from lineups in 24% of cases, 8% of the time from bodies and 27% from profile faces. Combinations of all lineups led to a probability of guilt of 0.9.

Across all these experiments identification accuracy for frontal views provided consistently high identification rates, and combinations of other person aspects with such frontal face portraits were most useful for diagnosing eyewitness accuracy. However, these results are curtailed by the rather poor identification accuracy for some of the person aspects. For example, correct identifications of voice lineups were obtained on only 27% of trials (Pryke et al., 2004), and this number was lower still for bodies and accessories, at 18% and 11%, respectively (Sauerland & Sporer, 2008).

It is clear from these findings that bodies and accessories are not recognised with any regularity. This is unsurprising since it has been well established that the face provides the primary visual means for person identification (Burton, Wilson, Cowan, & Bruce, 1999; O'Toole et al., 2011; Robbins & Coltheart, 2012; Bruce & Young, 1986) and that difficulties in processing faces seem to contribute directly to eyewitness errors (e.g., Bruce et al., 1999; Megreya & Burton, 2006a, 2008). It follows that face perception research could contribute greatly to the question of eyewitness accuracy.

1.6 The Role of Face Perception

1.6.1 Individual differences as a postdictor.

People can differ greatly in their ability to encode and remember unfamiliar faces (e.g., Bindemann, Avetisyan, & Rakow, 2012; Burton, White, & McNeill, 2010; Russell, Duchaine, & Nakayama, 2009), which seems to reflect an inherent and heritable ability (Schmalzl, Palermo, & Coltheart, 2008; Wilmer et al., 2010; Zhu et al., 2010). Consequently, different observers may not be equipped equally to act as good eyewitnesses. In a study that demonstrated individual differences in this task, Bindemann, Brown et al. (2012) showed participants a crime video and then asked them to identify the culprit in a lineup. This was followed by several trials of the 1-in-10 task (Bruce et al., 1999). Performance on the two tasks were positively correlated (see Figure 1.2), confirming that a person's ability to remember and recognise faces relates to their ability as an eyewitness, whereby observers who are particularly adept at processing faces are also more likely to make accurate eyewitness identifications. These differences are also not restricted to between subject variation.

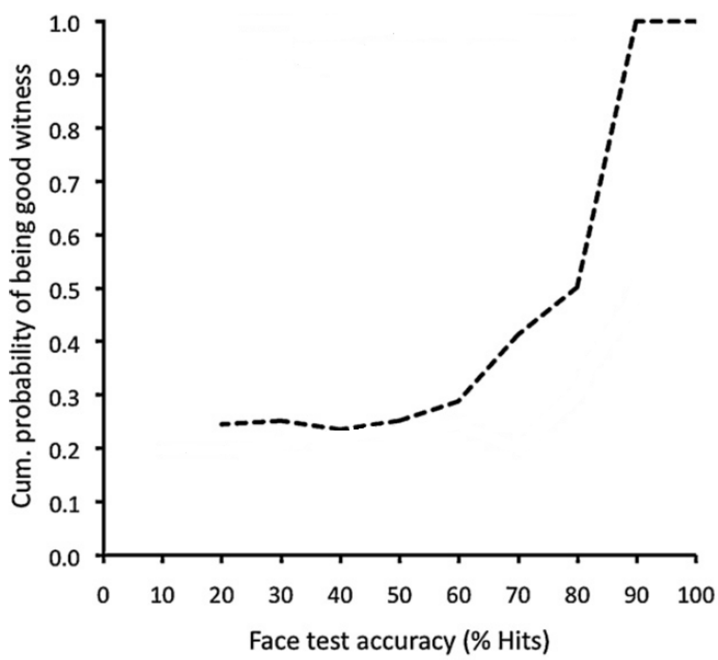

Figure 1.2 Correlation between probability of correctly responding to a lineup and percentage accuracy on an associated face test. Taken from Bindemann, Brown et al. (2012).

In an experiment requiring participants to match the same faces over a period of three days, Bindemann, Avetisyan et al. (2012) found within-subject variability whereby some participants were able to do this task every day but others' performance changed over the course of the test period (see Figure 1.3). These experiments offer evidence that individual differences in face processing ability exist. They also provide evidence for both inter- and intra-observer differences. Since experiments require accuracy to be tested more than once for each participant, and since it has previously been shown that individual differences relate to eyewitness accuracy (e.g., Bindemann, Brown et al., 2012) it follows that eyewitnesses themselves could be tested repeatedly. Specifically, existing multiple lineup techniques could be modified to only include faces (rather than bodies, bags, etc.), thus providing a repeated test of face recognition ability. An explanation of how instances of unfamiliar faces can provide independent tests of recognition is provided in this chapter and multiple face lineups will be considered further in Chapter 3. Prior to this, the thesis will examine another method that shows promise in the assessment of identification accuracy, repetition priming.

1.6.2 Repetition priming.

One established method in the theoretical study of face recognition that shows promise for the diagnosis of eyewitness accuracy is repetition priming (e.g., Bruce, Burton, Carson, Hanna, & Mason, 1994; Bruce & Valentine, 1985; Bruce & Young, 1986). However, so far it has not been applied to the assessment of eyewitness accuracy. In this method, observers are typically exposed to a set of famous faces in the initial *priming* phase of an experiment. When these faces are then *repeated* in a subsequent test phase, response times are facilitated in comparison with unprimed famous face identities.

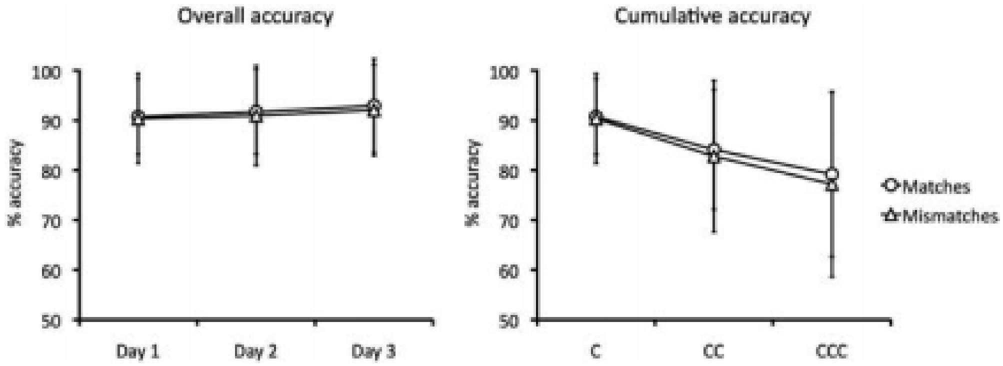

Figure 1.3 Overall accuracy and cumulative accuracy (whether correct on the first day, 0n the first and second day, or on all three days) in the test. Error bars show standard deviations around the means. Taken from Bindemann, Avetisyan et al. (2012).

This effect is face-specific in the sense that names or bodies cannot prime faces (Bruce & Valentine, 1985; Ellis, Young, Flude, & Hay, 1987) and it is typically found when recognition at test is measured using a familiarity judgement (Ellis, Young, & Flude, 1990). Moreover, while priming is strongest when the same image of a face is shown in the prime and test phases, indicating a partly perceptual basis of this effect, it is also found across different photographs and views of the same person (Ellis, Burton, Young, & Flude, 1997; Ellis et al., 1987). This cross-image priming effect is held to reflect the activation of internal cognitive representations of a known face, which are not tied to a particular photograph, but allow for the recognition of a person across a wide range of conditions (e.g., Bruce & Young, 1986; Burton, Jenkins, Hancock, & White, 2005; Burton, Jenkins, & Schweinberger, 2011). Consequently, repetition priming is held to occur in the cognitive system that stores and responds to facial identity.

Traditionally, the measurement of repetition priming focuses on reaction times to the famous faces shown at test, whereas the unfamiliar faces are only included to make up the task demands. Of course, eyewitnesses to crimes are rarely required to identify a familiar person. However, repetition priming has recently been applied to *un*familiar faces (Martin & Greer, 2011; Martin, Nind, & Macrae, 2009). Martin et al. (2010) observed facilitation of previously seen unfamiliar faces when participants were required to categorise these faces into male or female groups as quickly as possible. In a second part to the experiment the participants saw some of the same faces and some new ones from a different angle and were significantly faster to categorise the previously seen faces. The results demonstrate that repetition priming can work with unfamiliar faces, with short exposure times and with a change in image and context. It can also work in the absence of overt recognition. It would appear from this research that if an eyewitness has been sufficiently familiarised with an identity they should show repetition priming in a later test.

1.6.3 Familiar and unfamiliar face processing.

In this chapter much has been made of the difference between familiar and unfamiliar faces. It is a common belief that people are experts at recognising faces (see Bindemann, Attard & Johnston, 2014; Hancock et al., 2000). It is true that we are able to detect faces from a very young age (Simion, Turati, Valenza & Leo, 2006) and with great expertise (Bradshaw & Wallace, 1971; Sergent, 1984). However, the recognition of a face (i.e. the classification of an identity from this visual information) is an error-prone process. The prevailing conclusion that has emerged from the research is that we are experts at recognising *some* faces, specifically those we have become familiar with.

Face matching with unfamiliar faces elicits surprisingly poor performance and it is suggested the reason we feel we are so good at recognising faces is that we generalise our ability from the positive feedback from familiar faces (White, Kemp, Jenkins & Burton, 2014). These, can be recognised over a wide range of circumstances (Megreya & Burton, 2006a) even from very low quality images (Burton, Wilson et al., 1999), if the proportions are distorted (Sandford & Burton, 2014), or even if the identity is intentionally altered (Jenkins & Burton, 2011). In contrast to familiar face recognition, the identification of unfamiliar faces is rather difficult (Hancock et al., 2000). Figure 1.4 demonstrates the difference in difficulty between these two tasks. The photographs are matched for expression so the images in a column are similar but it is far easier to determine whether the top row shows the same person or two different people than it is in the bottom row.

An explanation for the differences between familiar and unfamiliar faces was provided by Bruce and Young (1986) when they introduced their functional model of face recognition. In this model they suggested that exposure to a face allows us to build an 'internal representation' of it. That is, a representation of what the face looks like based upon previous experiences. A stable internal representation of identity can be formed through experience of seeing the face under varying conditions. Once the face has become familiar, and the internal representation is strong, it becomes possible to identify this face under a very wide range of conditions including never before seen views.

It is rare for an eyewitness to make an incorrect identification when they are familiar with the perpetrator (e.g., Memon et al., 2011). However, when asked to identify an *unfamiliar* target (i.e. a person of which they have only limited perceptual experience, such as the brief exposure to a person at a crime scene) errors are far more common. Even under best-possible

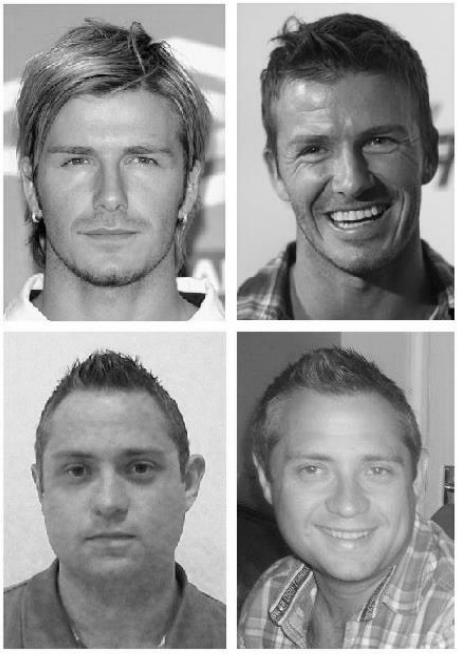

Figure 1.4 Illustration of the difficulty in matching unfamiliar faces compared to familiar faces. The top row shows two familiar faces, the bottom row shows two unfamiliar faces. In a typical face matching experiment the participant would be asked to choose whether the left and right images show the same person or two different people. This task is more difficult when an unfamiliar face is the target.

conditions, the identification of unfamiliar people can be very difficult (e.g., Bruce et al., 1999; Megreya & Burton, 2006a, 2008; Memon et al., 2011).

1.6.4 Variability of images in recognition.

The aims of this thesis are dependent on there being variability between images of the same face. It is perhaps counter to intuition to accept that some faces can be recognised on one occasion but not on another. Until Jenkins, White, Van Montford and Burton (2011) investigated this further, it was generally assumed that one image of a person contained sufficient information to recognise them repeatedly. In fact, as illustrated in Figure 1.5, people naturally exhibit a great deal of within-person variability in their facial appearance (e.g., Burton et al., 2005; Jenkins & Burton, 2008; Jenkins et al., 2011). As a consequence, even sophisticated automatic recognition systems may recognise some pictures of a person but will fail to recognise others (e.g., Jenkins & Burton, 2008; Phillips & O'Toole, 2014). For human observers, recognition failures can occur even when the differences between images are rather subtle. In the simple case of face matching, for example, identification errors are routinely encountered (Johnston & Bindemann, 2013). These errors persist under optimised conditions, such as when high-quality, same-day photographs are used, and in which faces are equated for expression, lighting and view (e.g., Burton et al., 2010; Megreya & Bindemann, 2009).

There is an opportunity to apply the same principles to eyewitness procedures. In the same way that different photographs of the same face can provide dissociable instances for person-identification (e.g., Burton et al., 2005; Burton et al., 2011; Jenkins & Burton, 2011) different lineups could offer the same opportunity for recognition from different images. Repeated identifications of one perpetrator in several lineups should be possible for only some eyewitnesses who have gained sufficient familiarity with an identity in the initial exposure. If this is the case this could be used as a postdictor of accuracy.

These matching tasks demonstrate that even seemingly similar images of the same person's face can provide dissociable instances for identification. Indeed, it has already been shown that different images can enhance the accuracy of other forensic tasks. Person identification from photo-identity documents, for example, can be improved by providing multiple face images of the same person for comparison (Bindemann & Sandford, 2011; White, Burton, Jenkins, & Kemp, 2014).

Figure 1.5 Illustration of variability within a face. There are only two identities presented here. Taken from Jenkins et al. (2011).

1.7 The Structure of this Thesis

The aim of this thesis is to investigate two new avenues of research into the investigation of eyewitness accuracy which are based upon theories of face recognition. Chapter 2 introduces repetition priming of unfamiliar faces as a method of measuring implicit recognition of the perpetrator's face. Across three experiments participants were required to identify perpetrators (targets) in criminal lineups. These experiments included target-present lineups (Experiment 1), target-absent lineups (Experiment 2) and both target-present and –absent lineups (Experiment 3). These lineups were followed by an implicit test of priming where the participants were required to categorise many faces which included all lineup identities from earlier in the experiment.

Chapter 3 builds on the recent research into variability of the face and examines the utility of multiple face lineups. Participants were required to identify a target over three lineups instead of the usual one (Experiment 4). The focus of this analysis was to compare one-off accuracy (identifying the target in any given lineup) with consistent accuracy (identifying the target repeatedly). To ensure familiarity was the cause of the differences in identification accuracy this experiment was replicated with participants who were familiar with the target (Experiment 5). This research was then extended to include target-absent lineups (Experiment 6). Here repeatedly accurate identifications *and* rejections of target-absent lineups were considered in the consistent accuracy. Experiment 7 investigated whether it was possible to learn the target's identity over the course of multiple presentations. In order to do this, the video exposure was forgone and instead participants in the cued condition had the target face highlighted in the first lineup, while those in the uncued condition had no assistance in identifying this person. Finally, the multiple lineup procedure was examined with sequential lineups (Experiment 8).

Chapter 4 investigates the utility of multiple lineups still further by testing them in a field setting. Furthermore, the possibility of consistent inaccurate identification was examined by providing lineups which included multiple instances of all faces and not just the target. First, participants were required to identify a person they had met incidentally in the street from multiple standardised photo lineups (Experiment 9). In the final experiment, this methodology was replicated with the added variable of photograph types (Experiment 10). Here, ID card photos and personal photos were added to the standardised images from Experiment 8 to increase variability between the images used.

Chapter 2

Repetition Priming of Unfamiliar Faces as an Index of Eyewitness Identification Accuracy

Introduction

Several postdictor variables, examining the accuracy of an observer's response to a lineup, were identified in the previous chapter. This chapter will introduce a new method for this purpose using repetition priming. This is a well-established effect where an initial presentation of a stimulus facilitates later categorisation of that same stimulus (see Scarborough, Cortese & Scarborough, 1977; Logan, 1990). This facilitation has been found in several areas of psychological research, such as recognition memory and lexical decision tasks (e.g., see Forster & Davis, 1984; Rugg, 1985; Neil, 1997), and repetition priming has also been applied to research in face recognition (e.g., Bruce et al., 1994; Bruce & Valentine, 1985; Bruce & Young, 1986). In this domain, a face that has been seen earlier in an experiment typically elicits a faster response time when categorised in a later step due to priming.

While repetition priming provides a good index of face recognition in psychological experiments, it is also a particularly sensitive measure. For example, repetition priming can survive many intervening items between exposure to a specific face identity at prime and test (Bruce & Valentine, 1985). This effect also persists after a change in context between prime and test phase (Bruce, Carson, Burton, & Kelly, 1998) and is found for faces that were initially viewed incidentally (Bruce et al., 1998; Ellis, Flude, Young & Burton, 1996) or peripherally (Bindemann, Jenkins, & Burton, 2007), or could only be seen partially (Brunas, Young, & Ellis, 1990; Johnston, Barry, & Williams, 1996). Repetition priming can also reveal prior exposure to a person even when an observer cannot remember this explicitly (Jenkins, Burton, & Ellis, 2002) or when faces were initially shown too briefly to be recognised overtly (Morrison, Bruce, & Burton, 2000). Thus, repetition priming appears to be a long-lasting, robust, and highly sensitive measure of whether a face identity has been seen previously.

These characteristics indicate that repetition priming might provide a useful method for assessing the identification accuracy of eyewitnesses. And another advantage of repetition priming has emerged recently that might be beneficial for this purpose. While such effects are traditionally measured to assess the recognition of familiar (i.e., famous) faces, (e.g., Bruce et al., 1998; Ellis et al., 1987; Ellis et al., 1990), repetition priming has now also been observed with unfamiliar faces (Martin et al., 2010; Martin & Greer, 2011; Martin et al, 2009). With only a single initial exposure to a static image of an unfamiliar face, these effects do not transfer to different instances of the same person (Martin et al., 2010), which indicates limited learning of the facial identities. However, these effects can generalise across different instances when the initial exposure to these identities is more extensive (Martin & Greer, 2011). Moreover, eyewitness identification is essentially a test of the *degree* of familiarity that an observer has gained with a target identity. Thus, eyewitnesses should only be able to identify a previously-unknown "unfamiliar" target from an identity lineup if sufficient familiarity with this person was gained during an earlier exposure, and repetition priming of the target should operate correspondingly.

In this chapter, repetition priming was used to assess recognition in a laboratory eyewitness paradigm. In this approach, observers were first shown video footage of two target persons, which effectively served as the priming phase. They were then provided with two photographic identity lineups, one for each target, and were asked to determine if the targets were present, and if so, to indicate the corresponding lineup face. Participants' gained familiarity was then investigated by measuring repetition priming of these identities in a subsequent test phase. For this purpose, observers were shown a block of famous and unfamiliar faces, which included all the faces from the lineups, and were asked to make speeded fame decisions (i.e., famous vs. not-famous). Repetition priming of the target faces was then compared with that of the non-target (foil) faces from the lineup and also with new, previously unseen unfamiliar faces. As observers are exposed to both targets and non-target foils prior to the test phase (i.e., in the preceding identity lineups), both categories of faces might show some repetition priming in comparison with previously unseen faces. However, in contrast to the lineup foils, the target faces were also primed in the initial video. The observers were therefore not only exposed to the targets for longer, but such moving footage also primes faces more effectively than static images (Lander & Bruce, 2004). The target faces should therefore show a more pronounced priming effect than the lineup foils.

The expectation is to find this effect in observers who correctly identified the targets from the preceding lineups. However, considering that repetition priming is evident even when observers cannot explicitly remember the prior exposure to a face (Jenkins et al., 2002; Morrison et al., 2000), a secondary question is whether priming of the target will also be found when an eyewitness has made a misidentification or judged the target to be absent from a lineup. In these instances, repetition priming would provide a covert recognition index that is more accurate than observers' explicit eyewitness identification responses. This question is explored by assessing eyewitness accuracy for target-present lineups in Experiment 1, then for target-absent lineups in Experiment 2 and, finally, for both types of lineup in Experiment 3.

Experiment 1

In this experiment, eyewitness accuracy was assessed with repetition priming for target-present identity lineups. To increase the available data points for analysis, observers were exposed to two target identities (one male and one female) in a video. The participants were then asked to select these targets from two separate identity lineups, which always comprised one of the targets and nine foil faces. In the final phase of the experiment, repetition priming was measured for all of the lineup faces (targets and foils). For this purpose, the observers were shown an extended set of unfamiliar faces, which included all lineup identities, and a corresponding number of famous faces, and were asked to classify these accordingly.

If participants encode the target identities during the screening of the video, then greater identity priming should be observed for these faces in comparison with all other unfamiliar faces. Of particular interest here is how such priming effects relate to the participants' explicit responses to the identity lineups. It would be expected that observers who manage to identify a target from a lineup will also show a robust repetition priming effect for this person. It is less clear whether such priming effects will be found when observers cannot identify the target. In such cases, the absence of a priming effect would indicate that a target's appearance was initially encoded or remembered insufficiently by an eyewitness. If this proves to be the case, then the existence of any repetition priming effects would simply correspond directly to observers' accuracy in the lineup task. However, considering that repetition priming has been demonstrated even when observers cannot explicitly remember prior exposure to a person (Jenkins et al., 2002; Morrison et al., 2000), it is also possible that such effects might be found here when observers cannot identify the correct target from a lineup. Thus, repetition priming could also provide a covert index of eyewitness accuracy that operates irrespective of observers' explicit identification decisions.

Method

Participants

Forty-nine undergraduate students (37 female, 12 male) from the University of Kent, with a mean age of 19.5 years (*SD* = 1.4), participated in this experiment as a condition of their course. All reported normal or corrected-to-normal vision.

Stimuli and Procedure

This experiment consisted of three sequential parts. Observers first watched a video of two target identities (part 1). Eyewitness accuracy for the targets was then assessed with two identity lineups (part 2). Finally, processing of the target identities was assessed again, via repetition priming in a speeded fame categorisation task (part 3).

Part 1: Video exposure to targets.

The stimulus materials for part 1 consisted of a video of a male and a female target person, who were depicted in conversation for 60 seconds. The faces of both targets were visible for the full duration of the video and could be seen across a range of views (e.g., frontal, ¾ and profile view). The video was presented at a size of 30 (W) x 17 (H) cm on a standard computer monitor and did not contain sound. For illustration, example stills from the video are depicted in Figure 2.1.

Part 2: Lineup identification of targets.

Following the screening of the video, observers were given two identity lineups, one for each of the targets, to provide an explicit test of eyewitness accuracy. Each of these identity lineups consisted of a photograph of a target's face and nine foil faces, which were shown simultaneously, alongside the target. The foil faces were taken from the Glasgow University

Figure 2.1 Example still frames from the video of the two target identities.

Face Database (Burton et al., 2010), and were chosen by the experimenters to be of the same sex and of similar age and appearance to the target in each lineup (for an illustration of these lineups, see Figure 2.2). In the lineups, each face was shown from a frontal view and with a neutral expression at a size of approximately 5 (W) x 7 (H) cm.

Observers were asked to study each lineup closely and to decide whether the male/female target was present or absent, and if present, to indicate which of the 10 was the target. Note that participants were given these unbiased instructions to allow for the rejection of the lineups when observers were unable to identify a face as the target. The target faces were, in fact, always present to maximise data collection for this type of lineup in part 3 (the repetition priming task). Participants indicated their responses by pressing the number key, on a standard computer keyboard, that corresponded to the lineup location of the target (e.g., “1” for face 1, “2” for face 2, etc., “0” for face 10) or by pressing

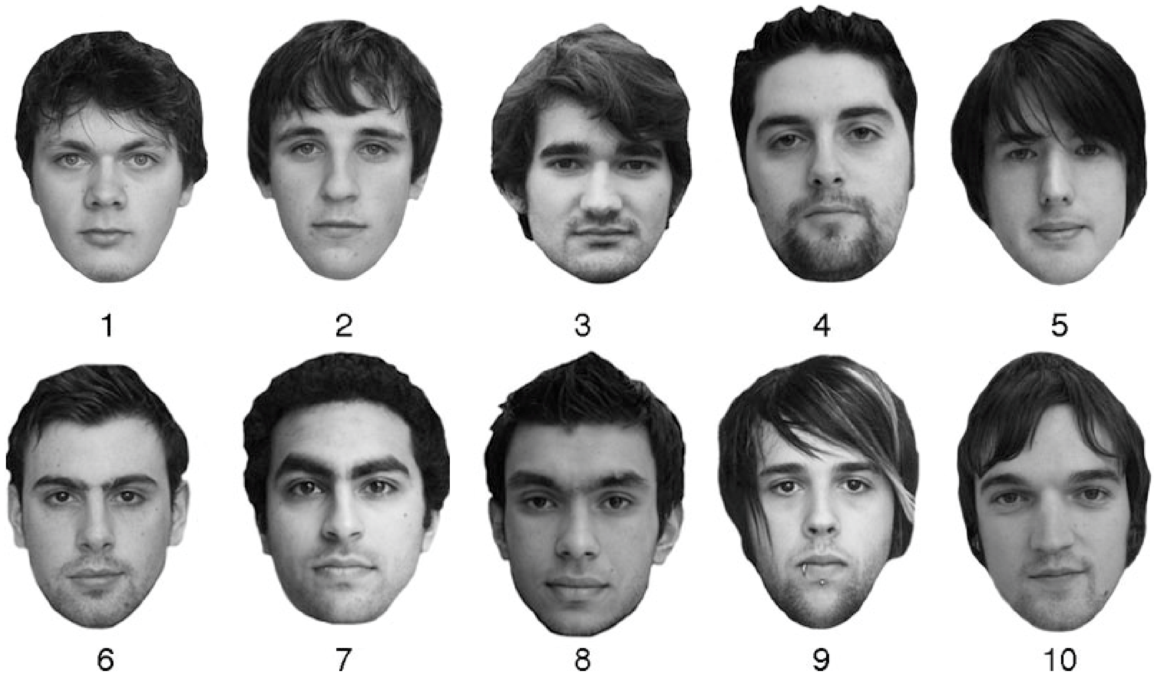

Figure 2.2 An example of a target-present identity lineup.

“A” if they believed the target absent from a lineup. They were asked to respond as accurately as possible and were told that there was no time limit for the task. The order of presentation of the male and female identity lineups was counterbalanced across participants.

Part 3: Repetition priming test.

In the final part of the experiment, repetition priming of the target identities was assessed with a speeded fame categorisation task. The stimuli for this part consisted of the 20 face identities from the lineups (1 male target, 1 female target, 9 male foils, 9 female foils) and a further 20 unfamiliar faces (10 male, 10 female), which were taken from the same face database as the foil identities (Burton et al., 2010). For the target and foil faces, the same photographs were used as had been seen in the preceding lineups. In addition, the photographed faces of 40 celebrities (20 male, 20 female) were used as famous stimuli, to make up the task demands. All of the famous and unfamiliar faces were depicted in a frontal view and with a neutral expression, and were shown in the centre of the screen at a size of approximately 8.5 (W) x 10 (H) cm.

In the experiment, each trial began with a fixation cross for 1000 ms, followed by a face stimulus, which was displayed until a response was registered. Participants were instructed to classify these faces as “famous” or “not famous”, as quickly and as accurately as possible, by pressing one of two possible keys on a computer keyboard with their index fingers. All participants completed 80 trials, comprising the 40 unfamiliar and 40 famous faces. These 80 faces were presented in a randomised order.

Results

Lineup Identification Accuracy

The first step of the analysis focused on assessing observers’ responses in the lineup task. In line with previous studies, these responses were broken down into hits (the identification of the correct lineup face as the target), misidentifications (identification of a wrong face as the target), and misses (the incorrect response that a target is absent from a lineup). The percentage of responses that fell into these categories was calculated and combined for the male and female lineups. These data show that eyewitness accuracy was generally low. Overall, observers recorded 41% hits, 40% misses, and 19% misidentifications.

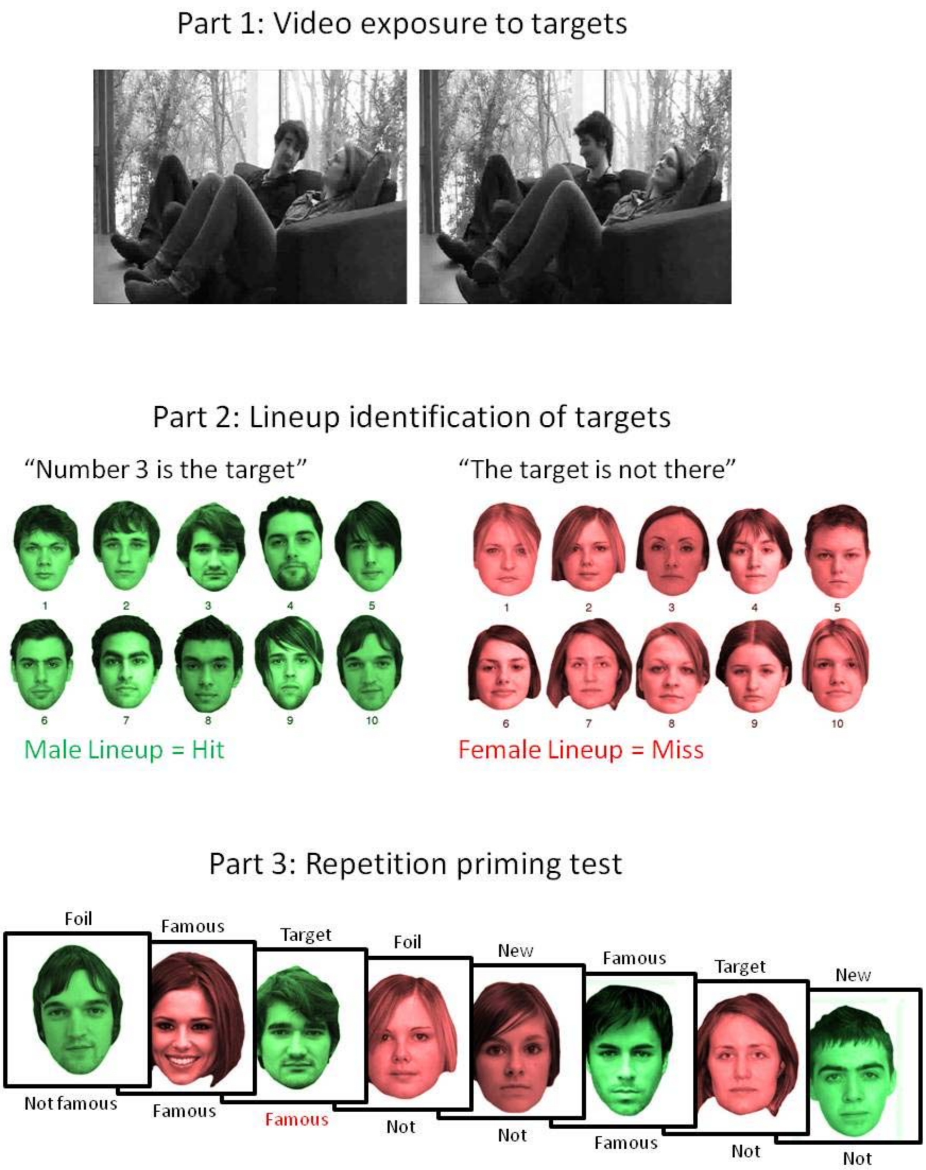

Figure 2.3 Graphical representation of the experimental procedure. Participants were required to watch a video introducing two identities in part 1. They were asked to identify them from lineups in part 2. In part 3 their task was to categorise 'famous' and 'not famous' faces. The 'not famous' faces in part 3 included both targets and all foils from part 2 as well as some new unfamiliar faces. An equal number of famous faces were also added to make up the task demands. The type of face is presented above each item, the response to it in the 'famous/not famous' task is presented below.

Accuracy	Targets	Foils	New
Hits	90.9 (29.2)	97.5 (4.4)	98.3 (4.4)
Misses	90.9 (29.2)	96.5 (10.2)	96.8 (8.7)
Misidentifications	100.0 (0.0)	97.7 (5.9)	97.9 (4.0)

Table 2.1. Accuracy (%) for the Targets, Foil Faces, and New Faces in the Fame Task of Experiment 1, Broken Down by Hits, Misses, and Misidentifications. Standard deviations are presented in parentheses.

Repetition Priming of Lineup Faces

The fame categorisation task (part 3) was then analysed to determine whether any repetition priming effects were found. Overall accuracy of observers' responses was calculated first, which showed that 98% (*SD* = 4.7) of unfamiliar and 84% (*SD* = 15.3) of famous faces were categorised correctly. This demonstrates that observers were complying with the task demands. In addition, a breakdown of the accuracy data is provided for the target faces from the lineup, the lineup foils, and the new unfamiliar face identities. These data are shown in Table 2.1, broken down for cases in which a hit, miss, or misidentification was registered to the identity lineups. As illustrated in Figure 2.3, the priming data in this analysis is drawn from part 3 but is *categorised* according to performance in part 2. To illustrate, Table 2.1 demonstrates that accuracy for the target faces when the preceding lineup (containing that face) registered a hit was 91%, however when this preceding lineup led to a misidentification, the accuracy in the subsequent priming task was 100%. Table 2.1 shows that accuracy was generally comparable across these conditions. Three one-factor within-subjects ANOVAs were conducted, which found no difference in accuracy between target, foil and new faces for

Response Times	Targets	Foils	New	Foils-as-target
Hits	1152 (480, N=37)	754 (220, N=351)	794 (203, N=394)	--
Misses	997 (645, N=36)	720 (222, N=340)	713 (152, N=378)	--
Misidentifications	939 (276, N=20)	783 (259, N=160)	729 (157, N=195)	--
Misidentifications II	939 (276, N=20)	776 (243, N=143)	729 (157, N=195)	889(423, N=17)

Table 2.2. Response Times (ms) for the Targets, Foil Faces, and New Faces in the Fame Task of Experiment 1, Broken Down by Hits, Misses, and Misidentifications. Standard deviations and sample size are presented in parentheses.

hits, $F(2,64) = 1.77$, $p = 0.18$, misses, $F(2,64) = 2.38$, $p = 0.10$, and misidentifications, $F(2,32) = 1.50$, $p = 0.24$. The accuracy data was therefore not analysed further.

The response times, which are the data of main interest here, were calculated next. Incorrect responses were excluded from this analysis and the median reaction times (RTs) were then calculated separately for the target faces, the lineup foils, and the new face identities. These data are also shown in Table 2.2, broken down for cases in which a hit, miss, or misidentification was registered in the preceding lineup task. Note that these data show the combined RTs, for male and female lineups. However, for the calculation of inferential statistics, the responses to faces from these lineups were coded separately. To illustrate, if an observer achieved a hit for the male identity lineup but recorded a misidentification for the female lineup, then the RTs for this person were calculated for hit trials from the response to the male target, the responses to the male foil faces, and the new male faces. Similarly, for misidentifications, the RTs were based on the female target, the responses to the female foil faces, and the new female faces. In this procedure, when observers achieve two lineup responses that fall into the same category (e.g., two hits), then these are counted as separate instances, on a between-subject basis. This procedure was adopted to maximise the available data points for analysis.

The data in Table 2.2 show an intriguing pattern. Irrespective of whether observers had initially registered a hit, miss or misidentification to an identity lineup, response times were slower for the target faces than the foils and new faces. To analyse these findings formally, three separate one-factor within-subject ANOVAs with the levels target, foil and new faces were conducted for hits, misses, and misidentifications. Note that some observers incorrectly classified the target as 'famous' in the test phase. In these cases, RT values were obtained for the foils and the new faces but not for the target, and so these data were excluded altogether from the analysis. This left data for 37 cases for hits (three excluded), 36 cases for misses (three excluded), and 19 for misidentifications (none excluded).

The ANOVA for hits showed a main effect of face type (i.e., target, foil, new face), $F(2,72) = 21.44$, $p < 0.001$. Tukey HSD test showed that responses were slower to the lineup targets than to the foils, $q = 8.00$, $p < 0.001$, and the new faces, $q = 8.04$, $p < 0.001$, while the foils and new faces did not differ from each other, $q = 0.03$. A similar pattern was observed for misses, $F(2,70) = 10.18$, $p < 0.01$, where responses were again slower for the targets than the foil faces, $q = 5.62$, $p < 0.001$, and the new faces, $q = 5.43$, $p < 0.001$, while foil and new faces did not differ from each other, $q = 0.19$. Finally, an effect of face type was also found for misidentifications, $F(2,36) = 5.29$, $p < 0.01$. In this category, responses were also slower to the targets than the new faces, $q = 4.47$, $p < 0.01$, but did not differ between target and foils, $q = 3.17$, or foils and new faces, $q = 1.31$.

Overall, these results therefore show that observers' response times can distinguish between the targets and the foil identities and previously unseen faces. This priming effect was found after observers had made a correct target identification in the identity lineups, a mistaken identification, or when they had erroneously indicated the target to be absent from the lineups.

Repetition Priming of Misidentified Foils

In the preceding analysis, the repetition priming data is split into response times to the target, the lineup foils and the new faces. This analysis has the advantage of making the hits, misses and misidentification conditions directly comparable. However, it also has the disadvantage of grouping together two types of faces in the foil category of the misidentification condition, namely the foil faces that were mistakenly selected by observers as the targets and the remaining, unselected foils. Considering that response times were slower to the target than the foils in the hit condition, it is possible that a similar effect is also found for the foil face that is mistakenly identified as the target. To explore this possibility, the responses in the misidentification condition were divided further, into foils that were mistakenly identified as the target and the remaining, unselected foils (see Misidentifications II in Table 2.2). These data comprised 19 cases but one of these, which did not yield a correct response time to the foil face that was misidentified as the target, was excluded from the analysis. The remaining data shows that responses were slower to the foil faces that were mistakenly identified as the target than the remaining foil faces and the new faces. However, a one-factor within-subjects ANOVA with the levels target, foil, new faces, and misidentified-as-target foils found no effect of face type, $F(3,51) = 1.09$, $p = 0.36$.

Discussion

This experiment examined whether repetition priming can provide an index of eyewitness accuracy for target-present lineups. A robust priming effect was discovered, whereby responses were slower to the target faces than the lineup foils and new unfamiliar faces. However, this effect was not only obtained when observers had previously identified the correct face from a lineup but also when a foil face was selected instead by mistake or when no lineup identification was made at all. This indicates that repetition priming can provide a covert index of eyewitness accuracy that operates even when observers' explicit lineup decisions are incorrect. These findings therefore converge with studies that show that repetition priming can reveal prior exposure to a person even when an observer cannot remember this explicitly (Jenkins et al., 2002; Morrison et al., 2000).

Contrary to these previous studies, it is notable that these priming effects were expressed here by a slowing in responses to the target faces. This *negative* priming effect is surprising and differs from repetition priming of familiar faces, which usually shows a facilitation of responses (e.g., Bruce et al., 1998; Bruce & Valentine, 1985; Ellis et al., 1987; Ellis et al., 1990). Negative priming has been observed for task-irrelevant unfamiliar faces (Khurana, Smith & Baker, 2000), which might reflect the inhibition of these stimuli (e.g., DeSchepper & Treisman, 1996). However, considering that the target faces were always task-relevant here, this seems an unlikely explanation for the results. Alternatively, the cause of this *slowing* effect may reflect the categorisation task with which priming was measured. This task normally requires familiarity decisions to famous and unfamiliar faces (i.e., familiar vs. unfamiliar judgments). This categorisation would have posed a problem in the current paradigm, in which the target faces are no longer strictly "unfamiliar" at the stage at which priming is measured but also cannot evoke the strong sense of familiarity of famous faces.

For this reason, a fame decision task was adopted (i.e. famous vs. not famous). However, it is possible that this task still produces some uncertainty in the categorisation of the target faces, as these faces might also trigger famous judgments due to their prior familiarisation in the initial video. It is already established that target responses can be slowed when stimuli provide conflicting information about identity and familiarity (for example, responses onto familiar names are slowed when accompanied by the faces of a different familiar person (Bindemann, Burton & Jenkins, 2005; Young, McWeeney, Ellis, & Hay, 1986)). It is conceivable that in the current experiment such conflicting information could be provided by a single stimulus whereby the previously seen face of a target is both familiar due to its prior exposure in the video but is unfamiliar in comparison with the famous faces of the repetition priming task. The 'negative priming' effect might therefore reflect a response conflict, whereby the familiarity that is gained with the target faces in the initial video might interfere with the speeded fame decisions of the repetition priming phase. It is also possible that the single measure of reaction time for the target (compared to nine such measures for the foils) could cause an issue, delaying the response time by chance.

The issue of negative priming will be returned to in Experiment 3. For now, another question must be investigated that Experiment 1 cannot address. So far, repetition priming has only been combined with target-present lineups, in which the identities from the video are always included. The next experiment seeks to explore these effects with identity lineups from which the targets are absent.

Experiment 2

In Experiment 1, repetition priming of the target identities was not only observed after a correct lineup identification had been made, but also when participants had previously selected the wrong face or had failed to select any face from a target-present identity lineup. In the next experiment, this task was repeated with lineups in which the target identities were not present. Such target-absent lineups require identification skills that appear to be dissociable from target-present lineups, whereby performance for one type of lineup is not related to the other (Megreya & Burton, 2007). Crucially, however, these absent lineups also allow the assessment of priming of a target when a face is not picked or when a wrong face is selected. If the results of Experiment 1 are robust then a priming effect for the target should be observed in Experiment 2 regardless of observers' initial lineup identification decisions.

Method

Participants

Forty-five undergraduate students (34 female, 11 male) from the University of Kent, with a mean age of 21.8 years (*SD* = 4.3), participated in this experiment as a condition of their course. All reported normal or corrected-to-normal vision. None had participated in Experiment 1.

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 1, except for the following changes. In the identity lineups, the male and female targets were replaced by two unfamiliar faces. As for the other lineup foils, these faces were selected from the Glasgow University Face Database (Burton et al., 2010) and were shown in a frontal view, with a neutral expression, and at the same size as all other lineup faces. Despite these changes, all observers were again given unbiased instructions by being told that the target could be present or absent from each lineup.

Despite the changes to the identity lineups, the target faces were again included in part 3 of the experiment, so that repetition priming could be measured for these identities. The number of faces that were shown in part 3 therefore increased slightly, to accommodate the two new face identities that replaced the targets in the lineups. In contrast to Experiment 1, each observers was therefore shown 82 faces in part 3, consisting of the two target identities (1 male, 1 female), the 20 faces from the target-absent lineups (10 male, 10 female), 20 new unfamiliar faces (10 male, 10 female), and 40 famous faces (20 male, 20 female). As before, the faces were presented in a randomised order and required speeded "famous" or "not-famous" decisions.

Results

Lineup Identification Accuracy

In line with previous studies, the responses to the target-absent lineups were broken down into correct rejections (the correct response that the target is absent from a lineup) and false positives (the mistaken identification of a lineup face as the target). Overall, observers made 52% correct rejections and recorded 48% false positives, which shows that eyewitness accuracy was low.

Repetition Priming of Lineup Faces

The responses from the fame categorisation task were analysed next. First, the general accuracy of observers' responses was calculated showing that 96% (*SD* = 4.8) of unfamiliar and 88% (*SD* = 13.6) of famous faces were categorised correctly. Once again, an overview of the accuracy data for the targets, foils and new unfamiliar faces is provided, broken down for cases in which a correct rejection or false positive was initially registered for the identity lineups (see Table 2.3). A one-factor within-subjects ANOVA found no difference in accuracy between the classification of the targets, foils and new faces for correct rejections, $F(2,60) = 0.78$, $p = 0.46$. For false positives, such an effect was found, $F(2,56) = 3.97$, $p < 0.05$, reflecting lower accuracy for the targets than the foils, $q = 3.59$, $p < 0.05$. Overall, however, accuracy for the targets was still high and did not differ from the new faces, $q = 3.30$, and between the foil and new faces, $q = 0.29$.

Accuracy	Targets	Foils	New
Correct rejections	93.5 (21.4)	97.7 (6.2)	94.5 (9.4)
False positives	84.5 (33.0)	97.4 (4.9)	96.4 (6.4)

Table 2.3. Accuracy (%) for the Targets, Foil Faces, and New Faces in the Fame Task of Experiment 2, Broken Down by Correct Rejections and False Positives. Standard deviations are presented in parentheses.

The response times were calculated next. As before, incorrect responses were excluded and the cross-subject medians for the targets, foils, and the new faces were calculated. These data are also provided in Table 2.4, broken down for correct rejections and false positives, and show that responses were slower to the targets than the foils and new faces when correct rejections were made. For the statistical analysis, once again responses to male and female lineups were treated as separate instances, on a between-subject basis, to maximise the available data points for analysis. However, a small number of cases had to be excluded from the analysis in which the target faces were incorrectly classified as famous. This left 44 cases for the analysis of correct rejections (three excluded) and 36 cases for false positives (seven excluded). For correct rejections, a one-factor within-subjects ANOVA revealed a main effect of face type (target, foil, new faces), $F(2,86) = 9.06$, $p < 0.001$. Tukey HSD test showed that observers responses were reliably slower to the targets than the foils, $q = 5.20$, $p < 0.01$, and the new faces, $q = 5.22$, $p < 0.01$, while response times for the foils and new faces did not differ, $q = 0.02$. Similar effects were obtained for false positives.

A one-factor within-subjects ANOVA of these data also found an effect of face type, $F(2,70) = 4.50$, $p < 0.05$, which arises from slower responses for the targets than the foil faces, $q = 3.66$, $p < 0.05$, and the new faces, $q = 3.69$, $p < 0.05$, while foil and new faces did not differ from each other, $q = 0.03$.

Response Times	Targets	Foils	New	Foils-as-target
Correct rejections	1127 (767, N=44)	763 (229, N=460)	758 (315, N=447)	--
False positives	905 (534, N=36)	794 (425, N=377)	790 (370, N=414)	--
Misidentifications II	905 (534, N=36)	813 (526, N=338)	790 (370, N=414)	872 (342, N=39)

Table 2.4. Response Times (ms) for the Targets, Foil Faces, and New Faces in the Fame Task of Experiment 2, Broken Down by Correct Rejections and False Positives. Standard deviations are presented in parentheses.

Repetition Priming of Falsely Identified Foils

As in Experiment 1, RTs were analysed separately for the lineup faces that were falsely identified as the target and for all other lineup foils. These data are also presented in Table 2.4 (see False positives II) and shows that observers were slower to respond to the falsely identified lineup faces than to the other foils, and the new faces. Note that these data comprise 43 cases, but 11 of these did not yield a RT for either the target (seven cases) or the face falsely identified as the target (four cases) when these faces were classified incorrectly as famous in the fame categorisation task. These 11 cases are therefore excluded from the following analysis. For the remaining data, a one-factor within-subjects ANOVA with the levels target, foil, new, and misidentified-as-target foil face once again found no effect of face type, $F(3,93) = 1.82$, $p = 0.15$.

Discussion

This experiment measured repetition priming for a target identity following the presentation of a target-absent lineups. In line with previous studies, identification accuracy was poor (e.g., Bruce et al., 1999; Megreya & Burton, 2006a, 2008), whereby the false identification of a lineup face as the target was almost as likely as the correct rejection of a lineup. Importantly, however, reliable repetition priming for the target identities was obtained, regardless of observers' responses in the lineup task. These results converge with Experiment 1 to show that repetition priming of a target face can be found irrespective of whether a correct or an incorrect identification decision was made initially to a lineup, and this effect occurs both in the presence of a target in a lineup (Experiment 1) or its absence (Experiment 2). This suggests that repetition priming provides a sensitive index of prior exposure to a target identity that operates independently of explicit identification decisions.

Experiment 3

In the final experiment, the relationship between repetition priming and eyewitness accuracy for target-absent and target-present lineups was examined within the same design. For this purpose, observers were again shown the initial video to provide exposure to the targets, but were then given a target-present lineup for one target identity and a target-absent lineup for the other. If the repetition priming effects of the preceding experiments are robust, it would be expected to observe similar results in the current study.

Here, the nature of the observed priming effects were also explored further. Repetition priming is typically expressed as a facilitation in reaction times, both when it is measured for famous identities (e.g., Bruce et al., 1998; Ellis et al., 1987; Ellis et al., 1990) and unfamiliar faces (Martin et al., 2010; Martin & Greer, 2011; Martin et al., 2009). The current findings deviate from these established effects, by showing a clear slowing in responses to the target faces. A possible cause of this *negative* priming effect may lie in the implementation of the categorisation task with which priming was measured here. Typically, this task requires 'familiar' versus 'unfamiliar' decisions to famous and unfamiliar faces, and the latter are not seen in the experiment prior to the test stage. This categorisation would pose a problem in the current paradigm, in which the target and foil faces from the identity lineups are no longer completely "unfamiliar" at the stage at which priming is measured but also do not possess the strong familiarity of already known, famous faces. To provide greater clarity over how these faces should be categorised, participants were asked to make 'famous' versus 'not-famous' decisions. However, it is possible that this categorisation task remains too similar to the traditional familiarity decisions. As a consequence, the familiarity that is gained with the target faces in the initial video might interfere with the speeded fame decisions, which could produce the slowing of responses that was observed in Experiment 1 and 2.

If this is the case, then a facilitatory priming effect might be observed with a task in which such response conflicts are avoided. To investigate this possibility, a further categorisation task was included in Experiment 3, in which participants were instructed to classify any faces that were encountered previously in the experiment (i.e., in the initial video or the identity lineup) as 'old' and any previously unseen unfamiliar faces (i.e., the new faces) as 'new'. Note that this task cannot rule out response conflicts entirely. For example, in cases of mistaken lineup identifications, observers' explicit memory for the target faces must be limited. In these cases, observers might therefore experience uncertainty as to whether these faces should be classified as ‘old’ or ‘new’, which could give rise either to facilitatory or negative priming effects. However, targets that have been identified correctly from a lineup should be classified unequivocally, and therefore without conflict, as ‘old’. If facilitatory repetition priming effects can be found for unfamiliar faces in this paradigm, then such effects might therefore be obtained for the target faces from hit trials in the old-new categorisation task.

Method

Participants

Eighty-four participants (74 female, 10 male) from the University of Kent, with a mean age of 19.4 years (*SD* = 3.4), participated in this experiment as a condition of their course. All reported normal or corrected-to-normal vision. None had participated in the preceding experiments.

Stimuli and Procedure

The stimuli and procedure were identical to the first two experiments, except for the following changes. The experiment now consisted of four sequential parts, comprising the screening of the target video (part 1), followed by the identity lineups (part 2). In contrast to the preceding experiments, observers were now always presented with one target-present and one-target-absent lineup in part 2, which were administered, one after the other, in a fully counterbalanced design.

In part 3, observers were then presented with an old/new categorisation task, which comprised the target identities (1 male, 1 female face), the foil faces for target-present and target-absent lineups (10 male, 10 female), and 20 new, previously-unseen unfamiliar faces (10 male, 10 female). These faces were the same images that were used in the fame task of the preceding experiments, and were presented at a size of approximately 8.5 (W) x 10 (H) cm. During the categorisation task, each trial began with a fixation cross for 1000 ms and was followed by a face stimulus, which was displayed until a response was registered. Participants were instructed to classify these faces as quickly and as accurately as possible, by pressing one of two possible keys with their index fingers. They were told explicitly that any faces that had been encountered previously in the experiment, in the video or the identity lineup, should be classified as "old", and any previously unseen faces as "new". All participants completed 42 trials in a randomised order.

The fourth and final part of the experiment then consisted of the same fame categorisation task that was employed in the first two experiments. As before, observers were shown 82 faces, one at a time, comprising the target faces (1 male, 1 female) the foil faces (10 male, 10 female), 20 new unfamiliar faces (10 male, 10 female), and 40 famous faces (20 male, 20 female). Participants were instructed to classify these faces as "famous" or "not famous" as quickly and as accurately as possible.

Results

Lineup Identification Accuracy

Eyewitness accuracy was, once again, error prone. For the target-present lineups, observers recorded 32% hits, 44% misses, and 24% misidentifications. Thus, less than a third of responses reflected the correct identification of the targets. For target-absent lineups, 63% correct rejections and 37% false positives were made.

Accuracy	Targets	Foils	New
Hits	92.6 (26.7)	46.3 (21.5)	81.9 (14.2)
Misses	48.6 (50.7)	55.7 (18.9)	75.4 (17.4)
Misidentifications	45.0 (51.0)	52.5 (22.9)	77.0 (19.8)
Correct rejections	32.1 (47.1)	54.0 (26.0)	80.6 (14.6)
False positives	32.3 (47.5)	56.5 (22.0)	69.4 (20.3)

Table 2.5. Accuracy (%) for the Lineup Targets, Foil Faces, and New Faces in the Old/New Task of Experiment 3, Broken Down by the Outcome of the Identity Lineups. Standard deviations are presented in parentheses.

Old/New Categorisation of Faces

Observers' responses from the old/new categorisation task were analysed next. First, the overall accuracy of observer's responses was calculated, which shows that 52% (*SD* = 18.1) of the target and lineup faces were categorised correctly as 'old', while 77% (*SD* = 14.3) of the remaining unfamiliar faces were classified correctly as 'new'. A breakdown of these data for target, foil and unfamiliar faces by accuracy in the lineup task (i.e., hits, misses, etc.) is provided in Table 2.5. These data show that observers were generally most accurate at classifying the unfamiliar faces as 'new', while performance for lineup foils was poor at around 50%. For the targets, memory was even worse, at between 30% and 50% accuracy. The notable exception here are cases in which a hit was registered. For these cases, accuracy was at over 90%, which is the highest accuracy level in any of the categories here. Overall, these data therefore show that memory for the targets and lineup faces was generally poor in the old/new task, except in cases in which a correct lineup identification had been made previously.

To analyse the observations formally, a series of one-factor within-subjects ANOVAs were conducted to compare accuracy for the target, foil and unfamiliar faces in each of the conditions. These ANOVAs showed an effect of face type for hits, $F(2,52) = 13.01$, $p < 0.001$, misses, $F(2,72) = 6.26$, $p < 0.01$, misidentifications, $F(2,38) = 5.23$, $p < 0.01$, correct rejections, $F(2,104) = 31.57$, $p < 0.001$, and false positives, $F(2,60) = 9.37$, $p < 0.001$. For all conditions, Tukey HSD test showed that accuracy for the target was lower than for the unfamiliar faces, all $qs \geq 3.56$, $p \leq 0.05$, except for hits, for which the target and unfamiliar faces did not differ from each other, $q = 2.66$. In addition, accuracy was lower for the foils than the unfamiliar faces in hits, misses and correct rejections, all $qs \geq 4.82$, $p \leq 0.01$, while the foils did not differ reliably from unfamiliar faces in misidentifications and false positives, both $qs \leq 3.35$. Finally, accuracy was higher for the targets than the foils in the hits condition, $q = 11.46$, $p < 0.001$, equivalent for targets and foils in misses and misidentifications, both $qs \leq 1.27$, and lower for targets than the foils in correct rejections and false positives, both $qs \geq 3.93$, $p \leq 0.05$.

Response Times	Targets	Foils	New	Foils-as-target
Hits	1230 (398, N=25)	1278 (404, N=109)	1055 (292, N=237)	--
Misses	1268 (486, N=18)	1180 (329, N=182)	1288 (871, N=302)	--
Misidentifications	2326 (3429, N=8)	1462 (1284, N=77)	1183 (457, N=171)	--
Correct rejections	2444 (2010, N=13)	1162 (366, N=243)	1160 (363, N=343)	--
False positives	1308 (689, N=9)	1059 (340, N=132)	1447 (1277, N=179)	--
Misidentifications II	2326 (3429, N=8)	1462 (1284, N=64)	1183 (457, N=171)	1322 (1100, N=13)
False positives II	1308 (689, N=9)	1061 (351, N=114)	1447 (1277, N=179)	1167 (482, N=18)

Table 2.6. Response Times (ms) for the Lineup Targets, Foil Faces, and New Faces in the Old/New Task of Experiment 3, Broken Down by the Outcome of the Identity Lineups. Standard deviations are presented in parentheses.

The response times for old/new decisions were calculated next. Incorrect responses were excluded from this analysis and the cross-subject median RTs were then generated for target, foil and unfamiliar faces across the conditions. These data are shown in Table 2.6. For the statistical analysis, responses to male and female lineups were once again treated as separate instances, on a between-subject basis. A series of one-factor within-subjects ANOVAs found no differences between response times to the target, foil and unfamiliar faces for misses, $F(2,34) = 0.87$, $p = 0.43$, misidentifications, $F(2,16) = 0.77$, $p = 0.48$, and false positives, $F(2,16) = 1.01$, $p = 0.39$, but an effect of face type was found for hits, $F(2,46) = 3.53$, $p < 0.05$, and correct rejections, $F(2,30) = 5.92$, $p < 0.01$. For hits, Tukey HSD test showed that response were slower to foils than to new faces, $q = 3.74$, $p < 0.05$, but did not differ between targets and foils, $q = 1.54$, and targets and new faces, $q = 2.20$. For correct rejections, responses were slower to targets than foils, $q = 4.30$, $p < 0.05$, and new faces, $q = 4.12$, $p < 0.05$, but did not differ between foils and new faces, $q = 0.18$.

In an additional step, RT data for the foils were split into faces that were mistakenly selected from a lineup as the target and the remaining lineup faces (see Misidentifications II and False Positives II in Table 2.6). A one-factor within-subjects ANOVA with the levels target, foil, unfamiliar faces and misidentified/falsely identified foils-as-targets also did not show an effect of face type in the misidentification condition, $F(3,18) = 0.68$, $p = 0.57$, and for false positives, $F(3,12) = 1.91$, $p = 0.18$.

Overall, the accuracy data therefore show that observers were more likely to remember the target than the lineup foils when a hit response had been made earlier, than when misses or misidentifications were recorded. The accuracy data of the old-new task therefore corroborates observers' responses for target-present lineups. In addition, in the target-absent condition, the target faces were remembered significantly less than the foils from the lineups. This contrast between the target-present and target-absent condition indicates that seeing the targets in the lineups, in addition to the initial exposure to these identities in the video, helps to improve observers' explicit memory for these faces to some extent in the old/new task. However, the response times for the old/new task, which is the data of main interest here, generally cannot dissociate the targets from lineup foils and new faces.

Famous/Not Famous Categorisation of Faces

Observers response for the famous/non-famous categorisation task were analysed next. Overall, 97% (SD = 4.8) of unfamiliar and 91% (SD = 9.2) of famous faces were categorised correctly. A breakdown of these data for target, foil and unfamiliar faces is provided in Table 2.7 and shows that accuracy was generally high regardless of the lineup response that was registered (e.g., hits, misses, etc.). A series of one-factor within-subjects ANOVAs of target, foil and unfamiliar faces found no differences for hits, $F(2,52) = 1.01$, $p = 0.37$, misses, $F(2,72) = 0.13$, $p = 0.88$, misidentifications, $F(2,38) = 0.23$, $p = 0.80$, and correct rejections, $F(2,104) = 0.69$, $p = 0.51$. However, an effect of face type was found for false positives, $F(2,60) = 4.37$, $p < 0.05$, which reflects lower accuracy for the target than the foil and unfamiliar faces, $q = 3.58$, $p < 0.05$ and $q = 3.66$, $p < 0.05$, respectively. The accuracy for foil and unfamiliar faces did not differ, $q = 0.08$.

Accuracy	Targets	Foils	New
Hits	92.6 (26.7)	97.9 (4.4)	97.8 (4.2)
Misses	97.3 (16.4)	98.2 (7.6)	97.6 (6.4)
Misidentifications	95.0 (22.4)	97.8 (5.8)	97.0 (6.6)
Correct rejections	98.1 (13.7)	96.9 (8.0)	96.0 (8.4)
False positives	83.9 (37.4)	97.1 (7.6)	97.4 (5.8)

Table 2.7. Accuracy (%) for the Lineup Targets, Foil Faces, and New Faces in the Fame Task of Experiment 3, Broken Down by the Outcome of the Identity Lineups. Standard deviations are presented in parentheses.

The data of main interest, observers' response times, were analysed next. Incorrect responses were excluded from analysis and the cross-subject median RTs were then calculated for all conditions (e.g., hits, misses, etc.), broken down by responses to targets, foils and unfamiliar faces. These data are also provided in Table 2.6 and shows that RTs were slower to the target faces than the lineup foils and the unfamiliar faces in all conditions. For the statistical analysis, responses to male and female lineups were treated as separate instances, on a between-subject basis. Once again, some cases had to be excluded from the analysis, in which the target faces were incorrectly categorised as famous. This left 25 cases for the analysis of hits (two excluded), 36 cases for misses (one excluded), 19 for misidentifications (one excluded), 52 for correct rejections (one excluded), and 26 for false positives (five excluded).

A series of one-factor within-subjects ANOVAs found an effect of face type for hits, $F(2,48) = 6.78$, $p < 0.01$, misses, $F(2,70) = 4.14$, $p < 0.05$, misidentifications, $F(2,36) = 3.73$, $p <$

0.05, correct rejections, $F(2,102) = 27.20$ $p < 0.001$, and false positives, $F(2,50) = 5.71$, $p < 0.01$. In all of these categories, Tukey HSD test showed that observers were slower to respond to targets than to foil faces, all $qs \geq 3.58$, $ps \leq 0.05$, while foils and unfamiliar faces did not differ, all $qs \leq 0.53$. In addition, response were also consistently slower to target than to unfamiliar in all categories, all $qs \geq 3.39$, $ps \leq 0.05$, except for misidentifications, for which these conditions did not differ, $q = 3.05$. Overall, these results therefore show that observers' responses in the fame categorisation task distinguish between target and foil identities regardless of the type of lineup decision that was originally made.

Repetition priming of misidentified and falsely identified foils

The RT data for the repetition priming task was also analysed further by splitting the lineup foils into faces that were mistakenly selected from a lineup as the target (i.e., the misidentified or falsely identified targets) and the remaining lineup faces (see Misidentifications II and False Positives II in Table 2.7). Note that these data comprise 20 cases for misidentifications and 31 cases for false positives but 6 cases that did not yield a correct response to the target were excluded (1 for misidentifications, 5 for false positives). A one-factor within-subjects ANOVA with the levels target, foil, unfamiliar faces and misidentified foils-as-targets show a marginally significant effect of face type in the misidentification condition, $F(3,54) = 2.77$, $p = 0.05$. Tukey HSD test showed a negative priming effect for the target face in comparison with the lineup foils, $q = 3.79$, $p < 0.05$, but none of the other comparisons between face types were significant, all $qs \leq 3.19$. An analogous ANOVA for false positives also found a main effect of face type, $F(3,75) = 3.07$, $p < 0.05$, but Tukey HSD found no reliable differences between the target, foils, unfamiliar faces, and foil faces that were falsely identified as targets, all $qs \leq 3.56$.

Response Times	Targets	Foils	New	Foils-as-target
Hits	946 (609, N=25)	690 (190, N=238)	666 (116, N=264)	--
Misses	1216 (1475, N=36)	770 (432, N=327)	799 (418, N=361)	--
Misidentifications	1049 (866, N=19)	686 (142, N=171)	738 (236, N=194)	--
Correct rejections	1085 (520, N=53)	736 (215, N=460)	725 (174, N=513)	--
False positives	946 (451, N=26)	739 (292, N=259)	740 (224, N=302)	--
Misidentifications II	1049 (866, N=19)	682 (128, N=152)	738 (236, N=194)	823 (295, N=19)
False positives II	946 (451, N=26)	728 (272, N=228)	740 (224, N=302)	881 (525, N=31)

Table 2.8. Response Times (ms) for the Lineup Targets, Foil Faces, and New Faces in the Fame Task of Experiment 3, Broken Down by the Outcome of the Identity Lineups. Standard deviations are presented in parentheses.

To determine if a priming effect might emerge for foils that were misidentified as targets with a bigger sample size, the data were combined for misidentified foils from target-present lineups in Experiment 1 and 3 and from target-absent lineups in Experiment 2 and 3. Two separate ANOVAs of these data showed a main effect of face type for target-present, $F(3,108) = 3.81$, $p < 0.05$, and target-absent lineups, $F(3,108) = 3.81$, $p < 0.05$. In both conditions, this effect reflects slower response to the targets than to foils and unfamiliar faces, all $qs \geq 3.76$, $ps \leq 0.05$. None of the other comparisons were significant, all $qs \leq 3.50$.

Repetition priming as a measure of individual eyewitness accuracy

So far, the results show a repetition priming effect for the target faces, independent of the lineup identification decisions that eyewitnesses had previously made. In a final step of the analysis, the extent to which repetition priming can provide such an index at the level of the individual was explored. For this purpose, the data from all three experiments were combined. This was done separately for the male and female lineups, and for each lineup type (i.e., target-present and -absent). For example, for the target-present female identity lineup, the mean RT and standard deviation for all unfamiliar female faces (i.e., the female target, 9 female foils, and 10 remaining unfamiliar female faces) were individually calculated for each observer. These means and standard deviations were then used to convert each individual's target RT into a z-score. Note that most observers therefore contribute two z-scores to this analysis, corresponding to the female lineup and the male lineup, unless they failed to register a correct response to one or both of the targets in the fame categorisation task.

These z-scores were converted into conditional probabilities according to Bayes' theorem, using the following formula:

$$P(A/B) = P(B/A)P(A) / P(B).$$

Where P(A) represents the probability that a correct lineup identification was made, P(B) represents the probability that a target RT has a z-score above 1.96 (thus denoting a score that is two standard deviations, i.e., significantly at $p < 0.05$, above an observer's mean RT for all unfamiliar male/female faces), and P(B/A) denotes the proportion of correct lineup identifications for which a z-score over 1.96 was recorded. The outcome of this formula (i.e., P(A/B)) provides a measure of the probability that an original lineup identification was correct if a significant z-score (of over 1.96) is obtained subsequently for a target face in the repetition priming test.

This probability was calculated for three separate instances. In the first instance, the focus was on cases in which a lineup identification was made (i.e., a hit, misidentification or a false positive). In this context, P(A/B) therefore provides a measure of the probability that a correct eyewitness identification was made given that a face was actually selected from a lineup. For these instances, P(A/B) = 0.23 * 0.38 / 0.15. This gives a probability of only 0.58 that a correct lineup identification has been made when a concurrent priming effect is found for the target face. However, the sensitivity of this approach is compromised because this analysis does not compare z-scores on a like-for-like basis for correct lineup identifications (i.e., hits) and incorrect identifications (i.e., misidentifications, false positives). Specifically, while this analysis calculates the z-scores for the targets on hit trials, this analysis also focuses on the z-scores of these same faces when misidentifications and false positives were made, despite the fact that a different face identity was selected in these cases as the target.

If the purpose of this analysis is to relate specific z-scores to a face that was previously selected from a lineup, to determine if this face was initially identified correctly (or incorrectly), then it is therefore important to contrast the z-scores for the target face on hit trials with the z-scores for the selected foil faces when a misidentification or false positive was recorded. For these instances, P(A/B) = 0.23 * 0.37 / 0.12, which gives a probability of 0.71 that a correct lineup identification has been made when a concurrent priming effect is found for the face that was selected from a lineup.

This analysis was also applied to cases in which no lineup identification was made (i.e., correct rejections and misses). For these instances, P(A/B) = 0.21 * 0.57 / 0.19, which gives a probability of 0.63 that a correct lineup rejection has been made when a concurrent priming effect is found for the target face. These analyses indicate repetition priming can also provide a probability index to estimate whether an observer has made a correct eyewitness identification decision.

Discussion

As in the preceding experiments, a robust negative priming effect for the target identities was found, and this effect was present irrespective of the lineup decision that was initially made. This experiment therefore provides further evidence that repetition priming can provide a covert index of eyewitness accuracy, even when observers are unable to identify a target overtly. In line with Experiment 1 and 2, the data for the repetition priming task were also analysed further by splitting the lineup foils into faces that were mistakenly selected from a lineup as the target and the remaining lineup faces. The response times to these selected foils were faster than for the actual target faces (i.e. they were not negatively primed), and did not differ from the other lineup faces and new unfamiliar faces. This is an important finding because it suggests that repetition priming effects are not found for the faces that observers select erroneously from identity lineups, but only for the *actual* target identities.

Experiment 3 also sought to determine whether the negative priming effects of the fame categorisation task can be converted into a facilitatory effect when an old/new face categorisation task is used. It was specifically predicted that such an effect should be most clearly visible for target faces that were also identified correctly from the lineups, but expected mixed results for mistaken or incorrect lineup identifications. It was found that accuracy for the target faces was indeed high, at over 90%, after a hit had been recorded and was lower than 50% in all other cases, which converges with the predictions. However, the RT data generally failed to show clear negative or facilitatory priming in this task.

While the current findings are unable to explain this outcome, there is one possibility. Such old/new decisions rely on episodic memory and therefore may be unsuitable for measuring repetition priming. Previous research has shown, for example, that priming is only found with tasks that require access to identity information of a face, such as familiarity and semantic decisions (Ellis et al., 1990). By contrast, priming is not found onto sex or expression decisions, even when the face identities were categorised according to the same criteria in the initial priming phase (Ellis et al., 1990). This indicates that priming is only found for decisions that require direct access to stored facial identity information and cannot be elicited by episodic memory alone. This suggests that the old/new task might, in fact, have been inappropriate to produce priming effects in Experiment 3.

In light of this shortcoming, it is notable that it is difficult to design a task that requires access to the cognitive system for facial identity processing but is not liable to a response conflict between already-known familiar faces, unfamiliar faces, and *familiarised* unfamiliar faces. Repetition priming can be obtained with unfamiliar faces with a sex decision task when external facial features, such as hairstyle, are removed from faces during stimulus encoding (Goshen-Gottstein & Ganel, 2000; Martin et al., 2009; Martin et al., 2010; Martin & Greer, 2011). However, this alternative would not be feasible in scenarios in which the whole face is visible. This problem therefore awaits solution but the old/new task does provide an interesting advantage here. The inclusion of this task shows that the negative priming effects of the fame task can survive the intervening presentation of the entire cohort of unfamiliar faces (targets, foils and new faces). This reinforces the reliability of the fame categorisation task for measuring these effects.

General Discussion

The aforementioned effects have never been explored before in an eyewitness paradigm but are not new in the face recognition literature. Prosopagnosic patients, who cannot recognise faces overtly, can show signs of covert recognition in priming tasks (Young, Hellawell, & de

Haan, 1988). Similar effects can be observed in normal subjects when prime faces are presented too briefly to allow overt recognition (Morrison et al., 2000) or when faces are presented under high attentional load so that explicit memory for these faces is reduced (Jenkins et al., 2002). Models of face recognition can also provide an explanation for these effects (e.g., Burton, Bruce, & Hancock, 1999; Burton, Bruce, & Johnston, 1990; Schweinberger & Burton, 2003). According to such models, the recognition of a face first requires the activation of a stored visual representation *and* an overt familiarity response at a subsequent processing stage. While the presentation of a face can be sufficient for the activation of stored visual representations, the accompanying activation of a familiarity response can remain at a sub-threshold level that is insufficient for triggering overt recognition (e.g., Burton, Bruce et al., 1999; Eimer, Gosling, & Duchaine, 2012; Morrison et al., 2000; see also Burton, Young, Bruce, Johnston, & Ellis, 1991; Young & Burton, 1999).

The current effects appear to arise from such sub-threshold familiarity responses. Accordingly, the stored representations that have been formed of the target faces during the initial exposure in the video might be sufficient for supporting repetition priming but can be insufficient for triggering overt recognition (e.g., Eimer et al., 2012; Jenkins et al., 2002; Morrison et al., 2000). As a consequence, it is possible for observers to fail to identify the sought-after target from a lineup whilst also showing priming response of the same identity.

A limitation of this paradigm is that the initial familiarity could have been gained from a previous exposure to the target entirely unrelated to the crime. It has been established that priming effects are robust (Bruce & Valentine, 1985) and can be triggered without overt recognition (Jenkins et al., 2002; Morrison et al., 2000). This combination of effects could mean that a person who appears to show priming to a perpetrator from a crime may in fact have been primed by a different occasion, possibly without their knowledge. This problem is not unique to the current study, all eyewitness procedures can be undermined by such a previous exposure. However, this methodological shortcoming is not resolved here.

Finally, this chapter examined the question of whether repetition priming can be used to predict accuracy at the individual level. A Bayesian analysis was conducted to convert individual z-scores into conditional probabilities of a correct identification. This analysis showed some promise but was not definitive in its diagnosticity. Further to this is the drawback that all differences in response latency identified in this chapter are based on mean scores. These means were not calculated in a like for like manner since the priming task required participants to respond to uneven numbers of faces in each category. For every one target response, participants were required to respond to nine foils (or ten in the target-absent conditions) and ten unfamiliar faces. To illustrate, the reaction times for the targets throughout this chapter are based on between participant medians, as are those for foils and new faces. The difference is that for each lineup, every participant contributes one face to the target average, but nine foils and ten new faces. Due to this imbalance, outlying response times for the target faces would have had much more effect on the mean score than in any other category. Here, the target faces reliably elicited slower reaction time confirming the effect to be genuine, but the possibility of an unrepresentative result is further cause for concern when considering application for predicting individual accuracy. Repetition priming appears to be a tool which can detect covert recognition but in its current state it is not sufficiently sensitive to be used at the individual level.

In summary, this chapter examined whether repetition priming can provide an index of accuracy in eyewitness scenarios. Overall, eyewitness identification accuracy was poor. For example, observers managed to identify the targets in only 41% of cases in Experiment 1 and mistakenly selected a different face on 19% of encounters. And in Experiment 2, the absence of the target from the lineups was noted on only 52% of trials, while observers made a false identification on 48% of trials. This poor accuracy was expected and is, in itself, not novel (e.g., Bruce et al., 1999; Megreya & Burton, 2006a, 2008). However, these experiments also revealed a consistent negative priming effect, whereby target responses were slowed in a subsequent categorisation task in comparison with lineup foils and new unfamiliar faces. Importantly, this effect was observed regardless of whether observers had initially managed to identify the correct target from a lineup, had misidentified a wrong person as the target, or had deemed the target to be absent. This indicates that repetition can provide a covert index of recognition that indicates prior exposure to a target even when observers cannot make such an identification explicitly.

The aim of this thesis is to explore methods for the diagnosis of eyewitness accuracy. To be ultimately useful a method must be able to diagnose accuracy for an individual. The Bayesian analysis adopted in this chapter demonstrated that repetition priming is not able to provide this level of diagnosticity. The next chapter will examine another possible method of postdicting accuracy, multiple face lineups.

Chapter 3

Examining Recognition Accuracy with Multiple Face Lineups

Introduction

Chapter 2 introduced repetition priming as a potential method for postdicting eyewitness accuracy. Measurement of speeded responses demonstrated a difference between the target and the other faces in the experiments. However, these differences were observed only in the means averaged over the sample. An examination of the individual scores did not provide a reliable method for distinguishing between right and wrong identifications and were, consequently, unable to predict accuracy at the individual level. Many observers would be necessary to use repetition priming as a postdictor of accuracy and this is not always a possibility.

It is common for multiple observers to be required in many existing postdicting methodologies (e.g., Sporer, 1992; Sauerland & Sporer, 2007, 2009; Charman & Cahill, 2012). However, in order to be of realistic application as a postdictor, a measure must be sensitive enough to discriminate between correct and incorrect identifications when only one witness is available. It has been shown that it is possible to diagnose accuracy at the individual level (e.g. Bindemann, Brown et al., 2012; Bindemann, Avetisyan et al., 2012; Megreya & Bindemann, 2013), but no evidence exists where participant accuracy for the target lineup has been tested directly. The purpose of this chapter is to introduce such a measure by considering the use of multiple trials.

Wells and Luus (1990) likened a criminal lineup to a laboratory experiment in which the hypothesis ‘that a perpetrator is present’ is tested by examining eyewitnesses’ responses to this task. However, there is an important conceptual difference between criminal lineup proceedings and laboratory experiments; the inclusion of multiple trials. In face recognition experiments, participants are never tested on only one trial. It is well understood that a single data point may be an outlying value and may not be typical. Despite this, criminal lineups and most research conducted in this field still rely on one observation for each participant. Several studies have recognised this shortcoming and have employed multiple lineups to test identification accuracy (Lindsay et al, 1987; Pryke et al., 2004; Sauerland & Sporer, 2008; Sauerland et al., 2013). However all of these studies have included person aspects other than the face to make up the additional lineups. These other aspects are consistently met with low identification accuracy and do not provide an optimum test of recognition since it is well-known that the face is the primary source of visual information when making an identification (Burton, Wilson et al., 1999; O'Toole et al., 2011; Robbins & Coltheart, 2012; Bruce & Young, 1986).

This chapter seeks to explore this face advantage further, by exploring a new variant of the multiple-lineup procedure. In this procedure, observers are required to identify a target from multiple lineups that are composed *only* of faces. This manipulation is logical given the comparatively high recognition accuracy for faces in previous multiple-lineup studies, but also has a strong theoretical grounding in the face perception literature. According to cognitive theories of face processing (e.g., Burton et al., 1990; Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000; Schweinberger & Burton, 2003), the successful recognition of familiar people, such as family, friends or colleagues, is highly robust and can be triggered by *any* instance of their face. The ultimate hallmark of accurate person identification is therefore the ability to recognise the same person's face *repeatedly*, across many different encounters.

In line with this theorising, eyewitness identification errors are made rarely when the perpetrator is someone that is already known to a witness (e.g., Memon et al., 2011). A different picture emerges when eyewitnesses are required to identify *unfamiliar* people, of which they have only limited perceptual experience, such as the brief exposure to a person at a crime scene. The identification of such people can be rather difficult, even under best-possible conditions (e.g., Bruce et al., 1999; Megreya & Burton, 2006a, 2008; Memon et al., 2011). Moreover, in contrast to familiar face recognition, the *repeated* identification of unfamiliar faces is also difficult. As a result, observers might recognise a person in one instance but fail to do so a few moments later (Alenezi & Bindemann, 2013; Bindemann & Sandford, 2011), or on one day but not on the next (Bindemann, Avetisyan et al., 2012).

Considering the well-documented difficulty of eyewitness identification (e.g., Wells et al., 2006; Wells & Olson, 2003), it is expected that accuracy for any of the individual lineups will be error-prone. It is less clear to what extent the repeated identification of the target face is possible. If a single identification reflects a robust recognition, the participant should be able to identify the target repeatedly. However, previous research suggest that this is not the case (Burton et al., 2005; Jenkins & Burton, 2008; Jenkins et al., 2011). It is likely therefore, that some observers will be able to identify faces repeatedly but others may not. The presence of this distinction between participants in the task will provide evidence that multiple face identifications can distinguish between accurate and inaccurate participants at the individual level.

Experiment 4

This experiment introduces a new means of assessing eyewitness accuracy. Participants were exposed to two target identities in a video. The participants were then required to select these targets from several identity lineups which always comprised one of the targets amongst an array of foils. The aim of this experiment is to investigate the extent people are able to identify faces repeatedly.

Method

Participants

Thirty undergraduate students (23 female, 7 male) from the University of Kent with a mean age of 20 years (*SD* = 2.3) took part in this experiment as a condition of their course. All participants reported normal or corrected-to-normal vision.

Stimuli and Procedure

There were two parts to this experiment. First participants watched a short video which featured two target identities (part 1). Next they were presented with three identity lineups to assess their recognition of these two people (part 2).

Part 1: Video exposure to targets

The stimulus materials for part 1 consisted of a video of a male and a female target, who were shown in conversation for 60 seconds. The faces of both targets were visible across a range of views (e.g., frontal, ¾ and profile view) throughout. The video was presented at a size of 30 (W) x 16.8 (H) cm on a standard computer monitor and did not contain sound. For illustration, example stills from the video are presented in Figure 3.1.

Part 2: Lineup identification of targets

Next, for each target, observers were given three identity lineups to provide separate tests of eyewitness accuracy. Each of these identity lineups consisted of a photograph of a target's face and nine non-target foil faces, which were shown simultaneously, alongside the target, composing two rows of five faces. No foil identities were repeated in more than one lineup. The foil faces were taken from the Glasgow University Face Database (Burton et al., 2010),

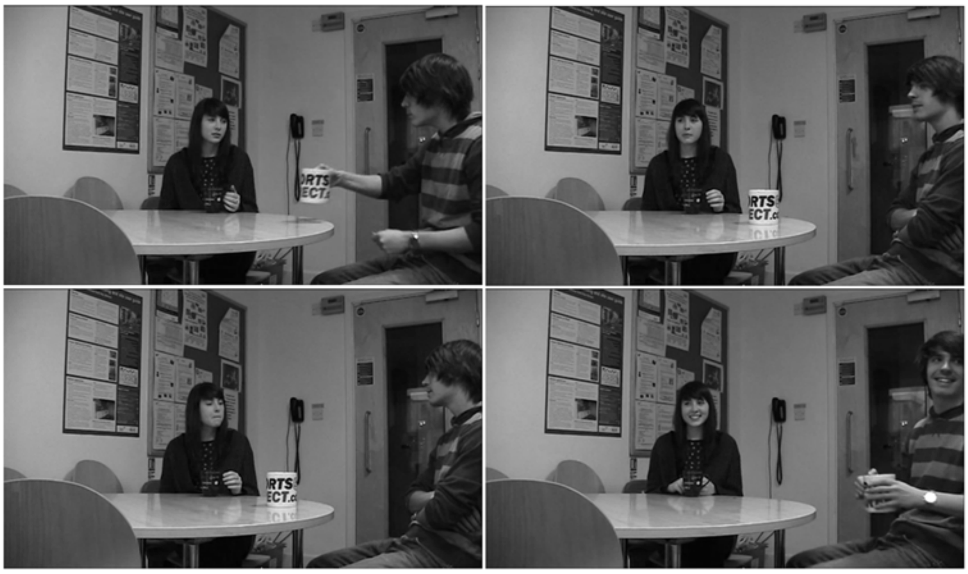

Figure 3.1 Screenshots from the target presentation video shown to participants in part 1.

and were chosen by the experimenters to be of the same sex and of similar age and appearance to the target in each lineup. The photographs were standardised by cropping clothing and background (for an example of these lineups, see Figure 3.2).

In the lineups, each face was shown from a frontal view and with a neutral expression at a size of approximately 5.5 (W) x 7.5 (H) cm. Whether the male or female lineups were shown first was counterbalanced between participants. The three lineups containing each target were presented in a random order.

Observers were asked to study each lineup closely and to decide whether the male/female target was present or absent. Although a deception, participants were told there was an equal chance that the target would be present or absent in each lineup to allow for the rejection of the lineups when observers were unable to identify a face as the target. Participants indicated their responses by pressing the number key, on a standard computer keyboard, that

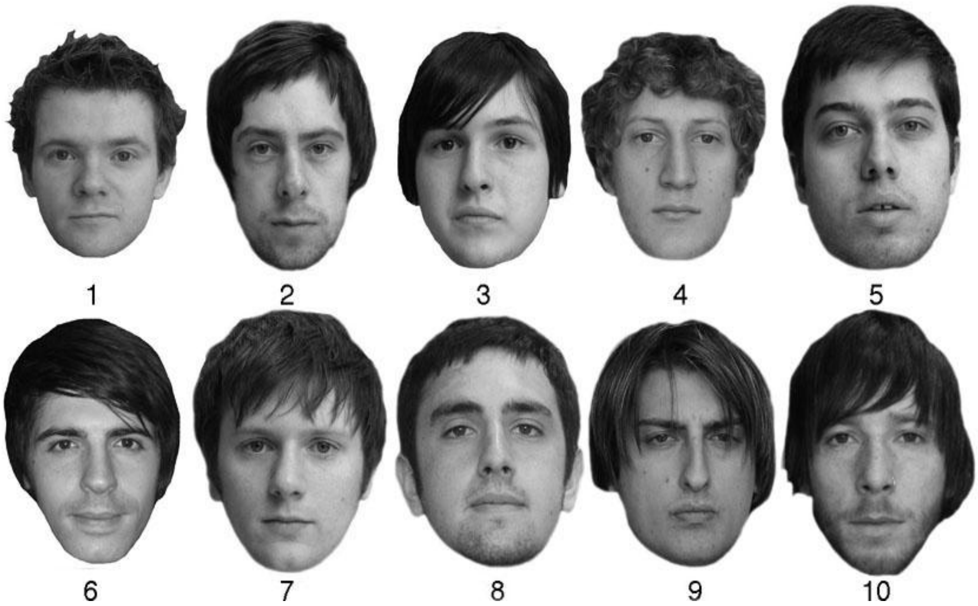

Figure 3.2 Example lineup from part 2 containing a target face and nine foils.

corresponded to the lineup location of the target (e.g., "1" for face 1, "2" for face 2, etc., "0" for face 10) or by pressing "a" if they believed the target absent from a lineup. They were asked to respond as accurately as possible and were told that there was no time limit for the task.

Results

In the first step of the analysis identification accuracy was calculated separately for each trial. Since all lineups contained the target there were three possible responses, a participant could score a hit (correctly identifying the target), a miss (incorrectly stating that the target was absent from the lineup), or a misidentification (incorrectly identifying a foil as the target). These data are presented in Table 3.1 with the order participants encountered each lineup preserved (i.e., target-present lineup 1 refers to the first lineup that was encountered within the trial sequence). These data show that observers were generally able to identify the target on between 42% and 63% of trials. The target was mistakenly declared absent in between

Target-present	Lineup 1	Lineup 2	Lineup 3
Hits	63.3	41.7	63.3
Misses	33.3	53.3	31.7
Misidentifications	3.3	5.0	5.0

Table 3.1 Percentage breakdown of Participants in Target Present Lineups (N=30) for Experiment 4.

32% and 53% of trials. Misidentifications occurred less frequently, in between 3% and 5% of trials.

These data show that the identification accuracy was rather error-prone in any of the individual lineups. However, the inclusion of multiple trials meant it was possible to examine the data further by considering consistent accuracy. In order to distinguish between occasionally and consistently accurate participants, the three lineups were scored in series. The results are shown in Figure 3.1 with the order the lineups were shown preserved. A consistent accuracy score was calculated by adding only the scores that were correct for that lineup *and* had also been correct in every other lineup to that point without interruption. For example, if a participant was correct on lineups 1 and 3 they would only be considered consistently accurate on lineup 1 because of the intervening error. In this way inaccurate eyewitnesses were systematically removed from the analysis, and by the last trial only those who were correct in all three lineups remained.

In Figure 3.1 one-off and consistent accuracy are, by definition, equal in lineup 1. Consistent accuracy refers to repeated correct identifications and in lineup 1 a correct identification must be considered consistent. There appears to be a small disparity between one-off (42%) and consistent accuracy (27%) in lineup 2, however a Chi-Square test contradicts this: χ^2 (1, N = 30) = 3.01, p = .06, phi = -.16. By lineup 3 the distance between the scores is greater and there is now a substantial difference between the one-off (65%) and consistent accuracy (20%): χ^2 (1, N = 30) = 24.86, p < .01, phi = -.46. These data demonstrate that hits on any one of the lineups occurred between 42% and 63% of occasions, but hits in consecutive lineups fell from 63% to 20%. This indicates that some participants who are able to identify a target on one or two lineups fail to do so consistently.

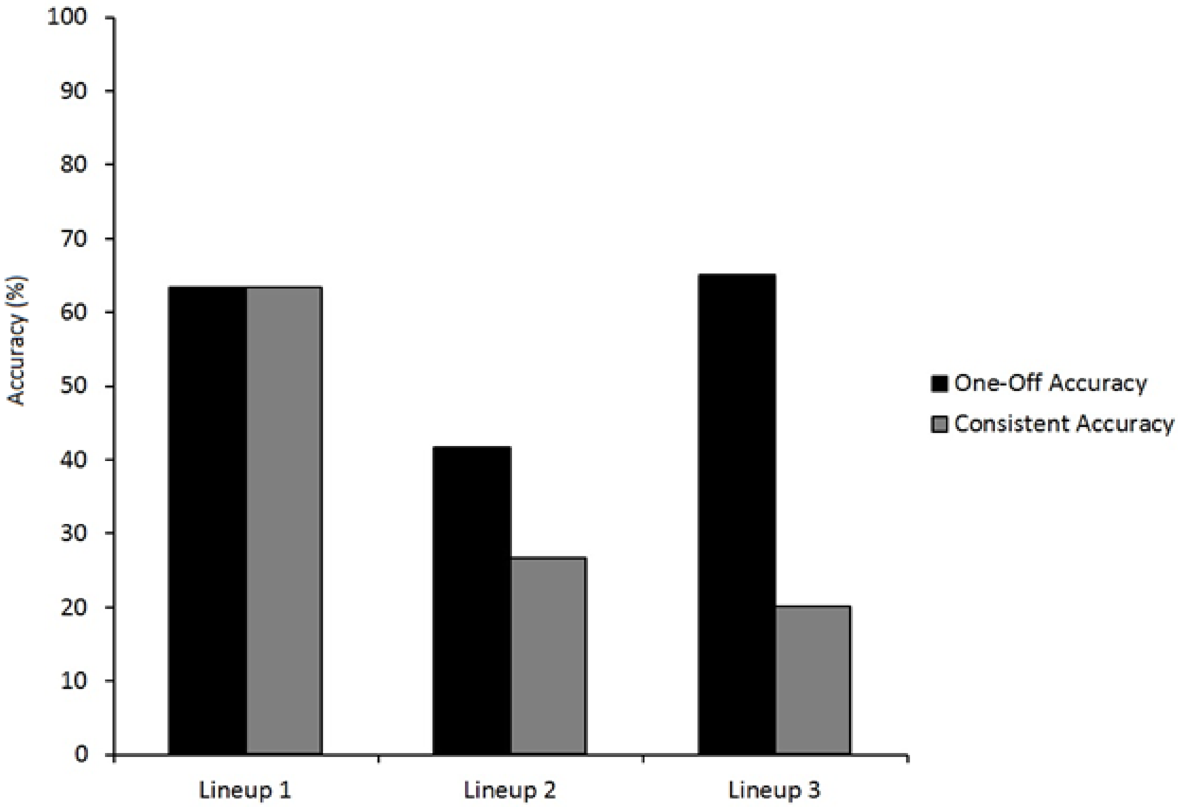

Figure 3.1 Percentage of eyewitnesses able to identify the target face as a one-off on a lineup and consistently in Experiment 4.

The progressive fall in consistent scores was examined next. In order to investigate the step-wise reduction, each lineup was compared to the next in pairs. There appears to be a difference between lineups 1 (63%) and 2 (27%), this was confirmed: χ^2 (1, N = 30) = 16.30, p < .01, phi = -.37. The next step down, between lineups 2 (27%) and 3 (20%), was not a substantial reduction in accuracy: χ^2 (1, N = 30) = .75, p = .26, phi = -.08. There was a drop over the course of the whole series, between lineups 1(63%) and 3(20%): χ^2 (1, N = 30) = 23.18, p < .01, phi = -.44. The difference in consistent accuracy between lineups 1 and 3 cannot be explained by any one consecutive pairing meaning the fall in accuracy is a cumulative effect of the multiple lineups.

Discussion

In this experiment participants were required to watch a video featuring two targets. They were then asked to identify these people in three target-present lineups. Accuracy on any one of these lineups fell between 42-63%. However, if responses were examined for consistency it emerged that only 20% of observers could identify the target in all lineups. The disparity between the accuracy for a single lineup and the consistent accuracy score for all three demonstrates how misleading a single measure of accuracy can be.

Since 20% of participants could repeatedly identify the target it would seem this task is achievable. However, it is possible that some of the remaining participants failed to identify the target repeatedly because of the images used rather than the task itself. To ensure that this is not the case, and that the different images used are identifiable as the same (target) identity, the next experiment tested whether the stimuli allowed repeated identifications under optimal circumstances. It replicated the methodology but tested participants who were familiar with the target faces. It is well-known that familiar face recognition is a highly accurate process (Burton, 2012; Burton et al., 1999) and if the stimuli and procedure are not artificially hard the repeated identification of familiar targets should not prove difficult.

Experiment 5

In Experiment 4 it was found that although many participants were able to identify a target in at least one lineup, far fewer were able to do so repeatedly. In order to ensure the difficulty was a product of the task and not of the images used, Experiment 4 was replicated using participants who were previously *familiar* with the two targets. If multiple identifications are possible without error under these circumstances, it will confirm that no observers were eliminated from Experiment 4 based on superficial aspects of the images used.

Method

Participants, Stimuli and Procedure

Nine postgraduate students (6 female, 3 male) with a mean age of 25.1 years (*SD* = 3.1) participated in this experiment. All participants had normal or corrected-to-normal vision and no participants had taken part in Experiment 4. The methodology of this experiment was exactly the same as that of the previous one except that the participants were already *familiar* with the two targets. All were friends or colleagues of the targets and had met them both several times over a period of at least two months. Participants were not familiar with any foil faces in the experiment.

Results

The data were prepared in the same way as in Experiment 4. Accuracy for both one-off and consistent responses was at 100% across all lineups. It is clear from this evidence that when participants are asked to identify familiar targets they are able to do this consistently and without error.

Discussion

This experiment tested the stimuli used in this chapter. When participants who were familiar with the targets were tested, their identification accuracy was perfect. This finding converges with previous research in familiar face recognition (e.g., Burton et al., 1990; Bruce & Young, 1986; Haxby et al., 2000; Schweinberger & Burton, 2003) and demonstrates that the task set for participants is achievable and the stimuli used are not unduly difficult. This is important because this procedure is designed to test acquired familiarity with the target faces. If it is not possible to identify them repeatedly even with maximal familiarity this test is not measuring the variable of interest. So far this chapter has shown that it is possible for some people to identify a previously seen face over multiple trials but as yet it has not been tested whether the absence of this face is also detectable. To examine this, a third experiment was conducted that included both target-present and target-absent lineups.

Experiment 6

Experiment 4 showed that some participants could identify a target multiple times while others could not. However, until now target-absent lineups have been neglected. It has been previously shown that ability to identify the correct face in a target-present lineup and to reject a target-absent lineup are dissociable (e.g., Bruce et al., 1999; Megreya & Burton, 2006a; Megreya & Burton, 2007). This means that even those participants who were able to identify the target in all three lineups of Experiment 4 may be unable to detect when the target is absent in another array. Eyewitnesses can make mistakes in both failing to identify a present target and in falsely identifying an innocent person as the perpetrator. In order to have confidence in the participant it is equally important that they are able to identify a target when they are present, and reject a target-absent lineup. Multiple trials must show sensitivity to both of these tasks.

In Experiment 4 participants were presented with three target-present lineups for each identity. This methodology was extended in the current experiment by including three target-absent lineups for each. Therefore, participants were presented with six lineups for each identity, three that included the target and three that did not.

Method

Participants

Thirty-five undergraduates (33 female, 2 male) with a mean age of 19 years (*SD* = 2.0) participated in this experiment as a condition of their course. All participants had normal or corrected-to-normal vision. No participants had taken part in Experiments 4 or 5 and all were unfamiliar with the targets and foils.

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 4 except for the addition of three target-absent lineups for each identity making a total of 12 lineups seen by each participant. The faces in these lineups were presented at approximately 5.5 (W) x 7.5 (H) cm to match the existing stimuli. No foil identity was repeated in more than one lineup. Whether the participant saw the male or female lineups first was counterbalanced and the sequence of the six lineups within each target group was randomised. Participants indicated their responses by pressing the number on a computer keyboard that corresponded to the lineup location of the target (e.g., “1” for face 1, “2” for face 2, etc.) or by pressing “a” if they judged the target to be absent from a lineup.

Results

As in Experiment 4, participants were scored for accuracy on each lineup. Since target-absent lineups were now included, there were two more possible responses, a correct rejection (correctly stating a target-absent lineup to be target-absent) and a false positive (incorrectly identifying a foil as the target in a target-absent lineup). The full breakdown of these scores can be seen in Table 3.2 with the lineups split by target-presence or –absence but with the presentation order preserved within these categories. The percentage scores in target-present lineups were similar to those of Experiment 4, hits ranging from 49% to 64%. Misses occurred on between 31% and 40% of trials, and misidentifications on between 4% and 16% of occasions. Together, these scores once again demonstrate that this task is error-prone.

Target-present	Lineup 1	Lineup 2	Lineup 3
Hits	48.6	64.3	55.7
Misses	35.7	31.4	40.0
Misidentifications	15.7	4.3	4.3
Target-absent	**Lineup 1**	**Lineup 2**	**Lineup 3**
Correct rejections	75.7	75.7	80.0
False Positives	24.3	24.3	20.0

Table 3.2 Percentage breakdown of Participants (N=35) for Experiment 6.

The scores for the target-absent lineups show that participants are similarly fallible at recognising when the target is not present. Correct rejections occurred on between 76% and 80% of trials and false positives ranged from 20% to 24%.

The data of most interest concern performance across the lineups. Figure 3.4 shows that one-off accuracy on any lineup (i.e. correct identifications and rejections combined) was fairly stable and ranged from 57% to 74% but consistent accuracy fell from 71% to 23%. There appears to be a small difference between scores for one-off accuracy (57%) and consistent accuracy (49%) in lineup 2: χ^2 (1, N = 35) = 1.03, p = .2, phi = -.09. The difference (63% and 37%) has grown larger by lineup 3: χ^2 (1, N = 35) = 9.26, p < .01, phi = -.26. Analysis of one-off and consistent accuracy for lineups 4, 5 and 6 all confirmed a substantial difference: χ^2 (1, N = 35) = 20.87, p < .01, phi = -.39; χ^2 (1, N = 35) = 24.15, p < .01, phi = -.42; χ^2 (1, N = 35) = 37.06, p < .01, phi = -.51. These statistics support the finding of Experiment 4, that the presentation of more than one lineup gives a different indication of accuracy than just one.

Next, the step-wise dropping of consistent accuracy across lineups was examined. Each lineup was compared to the next in the sequence. The drop in consistent accuracy appears to begin between lineups 1(71%) and 2(49%). This was confirmed: χ^2 (1, N = 35) = 7.62, p < .01, phi = -.23. There also appears to be a lowering between lineups 2(49%) and 3(37%), however this was not confirmed: χ^2 (1, N = 35) = 1.87, p = .12, phi = -.12. Lineups 3(37%) and 4(29%) appear to show a similarly small difference, this was supported by Chi-Square: χ^2 (1, N = 35) = 1.17, p = .18, phi = -.09. The differences between all pairs of lineups from this point were insubstantial: χ^2 (1, N = 35) =.14, p = .43, phi = -.03; χ^2 (1, N = 35) = .16, p = .42, phi = -.03. Finally, consistent accuracy for lineup 1 was compared to lineup 6: χ^2 (1, N = 35) = 33.14, p < .01, phi = -.49 demonstrating a difference between these scores. These results show that, as in Experiment 4, although there is a lowering of consistent accuracy over the course of the lineups, no particular trial is responsible for the drop.

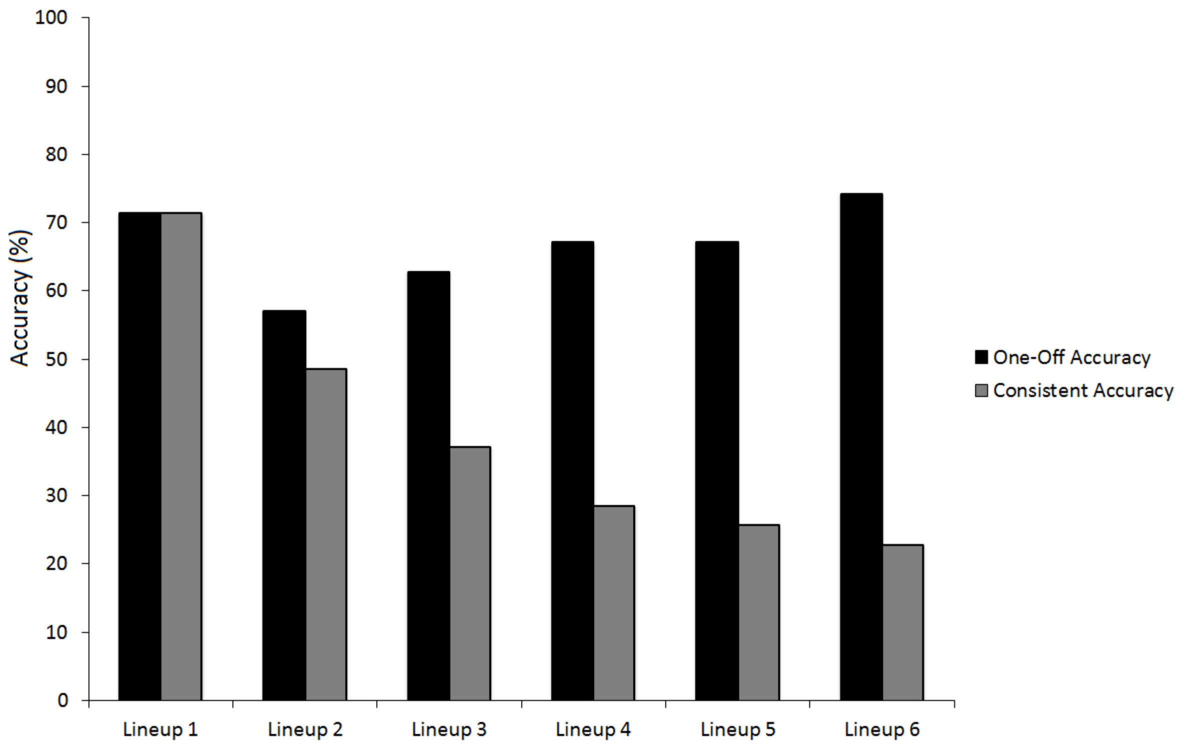

Figure 3.4 Percentage of eyewitnesses able to identify the target face as a one-off on a lineup and consistently in Experiment 6.

Discussion

Participants were shown three target-present and three target-absent lineups for each target. Similar to Experiment 4, 57% to 74% of people correctly responded to any one lineup, but this figure decreased to 23% when consistent accuracy over six lineups was examined. This provides evidence that although some eyewitnesses were able to pick the suspect from a lineup and/or reject a lineup that did not contain the suspect, some were able to do both of these tasks consistently. Based upon the stepwise reduction in consistent accuracy, it would appear that no one lineup is responsible for this effect. The 'filtering' of inconsistent participants is a function of the number of lineups rather than a particular set of stimuli.

A potential problem that has not been addressed to this point stems from the inclusion of different foils in each lineup. Since, unlike any other faces in the experiment, the targets appear three times it may be possible for a participant to recognise a repeated identity from the previous lineups rather than the initial video presentation. In the previous chapter it was found that no memory effects for the foils exist in a lineup experiment, but since here the targets have been presented more than once and no other faces have been similarly treated it is conceivable that a participant could be learning throughout the experiment. In order to address this issue a control experiment was conducted which omitted the initial presentation of the targets in the video and tested participants on their recognition of these identities based upon the lineups alone.

Experiment 7

Experiments 4 and 6 showed that using multiple lineups with an unfamiliar target gave a clearer indication of a participants' accuracy than a single lineup which alone can be misleading. By presenting an identity several times it was possible to independently test this observer over multiple instances and therefore examine their accuracy more thoroughly. However, the repetition of the target means there is a possibility that a participant could gain information about their appearance from the earlier trials. This could mean that a lineup appearing at the end of the trial sequence gives a different result than if it were presented first. For this reason, it is important to determine whether learning can take place during the recognition task. In order to test this possibility the current experiment was conducted without the initial target exposure.

In this experiment the target was always presented in the first lineup, and was either highlighted or was not marked differently in any way. In the non-highlighted (uncued) condition participants were required to guess the target in the first lineup, effectively setting their accuracy at 0%. It was expected that, under these circumstances, when an initial exposure had not occurred it would not be possible to consistently identify the target after such a demanding data load as the entire first array. The highlighted (cued) condition provided the identity of the target in the first lineup, setting accuracy at 100%. It was expected that the relatively poor information provided (i.e. a static image instead of a video exposure as provided in previous experiments) would make this task more difficult and repeated identification would not be possible here either.

Method

Participants

Forty-eight sixth-form students (12 female, 36 male) with a mean age of 17.8 years (*SD* = 5) participated in this experiment as a part of a visit to university premises. All participants had normal or corrected-to-normal vision. No participants had taken part in any preceding experiments in this series.

Stimuli and Procedure

Participants were presented with the same target-present and target-absent lineups as were shown in Experiment 6. However in this experiment, no video or distracter tasks were provided. The first lineup in this experiment was always target-present. There were two conditions, whether this lineup contained a cued target, or an uncued target. In the cued condition one face was indicated by a box around it and it was explained to the participant that this was the face of the target. It was only possible to select this face on the keyboard for this lineup. In the uncued condition no face was highlighted but it was explained that the target was present and that the participant must guess their identity in this lineup. It was not possible to reject the first lineup as target absent. In both conditions participants were informed that the target's face would be repeated throughout the lineups and that there would be a 50% chance in every trial that they would be present. Whether the male or female target was the cued or uncued identity was counterbalanced between participants. After the initial lineup, the other arrays were presented in a random order within their gender categories. Identities were chosen by pressing the number key corresponding to the face or by pressing the “a” key to declare the lineup target-absent.

Results

Uncued Target

In this condition, it was not possible to reject the first lineup, forcing participants to choose an identity at random. Table 3.3 gives the breakdown of responses in each lineup. Responses to the first target-present lineup were guesses so it is unsurprising to find accuracy was lower in this trial, at 8%. Accuracy in both target-present and –absent trials was also lower than has been observed in previous experiments due to the lack of a known target.

Target-present	Lineup 1	Lineup 2	Lineup 3
Hits	8.3	16.7	14.6
Misses	0.0	41.7	56.3
Misidentifications	91.7	41.7	29.2
Target-absent	**Lineup 1**	**Lineup 2**	**Lineup 3**
Correct rejections	52.1	52.1	52.1
False positives	47.9	47.9	47.9

Table 3.3 Percentage breakdown of Participants’ (N=48) responses following an uncued target presentation in Experiment 7.

As in Experiments 4, 5 and 6, lineups were scored for one-off accuracy and consistent accuracy. Since in the uncued condition participants were forced to guess who the target was in lineup 1, consistent accuracy was recorded from lineup 2 onwards. Therefore, it was possible for participants to be included in the sixth lineup consistent accuracy score without being correct in lineup 1. These scores can be seen in Figure 3.5.

Examination of consistent responding shows a rapid drop in accuracy. The step-wise reduction was first examined with a comparison of lineups 2(39%) and 3(16%). A Pearson Chi-Square test of significance confirmed that consistent accuracy fell between these lineups: χ^2 (1, N = 48) = 6.24, p < .05, phi = -.26. Lineups 3(16%) and 4(2%) were tested next. Again, a substantial drop was observed: χ^2 (1, N = 48) = 6.01, p < .05, phi = -.25. Between lineups 4(2%) and 5(0%) there was no significant finding: χ^2 (1, N = 48) = 1.01, p = .5, phi = -.10. There were no data points in lineups 5 and 6 so this analysis ends here. Finally, lineups 2(39%) and 6(0%) were compared to measure the overall fall in accuracy: χ^2 (1, N = 48) = 23.69, p < .01, phi = -.50 and a considerable difference was found here.

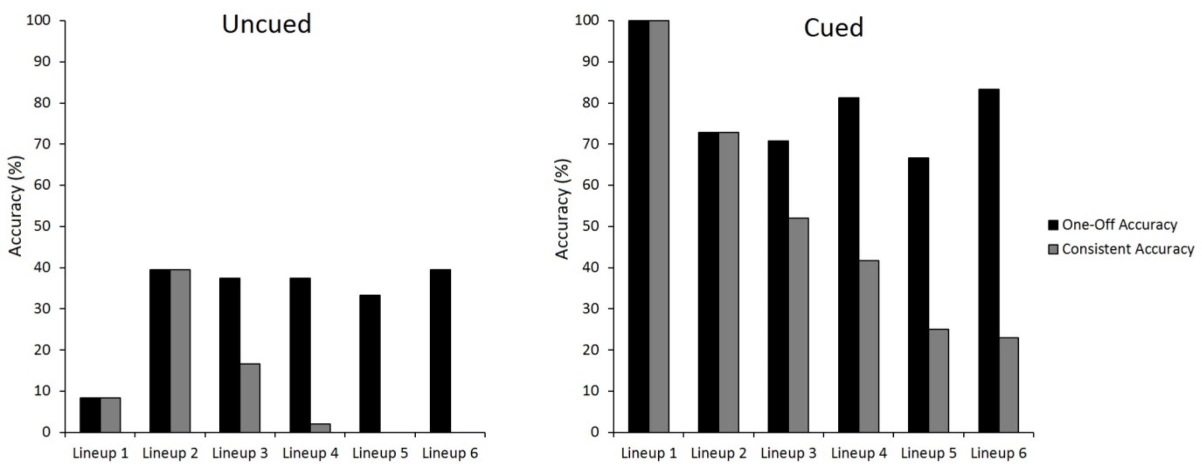

Figure 3.5 Percentage of eyewitnesses who correctly identified the target face as a one-off and consistently when the target was cued or uncued in lineup 1.

Cued Target

When the target was cued it was only possible to select the correct face in lineup 1, hence hit accuracy here is 100%. Figure 3.5 shows that one-off accuracy for the rest of the lineups ranged from 71% to 83% making it slightly higher than previous experiments in this chapter despite the limited information given in the initial exposure. A breakdown target-present and -absent scores can be seen in Table 3.4.

Target-present	Lineup 1	Lineup 2	Lineup 3
Hits	100.0	64.6	64.6
Misses	0.0	27.1	31.3
Misidentifications	0.0	8.3	6.3
Target-absent	Lineup 1	Lineup 2	Lineup 3
Correct rejections	83.3	79.2	83.3
False positives	16.7	20.8	16.7

Table 3.4 Percentage breakdown of Participants' (N=48) responses following a cued target presentation in Experiment 7.

Next, the progressive drop in consistent accuracy was analysed. There is a fall between lineups 1(100%) and 2(73%): χ^2 (1, N = 48) = 15.04, $p < .01$, phi = -.40. The next step, between lineups 2(73%) and 3(52%) also showed a drop: χ^2 (1, N = 48) = 4.44, $p < .05$, phi = -.22. However, the step between lineups 3(52%) and 4(42%) did not: χ^2 (1, N = 48) = 1.05, p = .21, phi = -.10. The reduction in consistent accuracy between the pairs of lineups 4(42%) and 5(25%) and 5(25%) and 6(23%) were also both insubstantial: χ^2 (1, N = 48) = 3.00, p = .07, phi = -.18 and χ^2 (1, N = 48) = .06, p = .5, phi = -.02. This supports the apparent 'levelling off' of the decline in consistent accuracy over the final three lineups. Finally lineup 1(100%) was compared to lineup 6(23%): χ^2 (1, N = 48) = 60.20, $p < .01$, phi = -.79 demonstrating a difference in accuracy across the whole condition.

Discussion

In order to ensure that participants were not able to detect the identities throughout the multiple lineup procedure they were presented with the testing phase without the initial video exposure. The results clearly show that participants were not able to learn the identity throughout the methodology as accuracy in the uncued condition fell to 0% by the fifth lineup. It would appear that when the target is not previously indicated there is too much information in the first lineup to recall the face that has been repeated in later trials. However, the cued condition provided some unexpected results. Here, it was predicted that participants would fail to identify the target's face repeatedly because of the limited visual information they had been exposed to (the target's face in the first lineup rather than the minute long video). However, this did not prove to be the case. Consistent accuracy dropped sharply from the 100% set in the initial lineup but only fell to 23%. This evidence suggests that the static image is enough for some participants to make repeated identifications.

A possible explanation is that although the amount of information about the target faces was reduced, the relevance of the information was increased. The difference in angle, size and lighting between the initial video in previous experiments and the lineups was much greater than the difference between the image participants studied in this experiment and the lineups. Furthermore, since a face was highlighted in the first lineup, the distractor faces could be ignored entirely making this condition comparable to the 1-in-10 task introduced by Bruce et al. (1999). This task has previously shown that recognition of an unfamiliar face from a static image is possible albeit an error-prone process (Bruce et al., 1999). Also, it has been demonstrated that multiple targets reduce accuracy (Megreya & Burton, 2006b), so the current task could be considered less demanding than the first experiment because here participants were only required to recall one face at a time rather than two.

Another possible explanation for the higher than expected accuracy is the inclusion of target-absent lineups in the methodology. With a target-absent lineup the participant's ability to reject the lineup is tested based upon the person they have familiarised themselves with *not* being present. If a participant is not familiar with anyone they can declare the lineup target-absent and be correct in more than half of cases (since the first target-present lineup is used to highlight the target leaving two target-present lineups and three target-absent lineups). As a result of this the accuracy across the six lineups may appear artificially high.

This chapter has considered the effect multiple lineups could have on simultaneously presented lineups. An established alternative to this is a sequential lineups methodology (Lindsay & Wells, 1985; Cutler & Penrod, 1988). A sequential lineup presents participants with each face individually rather than the whole array simultaneously. Eyewitnesses are required to decide, for each face, whether they are the target or not. This variant to the lineup procedure has been suggested to yield greater diagnosticity than conventional simultaneous lineups (Lindsay & Wells, 1985). Not only does this type of lineup mean that a participant must look at every face during their decision-making process, it also means that the eyewitness must make an absolute judgement for each face rather than comparing faces across the array. Moreover, this is now the standard type of lineup used by the police in the UK (PACE, 1984). If the utility of multiple lineups is to be evaluated it is necessary to also consider them sequentially.

Experiment 8

To this point, participants have been presented with multiple simultaneous lineups. Since a widely used alternative is a sequential presentation of identities, it is necessary to test the effectiveness of using multiple lineups with this variant. The use of sequential lineups as opposed to simultaneous lineups requires the participants to make an absolute judgement about each identity (Lindsay & Wells, 1985). They are no longer able to compare the faces directly to make a decision. In the current experiment participants were presented with a video of the targets which was followed by three sequential lineups.

Method

Participants

Forty undergraduates (26 female, 14 male) with a mean age of 21.3 years (*SD* = 3.9) participated in this experiment as a condition of their course. All participants had normal or corrected-to-normal vision. No participants had taken part in any preceding experiments in this series.

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 4 except in this experiment the lineup faces were presented one at a time. No target-absent lineups were included in this experiment since a sequential lineup procedure uses single identification decisions and each foil identity represents a target-absent decision. The faces in these lineups were presented at approximately 6 (W) x 8.5 (H) cm in the centre of the screen. Participants were not told how many faces would be in a lineup (in accordance with Lindsay, Lea & Fulford, 1991) but after the first ten faces a heading appeared at the top of the screen stating ‘Lineup 2’, and after the next ten ‘Lineup 3’ appeared. Participants were told there was an equal chance the target would or would not be present in each lineup. No foil identity was repeated more than once. Whether the participant saw the male or female lineups first was counterbalanced but the order of the lineups and the order of the faces within them were kept constant. Participants were required to press the 'y' key if they thought a face was the target and the 'n' key if they thought it was not.

Results

The sequential presentation of faces in this experiment meant that descriptive statistics were calculated differently. Misidentifications were not possible since a wrongly selected lineup face constituted a false positive. Within these constraints, hits ranged from 48% to 71% over the three arrays. Misses occurred on between 29% and 53% of trials. For the target-absent trials, correct rejections occurred often, between 95% and 98% of the time, and false positives on between 2% and 5% of trials. The full breakdown of scores is presented in Table 3.5.

Target-present	Lineup 1	Lineup 2	Lineup 3
Hits	62.5	47.5	71.3
Misses	37.5	52.5	28.8
Correct rejections	96.7	95.4	98.2
False Positives	3.3	4.6	1.8

Table 3.5 Percentage breakdown of Participants' (N=40) responses for a sequential lineup in Experiment 8.

Participant accuracy is plotted in Figure 3.6. A single mistake on any face constituted an incorrect lineup so a correct lineup comprises one hit and nine correct rejections. First, consistent accuracy was compared to one-off accuracy for each lineup. In lineup 2, consistent accuracy (23%) appears to be lower than one-off accuracy (38%) and this was confirmed: χ^2

(1, N = 40) = 4.33, $p < .05$, phi = -.17. Similarly, there appears to be a large difference between one-off (60%) and consistent accuracy (18%) for lineup 3. This was also confirmed: χ^2 (1, N = 40) = 32.82, $p < .01$, phi = .46.

Next, the step-wise fall in consistent accuracy was examined. There appears to be a drop between consistent accuracy for lineups 1(51%) and 2(23%). This was confirmed: χ^2 (1, N = 40) = 13.28, $p < .01$, phi = -.29. Next, consistent accuracy in lineups 2(23%) and 3(18%) was compared, however, this difference was not significant: χ^2 (1, N = 40) =.63, p = .28, phi = -.06.

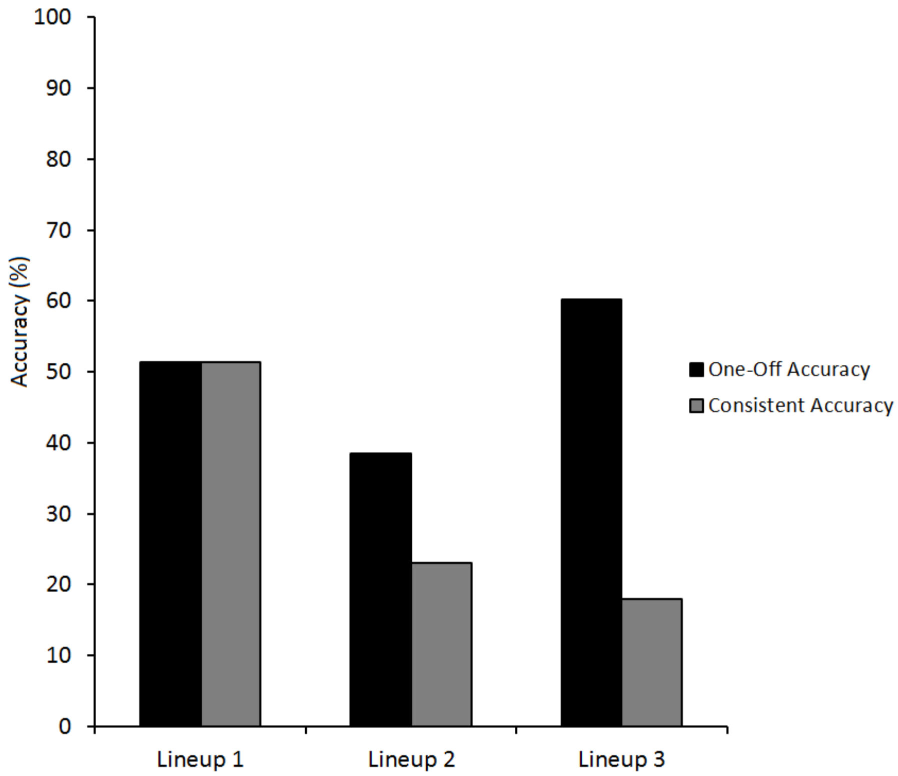

Figure 3.6 Percentage of eyewitnesses able to identify the target face as a one-off on a lineup and consistently in Experiment 8.

Finally, lineup 1(51%) was compared to lineup 3(18%): χ^2 (1, N = 40) =19.15, $p < .01$, phi = -.35. As in the simultaneous version of the methodology, sequential lineups elicit similar levels of accuracy for individual trials but there is a drop in consistent accuracy over the course of the three lineups. As before, the total reduction in consistent accuracy cannot be explained by any one lineup.

Discussion

Participant accuracy was examined with sequential lineups to elicit absolute judgements for the faces in the lineup rather than allowing a comparison between them. It is clear from Figure 3.6 that the same pattern of results as has been seen in the previous experiments of this chapter can be seen here also. One-off accuracy for any one of the lineups is comparable but consistent accuracy falls from the first lineup to the last. This gives evidence that multiple lineups can distinguish between consistently correct and inconsistent eyewitnesses in a sequential methodology as well as in a simultaneous one.

General Discussion

In this chapter the utility of multiple face lineups was tested as a tool for the discrimination between those who could identify a person once and those who could do this repeatedly. Experiments 4, 6 and 8 demonstrated that while some participants were able to make consistently correct decisions others were not. Accuracy on any single lineup fell between 38% and 74%. However when consistent accuracy over all lineups was considered, this figure was lower, ranging from 18% to 23%. It is clear from this evidence that accuracy derived from a single identification can be rather misleading. This method of assessing accuracy is a direct test of observer accuracy for *any particular exposure* in so much that one eyewitness could repeatedly identify a perpetrator and fail to do so in another instance if they had not gained a sufficiently good view of them.

An anticipated criticism is that it may seem predetermined that this result came to pass. Consistent accuracy, by definition, cannot increase from its initial level and one error in a lineup is enough to cause a reduction in that and all future trials, so consistent accuracy must be lower than the initial one-off value. While consistent accuracy as it is defined here cannot increase from its initial level, it does not necessarily have to fall. If an observer acts as a reliable eyewitness and is able to identify a target over many different circumstances, consistent accuracy will be the same in the first and last trial. This was shown in Experiment 5 with participants who were familiar with the targets. Here, accuracy was perfect throughout all trials highlighting these observers' testimony as reliable in this case. In fact, an erroneous assumption present in much related research is that this accuracy will remain the same for *all* eyewitnesses. This is the reason one trial has been considered acceptable as a measurement of accuracy in the past.

By using highlighted and un-highlighted targets with no exposure video, Experiment 7 showed that despite the varying degrees of accuracy in recognising a target from a static image, it was not possible to make consistent responses without *any* form of initial exposure to the target even if, by chance, the target was initially chosen. Consistent accuracy for trials where the target was not shown to participants fell to 0%. This is an important finding because if even one person was able to respond perfectly to every lineup without an initial exposure it would cast doubt into any future assessment of accuracy. Based on this evidence, consistent correct responses do not appear to be possible without this initial exposure.

However, an issue not covered in this chapter was whether an innocent lineup member could be repeatedly identified.

In the context of these experiments, an innocent suspect would be represented by an uncued target, since no initial exposure would have taken place. However, it is evident from this chapter that even highlighting a face is enough to lead to repeated recognition in some cases. There also exists evidence that after making an initial identification, observers may 'commit' to the chosen face even if it is not the target (Deffenbacher, Bornstein & Penrod, 2006), making them vulnerable to repeated incorrect identifications. It follows that in a multiple lineups methodology, if the wrong person was chosen in the first lineup, they may be identified again in subsequent lineups. Since only the target identity was ever repeated here, it was not possible to test this in the current series of experiments. The final experimental chapter examines this possibility while further exploring the use of multiple trials in a more realistic setting.

Chapter 4

Multiple Face Identifications in a Field Study

Introduction

In the experiments described in Chapter 3 participants were required to identify a target identity they had previously seen from six identity lineups. All lineups were entirely made up of faces with the target identity repeated in three of them. When asked to identify the target and to state when they were absent, accuracy for any one lineup fell between 42% and 80%. However, when consistent responses after all trials were considered accuracy was far lower, at 23%. In this chapter this will be investigated further in a field study.

Sauerland and Sporer (2008) used a field study methodology to test identification accuracy after an apparently inconsequential initial exposure to the target. Correct identifications occurred in 61% of portrait face lineups, with non-portrait face lineups providing considerably lower accuracy. The finding of most interest here, however, was that the probability of any individual identification being correct could be calculated based on the combination of responses for each lineup. To illustrate, an identification of a target's body gave a probability of 0.6 that the target had been chosen. When this was combined with a separate identification of a face, the probability rose to 0.9. Chapter 3 illustrated that multiple face lineups give different information when considered together than individually. By adopting a probabilistic assessment of the lineup combinations it should be possible to evaluate each participant's identification accuracy at the individual level.

In the previous chapter the only face presented in more than one lineup was that of the target. This meant that it was possible to test repeated recognition of the target amongst a large number of distractor identities (i.e. the other lineup faces). However, this did not allow for the possibility of multiple *in*correct responses. In order to determine the likelihood that a person has chosen the target when they have identified the same face multiple times, it is necessary to allow participants the opportunity to identify a non-target face repeatedly. Deffenbacher et al. (2006) recognised a circumstance where observers may ‘commit’ to an identity after they had chosen it and this phenomenon is of interest here.

Experiment 7 required participants to be presented with lineups that were not preceded by an initial exposure to the target. The uncued condition of this experiment revealed that it was not possible to make multiple correct decisions when no target had been presented. Following these findings it would be expected that a consistent set of responses should only be possible when the participant has chosen the target since this is the only time where they will have received prior exposure to the identity. However, the cued condition of Experiment 7 showed that some participants were able to identify a face after being shown the correct response in the first lineup despite not receiving any other kind of prior exposure. Since this was the case it is unclear if it would be possible for a participant to choose the wrong face in the first lineup and then continually identify this same person throughout the rest of the trials. In the worst case this could be an innocent suspect who will be wrongly prosecuted if identified.

The designation of innocent suspects poses problems in experimentation (e.g., Pryke et al., 2004; Sauerland & Sporer, 2008). In police investigations, suspects are arrested on the basis of their similarity to a witness’ description. However, it can be difficult to establish the perceived similarity of targets and suspects in advance. Different strategies for designating innocent suspects and lineup foils appear to influence eyewitnesses’ identification decisions (Lindsay, Martin, & Webber, 1994; Luus & Wells, 1991; Wells, Rydell, & Seelau, 1993), but the study of such strategies has also yielded inconsistent results (e.g., Darling, Valentine, & Memon, 2008; Tunnicliffe & Clark, 2000). In addition, people vary considerably in their ability to perceive the similarity of different identities. For example, even under highly optimised conditions, observers frequently demonstrate inter- and intra-individual variation in how they perceive the resemblance of faces in person identification tasks (e.g., Alenezi & Bindemann, 2013; Bindemann, Avetisyan et al., 2012; Bindemann & Sandford, 2011).

In light of these problems, a different method was adopted here. Instead of pre-selecting a designated suspect, this identity was defined a posteriori. Two contrasting approaches of innocent suspect designation were employed. For the first approach, the innocent suspect was defined as the foil identity that was selected *first* by an eyewitness in the multiple-lineup procedure. This approach minimises data loss by including all incorrect eyewitnesses in the analysis and provides a "worst case scenario" by comparing consistent target selections with the greatest possible number of the corresponding foil identifications.

In the second approach, the innocent suspect was defined as the foil identity that was selected most often as the target by all observers during the course of the experiment. This "worst foil" approach has also been adopted in previous research because it provides the highest number of suspect identifications when these are defined by only a single foil identity (e.g., Pryke et al., 2004; Sauerland & Sporer, 2008). In the current study, this is the more lenient approach as it inevitably provides less repeated suspect identifications for comparison with the target.

In this chapter, the multiple lineup procedure of Chapter 3 was re-examined with lineups that included repeated instances of all faces, and not just the target. Furthermore, the circumstances of the initial exposure were made more realistic by introducing a live encounter where the participant was initially not aware they were part of an experiment.

Experiment 9

To investigate the potential of a multiple-lineup procedure with repeated foil faces, a field experiment was conducted in which pedestrians in a city centre were approached by a target person under the pretense of requiring route directions to a local landmark. Shortly after this exchange had finished, these observers were approached by another experimenter and asked to attempt to identify the just-seen target. For this purpose, six successive identity lineups of faces were shown, comprising a mixture of three target-present and three target-absent lineups. The aim was to assess the extent to which observers could identify the target person repeatedly, or alternatively, whether they could identify a different face multiple times. A comparison of these possibilities should allow an insight into whether multiple lineups can provide a better index of eyewitness accuracy than a single lineup.

Method

Participants

Forty pedestrians in a city centre (23 female, 17 male), consisting of students and young professionals with a mean age of 22 years (*SD* = 4.7), took part in this experiment. These participants agreed to take part once they had been made aware of the true purpose of the initial interaction with the target and had provided informed consent to continue further. Approximately ¾ of people originally approached agreed to continue with the study. All participants reported normal or corrected-to-normal vision.

Stimuli

The faces of twelve people were used for the lineup construction. These consisted of the target and eleven filler identities. All of the fillers fitted the general description of the referring target (Wells et al., 1993), as determined in two pilot studies with 20 mock witnesses. For each identity, three colour face photographs were collected, which showed these persons in a frontal view with a neutral expression. These photographs were standardised by cropping clothing and background. These images were taken on the same day to eliminate transient differences in age, facial hair, and so forth (e.g., Bruce et al., 1999; Burton et al., 2010). All of the resulting face images measured approximately 5 (W) x 7.5 (H) cm.

These images were then used to construct three target-absent and three target-present lineups. Each lineup therefore consisted of six faces, which were arranged in two rows of three pictures. The target and filler faces were distributed across these arrays, so that none of the identities appeared more than once in any of the lineups and not more than once in any of the locations within a lineup. In addition, none of the lineups shared more than three of the 12 identities. However, each of the 11 filler identities appeared alongside the target at least once. Effective lineup sizes were calculated using Tredoux's *E*s and were determined to be between 3.6 and 5.1 identities (Tredoux, 1998, 1999). These lineups can be seen in Figure 4.1.

Procedure

The target, a 32-year-old Caucasian male, approached pedestrians in the centre of a Dutch town to ask for directions. In these interactions, the target wore the same clothing throughout the testing period and kept the conversation as similar across participants as possible. These interactions lasted approximately one minute. Typically, the approached pedestrian would look at the target several times during this time period. If the interaction did not follow this pattern, the pedestrians were not approached again for the subsequent identification task.

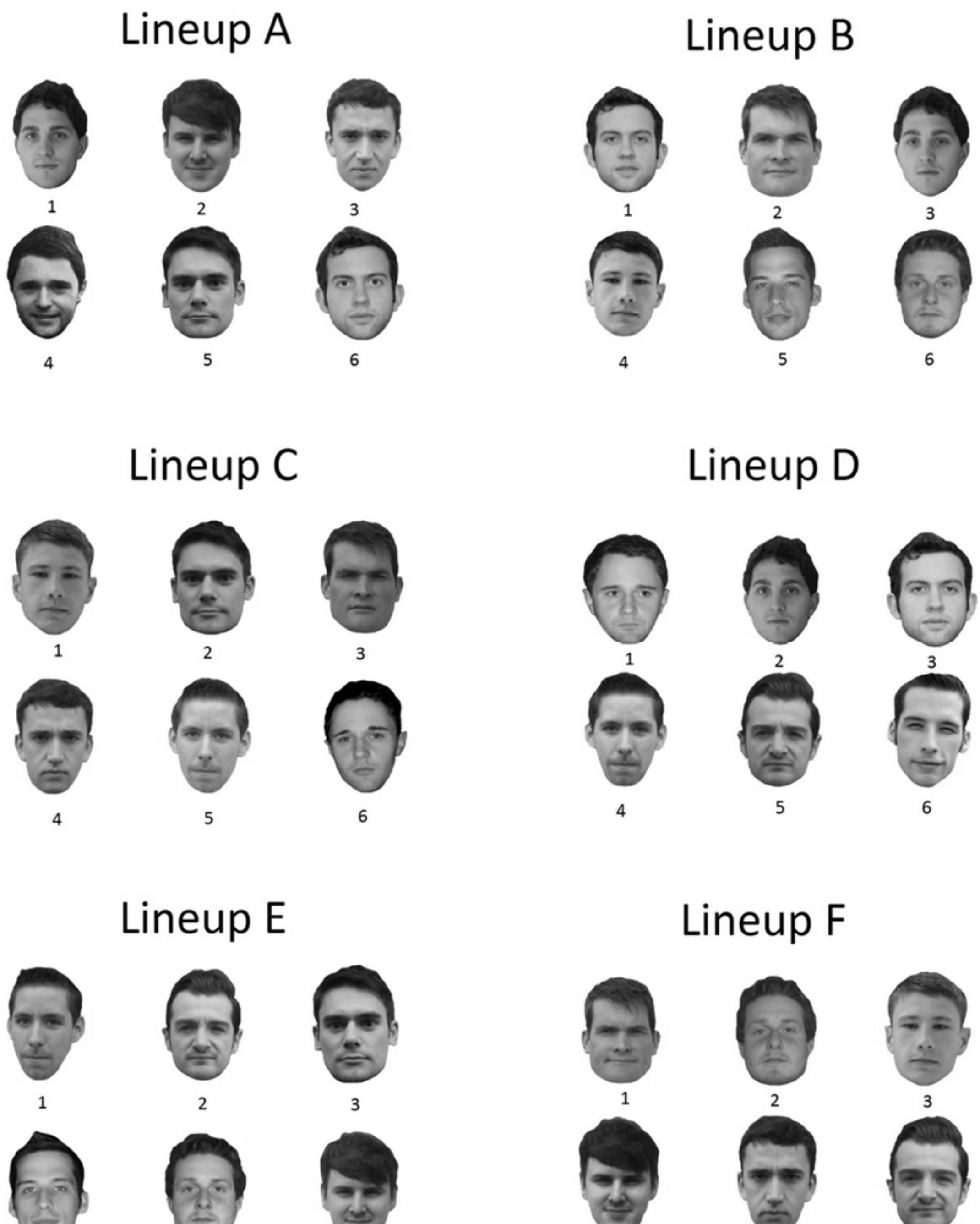

Figure 4.1. An illustration of the three target-present and the three target-absent lineups in Experiment 9. The arrays show target-present lineups on the left and target-absent lineups on the right.

After an interval of approximately one minute, these pedestrians were approached by an experimenter, who was positioned upstreet of the initial interaction with the target. At this stage, the purpose of the experiment was explained to the pedestrians and their consent for further participation was obtained. They were then presented with six successive lineups, which were shown in a random order. Participants were told there was an equal chance that the target would be present or absent in a lineup. They were asked to attempt to identify the target when he was present or to declare his absence when he was not. Once a lineup had been completed, it was moved out of view of the participant before the next lineup was presented. Accuracy of responses was emphasised so no time limit was given for the identification task.

Results

One-off and Consistent Target Identifications

In a first step of the analysis, identification accuracy was calculated separately for each of the six lineups. For the three target-present lineups, observers' responses were categorised either as hits (i.e., the correct identification of the target from a lineup), misses (the incorrect response that the target is absent), or misidentifications (the selection of a wrong face). For target-absent lineups, responses were classified as correct rejections (the correct response that a lineup does not contain the target) or false positives (the selection of a lineup face despite the target's absence). These data are presented in Table 4.1, grouped by target-present and target-absent trials. Note that the order in which these lineups were encountered is not preserved in the table.

These data show that observers identified the target on between 53% and 68% of lineups. Similarly, accuracy for target-absent lineups ranged from 53% to 70%. In turn, identification errors occurred with considerable frequency. For example, misidentifications were recorded on between 8% and 38% of target-present trials, whereas false positives were made on

Target-present	Lineup A	Lineup C	Lineup E	Overall
Hits	52.5	67.5	55.0	58.3
Misses	10.0	25.0	25.0	20.0
Misidentifications	37.5	7.5	20.0	21.7

Target-absent	Lineup B	Lineup D	Lineup F	Overall
Correct rejections	70.0	67.5	52.5	63.3
False positives	30.0	32.5	47.5	36.7

Table 4.1. Eyewitness Accuracy for Each of the Target-present and Target-absent Lineups in Experiment 9

between 30% and 48% of target-absent lineups. Moreover, over the course of the experiment, nine of the eleven foil identities were mistaken for the target.

These data show that eyewitness identification accuracy was generally error-prone for any of the individual lineups. However, the question of main interest is whether observers were consistent in their identification responses across multiple lineups. To address this question, the data were recoded into correct and incorrect responses irrespective of target-presence and were analysed in the exact order in which the six lineups were encountered by a participant. The percentage of participants that achieved a correct identification for any of the lineups is illustrated in Figure 4.2. Across all lineups, one-off accuracy averaged 61%. This shows that approximately two thirds of observers made a correct response, such as the identification of a target or the rejection of a lineup in its absence, to any of the lineups. In addition, accuracy also appeared to increase over the course of the multiple-lineup procedure, from a minimum of 45% in Lineup 1 to a maximum of 70% in Lineup 6. A Chi-square test showed that this increase was significant, $\chi^2(1, N = 40) = 5.12$, $p < .05$, phi = .25.

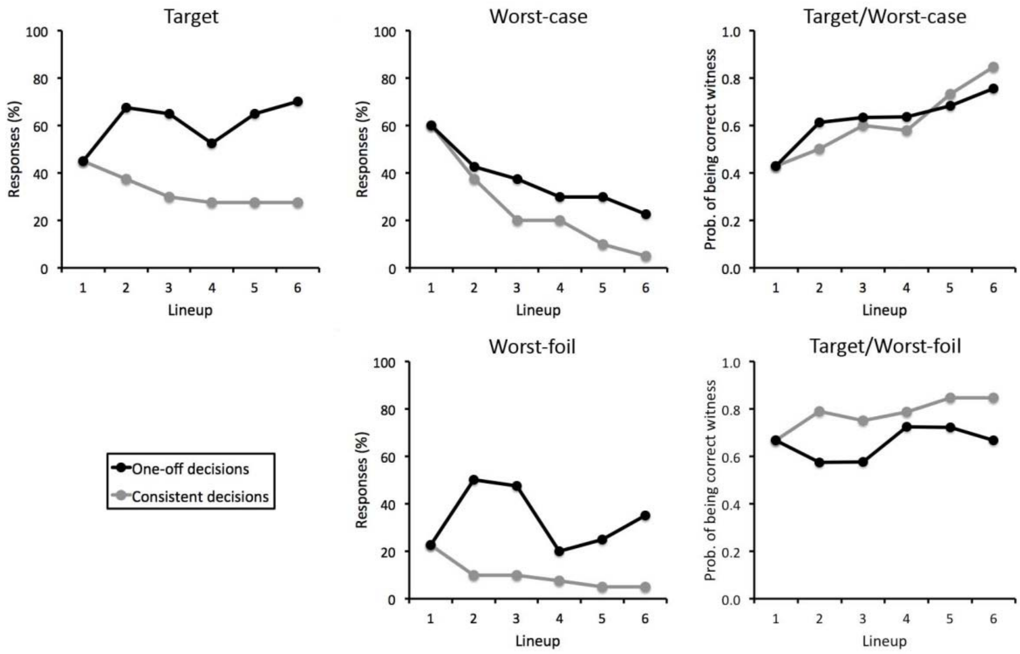

Figure 4.2. One-off and consistent responses for target selections (top left), the worst-case (top centre) and worst-foil analysis (bottom centre) in Experiment 9, and the probabilities that a correct target identification has been made (top and bottom right).

In addition to one-off accuracy, a consistent-accuracy score was determined for each lineup. This captures the extent to which observers made a correct response on the first of the lineups, and then carried on to do so without interruption on successive lineups. Figure 4.2 also shows the cross-subject means of the percentage accuracy of these responses. These data show that consistent accuracy declined gradually with each additional identity-lineup, from 45% in Lineup 1 to 28% in Lineup 6, but this drop in performance was not reliable, $\chi^2(1, N = 40) = 2.65$, $p = .08$, phi = -.18. In addition, however, a direct comparison also shows that one-off accuracy was reliably better in Lineup 6 than consistent accuracy, $\chi^2(1, N = 40) = 14.46$, $p < .01$, phi = -.43. Thus, whereas the majority of observers (70%) make a correct identification decision to the final lineup of this procedure, only a subset of these observers (28%) responded with consistent accuracy throughout.

"Worst Case" Non-Target Selections

An important contrast for these data is observers' consistency when an incorrect suspect identification was made. To create this contrast, participants' responses were recoded if they had selected a foil lineup member prior to any correct identification of the target. In these cases, the first foil that was selected by an observer was adopted as the suspect identity for that individual. Any prior and subsequent lineup responses were then recoded accordingly. For example, if observers previously or subsequently rejected a lineup in which this foil was not present, then this was treated as a correct rejection (regardless of the presence of the actual target identity). By recoding the data in this way, this analysis essentially seeks to mimic situations in which an innocent suspect is placed in a lineup instead of a target and is then selected by an eyewitness.

These data are also provided in Figure 4.2 and show that one-off foil identifications were initially high, at 60% in Lineup 1[1]. These responses then declined with each subsequent lineup to only 23% at the final trial. A similar pattern was observed for consistent foil identifications. However, the drop in performance across successive lineups was more marked in these scores, so that consistent foil selections were at only 5% by the last lineup. To analyse this drop in accuracy, Chi-square tests were conducted to compare one-off and consistent selections for the first and the sixth lineup. This showed that foil selections dropped significantly over the course of the experiment for one-off and consistent identification decisions, $\chi^2(1, N = 26) = 18.66$, $p < .01$, phi = -.60, and $\chi^2(1, N = 26) = 37.23$, $p < .01$, phi = -.85, respectively. However, a direct comparison of these measures showed that

[1] Note that a small number of absent responses to Lineup 1 ($N = 2$), which are correct in reference to the actual target (i.e., when this lineup does not include the target), are also correct in reference to foil identifications (i.e., when this lineup does not include the foil that is misidentified in a subsequent lineup as the target). These responses to the initial lineup are therefore included in the target *and* the worst-case non-target selections here.

fewer consistent foils selections were made by Lineup 6 than one-off selections, $\chi^2(1, N = 26) = 5.65$, $p < .05$, phi = -.33.

“Worst Foil” Non-Target Selections

In addition to the “worst case” analysis, which was based on *any* of the foil identities that were initially mistaken as the target, a “worst foil” analysis was also conducted. For this purpose, the most frequently chosen foil identity across the whole experiment was identified. All lineup scores were then recalculated by adopting this identity as the (innocent) target suspect. These data are provided in Figure 4.2 and show that foil selections were made by only 23% of observers in Lineup 1. This value fluctuated across subsequent lineups, ranging from 20% to 50% of lineup selections, and was at 35% by Lineup 6. By contrast, consistent foils selections gradually declined from 23% in Lineup 1 to just 5% in Lineup 6. These observations were confirmed with Chi-square tests, which showed no reliable change in one-off foil selections between Lineup 1 and Lineup 6, $\chi^2(1, N = 40) = 3.16$, $p = .06$, phi = .12, but a significant drop in consistent foil selections, $\chi^2(1, N = 40) = 5.17$, $p < .05$, phi = -.25. A direct comparison also showed that considerably fewer consistent than one-off foil selection were by Lineup 6, $\chi^2(1, N = 40) = 11.25$, $p < .01$, phi = -.38.

Likelihood of Correct Lineup Identifications

In a final step of the analysis, target and foil selections were compared directly to determine if the relative frequency of these identifications can be used to assess the accuracy of individual eyewitnesses. This was achieved by dividing target identifications by the total number of identifications for each lineup (i.e., target identifications / target + foil identifications), both for the one-off and the consistent accuracy measures. The resulting probabilities are shown in Figure 4.2.

For the worst-case analysis, these probabilities reveal that the likelihood of a correct target identification is just 0.43 when this is based on the responses for the first identity lineup, but this gradually rises to 0.76 when the one-off scores for targets and foils in the sixth lineup are compared. For consistent scores, this increase is even more marked, with a final probability of 0.85 that a participant has successfully chosen the actual target if they have identified the same face three times and also rejected the three lineups that do not include this person.

For the worst-foil analysis, the probabilities for one-off decisions that a correct eyewitness identification was made ranged from 0.57 to 0.72 across lineups. However, these probabilities were identical for Lineup 1 and Lineup 6, at 0.67, which indicates that the diagnosticity of eyewitness accuracy based on these one-off scores did not improve over the course of the experiment. By contrast, the probabilities for consistent eyewitness identifications again showed an improvement, from 0.67 in Lineup 1 to 0.85 in Lineup 6.

Taken together, these results show that the probability with which it is possible to estimate the accuracy of an eyewitness improves when performance is measured repeatedly, across multiple lineups. In this method, the diagnosis of eyewitness accuracy improves when one considers the consistency with which an eyewitness acts on the same identity across different lineups.

Discussion

In this experiment, participants were unknowingly introduced to a target and later asked to identify this person in a surprise recognition test to assess their eyewitness accuracy. Identification of the target was assessed repeatedly with six separate face lineups. However, in contrast with previous studies in this thesis, these lineups contained repeated instances of *all* faces, and not just the target. Eyewitness accuracy was generally poor. For example, correct identifications were made on only 58% of all trials, whereas filler faces were generally mistaken for the target on 22% of target-present and 37% of target-absent trials. This contrast between correct and incorrect responses is even more marked when some of the individual lineups are considered. For example, for Lineup A, correct target identifications accounted for only 53% of responses while mistaken identifications were recorded on 38% of trials (see Table 4.1). This demonstrates that a single lineup provides a poor index of eyewitness accuracy and is consistent with previous studies in this field (e.g., Wells et al., 2006; Wells & Olson, 2003).

A different picture emerged when the consistency of eyewitness responses across successive lineups was assessed. This showed that initial target identifications and correct rejections tended to be followed by further correct decisions in subsequent lineups. For example, whereas 35% of observers made a correct decision to Lineup 1, 28% also made such correct decisions for *all* six lineups (see Figure 4.2). In other words, these data suggest that of the group of observers who initially make a correct lineup decision, 80% also consistently selected the target from subsequent lineups and identified his absence. In contrast to these target selections, foil identifications were marked by the *inconsistency* of observers' responses. For example, whereas foil selections accounted for 60% of responses to Lineup 1, only 5% of observers acted on the *same* foil across all six lineups (see Figure 4.2). The majority of observers who mistake a filler face for the target therefore do not appear to base all subsequent decisions on this same identity.

This difference between target and filler identifications is particularly striking when the consistent-accuracy scores to these categories are compared directly to calculate the likelihood that a correct target identification has been made. This analysis shows that 85% of the observers who consistently act on the same facial identity across all six lineups, by identifying this person's presence or noting his absence are, in fact, correct eyewitnesses who have accurately identified the target. This differs substantially from the first lineup, which allows such inferences only with a probability of 0.37. This is an important result because it affirms that a single lineup provides limited insight into an individual's eyewitness accuracy. In turn, these findings demonstrate that it is possible to assess the accuracy of an eyewitness better by measuring the consistency with which identification decisions can be made. In the current experiment, this produces a remarkable shift, whereby the majority of initial responses are incorrect but the majority of consistent responses across all six lineups indicate a correct person identification.

Despite these promising results, there was a subset of participants (5%) who consistently selected the same foil identity across all six lineups. Such consistent foil identifications might have been caused by the limited variability of the face photographs that were used for the lineup displays, which comprised very similar same-day photographs for each identity. While the similarity of different face images facilitates identification (e.g., Terry, 1994; Davis & Valentine, 2009; Megreya, Sandford, & Burton, 2013), an underlying assumption of this method is that each lineup should provide a relatively independent test of person recognition. This assumption might have been violated if observers were able to identify foils repeatedly due to the superficial similarity of the images across different lineups rather than recognition of the face itself. If this was the case, then such foil identifications should be eliminated by introducing more variability in face photographs across the different lineups.

Experiment 10 examines this possibility by utilising three different types of photographs for each facial identity, comprising a standard image, a picture from a photo-ID, and an uncontrolled photograph from the profile of a social networking site. The additional variation that is introduced by these image categories should increase the difficulty of consistently identifying the same foil, all of which are completely unfamiliar to participants (e.g., Burton et al., 2005; Jenkins & Burton, 2011; Jenkins et al., 2011). By contrast, recognition of the familiarised target faces should be less susceptible to such image variation (e.g., Burton et al., 1999; Clutterbuck & Johnston, 2002; Jenkins & Burton, 2011). As a result, consistent foil selections should decrease, but such target identifications be less affected.

Experiment 10

This experiment aims to replicate the multiple-lineup procedure of Experiment 9 with face images that introduce greater variation in the appearance of the target and the fillers. This should reduce superficial similarities in the appearance of these identities across lineups. Due to the increased difficulty that such variation in facial images should provide for identification (e.g., Burton et al., 2005; Jenkins & Burton, 2011; Jenkins et al., 2011), lower general accuracy on this task is expected than was observed in Experiment 9. This should particularly affect the repeated selection of the unfamiliar filler faces, which were not encountered prior to the lineups. By contrast, observers should be able to tolerate such variation better in the appearance of the familiarised target faces (e.g., Burton et al., 1999; Megreya & Burton, 2006a). This should enhance the diagnosticity of the multiple-lineup procedure by eliminating consistent foil identifications.

Method

Participants

Forty pedestrians in a city centre (23 female, 17 male), consisting of students and young professionals with a mean age of 20 years (*SD* = 4.5) took part. None had participated in the preceding experiment. However, as in Experiment 9, these participants agreed to take part once they had been made aware of the true purpose of the initial interaction with the target, and had provided informed consent to continue further. All participants reported normal or corrected-to-normal vision.

Stimuli and Procedure

The lineups were constructed using the same target and filler identities as Experiment 9. Three photographs were included for each of these identities, which comprised a standardised photograph (from Experiment 9), a photograph from a student identity-card, and a profile picture from a popular social networking site. The standardised and social face images measured approximately 5 (W) x 7.5 (H) cm but the dimensions of the identity-card images were smaller, at 2.5 (W) x 3.5 (H) cm. Examples of these stimuli can be seen in Figure 4.3. The procedure was identical to Experiment 9.

Results

One-off and Consistent Target Identifications

The data were analysed analogous to Experiment 9. Once again, eyewitness accuracy was generally poor (see Table 4.2). For example, correct identifications of the target occurred on only 39% of target-present trials, whereas correct rejections accounted for 61% of responses. In addition, mistaken identifications of foil faces occurred on 12% of target-present and 39% of target-absent trials.

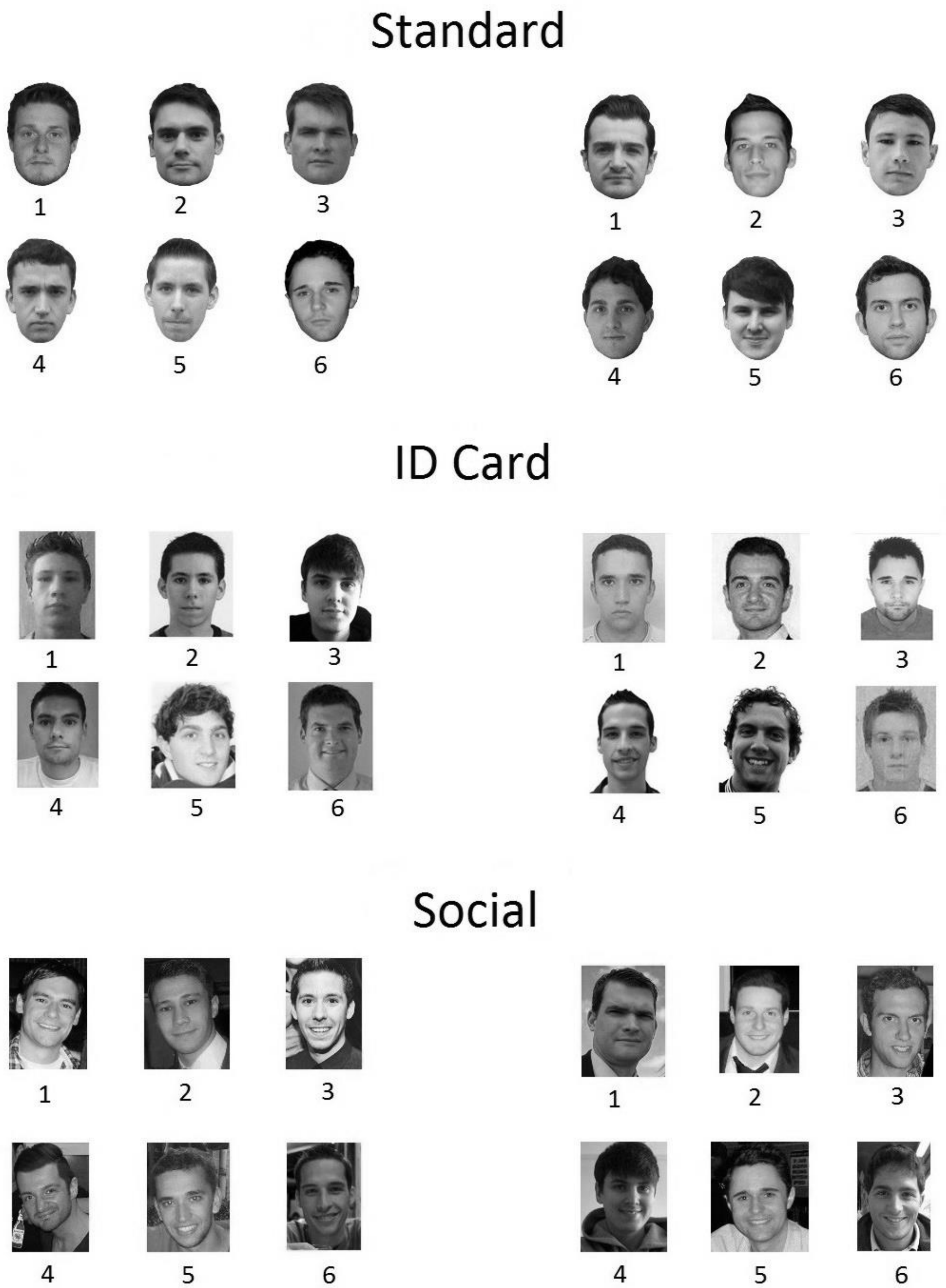

Figure 4.3. An illustration of the three target-present and the three target-absent lineups in Experiment 10. The stimuli consisted of standardised photographs, images from photo-identity cards, and profile pictures from a social networking site. The arrays show target-present lineups on the left and target-absent lineups on the right.

Target-present	Standard	ID Card	Social	Overall
Hits	45.0	32.5	40.0	39.2
Misses	42.5	52.5	52.5	49.2
Misidentifications	12.5	15.0	7.5	11.7

Target-absent	Standard	ID Card	Social	Overall
Correct rejections	47.5	75.0	60.0	60.8
False positives	52.5	25.0	40.0	39.2

Table 4.2. Eyewitness Accuracy for the Target-present and Target-absent Lineups in Experiment 10

The data of most interest concern performance across the different lineups (see Figure 4.4). These data show that one-off accuracy for the target (correct identifications and rejections combined) ranged from 48% to 55% across all six lineups, with no difference between the first (48%) and the last lineup (55%), $\chi^2(1, N = 40) = 0.45$, $p = .33$, phi = .08. In contrast, consistent accuracy fell significantly, from 48% to 5%, over the course of the experiment, $\chi^2(1, N = 40) = 18.66$, $p < .01$, phi = -.48. In addition, a direct comparison for Lineup 6 showed that one-off accuracy scores were higher than consistent accuracy, $\chi^2(1, N = 40) = 23.81$, $p < .01$, phi = -.55.

"Worst Case" Non-Target Selections

As in Experiment 9, foil identification was analysed by recoding the data according to the first foil that was selected by an observer. One-off foil selections exceeded target identifications in Lineup 1, at 45%, and fell to 20% over the course of the lineups, χ^2 (1, $N = 18) = 13.85$, $p < .01$, phi = .62.

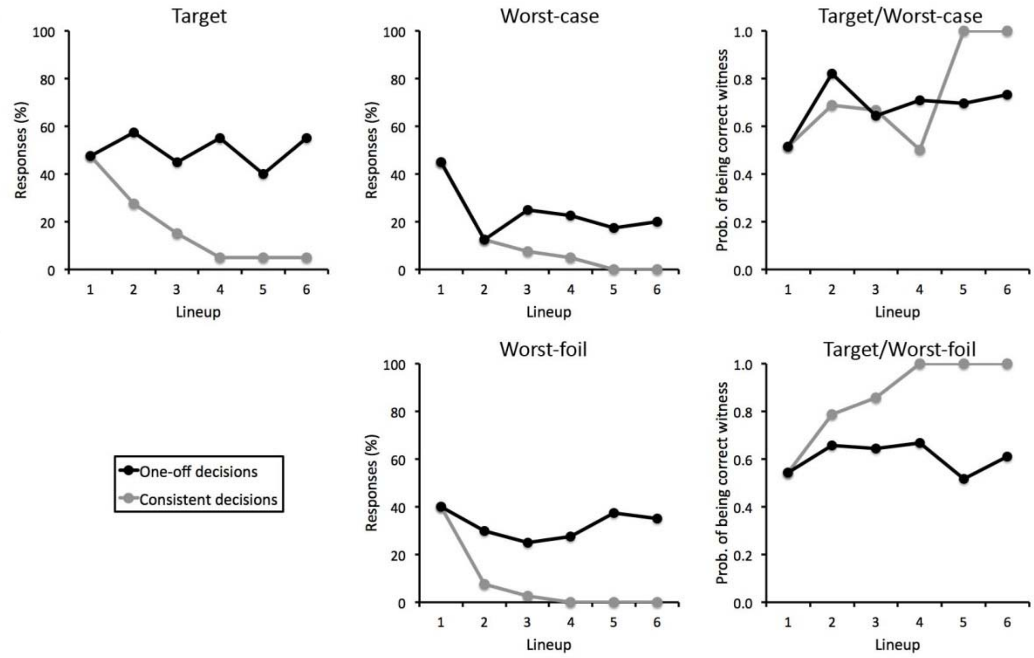

Figure 4.4. One-off and consistent responses for target selections (top left), the worst-case (top centre) and worst-foil analysis (bottom centre) in Experiment 10, and the probabilities that a correct target identification has been made (top and bottom right).

The drop in consistent foil selections was even more marked, falling from 45% to 0% between the first and the final lineup, χ^2 (1, N = 18) = 36.00, $p < .01$, phi = -1.00. This difference was confirmed by a direct comparison, which showed that consistent foil selections were lower than one-off selections by Lineup 6, χ^2(1, N = 18) = 10.29, $p < .01$, phi = -.54. These data indicate that none of the participants acted consistently on the same foil across all six lineups.

“Worst Foil” Non-Target Selections

The “worst foil” analysis, which is based on the most frequently selected foil identity, showed a similar pattern (see Figure 4.4). The percentage of one-off foil selection were similar for Lineup 1 (40%) and Lineup 6 (35%), χ^2(1, N = 40) = .21, p = .4, phi = -.05, and ranged between 25% to 40% across all lineups. By contrast, consistent foil selections declined rapidly from 40% in Lineup 1 to 0% in Lineup 4, $\chi^2(1, N = 40) = 20$, $p < .01$, phi = -.5. Again, a direct comparison showed that consistent foil selections were lower than one-off selections by Lineup 6, $\chi^2(1, N = 40) = 16.97$, $p < .01$, phi = -.46.

Likelihood of Correct Lineup Identifications

To compare the consistency of target and foil selections directly, the response data for these categories were again converted into probabilities that reflect the likelihood that a correct eyewitness identification was made (see Figure 4.4). For the worst case analysis, these probabilities show that the likelihood of a correct eyewitness responses stands initially at 0.51 for Lineup 1 and rises to 0.73 by Lineup 6 for one-off decisions. For the consistency scores, on the other hand, this probability improves to 1.00 by Lineup 5. The worst foil analysis shows a similar pattern. The initial probability of a correct eyewitness decision is at .54 and, for one-off decisions, still stands at only 0.61 by Lineup 6. By contrast, this increases sharply for consistent decisions to 1.00 by Lineup 4. Taken together, these data confirm that the probability with which it is possible to estimate the accuracy of an eyewitness improves when performance is measured across multiple lineups. In Experiment 10, which introduced greater variability among face photographs across lineups, this allows for the accurate diagnosis, with a perfect probability of 1.00 of eyewitness accuracy.

Discussion

This chapter explored a method for the assessment of eyewitness identification accuracy. Specifically, eyewitness accuracy across six lineups was tested, to explore the consistency with which identification decisions to target and foil faces are made. In Experiment 9, eyewitness accuracy was generally low, with 58% correct selections in target-present lineups and 63% correct rejections in target-absent trials. Once again, these results are consistent with previous research by showing that eyewitness identification is a difficult and error-prone task (e.g., Bruce et al., 1999; Megreya & Burton, 2008; Memon et al., 2011).

A different picture emerged when the consistency of identification responses was assessed. This showed that eyewitnesses who initially made a correct lineup identification were more likely to follow this up with further correct responses to subsequent lineups, both in the presence and absence of a target. By contrast, the majority of observers who initially identified a foil face as the target did not exhibit the same behaviour. A comparison of the consistency scores for these two groups allowed the calculation of the probability that a correct lineup identification had been made. When this calculation was based on the first identity lineup, this probability was only 0.43 for the worst-case and 0.67 for the worst-foil analysis. By contrast, this rose to 0.85 for both measures for observers who consistently acted on the same identity. This shows that the application of multiple lineups can provide a much-improved index of eyewitness accuracy.

Despite these promising results, Experiment 9 also revealed a subset of observers who identified the same foil identity consistently. This finding was attributed to the high similarity of face photographs that were employed across lineups in Experiment 9 (see Figure 4.1), which may have facilitated repeat-identifications of the foils (and targets). To address this concern, more within-person variability was introduced in the face photographs in Experiment 10, by combining standardised photographs with images from photo-identity cards and social networking sites. This manipulation generally reduced target and foil identifications. Crucially, however, only *consistent* decisions to the same foil identities were eliminated entirely. As a consequence, this procedure could detect correct target identifications with a certainty (i.e., a probability of 1.00) when eyewitness consistency was assessed over six lineups.

This was explored further using an approach that examined both an observer's ability to detect a target in a lineup and the rejection of a lineup in its absence. It is now established that these lineup types test dissociable aspects of person identification (e.g., Megreya & Burton, 2007) and both are important for the assessment of eyewitness accuracy (i.e., for detecting a perpetrator in a lineup; and for spotting when the perpetrator is not there, as might be the case when a lineup contains an innocent suspect). Despite this, conventional identification procedures, which are based on only a single lineup, make it impossible to apply both lineup types on a within-subject basis. This combination is an advantage of the multiple-lineup procedure.

To simplify the results, these lineup types have not been analysed separately here. This makes good sense as one could achieve high accuracy on target-absent lineups by making "default" absent responses whenever there is uncertainty as to whether a target is present in a lineup. Crucially, such a bias cannot undermine this procedure when the analysis of both lineup types is combined, as this should also result in low accuracy (i.e., increased misses) for target-present lineups. In this sense, these lineup types are clearly complimentary and need to be considered in unison to provide the most informative index of eyewitness accuracy.

General Discussion

This multiple-lineup method could be refined further by defining foil identifications more strictly. Outside of the laboratory, an identity lineup always includes a suspect, but this person may be the sought-after perpetrator of a crime or an innocent. The purpose of a lineup is essentially to determine whether a witness will select the suspect, thereby seemingly confirming them as the target, or will not choose this identity. The remaining faces act as "fillers" that are only there to complete the lineup and that would not be charged if they were selected by an eyewitness. To determine the extent to which observers might repeatedly identify the same non-target face in a multiple-lineup procedure, one could therefore replace the target with another identity that acts as a designated innocent suspect. By comparing repeated target identifications with selections of the innocent suspect, it would then be possible to determine whether the consistency of observers' responses across multiple lineups can dissociate correct from incorrect eyewitness identifications.

The current experiments did not include such a designated innocent suspect. Instead, such "suspects" were defined after test as either the most frequently chosen between-subjects (worst foil) or selected individually for each observer (worst case), based on the first foil selection that was made (provided that the target was not identified first; see Results section of Experiment 9). This approach was adopted because the selection of innocent suspects poses its own problems in experimentation. For example, if an innocent suspect is selected in advance that is, perhaps by chance, less similar to the target than other lineup members then these suspect identifications might be extremely low. As a consequence, it would not have been possible to assess the merit of a multiple-lineup procedure properly. Moreover, foil selections often encompass many different lineup identities, even when these share only a general description. In Chapter 4, for example, all of the filler faces were identified at least once as the target between Experiments 9 and 10. For these reasons, foil (i.e., innocent suspect) identifications were based on the filler identity that was selected first by an eyewitness. This approach minimises data loss by including all incorrect eyewitnesses in the analysis. It has the added advantage of providing a "worst case scenario" by comparing consistent target selections with the greatest possible number of the corresponding foil identifications.

A final point to consider here is the number of eyewitnesses considered consistent. In Experiment 9, only 11 eyewitnesses chose the correct face repeatedly. In Experiment 10 this had fallen to only 2 eyewitnesses. On face value, this appears to be a limitation of the current method but, on the contrary, this actually demonstrates a strength. If only a small number of eyewitnesses are able to make a reliable identification it follows that the same small proportion are reliable in other eyewitness paradigms. The difference between previous methods and the one employed here is that the current method is able to detect this unreliability. It has been known for a considerable time that eyewitness reliability is a problem (e.g., Slater, 1994; Wright & McDaid, 1996; Memon et al., 2011). Until now the extent of this problem has been underestimated. This issue will be considered further in the next chapter.

Chapter 5

Summary, Conclusions and Future Research

5.1 Summary and Conclusions

This thesis investigated new methods of postdicting eyewitness accuracy. The introduction began by summarising the problem of incorrect identifications. It then described several variables that have been suggested as suitable to distinguish a correct from an incorrect identification. These include analysing the speed and confidence with which lineup identifications are made (e.g., Sauerland & Sporer, 2007, 2009), deducing eyewitness accuracy from observers' explicit memory for lineup foils (Charman & Cahill, 2012), or from their general ability to process unfamiliar faces (Bindemann, Brown et al., 2012). All of these methods attempt to dissociate eyewitnesses who have made a correct lineup identification from those who have not, but can only do so with limited accuracy.

Chapter 2 introduced repetition priming as a new method of diagnosing the accuracy of eyewitnesses. In the context of criminal identifications, the person previously seen by the eyewitness acts as the stimulus they have been primed towards. A task which requires classification of several faces including the target should show a difference in reaction time if the target has been seen previously. Eyewitnesses are almost solely required to identify people that they have very limited perceptual experience and support for using repetition priming to assess eyewitnesses is offered by its relatively recent application to unfamiliar faces (Martin & Greer, 2011: Martin et al., 2009).

Within this paradigm, participants were presented with a video which introduced an identity. They were asked to select this target from a lineup and then given a repetition priming task (classifying faces as famous or non-famous), which included a bank of famous faces and another of non-famous faces. Within the non-famous faces were the target identity, the foils from the lineup and additional never before seen unfamiliar faces. In Experiment 1, average scores demonstrated the target face elicited negative priming when compared to foils (i.e., a longer reaction time than responses to the other unfamiliar faces). This finding was independent of accuracy in identifying the face in the preceding lineup. In other words, the repetition priming task made it possible to detect covert recognition in the absence of overt recognition. Experiment 2 provided participants with a target-absent lineup instead of the array used previously. Even without the target appearing in the lineup, the subsequent priming task showed a slower response to the perpetrator than to the other faces. In the final experiment in this chapter participants were provided with both a target-present and a target-absent lineup. Furthermore, a second priming task was provided (judging the faces to be old or new to the experiment) after the original famous/not famous task. This second priming task was designed to elicit positive priming (i.e. a speeded response time) and would confirm the delayed response in Experiments 1 and 2 as negative priming. Although the covert recognition was seen once again in this experiment, it was only evident in the famous/not famous task. The old/new task did not demonstrate any difference between the stimuli.

Repetition priming is typically characterised by a facilitation in responses (e.g., Bruce et al., 1998; Ellis et al., 1987, Ellis et al., 1990), however here a slowing in responses for the primed target faces was consistently observed across all three experiments. The cause of this negative priming effect might lie in the implementation of the categorisation task, which required that all unfamiliar faces were classified as “not famous”. This is problematic insofar that the primed targets do not possess the strong familiarity of the already-known famous faces, but are also not strictly unfamiliar to the observers. Familiarity is processed faster than personal semantic information and names (e.g., Young et al., 1986; Young, McWeeney, Hay, & Ellis, 1986), so if the target faces produce a sense of familiarity due to their exposure in the initial video, then this might have been sufficient to interfere with the speeded fame decisions. In an attempt to determine if this could explain the negative priming of the target faces, a second categorisation task was included in Experiment 3, in which observers were required to classify the unfamiliar faces as "old" (i.e., seen before in the experiment) or "new". The accuracy data for this task was generally consistent with the outcome of the identity lineups (i.e. the target was identified as often as they were classified as 'old'). However, the response times failed to produce any clear priming effects. The negative priming effect was consistent throughout the three experiments but facilitation was inconsistent.

The negative priming effects found in Chapter 2 are based on medians from different sample sizes. For every response time recorded for the target, nine fillers had been recorded. This imbalance means that this method is prone to outlier bias in target response time. Furthermore, despite the consistent negative priming effect, this method does not provide an adequate procedure to assess eyewitnesses at the individual level. A Bayesian analysis exploring the use of repetition priming to determine accuracy for each individual eyewitness demonstrated that in its current form the likelihood of an accurate identification can be predicted to a maximum of only 71%. Repetition priming can offer a probability index but until this can be improved individual accuracy data are of limited use.

In order to address the shortcomings of the repetition priming methodology, Chapter 3 explored a more direct test of eyewitness accuracy. A multiple lineup approach was introduced in which participants were shown a target exposure video followed by several lineup trials rather than just one (as would typically be presented to them in a criminal investigation). Contrary to previous uses of this technique, these arrays were *all* comprised of faces rather than a single face lineup being supplemented by others including personal aspects such as clothing, bodies, accessories or voices (see Lindsay et al, 1987; Pryke et al., 2004; Sauerland & Sporer, 2008; Sauerland et al., 2013). This approach was taken for two reasons; because non-portrait face lineups are consistently met with low accuracy, and because of the observation that there exists considerable variation between different instances of the same face (Jenkins et al., 2011). This variation means that different portrait images of a person can provide relatively independent tests of recognition. Within this framework, a familiar face could be recognised in all instances over several lineups, whereas this would be difficult with an unfamiliar face. Therefore, if an eyewitness had gained sufficient familiarisation with the target during the exposure period they should be able to recognise them repeatedly.

Experiment 4 presented participants with target-present lineups to examine their identification accuracy. Fewer participants could consistently identify a target across all three lineups than could identify them in one or two. No one particular array was responsible for the mistakes (errors occurred in all three lineups). This indicates that each lineup does represent a relatively independent test of identification accuracy, and that this procedure is sensitive enough to differentiate between observers. Experiment 5 was a replication using participants who were familiar with the target identities. This variant was conducted to ensure the multiple lineup procedure was a fair test and that it was not superficial aspects of the images that had been the cause of failures to repeatedly identify the targets. Accuracy was perfect in this experiment demonstrating that if sufficient familiarity is acquired this task is possible. Experiment 6 added target-absent lineups to the procedure. It was found that accuracy in any one lineup (target-present or –absent) fell between 57% and 74% but some participants were able to respond perfectly across all six lineups.

By providing six lineups, participants were tested repeatedly at both identifying a target and rejecting a lineup when they were absent. This method is justified by the previous work in the field, but also makes intuitive sense. Pozzulo and O'Neill (2012) found that mock-jury members perceived witness' identifications as more reliable if they had previously identified the same person at an earlier date. It follows that real jurors would put greater trust in an eyewitness who is able to consistently identify a person over multiple lineups compared to one who is accurate only sometimes. This was extended to sequential lineups in Experiment 8. Once again, accuracy on any one lineup was comparable but some participants were able to identify the target consistently while other could not. The data gained in Experiment 8 showed that the same pattern of results can be achieved with a sequential methodology as with a simultaneous lineup procedure when using multiple trials. This is of interest because although some nations use simultaneous lineups as standard, sequential lineups are the preferred method of identification used by the police in the UK (PACE, 1984). The consistent pattern of results observed in this chapter illustrate the usefulness of multiple lineups in either setting.

Experiment 7 explored the possibility that observers could learn the identities of the targets whilst looking at the lineups. Since the foils in this chapter were different in every lineup, the targets were presented more frequently than any other face. It was predicted that consistent identification would be impossible without prior exposure to the target and Experiment 7 tested this by omitting the presentation of a video but by either highlighting the targets (cued condition) or not giving any indication of their identity (uncued condition) in the first lineup. This effectively set participant accuracy at 100% or 0% but did not provide participants with an information rich exposure to the target, but rather, a static image. In the uncued condition accuracy on any one lineup was low and no participants were able to identify the target repeatedly. In the cued condition, however, accuracy for any one lineup was high and 25% of participants were able to respond consistently to the lineups. It would appear from these findings that if a target is initially highlighted it is possible to respond consistently to them even though only a single static image was ever seen. This finding could prove to be problematic if the eyewitness selected an innocent suspect by chance in the first lineup. It would appear from the results of Experiment 7 that in this event a participant may be able to identify the same person repeatedly. Added to this is evidence that observers can 'commit' to an identity after selecting it once (Deffenbacher et al., 2006). These findings in combination could seriously undermine the current methodology by allowing the possibility of multiple identifications of an innocent suspect. However, this chapter could not examine cases of consistent *non*-target identifications since the only repeated identity was that of the perpetrator. This means multiple *misidentifications* were not considered. This problem was addressed in the final experimental chapter.

Chapter 4 further investigated the utility of multiple lineups by including lineups that repeated all identities, not just the targets. It was also conducted as a field experiment to increase ecological validity. In Experiment 9, participants were approached in the street by a target asking for directions. After this exposure and a short delay, they were approached by another experimenter who asked them to identify the target in multiple lineups. As was found in Chapter 3, accuracy was variable for one-off identifications but a select number of participants were able to identify the target repeatedly. The most important analysis in this experiment is the comparison between repeated identification of the target, and repeated *mis*identification of a foil. 28% of participants identified the target repeatedly, whereas only 5% repeatedly identified another face leading to a probability of .85 that the target had been selected if responses were consistent throughout. This probability was increased to 1.00 in

Experiment 10 where the different images of the faces were more varied. In this experiment, the standardised image lineups used in Experiment 9 were supplemented by photo-ID photographs and social media profile images. When these different images were used 5% of participants could identify the target repeatedly and 0% a different face.

Experiment 10 indicates that variability in the appearance of the same person's face is important to the success of a multiple-lineup paradigm. At this stage, however, a principled method to establish when these requirements are met in advance of the administration of identity lineups cannot be provided. It might be possible to achieve this with a simple sorting procedure, whereby observers are asked to group individual face photographs into relevant identities, prior to the lineup construction (e.g., Bruce et al., 1999; Burton et al., 2010; Jenkins et al., 2011). If such sorting can *not* be done accurately by independent observers who are unfamiliar with the lineup identities, then the selected images might provide sufficient within-person variability for a multiple-lineup procedure. The image categories that were used in Experiment 10 may provide a good basis for such a sorting procedure and might be readily available in many instances. For example, standardised images of suspects are already recorded routinely in police investigations, while existing images from photo-identity documents might be accessible on file, and the use of social networking websites is widespread. Alternatively, such social face photographs could be replaced with CCTV stills of the suspect from inside police stations, which are already an accepted image source for lineups in the UK (see PACE, 1984).

There are a number of differences between the experiments presented in Chapters 3 and 4. For example in Chapter 3 no foil was repeated in more than one lineup although the targets appeared multiple times. This meant that if a participant did not remember the target, all the faces in all lineups would have appeared new. If they became aware of the repetition of the target this could lead to increases in target choices. For this reason Chapter 4 included repeated foils so this strategy was no longer available. The size of the lineup was also changed. Chapter 3 included lineups of ten faces, while Chapter 4 included lineups of six. There is some evidence to suggest that lineup size can affect identity rates (Bindemann, Sandford, Gillatt, Avetisyan & Megreya, 2012). However, the same key task difficulty of identifying the target exists in both lineup sizes. It was this task that was important to the current research, and not the identification rates themselves. Despite this focus, the differences in procedure between the two chapters may be a cause of concern but a positive in the current research is that despite these differences the same pattern of results were obtained.

Conversely, the participants' emotional responses to the stimuli were carefully controlled for. It has been previously shown that identification rates are affected by personal involvement in a crime (Hosch et al., 1984) and there is some evidence to suggest that accuracy could be affected by the emotions of the target faces (Jackson, Wu, Linden & Raymond, 2009; Surguladze et al., 2004), however, this research field is limited to clinical samples and the findings are mixed. Despite these mixed results it makes sense that there are more variables involved in a realistic criminal setting than a neutral test of recognition. There are necessarily more distracting elements to a crime scenario than a simple conversation. The advantage of using a conversation as the initial stimulus instead of a crime is that it is possible to measure recognition without the inclusion of these extraneous variables. When introducing a new paradigm a pure test of memory must be the first step since any the memory will be just as affected under more emotionally charged circumstances.

Traditionally, eyewitness identification evidence has been obtained with *live* lineups, in which a suspect is placed among other people. However, this approach has been replaced with video-based identity parades, such as the VIPER system in the UK (n.d.), which are administered by computer. As the success of a multiple-lineup method should depend on providing different instances of the same perpetrator in each of the lineups, it would be much easier to arrange this with a computer-based method than the traditional live lineups. In practice, a procedure that tests the same eyewitness repeatedly therefore does not have to present a fundamental change in the administration of identity parades (i.e., live *versus* computer-based), but only in the procedure in which this format is administered (only once *or* repeatedly). Despite this promise, these findings are clearly preliminary. Future investigations need to assess whether this procedure works equally well under more ecologically valid conditions, for example, when forensically-relevant delays are introduced between exposure and test (Shapiro & Penrod, 1986). These investigations are important since there are inferential difficulties associated with using short-term memory tasks, such as those employed in this thesis, to investigate eyewitness accuracy, when this task typically requires retaining a face in memory for much longer periods of time. It has been previously shown that under longer delays accuracy is reduced (e.g. Krouse, 1981; Bindemann, Avetisyan et al. (2012). However, emerging research has identified some individuals (referred to as super-recognisers) who are able to recognise a face after very long delays (Russell et al., 2009;). It is possible that those who were able to repeatedly recognise the targets in the current experiments are some kind of super-recognisers, in which case accuracy with a longer delay should remain stable. However, if this is not the case there may be a further drop in performance and this may render this method useless. This cannot be determined as yet and requires further investigation.

These predictions are based on models of face recognition that state face recognition units (FRUs) are formed as faces become familiar and incorporate the exposure into a robust representation of the face (Bruce & Young, 1986; Burton et al., 1999). This representation acts like a face average (as introduced by Burton et al., 2005) and can be used to identify a previously seen face from novel viewpoints due to the focus on the salient points of recognition and removal of superficial artifacts not present in all views. In this way internal representations can code for variation within a face. The current experiments have taken advantage of this by testing participants with varied images. The current interpretation of results is that participants here are forming a primitive FRU in their initial exposure to the target and this stored representation facilitates the repeat recognition of the target over the foils.

An as yet unanswered question is how many lineups are necessary to dissociate between occasionally and consistently accurate eyewitnesses. The maximum number of lineups seen by any one participant was six, three target-present and three target-absent, but this number was arbitrary and is not necessarily optimal. In all experiments (bar Experiment 7 where no initial exposure was provided) the same 'stepwise' drop in consistent accuracy can be seen with no apparent 'leveling off' suggesting that further lineups may give an additional benefit in dissociating between eyewitnesses. If, as predicted, some eyewitnesses have gained the necessary familiarity with an identity they should be able to identify the target in any presentation, extremely uncharacteristic views notwithstanding. If this is the case there will come a point where additional lineups do not provide any further improvement but this point is not identifiable from the current data. Of course the identity of the target may dictate the number of lineups required, and this may prove to be variable.

A potential criticism of this multiple-lineup paradigm relates to the *actual* number of eyewitnesses that remain useful after the completion of the procedure. For example, 18 of the 40 observers made a correct decision to the first lineup in Experiment 9 but only 11 managed to do so consistently. In Experiment 10, these numbers were lower still as only 2 of the 19 observers, who initially made correct decision, also acted consistently on the target across all six lineups. The repeated assessment of eyewitness accuracy can therefore lead to the exclusion of a great number of observers that would otherwise appear to be good eyewitnesses by current standards. While this data loss could be reduced by decreasing the number of repeat-identifications, it raises the question of how a "good eyewitness" should be defined more generally. In the study of person recognition in cognitive psychology, the repeated identification of the same person would not be considered a problem for the recognition of family, friends and other acquaintances. In fact, cognitive theories have stipulated for considerable time that the recognition of a familiar person should be triggered by any image of a face (e.g., Bruce & Young, 1986; Burton et al., 2005; Schweinberger & Burton, 2003). According to these theories, the repeated identification of the same person from different images is therefore a basic requirement to confirm that genuine familiarity exists.

Viewed in this way, it could be argued that the reduction of "usable" eyewitnesses in a multiple-lineup procedure should, in fact, be considered a data *gain*. Whereas a single lineup provides a greater pool of eyewitnesses that may be accurate, it can only provide limited information about the actual identification accuracy of a specific individual. In the current experiments, for example, eyewitnesses' responses to a single lineup translated into a probability of between 0.43 and 0.67 that a correct identification had been made across both experiments. These probabilities are such that it is difficult to rely on *any* particular eyewitness if responses to only a single lineup are considered. By contrast, multiple-lineups provided an index that far exceeded the utility of a single lineup, by determining eyewitness identification with near-perfect (a probability of 0.85 in Experiment 9) or perfect accuracy (a probability of 1.00 in Experiment 10). Multiple-lineups therefore offer precision over inclusiveness, but the inclusiveness of the single lineup might also be a fallacy if one wishes to genuinely assess the identification accuracy of *individual* eyewitnesses. The removal of observers that a multiple-lineup procedure necessitates might therefore lead to a more realistic presentation of actual eyewitness accuracy.

In turn, however, it is also possible that a multiple-lineup procedure introduces confounds that eliminate eyewitnesses who, at least initially, might have had a good memory for the perpetrator. This could occur if the faces of additional lineups interfere with the stored representation of the target identity. It has been shown, for example, that identification of mugshots decreases correct identifications and increases false alarms in a subsequent identity lineup (for a review, see Deffenbacher, Bornstein, & Penrod, 2006). The mere exposure to such intervening faces might be sufficient to interfere with the memory of a target (Perfect & Harris, 2003), though studies of this phenomenon have produced mixed results (e.g., Cutler, Penrod & Martens, 1987; Dysart, Lindsay, Hammond & Dupuis, 2001). However, if such interference is robust, then this could serve to eliminate good eyewitnesses over the course of a multiple-lineup procedure. These are important avenues for further research.

The high number of correct decisions on the first lineup introduces another important point of discussion. Subsequent lineups in Chapters 3 and 4 demonstrate that the first lineup often provides misleading information about how many eyewitnesses are accurate. In Chapter 4 the probabilities of a correct identification demonstrate that the first lineup was consistently amongst the least informative of any of the multiple lineups (see Figures 4.2 and 4.4). Now consider that currently the police *only* provide the first lineup when eyewitnesses are involved in their investigations. The need to develop and adopt an approach that utilises multiple lineups becomes clear.

5.2 Future Research

In conclusion, this thesis has applied existing methodologies in the cognitive psychology domain to the problem of assessing eyewitness accuracy. It has provided two new approaches to this problem in repetition priming and multiple face lineups. A problem that has emerged from these approaches is that a lot of eyewitnesses are lost throughout these procedures since they do not perform consistently well. A possible solution to this problem could exist within a combination of the repetition priming paradigm and the multiple lineups of the later chapters. Participants could be presented with a similar video as was provided in Chapter 2, but this would be followed by multiple lineup arrays, some target-present, some -absent, as in Chapter 3. After accuracy has been recorded for these lineups, a repetition priming task should be presented which requires the participant to categorise *all* previously seen images (including all instances of the target) along with suitable filler items. Such an adaptation would provide more reaction time data points for the target responses in the priming task, making the comparison between target and non-target response time more conducive to observing differences at the individual level due to the greater statistical power. This procedure could also take on elements of Chapter 4 and utilise repeated foil identities in the multiple lineups. Although overt perception of multiple presentations of non-target faces was ruled out in Experiment 7, there may be a covert effect of multiple presentations that could be identified by a priming task. In other words, although accuracy data is unaffected by multiple presentations of the same face, there may be a sub-threshold level of recognition both for the target and the foils which may be evident using response times. By considering repeated covert responses some of the eyewitnesses who were unable to identify the same person overtly and disregarded may become useful again. It may also help gain an understanding of the processes of the inconsistent eyewitnesses (i.e. those who can identify the target sometimes but not at other times).

In turn, the multiple lineups paradigm could be extended to include sequential lineups (i.e. arrays presented one face at a time, requiring a ‘yes’ or ’no’ decision as to whether each is the target) as in Experiment 8. Here, repeated foils should be introduced into the procedure. Each trial in a sequential lineup provides a test of absolute recognition. Due to this, intra-target variations in appearance would test the participants more stringently without offering the opportunity to compare the face with the distracters. This is perhaps the most conservative test for preventing false positive identifications within a multiple lineup procedure because of the difficulty in identifying the same face multiple times recorded in Chapters 3 and 4. Any eyewitnesses who could identify the same person repeatedly and reject all other identities in a sequential methodology would be expected to be accurate.

Moving a step away from the practical use of these procedures, investigation into follow-up tests could allow researchers to identify the driving force behind success in lineup identification. It has been suggested in this thesis that a gained familiarisation with the target is the cause of correct responses to the target identity. However, in a study using the 1-in-10 task requiring participants to identify a given face in an array of 10 distractors, Bindemann, Brown et al. (2012) identified a positive correlation between this task and accuracy on an identity lineup. This finding is suggestive of pre-existing individual differences concerning ability to perform this task. A manipulation of initial exposure duration would be expected to affect the identification accuracy if familiarisation is the factor of importance since it will be varied under these circumstances. Conversely, if individual differences are the driving force the accuracy should be less affected by a reduced exposure duration.

References

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

Attard, J., & Bindemann, M. (2014). Establishing the duration of crimes: An individual differences and eye-tracking investigation into time estimation. *Applied Cognitive Psychology, 28*, 215-225. doi:10.1002/acp.2986

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis, *Law and Human Behavior, 25*, 475-491. doi:10.1023/A:1012840831846

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior, 29*, 279-301. doi:10.1007./s10979-00503617-y

Bindemann, M., Attard, J., & Johnston, R.A. (2014). Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychology, 1,* 986-903. doi:10.1080/23311908.2014.986903

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied, 18*, 277-91. doi:10.1037/a0029635

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

Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition, 98*, 177-197. doi:10.1016/j.cognition.2004.11.004

Bindemann, M., Jenkins, R., & Burton, A. M. (2007). A bottleneck in face identification: Repetition priming from flanker faces. *Experimental Psychology, 54*, 192-201. doi:10.1027/1618-3169.54.3.192

Bindemann, M., & Sandford, A. (2011). Me, myself and I: Different recognition rates for three photo-IDs of the same person. *Perception, 40*, 625-627. doi:10.1068/p7008

Bindemann, M., Sandford, A., Gillatt, K., Avetisyan, M., & Megreya, A. (2012). Recognizing faces seen alone or with others: Why are two heads worse than one? *Perception*, *41*, 415-435. doi:10.1068/p6922

Bornstein, B. H., Deffenbacher, K. A., Penrod, S. D., & McGorty, E. K. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime and Law, 18*, 473-490. doi:10.1080/1068316X.2010.508458

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*, 691-695. doi:10.1037//0021-9010.72.4.691

Bradshaw, J. L., & Wallace, G. (1971). Models for the processing and identification of faces. *Perception and Psychophysics, 9*, 443-448. doi:10.3758/BF03210249

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30*, 31-50. doi:10.1007/s10979-006-9002-7

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8*, 46-58. doi:10.1037//1076-898X.8.1.44

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*, 105-116. doi:10.1111/j.2044-8295.1982.tb01795.x

Bruce, V., Burton, M., Carson, D., Hanna, E., & Mason, O. (1994). Repetition priming of face recognition. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp.179-201). Cambridge, Mass: MIT Press.

Bruce, V., Carson, D., Burton, A. M., & Kelly, S. (1998). Prime time advertisements: Repetition priming from faces seen on subject recruitment posters. *Memory and Cognition, 26*, 502-515. doi:10.3758/BF03201159

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

Bruce, V., & Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology, 76*, 373-383. doi:10.1111/j.2044-8295.1985.tb01960.x

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327. doi:10.1111/j.2044-8295.1986.tb02199.x

Brunas, J., Young, A. W., & Ellis, A. W. (1990). Repetition priming for incomplete faces: Evidence for part to whole completion. *British Journal of Psychology, 81*, 43–56. doi:10.1111/j.2044-8295.1990.tb02344.x

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science, 23*, 1-31.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*, 361-380. doi:10.1111/j.2044-8295.1990.tb02367.x

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

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology, 102*, 943-958. doi:10.1111/j.2044-8295.2011.02039.x

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching test. *Behaviour Research Methods, 42*, 286-291. doi:10.3758/brm.42.1.286

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

Burton, A. M., Young, A. W., Bruce, V., Johnston, R. A., & Ellis, A. W. (1991). Understanding covert recognition. *Cognition,* 39, 129-166. doi:10.1016/0010-0277(91)90041-2

Charman, S. D., & Cahill, B. S. (2012). Witnesses' memories for lineup fillers postdicts their identification accuracy. *Journal of Applied Research in Memory and Cognition*, *1*, 11–17. doi:10.1016/j.jarmac.2011.08.001

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior 29*, 575-604. doi:10.1007/s10979-005-7121-1

Clark, S. E, & Tunnicliffe, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behaviour, 25*, 199-216. doi:10.1023/acp.891

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception, 31*, 985-994. doi:10.1068/p3335

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281-290. doi:10.1037/0021-9010.73.2.281

Cutler , B. L. Penrod, S. D. (1995). *Mistaken identification: Eyewitnesses, psychology and the law.* New York: Cambridge University Press.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). Improving the reliability of eyewitness identification: Putting context into context. Journal of Applied Psychology, 72, 629-637. doi:10.1037/0021-9010.72.4.629

Darling, S., Valentine, T., & Memon, A. (2008). Selection of lineup foils in operational contexts. *Applied Cognitive Psychology, 22*, 159–169. doi:10.1002/acp.1366.

Davis, J. P., & Valentine, T. (2009). CCTV on trial: Matching video images with the defendant in the dock. *Applied Cognitive Psychology, 23*, 482-505. doi:10.1002/acp.1490

Deffenbacher, K. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*, 243-260. doi:10.1007/BF01040617

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behaviour, 30*, 287-307. doi:10.1007/s10979-006-9008-1

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*, 687-706. doi:10.1007/s10979-004-0565-x

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 27-47. doi:10.1037/0278-7393.22.1.27

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups – partial remembering. *Journal of Police Science and Administration, 1*, 287-293.

Dysart, J. E., Lindsay, R. C. L., Hammond, R., & Dupuis, P. (2001). Mug shot exposure prior to lineup identification: Interference, transference, and commitment effects. Journal of Applied Psychology, 86, 1280-1284. doi:10.1037/0021-9010.86.6.1280

Eimer, M., Gosling, A., & Duchaine, B. (2012). Electrophysiological markers of covert face recognition in developmental prosopagnosia. *Brain, 135*, 542-554. doi:10.1093/braib/awr347

Ellis, A. W., Burton, A. M., Young, A. W., & Flude, B. M. (1997). Repetition priming between parts and wholes: Tests of a computational model of familiar face recognition. *British Journal of Psychology, 88*, 579–608. doi:10.1111/j.2044-8295.1997.tb02659.x

Ellis, A. W., Flude, B. M., Young, A. W., & Burton, A. M. (1996). Two loci of repetition priming in the recognition of familiar faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 295–308. doi:10.1037/0278-7393.22.2.295

Ellis, A. W., Young, A. W., Flude, B. M., & Hay, D. C. (1987). Repetition priming of face recognition. *Quarterly Journal of Experimental Psychology, 39*, 193–210. doi:10.1080/14640748708401784

Ellis, A. W., Young, A. W., & Flude, B. (1990). Repetition priming and face processing: Priming occurs within the system that responds to the identity of a face. *Quarterly Journal of Experimental Psychology, 42*, 495-512. doi:10.1080/14640749008401234

Fisher, R. P., Geiselman, R., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of the actual victims and witnesses of crime. *Journal of Applied Psychology, 74*, 722-727. doi:10.1037/0021-9010.74.5.722

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 680-698. doi:10.1037/0278-7393.10.4.680

Goshen-Gottstein, Y., & Ganel, T. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgement task: Evidence for a common route for the processing of sex and identity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1198-1214. doi:10.1037//0278-7393.26.5.1198

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology, 99*, 45-56. doi:10.1348/000712607X199981

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*, 330-337. doi:10.1016/S1364-6613(00)01519-9

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223-233. doi:10.1016/S1364-6613(00)01482-0

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15*, 445-464. doi:10.1002/acp.718

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 and Human Behaviour, 38*, 94-108. doi:10.1037/lhb0000060

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior, 36*, 257-265. doi:10.1037/h0093959

Hosch, H. M., Leippe, M. R., Marchioni, P. M., & Cooper, D. S. (1984). Victimization, self-monitoring, and eyewitness identification. *Journal of Applied Psychology, 69*, 280-288. doi:10.1037/0021-9010.69.2.280

Huff, C. R. (1987). Wrongful conviction: Societal tolerance of injustice. *Research in Social Problems and Public Policy, 4*, 99-115.

Innocence Project (1992). Retrieved from http://www.innocenceproject.org/ on 10th October 2013.

Jackson, M. C., Wu, C. Y., Linden, D. E. J., & Raymond, J. E. (2009). Enhanced visual short-term memory for angry faces. *Journal of Experimental Psychology – Human Perception and Performance, 35*, 363-374. doi:10.1037/a0013895.

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science, 319*, 435-435. doi:10.1126/science.1149656

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society - Biological Sciences, 366*, 1671-1683. doi:10.1098/rstb.2010.0379

Jenkins, R., Burton, A. M., & Ellis, A. W. (2002). Long-term effects of covert face recognition. *Cognition, 86*, B43–B52. doi:10.1016/S0010-0277(02)00172-5

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

Johnston, R. A., Barry, C., & Williams, C. (1996). Incomplete faces don't give the whole picture: Repetition priming from jumbled faces. *Quarterly Journal of Experimental Psychology, 49*, 596–615. doi.org/10.1080/027249896392513

Johnston, R. A., & Bindemann, M. (2013). Introduction to forensic face matching. *Applied Cognitive Psychology, 27*, 697-699. doi:10.1002/acp.2963

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 1304-1316. doi:10.1037//0278-7393.22.5.1304

Khurana, B., Smith, W. O., & Baker, M. T. (2000). Not to be and then to be: visual representation of ignored unfamiliar faces. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 246-263. doi:10.1037//0096-1523.26.1.246

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology, 66*, 651-654. doi:10.1037/0021-9010.66.5.651

Lander, K., & Bruce, V. (2004). Repetition priming from moving faces. *Memory and Cognition, 32*, 640-647. doi:10.3758/BF03195855

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84*, 315-321. doi:10.1037/0021-9010.84.3.315

Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*, 741-745. doi:10.1037/0021-9010.76.5.741

Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18*, 527-541. doi:10.1007/BF01499172

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioural Science, 19*, 464-478. doi:10.1037/h0079998

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4*, 303-313. doi:10.1007/BF01040622

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentations. *Journal of Applied Psychology, 70,* 556-564. doi:10.1037/0021-9010.70.3.556

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22,* 1-35. doi:10.1016/0010-0285(90)90002-L

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about weapon focus. *Law and Human Behavior, 11*, 55-62. doi:10.1007/BF01044839

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of auto-mobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585-589.

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15*, 43-57. doi:10.1007/BF01044829

Luus, C. A. E., & Wells, G. L. (1994). Eyewitness identification confidence. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 348-361). Cambridge: Cambridge University Press.

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

Malpass, R. S., & Devine, P. G. (1980). Realism and eyewitness identification research. *Law and Human Behavior, 4*, 347-358. doi:10.1007/BF01040626

Malpass, R. S., & Devine, P. G. (1983). Measuring the fairness of eyewitness identification lineups. In S. M. A. Lloyd-Bostock and B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 81-102). Chichester: Wiley.

Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmet, N., Melsom, E. I., & Lindsay, R. C. L. (2012). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior, 36*, 513-526. doi:10.1037/b0093937

Martin, D., Cairns, S. A., Orme, E., DeBruine, L. M., Jones, B. C., & Macrae, C. N. (2010). Form-specific repetition priming for unfamiliar faces. *Experimental Psychology, 57*, 338-345. doi:10.1027/1618-3169/a000040

Martin, D., & Greer, J. (2011). From view-dependent to view-invariant repetition priming for unfamiliar faces. *Quarterly Journal of Experimental Psychology, 64*, 217-223. doi:10.1080/17470218.2010.541266

Martin, D., Nind, L. K., & Macrae, C. N. (2009). Lateralized repetition priming for unfamiliar faces. *Experimental Psychology, 56*, 147-155. doi:10.1027/1618-3169.56.3.147

Megreya, A. M., & Bindemann, M. (2009). Revisiting the processing of internal and external features of unfamiliar faces: The headscarf effect. *Perception, 38*, 1831-1848. doi:10.1068/p6385

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

Megreya, A. M., & Burton, A. M. (2006a). Unfamiliar faces aren't faces: Evidence from a matching task. *Memory and Cognition, 34,* 865–876. doi:10.3758/BF03193433

Megreya, A. M., & Burton, A. M. (2006b). Recognising faces seen alone or with others: When two heads are worse than one. *Applied Cognitive Psychology, 20*, 957-972. doi:10.10002/acp.1243

Megreya, A. M., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity based dissociation. *Perception and Psychophysics, 69,* 1175–1184. doi:10.3758/BF03193954

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology, 14*, 364-372. doi:10.1037/a001.3464

Megreya, A. M., Sandford, A., & Burton, A. M. (2013). Matching face images taken on the same day or months apart: The limitations of photo ID. *Applied Cognitive Psychology, 27*, 700-706. doi:10.1002/acp.2965

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3-35. doi:10.1037//1076-8971.7.1.3

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime and Law, 17*, 711-729. doi:10.1080/10683160903524333

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

Morrison, D. J., Bruce, V., & Burton, A. M. (2000). Covert face recognition in neurologically intact participants. *Psychological Research*, *63*, 83–94. doi:10.1007/s004260000037

Münsterberg, H. (1908). *On the witness stand.* New York: Doubleday, Page and Co.

Murray, D. M., & Wells, G. L. (1982). Does knowledge that a crime was staged affect eyewitness performance? *Journal of Applied Social Psychology, 12*, 42-53. doi:10.1111/j.1559-1816.1982.tb00847.x

Navon, D. (1992). Selection of lineup foils by similarity to the suspect is likely to misfire. *Law and Human Behavior, 16*, 575-593. doi:10.1007/BF1044624

Neil, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 1291-1305. doi:10.1037/0278-7393.23.6.1291

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259. doi:10.1037/0033-295X.84.3.231

O'Toole, A. J., Phillips, P. J., Weimer, S., Roark, D. A., Ayyad, J., Barwick, R., & Dunlop, J. (2011). Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. *Visio Research, 51*, 74-83. doi:10.1016/j.visres.2010.09.035

Perfect, T. J., & Harris, L. J. (2003). Adult age differences in unconscious transference: Source confusion or identity blending? Memory & Cognition, 31, 570-580. doi:10.3758/BF03196098

Police and Criminal Evidence Act (PACE) - Code D. (1984). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117599/pace-code-d.pdf on 7th July 2014.

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

Pozzulo, J. D., & O'Neill, M. C. (2012). Juror decision making when a witness makes multiple identification decisions. *Journal of Applied Social Psychology, 42*, 1192-1217. doi:10.1111/j.1559-1816.2011.00883.x

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. *Journal of applied Psychology, 89*, 73-84. doi:10.1037/0021-9010.89.1.73

Read, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology, 9*, 91-121. doi:10.1002/acp.2350090202

Robbins, R. A., & Coltheart, M. (2012). The effects of inversion and familiarity on face versus body cues to person recognition. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1098-1104. doi:10.1037/a0028584

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22*, 642-647. doi:10.1111/j.1469-8986.1985.tb01661.x

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin and Review*, *16*, 252-257. doi:10.3758/PBR.16.2.252

Sandford, A., & Burton, A. M. (2014). Tolerance for distorted faces: Challenges to a configural processing account of familiar face recognition. *Cognition, 132*, 262-268. doi:10.1016/j.cognition.2014.04.005

Sauer, J. D., Brewer, N., & Wells, G. L. (2008). Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential lineups? *Legal and Criminological Psychology, 13*, 123-135. doi:10.1348/135532506X159203

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime and Law, 13*, 611-625. doi:10.1080/10683160701264561

Sauerland, M., & Sporer, S. L. (2008). The application of multiple lineups in a field study. *Psychology, Crime and Law, 14*, 549-564. doi:10.1080/10683160801972519

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied, 15*, 46-62. doi:10.1037/a0014560

Sauerland, M., Stockmar, A. K., Sporer S. L., & Broers, N. J. (2013). The reliability of identification evidence with multiple lineups. *The European Journal of Psychology Applied to Legal Context, 5*, 49-71.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 1-17. doi:10.1037/0096-1523.3.1.1

Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology, 2*, 99-117. doi:10.1348/174866407X256554

Schweinberger, S. R., & Burton, A. M. (2003). Covert recognition and the neural system for face processing. *Cortex, 39*, 9-30. doi:10.1016/S0010-9452(08)70071-6

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75*, 221-242.

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

Simion, F., Turati, C., Valenza, E., & Leo, I. (2006). The emergence of cognitive specialization in infancy: The case of face preference. In M. Johnson & M. Munakata (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 189-208). Oxford University Press: Oxford.

Slater, A. (1994). *Identification parades: A scientific evaluation.* London: Police Research Group, Home Office.

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

Sporer, S. (1992). Postdicting eyewitness accuracy – Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22*, 157-180. doi:10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33. doi:10.1037//0021-9010.78.1.22

Sporer, S. L. (2001). The cross-race effect: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law, 7*, 170-200. doi:10.1037//1076-8971.7.1.170

Sporer, S., Penrod, S., Read, D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*, 315-327. doi:10.1037//0033-2909.118.3.315

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413-424. doi:10.1007/BF02352267

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

Surguladze, S. A., Young, A. W., Senior, C., Brebion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology, 18*, 212-218. doi:10.1037/0894-4105.18.2.212

Terry, R. L. (1994). Effects of facial transformations on accuracy of recognition. *Journal of Social Psychology, 134*, 483-492. doi:10.1080/00224545.1994.9712199

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behaviour, 22*, 217-237. doi:10.1023/A:1025746220886

Tredoux, C. G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology, 13*, 9-26. doi:10.1002/(SICI)1099-0720(199911)13:1+<S9::AID-ACP634>3.0.CO;2-1

Trollestrup, Turtle & Yuille (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-160). Cambridge: Cambridge University Press.

Tunnicliffe, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions. *Law and Human Behavior, 24,* 231-258. doi:10.1023/A:1005463020252

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17*, 969-993. doi:10.1002/acp.939

VIPER. Retrieved March 26, 2015, from http://www.viper.police.uk/.

Weber, N., & Brewer, N. (2003). The effect of judgement type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology, 88*, 490-499. doi:10.1037/0021-9010.88.3.490

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12 second rule. *Journal of Experimental Psychology: Applied, 10*, 139-147. doi:10.1037/1076-898X.10.3.139

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

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48*, 553-571. doi:10.1037//0003-066X.48.5.553

Wells, G. L. (2014). Eyewitness identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Current Directions in Psychological Science, 23*, 11-16. doi:10.1177/0963721413504781

Wells, G. L., Leippe, M. E., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and human Behavior, 3*, 285-293. doi:10.1007/BF01039807

Wells, G. L., & Luus, E. (1990). Police lineups as experiments: Social methodology as a framework for properly-conducted lineups. *Personality and Social Psychology Bulletin, 16*, 106-117. doi:10.1177/0146167290161008

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest, 7*, 45-75. doi:10.1111/j.1529-1006.2006.00027.x

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp 155-170). New York: Cambridge University Press.

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., Rydell, S. M., & Seelau, E. P. (1993). The selection of distracters for eyewitness lineups. *Journal of Applied Psychology, 78,* 835-844. doi:10.1037/0021-9010.78.5.835

White, D., Burton, A. M., Jenkins, R., & Kemp, R. I. (2014). Redesigning photo-ID to improve unfamiliar face matching performance. *Journal of Experimental Psychology: Applied, 20*, 166-173. doi:10.1037/xap0000009

White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback for facial image comparison. *Psychological Bulletin Review, 21*, 100-106. doi:10.3758/s13423-013-0475-3

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

Wogalter, M. S., Marwitz, D. B., & Leonard, D. C. (1992). Suggestiveness in photospread line-ups: Similarity induces distinctiveness. *Applied Cognitive Psychology, 6*, 443-453. doi:10.1002/acp.2350060508

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology, 10*, 75-84. doi:10.1002/(SICI)1099-0720(199602)10:1<75::AID-ACP364>3.0.CO;2-E

Wright, D. B., & Skagerberg, E. M. (2007). Post-identification feedback affects real eyewitnesses. *Psychological Science, 18*, 172-178.

Young, A. W., & Burton, A. M. (1999). Simulating face recognition: Implications for modelling cognition. *Cognitive Neuropsychology, 16*, 1-48. doi:10.1080/026432999380960

Young, A. W., Hellawell, D., & De Haan, E. H. F. (1988). Cross-domain semantic priming in normal subjects and a prosopagnosic patient. *Quarterly Journal of Experimental Psychology: Applied, 40*, 561-580. doi:10.1080/02724988843000087

Young, A. W., McWeeney, K. H., Ellis, A. W., & Hay, D. C. (1986). Naming and categorizing faces and written names. *The Quarterly Journal of Experimental Psychology, 38*, 297-318. doi:10.1080/14640748608401599

Young, A. W., McWeeney, K. H., Hay, D. C., & Ellis, A. W. (1986). Access to identity-specific semantic codes from familiar faces. *The Quarterly Journal of Experimental Psychology, 38*, 271-295. doi:10.1080/14640748608401598

Yuille, J.C., & Cutshall, J.L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71*, 291-301. doi:10.1037//0021-9010.71.2.291

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., Dong, Q., Kanwisher, N., & Liu, J. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology, 20*, 137-142. doi:10.1016/j.cub.2009.11.067