Multiple Lineup Identification Procedure:

Utility with Face-Only Lineups

by

NATALIE KALMET

A thesis submitted to the Department of Psychology

in conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

September, 2009

Copyright © Natalie Kalmet, 2009

Abstract

Pryke, Lindsay, Dysart, and Dupuis (2004) investigated a novel method of lineup administration where participants made identifications from multiple lineups showing faces and bodies or playing recorded voices. Identifications from these multiple lineups was diagnostic of guilt; that is, the more lineups a person was selected from the more likely it was that the selected person was actually seen by the witness (as opposed to an innocent suspect; Pryke et al., 2004). The current studies expanded on this procedure and assessed how well the multiple lineup method works when each of the lineups for a target show faces of the same lineup members, with each lineup showing the members facing one of three angles. In Experiment 1, participants (*n* = 72) saw the targets in the same three views that were shown in the lineups and were asked to make lineup decisions for each of the three lineups. In Experiment 2, participants (*n* = 96) saw the targets in only one view, which did not always match the views seen in the lineups. Again, participants made lineup decisions for each of the three lineups. For both studies, when the data were collapsed across targets, the procedure was diagnostic in that more selections were associated with a higher probability of guilt (operationalized as being the previously seen target). However, the effectiveness of the procedure varied across targets such that in some cases multiple selections were no more diagnostic of guilt than single selections. Pryke et al., (2004) reported that multiple identifications were highly diagnostic of guilt but relatively rare. In the current studies, most participants made multiple identifications of the targets, probably because all of the lineups used photos of faces. Results for assessments of confidence-accuracy and advantages for certain lineup angles were generally mixed and often differed between the two studies. In all, the most pertinent assessments of utility (diagnosticity and percentage of participants making multiple identifications) showed promise for using multiple lineups of faces.

Acknowledgements

My Supervisor, Rod Lindsay, for his support and guidance; my committee members, Lee Fabrigar and Kevin Munhall, for their input and advice; Daniel Stober for assistance with data collection and programming; Jamal Mansour, Michelle Bertrand, Jen Beaudry, and Liz Whaley for assisting with data collection and general support; Jeremy Leveque for unconditional support and for understanding the thesis would be done one day.

Social Sciences and Humanities Research Council of Canada for providing me with a scholarship and funding Rod Lindsay, which made the bulk of the research here possible. Natural Sciences and Engineering Research Council of Canada and National Institute on Deafness and Other Communication Disorders for providing Kevin Munhall with funding, which aided in the creation of some stimuli used in the thesis.

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter 1: Introduction

In the criminal justice system, eyewitnesses provide evidence about criminal events. In addition to providing information about what occurred, eyewitnesses often also provide identifications of suspects, which is direct evidence that an accused person (suspect) was present during the crime (Wells, 1984). Despite the perceived strength of eyewitness identifications, eyewitnesses are not infallible. Clear examples that misidentifications can occur include cases where convicted persons were later exonerated with DNA evidence. The Innocence Project (n.d.) states that of the DNA exonerations in the United States (over 220 cases), 75% involved eyewitnesses who misidentified the defendant. In one case, five eyewitnesses wrongly identified the defendant (Wells et al., 1998). In all, incorrect identifications are the leading cause of wrongful convictions in DNA exoneration cases (Innocence Project, n.d.).

The problem of misidentification is only enhanced by the fact that people have difficulty discriminating between eyewitnesses who are accurate and eyewitnesses who are not. For example, Lindsay, Wells, and O'Connor (1989) showed mock jurors video testimony of eyewitnesses (other participants who had witnessed a staged event), which included questioning from real prosecutors and defense attorneys. In one video, the defense attorney asserted that the case would be thrown out by a judge because his cross-examination had completely discredited the witness. The prosecutor agreed with the defense. Of the mock jurors watching this video, nearly 75% stated the defendant was guilty. Overall, the percentage of juror-participants who voted guilty was not significantly different between cases in which the witness had correctly (68%) and incorrectly (70%)

identified defendants. So, not only can eyewitnesses misidentify people, but jurors may also not be able to catch these misidentifications. As the consequences for wrongly identified people are serious (e.g., imprisonment), it is important to minimize misidentifications. At the same time, many cases could not be prosecuted without identification evidence so it is equally important to retain correct identifications.

Two types of variables influence eyewitness memory, including identifications: estimator and system variables (Wells, 1978). Estimator variables are factors that can influence memory accuracy but cannot be manipulated by actors in the criminal justice system (e.g., police), and include perpetrator and witness ethnicity, lighting conditions, and amount of violence (Wells, 1978). While knowledge of these variables can indicate whether there could be problems with witness memory in certain situations (Wells, 1978), they cannot be manipulated and used as a way to minimize mistaken identifications. In contrast, system variables are factors that could be controlled by the criminal justice system and could be manipulated in order to minimize mistaken identifications (Wells, 1978). For example, different types of identification procedures could influence the rates at which eyewitnesses correctly and falsely identify suspects.

When people witness crimes, it is not unusual for the police to ask them to identify the person(s) who committed the crime. One common procedure used in identifications is a lineup, which generally includes a single suspect among multiple foils (i.e., lineup members that are known innocents; Lindsay & Wells, 1985). Witnesses decide whether or not the perpetrator is present – and, if present, which lineup member it is. There are currently two main types of lineups: simultaneous and sequential. Simultaneous lineups present all lineup members at the same time. Lindsay and Wells (1985) argued that simultaneous lineups encourage witnesses to compare lineup members to each other, frequently leading them to select the lineup member most similar to their memory of the criminal. Wells (1984) referred to this approach as a relative judgment strategy. The use of relative judgments is problematic if the perpetrator is absent from the lineup as it leads to high rates of false positive selections, and thus false identifications in real world cases (Wells, 1984).

Lindsay and Wells (1985) created the sequential lineup method, which was intended to reduce reliance on relative judgments by making comparison of lineup members difficult and obviously ineffective as an identification strategy. They hoped their new technique would promote reliance on an absolute judgment strategy where decisions are based solely on the degree of match between each lineup member and the witness' memory of the perpetrator (Lindsay & Wells, 1985; Wells, 1984). The sequential lineup procedure presents lineup members one at a time (sequentially) and requires that the person administering the lineup does not know who the suspect is (blind testing) in addition to requiring that witnesses must make decisions about each lineup member before moving on, can only see each lineup member once, cannot change their decisions once made, and cannot know the number of lineup members to be shown (Lindsay & Wells, 1985).

When simultaneous and sequential lineups were compared, there was no significant difference in the rate of selecting the guilty suspect (i.e., the target in target-present lineups), but selections of innocent suspects (from target-absent lineups) was much less likely with sequential lineups than with simultaneous ones (Lindsay & Wells, 1985). In fact, a higher percent guilty was found for sequential (74.63%) than for simultaneous lineups (56.70%; Lindsay & Wells, 1985). Percent guilty is a descriptive statistic and is 100 times the ratio of witnesses making suspect identifications from target-present lineups to witnesses making suspect identifications from both target-present and target-absent lineups (i.e., Pryke, Lindsay, Dysart, & Dupuis, 2004). Percent guilty above 50% indicate that the identified suspect is likely the target (100% means that the identified suspect is always the target). Percent guilty below 50% indicate that the identified suspect is not the target (0% means that the identified suspect is never the target). Percent guilty of 50% indicates that the identification is not diagnostic of either guilt or innocence. A meta-analysis of studies testing the two lineup methods confirmed a significantly smaller proportion of participants identified a designated innocent suspect in target-absent lineups using the sequential method (9%) than the simultaneous one (27%; Steblay, Dysart, Fulero, & Lindsay, 2001). However, there was also a significantly smaller percentage of participants who identified guilty suspects from target-present lineups using sequential lineups (35%) compared to simultaneous lineups (50%). On average, the percent guilty was higher for sequential (79.54%) than simultaneous (64.94%) lineups.

Although innocent suspects are less likely to be misidentified with sequential lineups, the rate of false identifications is still very high, especially as many innocent suspects in the real world can be – and have been – prosecuted and imprisoned (Dupuis & Lindsay, 2007). In addition, there are serious disadvantages to the sequential method. For example, the procedures necessary for it to work properly may not be followed in reality (Dupuis & Lindsay, 2007) and it can promote increased identifications of innocent suspects when used with children (Pozzulo & Lindsay, 1998). Dupuis and Lindsay (2007) recommended that research on alternate lineup methods be pursued in order to further minimize identifications of innocent suspects and maximize identifications of guilty suspects. Examples of such alternatives include larger lineups (Levi, 2007), allowing people to make multiple selections of lineup members from large, sequential lineups (Levi, 2006), and using multiple lineups (Melara, DeWitt-Rickards, & O'Brien, 1989; Pryke et al., 2004).

Early examples of research using multiple lineups were published by Melara et al. (1989) and Lindsay, Wallbridge, and Drennan (1987). Melara et al. (1989) paired photos with voices, so that participants were simultaneously presented with both a photo and a voice clip for any particular lineup member, and found that identification accuracy increased with the simultaneous presentation of photos and voices (as opposed to only seeing photos or only hearing voices). Another example is Lindsay et al. (1987), who showed participants a face lineup followed by a clothing lineup. Overall, the percent guilty for face lineups was significantly higher if participants also identified the target clothing (93%) than if they did not identify the target clothing (66%; Lindsay et al., 1987). In short, face identifications were much more diagnostic of guilt when the suspect clothing was also identified.

Pryke et al. (2004) substantially expanded on this lineup method by assessing whether identifications of a suspect from several independent lineups (faces, voices, and bodies) would be more diagnostic of the suspect's guilt. Importantly, Pryke et al. argued that the lineups used were independent because people who had not seen the target were unable to match the various features by examining the lineups. When lineups are independent, the likelihood of a witness consistently selecting an innocent lineup member by chance is minimal as the witness would not have memory traces for multiple aspects of an innocent suspect (Pryke et al., 2004). The likelihood of identifying a guilty suspect from multiple independent lineups is higher as witnesses would likely have memory traces for multiple aspects of a guilty suspect (Pryke et al., 2004). For example, fair lineups that show faces only will not provide much information about fair lineups that show bodies only, thus minimizing a person's ability to simply match lineup members across arrays.

To provide a stringent test of diagnosticity, the innocent suspect was chosen post hoc as the lineup member most often misidentified as the target from the target-absent face array (referred to as the "worst-case scenario"; Pryke et al., 2004). In Pryke et al.'s (2004) Experiment 1, participants examined a face lineup, a body lineup, and a voice lineup (in that order), where the target was either present in all three lineups, or absent from all three. Using percent guilty, if participants rejected all three lineups (i.e., 0 suspect identifications made), this result was indicative of the suspect being innocent (25% guilty), as it was below 50%. If participants selected the suspect from only one lineup, this was only slightly indicative of guilt (58% guilty), as it was only slightly higher than 50%. However, selecting a suspect from two or three lineups was more indicative of guilt, as percent guilty increased to 89% and 91%, respectively.

In Pryke et al.’s (2004) Experiment 2, participants saw face, body, voice, and clothing lineups (in that order), with the target again either present in all lineups or absent from all. If zero suspect identifications were made, the percent guilty was very low (26%), but was somewhat higher if one suspect identification was made (63%). However, the percent guilty were extremely high when two (91%), three (100%), or four (100%) suspect identifications were made. Percent guilty of 100% indicate that *only* guilty suspects were identified under those conditions. Pryke et al.'s (2004) findings suggest that the multiple identification procedure may provide identification evidence that is self-calibrated; that is, the more times the suspect is selected, the stronger the evidence is that the suspect is guilty. The multiple identification procedure would provide potentially more useful evidence for the police and the courts than what is currently available from identification procedures.

A limitation of the Pryke et al. (2004) results however, was that multiple identifications of suspects were relatively rare, as voice and body identifications were much less frequent and less accurate than face identification. Although adding the clothing lineup increased multiple selections, clothing is not unique to individuals and using clothing identifications could be controversial in court.

Boyce, Lindsay, Dupuis, and Beaudry (2007) attempted to overcome the limitations of Pryke et al., by adding additional person lineups. Participants saw a total of five lineups: face-forward, body-forward, voice, body-profile, and face-profile. The profile and forward lineups were not independent for either the face or body lineups (as tested in pilot work). As in the previous work, the target was either always present in or always absent from the set. The innocent suspect was chosen using the same worst-case method as Pryke et al. (2004). In Experiment 1, even though some lineups were not independent, Boyce et al. (2007) had promising results, with lower percent guilty for zero (15%) and one (47%) identifications and much higher percent guilty for two (85%), three (85%), and four (82%) identifications (no one made five suspect identifications).

In Experiment 2, Boyce et al. (2007) used the same lineups, but compared the utility of the multiple lineup technique when the lineups were presented sequentially versus simultaneously. In general, sequential lineups provided better evidence than simultaneous lineups. For simultaneous lineups, percent guilty was low for zero (0%) and one (19%) identifications, increased with two identifications (80%), but then decreased substantially when three (58%) or four (57%) identifications were made. No participant identified a suspect from all five lineups using the simultaneous procedure. With sequential lineups, percent guilty was low for zero identifications (18%), increased to about chance levels with one (47%) or two (55%) identifications, increased further with three identifications (90%), and achieved perfect diagnosticity with four (100%) and five (100%) identifications.

Of particular interest for the current studies, percent guilty was 100% with sequential lineups when suspects were identified from both the face-forward and face-profile lineups (regardless of how many other identifications were made). As well, the highest percent guilty in Experiment 1 was when the suspect was identified from both the face-forward and face-profile lineups (95%). Both results are promising, as the forward and profile lineups lack independence.

Subsequent research by Sauerland and Sporer (2008) examined the multiple lineup procedure in a field setting, where participants were approached, asked for directions, and then asked to identify who they had been speaking to shortly thereafter. Participants were shown two face arrays (¾ and profile poses), a body lineup, and a bag lineup, with the order always being face-¾, body, bag, and then face-profile. Unlike the other multiple lineup studies, lineups were not always either target-present or target-absent: if the first two lineups were target-present, the latter-two were target-absent (and vice versa). As face-forward lineups were not included in this study, the worst-case scenario was applied to the face-¾ lineup. Because half of the lineups included the target and half did not, percent guilty could only be calculated for zero, one, or two identifications, and had to be calculated separately for the face-¾ and body lineups, and the bag and face-profile lineups. For the face-¾ and body lineups, percent guilty was low when zero suspect identifications were made (33%), and rose when the suspect was identified in either one (76%) or both of the lineups (74%). For the bag and face-profile lineups, the percent guilty was just below chance with zero suspect identifications (43%), and again rose when the suspect was identified in one (82%) or two (89%) of the lineups.

Although the results from Pryke et al. (2004), Boyce et al. (2007), and Sauerland and Sporer (2008) are promising, some findings do not support the utility of the multiple lineup strategy. In particular, relatively few participants make identifications of guilty suspects from at least two lineups. In Pryke et al.'s (2004) Experiment 1, 54% of participants identified the (guilty) suspect at least twice, but only 13% selected the guilty suspect from all three lineups. In Experiment 2, 50% identified the guilty suspect from at least two lineups, but only 10% identified the target from three lineups and only 7% identified him from all four. In Boyce et al.'s (2007) Experiment 1, 64% of participants made multiple identifications of the target; 22% made identifications from three lineups, and only 9% did so from four lineups. In their Experiment 2, when simultaneous lineups were used, 86% of participants made multiple identifications, though much fewer did so with three (18%), four (4%) or all five (0%) lineups. With sequential lineups, 63% of participants made multiple identifications, with fewer doing so from three (27%), four (10%), or five (3%) lineups. In Sauerland and Sporer (2008), only 12% of participants identified the target from both face-¾ and body lineups and only 2% identified the target from both bag and face-profile lineups.

As Boyce et al. (2007) found, sequential lineups promoted accuracy, but multiple identifications were more likely with simultaneous lineups. Thus, the challenge is to find a balance between accuracy and number of identifications, wherein the percentage of participants making two or more identifications is increased, without also substantially damaging the utility (i.e., diagnosticity) of the multiple lineup method; this is the major goal of the research reported here. The procedure used in this thesis involves showing participants multiple lineups where all the lineups consist of faces with each lineup including the same lineup members, but from different angles. One potential problem with this approach is that it deviates the most from Pryke et al.'s (2004) concern that the lineups should be independent, as lineups that are not independent may increase the likelihood of multiple false identifications of innocent suspects. In the studies conducted below, if an innocent suspect is selected in one lineup, seeing that lineup member in another face lineup may make it more likely that the participant will identify the same person multiple times, as viewing one lineup will provide information that could lead participants to choose the same lineup member from subsequent lineups.

Research into carry-over (or commitment) effects does substantiate this worry about using lineups that are not independent. For example, if participants are shown photos of foils multiple times or perform mugshot searches, the likelihood that those foils/mugshots will be incorrectly identified as the suspect in a later lineup increases – especially when a foil was earlier misidentified as the target (Deffenbacher, Bornstein, & Penrod, 2006; Haw, Dickinson, & Meissner, 2007). If the commitment effect occurs when using multiple face lineups, the benefits of the multiple lineup approach – as found in Pryke et al. (2004), Boyce et al. (2007), and (to a lesser extent) Sauerland and Sporer (2008) – may disappear, as misidentifications of innocent suspects will be more likely to happen in two or more lineups (thus reducing the diagnosticity of the multiple lineup method).

Yet, some evidence indicates that using lineups that are not independent will not necessarily result in greater numbers of innocent suspects being identified. As Burton and Jenkins (in press) note, people are not necessarily accurate when matching faces – particularly when the faces are of unfamiliar people – even when matching photos against people who are physically present. In Bruce et al. (1999), participants were relatively poor when matching pictures of unfamiliar others that were presented with similar – but by no means identical – fillers. In fact, even when pose and facial expression were the same between exposure and test, the target was not selected from the array 30% of the time; one particular target was correctly identified by only 20% of the participants (Bruce et al., 1999). As well, when Boyce et al. (2007) included both face-front and face-profile lineups, the lineups were actually beneficial – the highest percent guilty was found when participants identified the suspects from both of these lineups. Thus, even when only face lineups are used with the multiple lineup procedure, it may be difficult for participants to simply match lineup members across the lineups (especially as pose differs), thus preserving the diagnosticity of the multiple lineup method. Also, people may recognize a person selected in a previous lineup but realize from the different perspective that the previously selected person is not the target.

When choosing which angles to include in the multiple lineups in the current analyses, an array showing lineup members face-front was important, as it is a pose used throughout eyewitness identification literature – which reflects its widespread use by police forces. An array showing lineup members in profile was also included as it has also appeared in eyewitness literature – including other multiple lineup research (e.g., Boyce et al., 2007; Sauerland & Sporer, 2008).

Research in the face recognition area has indicated that the face-¾ view has an advantage over other views (Bruce, Valentine, & Baddeley, 1987; Krouse, 1981; Logie, Baddeley, & Woodhead, 1987), particularly in terms of recognizing unfamiliar others (Bruce et al., 1987). This advantage in recognition could be because some features used when telling others apart are missing in other views, but are all present in the face-¾ view (Krouse, 1981), or because it combines features seen in face-forward and face-profile views into one view (Laughery, Alexander, & Lane, 1971; Liu & Chaudhuri, 2002). For these reasons, the ¾ view was included in a third array.

However, other researchers have not found a face-¾ advantage (Hill, Schyns, & Akamatsu, 1997; Laughery et al., 1971) – especially when the view in the exposure matches the view during the test, or when the difference in angle of rotation between exposure and test is taken into account (Liu & Chaudhuri, 2002). For example, regardless of what view was seen during exposure, as the difference between the exposure and test views increases, recognition decreases (Hill et al., 1997). Others have noted that face-forward and face-¾ views have similar recognition accuracy when view changes between exposure and test (while face-profile leads to worse accuracy; Stephan & Caine, 2007), or that an advantage exists for an exposure angle of 22.5°, rather than for a ¾ angle (i.e., 45°; Laeng & Rouw, 2001).

Although some researchers find no advantage with a face-¾ view, there are still reasons to include this view in the research conducted here. In particular, a face-¾ lineup is a mid-point between face-forward and face-profile views, such that if exposure shows targets only from the front or profile, there may be enough overlapping information in a face-¾ lineup to increase the likelihood of making at least two correct identifications. As well, although most recent research argues against the face-¾ advantage, the research does not seem to indicate that there is a disadvantage to the ¾ pose when compared to the other views (in fact, it is the profile angle that seems to have a recognition disadvantage, particularly when exposure and test views differ; Stephan & Caine, 2007). Finally, the absolute accuracy of the identification decisions from the individual lineups is less important than the diagnosticity of the overall procedure. Three face lineups are expected to result in more identifications than a face lineup, a body lineup, and a voice lineup, thus presenting greater opportunity to demonstrate the possible utility of the multiple lineup strategy.

In addition to the presenting participants with three face lineups, the research conducted here also measured confidence for each lineup decision in order to assess the relationship between confidence and accuracy using the multiple lineup strategy. This relationship is an important one to assess, as people often rely on confidence when determining accuracy – even though confidence is not necessarily a strong indication of accuracy (Lindsay, 1994; Wells et al., 1998). There is a small correlation between confidence and accuracy; however, it is neither strong nor stable enough to clearly indicate when an identification is accurate (Lindsay, 1994). In one meta-analysis, the correlation between confidence and accuracy was $r = .25$, with a 95% confidence interval that excluded 0, but varied from .08 to .42 (Bothwell, Deffenbacher, & Brigham, 1987). Another meta-analysis found a similar confidence-accuracy correlation of $r = .28$ (weighted), with a 95% confidence interval of .15 to .41 (Sporer, Penrod, Read, & Cutler, 1995). This meta-analysis also found that the (weighted) confidence-accuracy correlation was consistently stronger for those who made identifications ($r = .37$) than for those that did not choose lineup members ($r = .12$; Sporer et al., 1995); a distinction also found in Sauerland and Sporer (2009). The confidence-accuracy relationship can vary depending on how quickly lineup decisions are made (Sauerland & Sporer, 2009) and whether or not the witness and perpetrator are of the same ethnicity (Wright, Boyd, & Tredoux, 2001).

The confidence-accuracy relationship has not been assessed in the multiple lineup context. Thus, the current set of studies addressed how, or if, confidence and accuracy vary together when identifications are made from multiple lineups. Of particular interest is whether there is a confidence-accuracy relationship even when the lineups used are not independent, as commitment effects may make it more likely that identifications are made with confidence, regardless of whether or not the decision is correct.

The two studies described below extend prior research by examining the multiple lineup method using only face views in the lineups and by examining the relationship between confidence and accuracy as it exists with this lineup method. In addition, the utility of the multiple lineup method when target exposure is limited was investigated. In the prior research, the exposure participants had to the targets was not directly manipulated but varied naturally with multiple angles shown for different amounts of time using either live or taped exposure. As well, the multiple lineup research noted earlier used only between-subjects designs. The current research used a mix of within- and between-subjects designs (Experiment 1) or a within-subjects design only (Experiment 2). If within-subjects designs provide equivalent results, this will be helpful to researchers, as within-subjects designs require fewer participants.

Chapter 2: Experiment 1

The first study was designed to assess the multiple lineup procedure using a best-case exposure scenario, wherein participants see the target in the same three views shown later in the lineups. In this case, exposure showed each target from the front, ¾, and profile views. The experiment was designed to allow for some comparison of results when participants were naïve versus knowledgeable about the experiment's purpose. Many between-subject designs in the multiple lineup research have ensured during the exposure period that participants were not aware that they would be making identifications of the target later. However, within-subjects designs would mean participants would be aware that they may be making identifications later (as they would have done so for previous targets). Thus, knowing how prior knowledge can impact results will help put results of between- and within-subjects designs into context. In all, the question here is how similar the results would be for these two types of designs.

Four hypotheses were tested in Experiment 1. First, it was hypothesized that the percent guilty would increase as the number of lineups a lineup member was identified in increased. In particular, the percent guilty should be low when the suspect was identified from zero lineups, higher when the suspect was identified from one lineup, and increase further with identification from two or all three lineups. Second, it was hypothesized that participants would be more accurate when identifying suspects from the face-¾ lineup than from the other lineups. Although recent face recognition literature states that there is no advantage for the ¾ view, as noted earlier, target exposure included all three pertinent views (front, ¾, and profile). As participants saw the targets in the ¾ view and because it is the midpoint between the front and profile views (meaning that the facial features seen in ¾ overlap with those seen in the other two views and the difference in angle between ¾ and the other two views is minimal), participants may be most accurate with this lineup.

Third, it was hypothesized that the diagnosticity of weighted confidence – where confidence in suspect identifications from all three lineups was taken into account – would increase as weighted confidence increased. In short, higher values of weighted confidence (indicating multiple suspect selections from lineups and high confidence in those selections) should occur more often when the target was present in the lineups than when the target was absent. The fourth hypothesis was that the above results would differ depending on whether or not participants were aware of the study's purpose. Results may differ when participants know the purpose of the research, as they may pay more attention to the targets and thus form stronger memory traces for the target faces.

Method

Participants

Participants were 72 Queen's undergraduates (52 female, 20 male). The average age of participants was 18.72 (*SD* = 2.00), with ages ranging from 18 to 29. The majority of participants were of European ethnicity (76.39%), with Asian (15.28%), Middle Eastern (2.78%), African (1.39%), First Nations (1.39%), and multiple or mixed (2.78%) ethnicities also represented. All participants were recruited from the Psychology 100 subject pool and were compensated with either credit or $5. Participants had to be naïve to the study's purpose, so participants who had previously taken part in an eyewitness study were prevented from booking an appointment via the subject pool system.

Data Analysis

The main dependent variable in Experiment 1 was the identification of a suspect from each lineup. In target-present (TP) lineups, this was the guilty suspect (i.e., the target); in target-absent (TA) lineups, this was the innocent suspect (i.e., one of the foils), as the target was not present in these lineups. Innocent suspects were designated *post hoc* using a ‘worst case’ scenario where the innocent suspect was the foil chosen most often from each of the three TA lineups for a target. This designation expands on Pryke et al. (2004), where the innocent suspect was the foil chosen most often from the face-front TA lineup. As the lineups in this study were not independent, it was thought that including identifications from all three lineups would provide a more stringent ‘worst case’ scenario for this assessment of the multiple lineup procedure. The calculations regarding percent of participants who made multiple identifications were based on what percent of participants made 0, 1, 2, or 3 suspect identifications, separately for TP and TA lineups.

For the first target that participants were exposed to, lineup decisions were the *uninformed* condition, as participants were unaware of the study’s nature when they saw this target's video. Data from one participant who indicated awareness of the study's purpose prior to seeing the first target were not included in the uninformed analyses, but were retained for the informed analyses. All six targets were represented equally often (12 times) in this position, with half of the lineup sets for each target being TP and half being TA. For the subsequent five targets each participant saw, lineup decisions were the *informed* condition, as participants were now aware of the study’s nature. As each participant provided data on multiple targets (i.e., the data lack independence) inferential analyses were not run on the informed data overall. Across participants, each target was seen 60 times in the informed condition (and then 12 times in the uninformed condition). Again, a target's lineups were TP half of the time and TA half of the time.

Materials

Targets/Exposure. The six targets (3 male, 3 female) used in the study were of European descent and were approximately in their mid-twenties. Participants were exposed to the targets through six videos (one for each target); each video was approximately 3 seconds long and showed the target from the shoulders up. In each video, the targets rotated 180° clockwise, from facing the front to facing the back, so that the participants were exposed to the three key views for each target. Aspects of the video, including speed of rotation and the exposure time for each key view, were not strictly controlled (see Discussion for limitations of the procedure).

Lineups. Six lineups were created for each target. Three lineups were TP and three were TA. The lineups for each target were facing forward (*face-front*), in profile (*face-profile*), and in the ¾ view (*face-¾*). Angle was not strictly controlled when foil photos were taken, so there was some variation among lineup members, particularly in instances where only a small number of foils matched a target's general similarity. Across all participants, exposure to each target was followed with TP and TA lineups an equal number of times, in both the uninformed and informed conditions.

Each lineup array contained six members and showed photos of these faces from the neck up. The photos were presented simultaneously in two rows of three. In TP lineups, one photo was of the target and five photos were of foils. In TA lineups, all six photos were of foils. Foils for the lineups were selected from a pool of available photos and possessed a general similarity to their respective targets (e.g., same sex, hair colour and style, ethnicity). Foils for each target were chosen based on what Turtle, Lindsay, and

Wells (2003) referred to as the “iterative technique.” The first foil chosen was one that closely resembled the target, the second foil chosen was one that closely resembled the first foil, and so on until six foils were chosen. The sixth foil was the one that replaced the target in the TA lineups. The iterative process was used so that the foils were not all identical to the target (in that most were matched for similarity to another foil), but also did not stand out as being very different (as they all shared general similarities). The three lineups for a target always showed the same foils and no foils overlapped between targets. In order to minimize matching based on lineup position, placement of each lineup member was different in each of the three lineups. All photos were edited so that the backgrounds were a single colour shared by all lineup members. See Appendix A for sample lineups.

For each lineup array, participants were presented with eight possible decisions: 1, 2, 3, 4, 5, 6, “I don’t know,” and “Not there.” The numbers corresponded to lineup positions – each photo had a number above (1, 2, or 3; top row) or below it (4, 5, or 6; bottom row). If the participant thought that a lineup member was the target, they were directed to select the number that corresponded to that photo. If the participant was unsure whether or not the target was in the lineup, they were directed to select “I don’t know.” If the participant thought that the target was absent from the lineup, they were directed to select “Not there.” Once a decision was made, participants could not go back or change their decision.

Confidence Scale. To measure confidence in lineup decisions, participants were asked to rate how confident they were on a scale from 0% (not at all confident) to 100%

(extremely confident) immediately after each lineup decision. If participants selected "I don't know," the confidence scale did not appear. The confidence rating was excluded for this lineup decision, as it would be confusing to ask participants to rate how confident they were that they did not know whether or not the target was in the lineup.

Intervening Tasks. Two intervening tasks were used to provide delays between exposure video and a target's first lineup, and between the lineups in a set. A delay was intended to reduce any tendency for participants to match across the lineups rather than deciding based on their memory of the targets. As well, including some delay is forensically relevant, as eyewitnesses experience delays between seeing a perpetrator and viewing a lineup. The first of the two intervening tasks was an anagram task, where participants unscrambled letters to form a word. This task was always presented between the exposure video and the first lineup. The second task was a visual search task, where participants searched for an "L" located in a field of "T"s. This task was presented after each of the three lineups. When participants solved an anagram or visual search, a new anagram or visual search would appear on the screen. Both tasks lasted for 30 seconds before the computer automatically moved on to the next portion of the study.

Procedure

Upon entering the laboratory, participants were led to a computer and instructed to follow the directions on the screen. Assignment to study condition was random for each participant. When participants began the study, they were prompted to enter demographic information (age, ethnicity, sex) and were then asked to press a button in order to continue and view a video (the first of the target exposures). As participants were unaware of the study's purpose prior to seeing the video, lineup decisions for the first exposure constitute the uninformed condition. Each target appeared 12 times in this condition, with the distribution across participants randomized. After viewing the first target exposure, participants read the letter of information and consent and were only then informed that the study involved eyewitness identification. If participants refused consent the study would have immediately stopped. All participants consented.

After consenting, participants were given an anagram task followed by either three target-present or three target-absent lineups. The lineups were presented in the following order: face-front, face-profile, and face-¾. The face-front lineups were presented first, as this view is most commonly used in lineups. The face-¾ lineups were presented last, as the information available in the ¾ view overlaps more with the information in the face-front and face-profile views (as noted earlier). This overlapping information could influence the face-front or face-profile lineups if they were presented after, so the ¾ lineup was presented last to negate any influence it had on the other lineups.

After the anagram task, participants saw the first target's face-front lineup. Once a lineup decision was made, a confidence rating was given (if applicable). Next, a visual search task appeared. Following this, the face-profile lineup was presented and participants again made a decision, rated their confidence, and performed a visual search task. At this point, the face-¾ lineup was shown and again participants made a decision, rated confidence, and performed a visual search task. Participants were given the visual search task after the third lineup in order to provide some delay between the one set of lineups and the next target's exposure video. After each lineup set, participants were asked whether they recognized anyone (other than the person seen in the exposure video), to account for unintended recognition of lineup members.

The remaining trials constituted the “informed” conditions. After finishing the procedure for the first target, participants saw the exposure for the second target. The process and order of events for the second target (as well as the third, fourth, fifth, and sixth) were the same as those for the first target. The order in which these last five targets were presented was randomized. After the last target, participants were asked if they knew that they would make lineup identifications before they began the study. This question was included to ensure that participants were truly naïve to the nature of the study prior to being exposed to the first (i.e., uninformed) target. Subsequent to this last question, participants were debriefed and the study ended.

Results

Analyses for both uninformed and informed conditions were run using all data, as well as excluding lineups where members were recognized. Reported analyses are those that included all data, though results without lineup recognition are also included when they differ substantially from the ‘all data’ analyses, to account for when recognition of lineup members has an impact on the results. Differing substantially was operationally defined as percent guilty differing by more than 10%, patterns in percent guilty over number of lineup identifications changed, or significance of analyses differed. Calculations of the percent of participants identifying the target in 0, 1, 2, or 3 TP lineups were compared against the percentages found in Pryke et al. (2004), Boyce et al. (2007) and Sauerland and Sporer (2008) to see whether the percentage of participants making multiple identifications of guilty suspects is higher in the current study (see General Discussion).

One participant indicated awareness that the study involved eyewitness identifications before taking part. This participant's responses for the uninformed condition were excluded from the analyses, but were retained for the informed condition. Information included in any Table below is of analyses using all of the data; any differences when analyses are done excluding lineups where a participant recognized a lineup member are noted in the text. Significance was determined using $\alpha = .05$.

Percent Guilty

To address the first hypothesis, percent guilty was calculated based on the same method used by Pryke et al. (2004) – number of participants making guilty suspect identifications divided by the number of participants making any suspect identification (guilty and innocent) – for each number of identifications (0, 1, 2, and 3). Inferential statistics for this analysis were chi-square analyses on frequencies; inferential statics were only run on the uninformed data. Chi-square analyses were not run on informed data (overall) due to a lack of independence and were not run on the by-target frequencies, as the cells did not meet the minimum required expected cell frequencies. For these latter two sets of data, the results will be discussed descriptively. Frequencies, percent guilty, and percent of participants identifying the guilty suspect appear in Table 1.

When uninformed, participants identified the target from two or more lineups 50% of the time, with almost 28% identifying the target from all three lineups. Percent guilty increased from zero (33%) to one (47%) to two (80%) identifications, and leveled off at three identifications (77%). Analyses that excluded lineups with recognized lineup members produced a lower percent guilty for one identification (33%) and a higher percent guilty for three identifications (91%). Results for all data, $\chi^2(3, N = 71) = 11.09$, $p = .01$, and excluding lineup recognition were both significant, $\chi^2(3, N = 58) = 13.88$, $p < .01$, indicating that the pattern in the results is likely due to a systematic relation between number of lineup selections and accuracy rather than random variation.

Table 1

Frequencies, Percent Guilty, and Percentage of Participants Making Identifications Collapsed Across Target

Target		Number of Identifications Made			
		0	1	2	3
Uninformed	TP	11 (30.00%)	7 (19.44%)	8 (22.22%)	10 (27.78%)
	TA	22 (62.86%)	8 (22.86%)	2 (5.71%)	3 (8.57%)
	Percent Guilty	33.33%	46.67%	80.00%	76.92%
Informed	TP	38 (21.11%)	27 (15.00%)	37 (20.56%)	78 (43.33%)
(Overall)	TA	104 (57.78%)	32 (17.78%)	30 (16.67%)	14 (7.78%)
	Percent Guilty	26.76%	45.76%	55.22%	84.78%

Note. TP = Target-Present lineup data; TA = Target-Absent lineup data. Percent of participants making that number of suspect identifications are bracketed.

When data from the informed condition were considered overall (i.e., collapsed across targets), the pattern in general was similar. Percent guilty was low with zero identifications (27%), rose to approximately chance for one (46%) and two (55%) identifications, and increased again with three (85%) identifications. The one substantial difference between the two conditions is with two identifications, where the percent guilty is much lower for informed participants (55%) than with uninformed participants (80%). A larger percentage of participants identified the target in at least two lineups (64%), as well as in all three lineups (43%).

In examining informed targets individually, it is clear that the expected pattern of accuracy, an increase in percent guilty as number of lineup selections increases, was not consistent for all targets (see Appendix B). Only four of the six targets produced equal or higher percent guilty with each increase in number of selections.

Lineup Angle

Proportion of suspect identifications was examined as a function of Target presence (present, absent) and Lineup angle (front, profile, ¾), for both the uninformed and informed data (see Table 2 for means and Appendix C for statistics). Target was between-subjects in the uninformed condition, as the uninformed data comprises lineup decisions for only one target; it was within-subjects for the informed data, as the informed data comprises lineup decisions for multiple TA and TP lineups.

With uninformed data, no significant differences were found in proportion of correct responses between the face-front, face-profile, and face-¾ lineups, either overall (main effect F, $df = 2, 138$, $p = .74$) or when target presence/absence was factored in (interaction F, $df = 2$, $p = .14$). The proportion of participants who identified the suspect was significantly higher in the TP lineups (.49) than in the TA lineups (.20; $F(1, 70) = 11.45$, $p < .01$).

For the informed condition, the proportion of correct responses was again significantly higher in TP (P = .62) lineups than in TA (.26) lineups, $F(1, 71) = 106.39$, p

< .01. However, the proportion of correct responses differed significantly across the front (.40), profile (.43), and ¾ (.49) lineups, $F(2, 142) = 6.76$, $p < .01$. Simple effects analysis revealed that the proportions of suspect identifications in face-¾ lineups were significantly higher than both face-front, $t(71) = -3.50$, $p < .01$, and face-profile lineups, $t(71) = -.260$, $p = .01$.

Table 2

Proportions of Suspect Identifications in Face-front, Face-Profile, and Face-¾ Lineups

Condition		*N*	Lineup Angle		
			Face-Front	Face-Profile	Face-¾
Uninformed	Present	36	.50 (.51)	.53 (.51)	.44 (.50)
	Absent	35	.14 (.36)	.17 (.38)	.28 (.47)
	Total	71	.32 (.47)	.35 (.48)	.37 (.49)
	Percent Guilty		78.12%	75.71%	61.11%
Informed	Present	72	.61 (.26)	.64 (.30)	.62 (.26)
	Absent	72	.20 (.28)	.21 (.27)	.36 (.28)
	Percent Guilty		75.31%	75.29%	63.26%

Note. Scores are the proportions of correct responses averaged across participants. Standard deviations are in brackets.

The interaction of target presence and lineup angle on accuracy was significant with the informed condition, $F(2, 142) = 8.42$, $p < .01$ (see Appendix C). The *t*-tests for the interaction showed no significant differences among the three angles in the TP lineups (p's = .33 to .71), which is unsurprising as the proportions only vary from .61 to .64. In the TA lineups, significantly larger proportions of suspect identifications were found with the face-¾ (.36) lineups than with either the face-front (.20), $t(71) = 4.59$, $p < .01$, or face-profile (.21) lineups, $t(71) = 4.10$, $p < .01$. In short, the ¾ lineup did significantly increase suspect identifications in the informed condition, but only for the innocent suspect.

When examining the percent guilty for number of identifications against percent guilty for each type of lineup, the uninformed percent guilty for three identifications (77%) is not much different than that for the face-front (78%) or face-profile (76%), although the percent guilty for two identifications is slightly higher (80%). In contrast, while the informed (overall) percent guilty for two identifications (55%) is much lower than that for each of the face identifications, the percent guilty for three identifications (85%) is higher, even when compared to the face-front and face-profile lineups (75% for both). For both the uninformed and informed data, the percent guilty for the ¾ lineup performed rather poorly: 61% and 63%, respectively.

Confidence-Accuracy Relationship

The confidence-accuracy analyses were computed by first summing the confidence provided for each suspect identification (by-target), in order to provide a weighted confidence score. In all, weighted confidence for each target could vary from 0 (indicating no suspect identifications, or 0% confidence in any suspect identifications) to 300 (indicating suspect identified all three times with 100% confidence). Weighted confidence was divided into ranges from 1-60%, 61-120%, 121-180%, 181-240%, and 241-300% (0% was excluded as no suspect was identified with a confidence of 0%).

Frequencies for the number of participants with weighted confidence scores falling in each range were then calculated for both TP and TA lineups (see Table 3). Inferential statistics (chi-square analyses of frequencies) were not run for these data, as informed data (overall) were not independent, and neither the uninformed nor by-target informed data produced expected cell values of 5 per cell for all cells.

Table 3

Frequencies for Weighted Confidence Scores Collapsed Across Target

Target		Weighted Confidence Score				
		1 – 60%	61 – 120%	121 – 180%	181 – 240%	241 – 300%
Uninformed	TP	11	5	10	1	0
	TA	8	5	0	0	0
	Percent Guilty	57.89%	50.00%	100.00%	100.00%	-
Informed	TP	13	27	27	35	40
(Overall)	TA	22	28	14	9	3
	Percent Guilty	37.14%	49.09%	65.85%	79.54%	93.02%

Note. TP = Target-present lineup data; TA = Target-absent lineup data. Dashes indicate no percent guilty can be calculated, as there are no observations.

For the uninformed condition, weighted confidence was relatively low for identifications in TA lineups (no score was above the 61 – 120% range). For TP lineups data fell into all but the highest category (241 – 300%). Percent guilty was near chance at lower confidence levels (1 – 60% and 61 – 120%) but 100% for both of the higher confidence ranges (121 – 180% and 181 – 240%). For the informed condition, some responses fell in each of the weighted confidence ranges for both TA and TP lineups. Percent guilty steadily increased, from 37% (1-60), to 49% (61-120), to 66% (121-180), to 80% (181-240), to 93% (241-300). The pattern of percent guilty increasing steadily as weighted confidence increased was not consistent when the targets were examined individually (see Appendix D). Only three targets showed an increase in percent guilty with each increase in weighted confidence.

Discussion

The first hypothesis addressed by this study was that percent guilty would increase as the number of lineups a suspect was identified in also increased. Although this hypothesis was supported by the data for some targets in the informed analyses, the results were by no means unequivocal. Looking at the uninformed data, the pattern was somewhat weak, particularly when compared to the informed (overall) data – although percent guilty increased from 0 (33%) to 1 (47%) to 2 (80%) identifications, it did not increase further with 3 identifications (77%). In contrast, the pattern for the informed (overall) data provided good (albeit not exceptional) support for the hypothesis, as percent guilty again increased relatively steadily, with the highest percent guilty (85%) occurring with 3 identifications. However, this latter pattern did vary when broken down by target. For four targets, the percent guilty was clearly highest when three suspect identifications were made. For a fifth target, percent guilty was also highest with three identifications (67%), although this percent guilty is neither very strong nor much higher than the percent guilty for one or two identifications (64% for each).

In all, while collapsing across targets provided moderate to good support for this hypothesis (i.e., uninformed condition, and informed overall), when examined by-target, support was much more variable. In particular, while the procedure seems to work well overall, it clearly works less well with particular targets – which is the unit of analysis used by police forces. Whether it is some feature of the target or the members in that target's lineup, there are factors that can mitigate how well the procedure works as a diagnostic tool for specific targets.

In terms of the second hypothesis, that there would be greater accuracy when identifying suspects from a face-¾ lineup, the results indicate that this is not the case. In examining the informed data, the face-¾ lineup did have a larger proportion of participants making suspect identifications than the other lineups. However, this difference seems to be driven by an increase in suspect identifications in TA lineups – that is, identifications of innocent suspects. When a target was present in a lineup, the proportion of participants making a suspect identification did not differ across the three lineups. When a target was absent however, significantly more participants identified the innocent suspect in the ¾ view. Although the interaction with the uninformed data was not significant (possibly because the between-subjects element made it a less powerful test than the completely within-subjects informed analysis) a similar pattern is evident in that data. The poorer diagnosticity also was revealed in the relatively lower percent guilty for the ¾ lineup for both uninformed (61%) and informed data (63%). In short, no face-¾ advantage was found – instead, a potential disadvantage was found. There are confounds with these results (e.g., lineup order) which will be addressed in the limitations.

The third hypothesis predicted that the diagnosticity of weighted confidence would increase as weighted confidence scores increased. Once again, support for the hypothesis was mixed. With the uninformed data, the percent guilty hovered around chance (57%, 50%) when weighted confidence ranged from 1-120%, and then increased sharply to 100% when confidence ranged from 121-240% (no participant had weighted confidence from 241-300%). When informed (overall) data was examined, the percent guilty did gradually increase from 37% (1-60% confidence) to 93% (241-300% confidence), thus supporting the hypothesis. However, the informed data again varied by-target. For several targets, the percent guilty gradually increased and either eventually hit 100% or came close to it (e.g., 92%). For other targets, the percent guilty was much more variable (see Appendix D). In all, weighted confidence did provide some measure of diagnosticity, particularly when data was examined across targets, though a gradual increase was not seen in the uninformed data.

The fourth hypothesis – that identification accuracy would differ when participants are aware of the study's purpose as compared to when they are naïve to the study's purpose prior to viewing a target – was supported in part. In terms of the utility of the multiple lineup procedure, the two conditions had similar percent guilty for zero, one, and three identifications. However, the pattern of percent guilty was cleaner with the informed data, as the percent guilty for two identifications (80%) was substantially higher than the percent guilty seen in the uninformed data (55%). In the weighted confidence analyses, the pattern found with uninformed data differed substantially from that of informed data, in that it hovered around chance (i.e. approximately 50%) for weighted confidences from 1-120%, but then increased sharply to 100% for weighted confidences from 121-240%. In contrast, the percent guilty for informed data (overall) increased relatively gradually as confidence increased.

In regards to the by-lineup analyses, although the patterns of significance differed between the uninformed and informed data (particularly in terms of the interaction), the difference between the conditions may be due more to different statistical power than to a fundamental difference. In particular, when examining the proportions of suspect identifications (Table 2), although not significant, the proportion of participants making a suspect identification from TA face-¾ lineups was generally twice that of face-front and face-profile lineups – which matches the pattern found with the informed data. As well, the percent guilty in the ¾ lineup was relatively similar between the uninformed (61%) and informed (overall) data (63%). Thus, the ¾ lineup seemed to perform relatively poorly for both the uninformed and informed conditions.

Surprisingly, participants were not overwhelmingly more accurate when informed about the study's purpose. One expectation was that participants would perform better when aware they would be making later identifications, as they would pay attention to the exposure video in a way that would facilitate recognition. However, percent guilty for two identifications was much worse when participants were informed, and percent guilty for weighted confidence never reached 100% as it did with the uninformed data. Lastly, although percent guilty for three identifications was higher when participants were informed, it was not a substantial difference (using the earlier criterion) over the same percent guilty when participants were uninformed.

There are potential problems when comparing the uninformed and informed results. In particular, the two conditions do not differ solely based on whether the participants were naïve or knowledgeable about the study's purpose. Because participants read the Letter of Information (LOI), gave consent, and had the anagrams task between seeing the first video and the first lineup, the amount of delay was longer than it was in the informed condition, where participants had only the anagrams task between exposure and lineup. As well, the amount of delay was likely more variable, as it depended on how long individual participants took to read through the LOI and provide consent. The longer, more variable delay – and essentially, focus on an additional task – may have had an impact on participants' subsequent identifications and confidence ratings. In order to provide a more stringent comparison of uninformed and informed conditions, the amount of delay between the two conditions needs to be more strictly controlled. Another issue is the possibility of interference, where performance may be negatively affected by having seen previous stimuli and lineups. Yet, even though choosing rates for each of the lineups were consistently higher when participants were informed, the overall accuracy was very similar to that seen when they were uninformed. As well, such interference could explain why participants had lower percent guilty for two suspect identifications when informed, although not why percent guilty was higher for three suspect identifications (when informed).

Beyond this difference in delay, there are three additional, notable limitations in Experiment 1. First, the lineups were always presented in the same order: face-front, face-profile, face-¾. Although there were reasons for this order, it could be the case that results would differ if the lineup order was counterbalanced – particular lineup orders may be particularly poor (or beneficial) when using the method with face-only lineups. Additionally, the difference in performance across the individual lineups (e.g., the particularly poor ¾ lineup) may also be due to this order effect, which in turn may impact how well the procedure performs relative to the individual lineups. Thus, lineup order needs to be taken into account.

Second, participants saw the targets in all three key views during the exposure, which represents a best-case scenario. However, actual witnesses may not see a perpetrator from all pertinent views – it is plausible that they may see a perpetrator in only one view, or even in a view that is not included in the lineups. In such cases, presenting multiple face lineups may be less helpful and may actually be harmful. As such, limited exposure (in terms of angles seen) should also be taken into account to assess how seeing a target in only one pose impacts the utility of the procedure.

Third, various aspects of the exposure stimuli were not tightly controlled. For example, targets did not necessarily rotate at the same speed, faces were not always static while rotating (i.e., some targets moved), and the exposure time for each key angle was not exactly equated. This variation is due in part to the fact that the exposure stimuli were originally created to be used as part of kinetic lineups and were adapted for use in Experiment 1. However, given that the videos are only 2-3 seconds long, the variation in speed between targets was small. Although problematic, variation with these aspects also is not necessarily damaging as real-world settings would likely include even greater variation, and because it is the cumulative usefulness of all three lineups which is most important with the multiple lineup procedure. Regardless, potential problems related to the faces remaining static and exposure time per view should be addressed.

Chapter 3: Experiment 2

Experiment 2 addresses the three notable limitations outlined in Experiment 1 – that is, lineup order, exposure, and stimuli variation. In terms of lineup order, Pryke et al. (2004), Boyce et al. (2007), and Sauerland and Sporer (2008) all used fixed lineup orders in their studies. However, as noted above, lineup order could potentially influence accuracy due to order effects (e.g., a face-¾ lineup influencing later face-front and face-profile lineups). To see whether the multiple lineup method has utility regardless of what order the lineups are shown in, the three lineups were shown in different orders across participants. Each possible permutation in order was employed and counterbalanced. Although each lineup order was not directly assessed and analyzed, collapsing across all possible orders allows for some indication of whether the multiple lineup method is useful above and beyond any effects of lineup order.

Experiment 2 also addresses issues related to how well the multiple lineup method works if participants have limited exposure to the target in terms of the angle seen. While targets in Experiment 1 were shown in all three pertinent views, targets in Experiment 2 were shown facing only one angle – including an angle that was not represented at all in the lineups. Such limited exposure would reflect situations where actual eyewitnesses are only able to view a perpetrator from one angle, or see a perpetrator in a view that is not shown in later identification procedures. There are other reasons why exposure to only one angle should be assessed, particularly because face recognition research shows that recognition of strangers depends on whether or not the view seen during exposure is the same as that seen during testing. The best recognition occurs when the views in exposure and test correspond; if the views do not correspond, this change has a negative impact on recognition (Bruce, 1982; Bruce et al., 1999; Busey & Zaki, 2004; Carbon & Leder, 2006; Hill et al., 1997; Jeffery, Rhodes, & Busey, 2006; Krouse, 1981; Liu & Chaudhuri, 2002; Logie et al., 1987). As such, the targets in Experiment 2 were shown facing one of four angles during exposure (face-front, face-¾, face-profile, or face-¼), though lineups included only three of these angles (face-front, face-¾, face-profile).

Lastly, Experiment 2 addresses potential limitations regarding stimulus variation, particularly exposure time for each angle and targets remaining still during the video. Targets remained stationary while simultaneously being filmed from several angles by stationary videocameras. Natural speaking movement did exist but was minimal and targets continued to face only one direction. As well, because targets only faced one direction in the video, the amount of exposure to that viewing angle could be controlled more tightly. Manipulation of these stimulus aspects still was not complete, but was stricter than in Experiment 1.

In regards to the face-perception literature, it was hypothesized that percent guilty may increase as the number of lineup identifications increase, though not to the same extent as in Experiment 1. Because targets are seen in only one angle, ability to recognize them from multiple lineups may be decreased. Related to this first hypothesis (and to the face recognition literature) the second hypothesis was that accuracy would be greatest for the lineups that corresponded to the view seen in the exposure video. The third hypothesis related to the confidence-accuracy relationship. It was predicted that confidence would generally increase as accuracy increased. Although this same hypothesis was not entirely supported by all of the data in Experiment 1, most analyses did show support for it, in that higher levels of weighted confidence were often more indicative of the guilty suspect (i.e., target) being identified. Lastly, a potential issue with seeing targets from only one view is that if recognition is poorer when exposure and test views do not match, it becomes important in applied settings to know what angle the perpetrator was facing so that the most accurate recognition test can be applied (i.e., a lineup using a corresponding view). Given that some features of exposure, such as distance, can be recalled incorrectly in eyewitness studies (Lindsay, Semmler, Weber, Brewer, & Lindsay, 2008), participants may also incorrectly recall other features of exposure, such as what angle the target was facing. As such, the fourth question examined to what extent participants were inaccurate when recalling the angle another person was facing.

Method

Participants

Participants were 96 Queen's undergraduates (79 female, 17 male). The average age of participants was 18.51 (*SD* = 1.05), with ages ranging from 18 to 25. As in Experiment 1, the majority of participants were of European ethnicity (70.83%), with Asian (20.83%), Middle Eastern (3.12%), African (3.12%), and multiple (1.04%) ethnicities also represented. Ethnicity data for one participant was not recorded. All participants were recruited via the Psychology 100 subject pool and were compensated with either course credit or $5.

Data Analysis

As in Experiment 1, the main dependent variable of interest was suspect (guilty or innocent) identifications. Innocent suspects were again chosen post hoc using the foil identified most often from the set of three target-absent (TA) lineups, for the same reasons as stated in Experiment 1. For one target, three foils were chosen an equal number of times across the lineups. For this target, the foil chosen most often from all three lineups was selected as the innocent suspect, as this was in keeping with a worst-case scenario. Reported percentages of participants who make identifications was performed using the same method as in Experiment 1, with the results compared to those of Pryke et al. (2004), Boyce et al. (2007), Sauerland and Sporer (2008), and Experiment 1 (see General Discussion for these comparisons).

One between-subjects condition in Experiment 2 was the presence/absence of a given target in the respective lineup sets. Exposure angle of any one target and order of lineups varied between participants for counterbalancing purposes.

Materials

Targets/Exposure. For the main portion of the study, eight targets (four male, four female) were used; all were of European descent and approximately in their mid-twenties. Exposure to the targets was through eight videos (one for each), each of which showed targets from the shoulders up. The videos were approximately three seconds long and showed a single target speaking a scripted phrase in one of four stationary positions: facing forward (*face-front*), in the ¾ view (*face-¾*), in profile (*face-profile*), and in the ¼ view (i.e., 135 degrees from face forward; *face-¼*). As the videos were originally made with the targets facing the opposite direction from the lineups (i.e., the right side of the face was shown, not the left), the videos were flipped so that the targets in the exposure videos were facing the same direction as the lineup members. Because the videos had to be flipped, targets were restricted to those with the most symmetrical faces, hair styles, and those with the fewest distinguishing marks (e.g., moles, scars) present in the exposure video. As no one is completely symmetrical, this could be problematic as participants were not exposed to the exact view seen during test.

Each participant was exposed to all eight targets and all four target views were seen twice by participants – once with a male target and once with a female target (with one of these targets having target-present lineups and the other having target-absent lineups). Which target was seen in which view was counterbalanced across participants. See Figure 1 for examples of exposure angles paired with matching lineup views.

For the portion of the study investigating how well participants recall the angle a target was facing during exposure, four additional targets were used. To prevent confusion with the targets in the earlier portion of the study, these targets looked very different: bald, European male; European male with facial hair and glasses; South Asian female with headband; elderly, European female. As with the other exposure videos, the videos used in this portion of the study were around three seconds long and showed a target from the shoulders up speaking in one of the four stationary positions. In this portion of the study, each participant saw each of the four views once, with which target was seen in which view counterbalanced across participants.

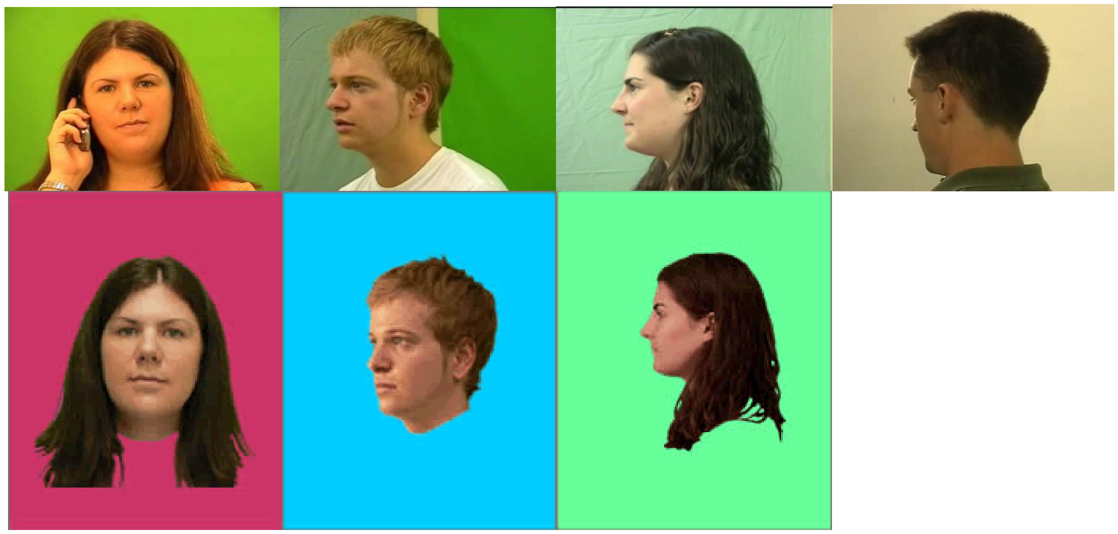

Figure 1. *Examples of exposure angles with matching lineup photo beneath (the ¼ view was not represented in a lineup).*

Lineups. As in Experiment 1, each target had six lineups, with three lineups being target-present (TP) and three being TA. For each target, the three lineups seen by participants were either all TP or all TA. For the eight targets viewed, four of the targets (two male, two female) had TP lineups and the other four (two male, two female) had TA lineups. Across participants TP and TA lineups for each target were presented an equal number of times. The three lineups viewed by participants for each target consisted of the face-front, face-profile, and face-¾ lineups, with angle in photo not strictly controlled (as in Experiment 1) which allowed for some variation. The order of presentation for the three lineups was counterbalanced between participants and was always the same for any one participant. Although face-¼ views were included in the exposure videos, they were not included in the lineups in order to assess how accurate participants would be when the view they saw does not correspond to any of the multiple lineups that were presented.

Each lineup array shown to participants contained six photos showing lineup members from the neck up. Lineups were simultaneous and the photos were presented in two rows of three. For TP lineups, the lineup members consisted of the target and five foils; for TA lineups, all six lineup members were fillers. As in Experiment 1, foils were taken from a pool of available photos and were chosen based on their general similarity to their respective targets (e.g., same sex, ethnicity, hair colour and style). The same iterative procedure described in Experiment 1 was used for Experiment 2 and again, the same foils were always used for a particular target and never overlapped between targets. As well, as in Experiment 1, no lineup member was seen in the same position across the three lineups in a set (in order to minimize matching) and backgrounds were edited to be uniform within a lineup. In terms of lineup decisions, participants were provided with the same options described in Experiment 1, and again could not go back or change their response after a decision was made. See Appendix A for sample lineups.

Angle Recall. In the angle recall portion of the study, participants were instructed that they would describe each target (e.g., age, ethnicity, sex, hair colour) immediately after seeing that target's video. These questions, along with a subsequent distracter task (anagrams), were to provide some delay between seeing the videos and being asked to recall what angle the target was facing. All four videos were seen and the respective responses to the descriptive questions given prior to testing for memory for pose (participants were unaware they would be tested on angle recall until this point). When asked to recall what angle each target was facing in the videos, participants were provided with a brief description of the target they should respond to (e.g., "male, European, bearded, and glasses”). Participants were also presented with an array consisting of five computer-generated images of faces, where the images represented a person facing-front, facing-¾, facing-profile, facing-¼, and facing the back (i.e., 180°; see Figure 1). Instructions directed the participants to make a decision (1, 2, 3, 4, or 5) regarding which image in the array most closely matched the direction the target was facing in the video. Each numbered decision corresponded to one of the positions represented in the array.

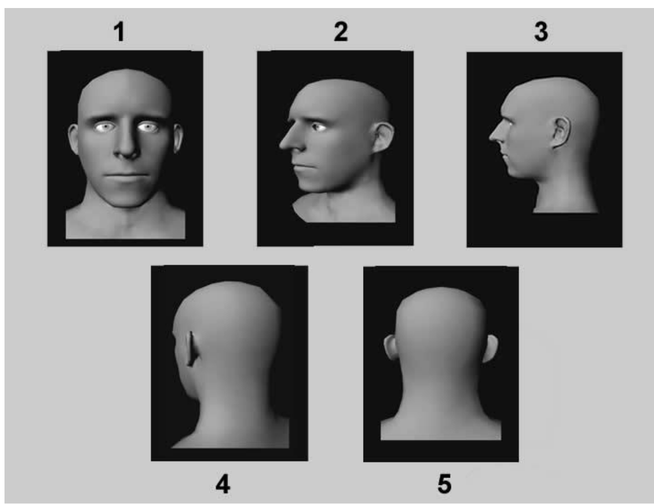

Figure 2. *Array of computer-generated faces used as the stimuli for the angle recall analyses.*

Intervening Tasks. The same two intervening tasks described in Experiment 1 were used: anagrams and visual searches. The explanation for and placements of the tasks are also the same as in the prior experiment, except that no distracter task was given after the third lineup in a set due to time constraints. As in Experiment 1, when an anagram or visual search was solved, a new anagram or visual search would appear. In the lineup portion of the study, the anagram task lasted for 60 seconds, while the visual search tasks lasted for 30 seconds. In the recall of angle viewed portion, the anagram task lasted for 30 seconds. After the 30 or 60 seconds elapsed, participants were automatically moved on to the next portion of the study.

Procedure

Participants were seated at a computer and instructed to follow the directions that appeared on the monitor. All participants were randomly assigned to a condition. When the program started, participants read the letter of information and were asked if they provided their consent. If participants had refused to consent (none did), the study would have immediately ended. After granting consent, the study continued and the participants were asked to enter their demographic information (sex, age, and ethnicity).

After entering their demographics, participants were taken to the next screen where they viewed the first target exposure. For each of the eight targets in the identification portion of the study, the tasks occurred in the following order: exposure video, anagram task, first lineup, visual search, second lineup, visual search, third lineup, recognition question (asking whether any lineup members were recognized, apart from the person seen in the video). The order that the eight targets were presented was randomized, to minimize order effect, albeit with some caveats. For example, targets facing the same angle in exposure views were not presented sequentially, to prevent participants from making lineup decisions based on the likelihood of target presence or absence based on the previous target. To clarify, if a participant saw a face-¾ video with TP lineups, and then saw another face-¾ video, they may assume that the target is not in the lineups, as they were in the prior set of lineups. Exposure angle of each target was counterbalanced across participants, in that each target was seen in the four exposure angles an equal number of times. As well, the order in which the three lineups in a set were presented was counterbalanced across participants (but was consistent for any one participant).

For the portion of the study examining recall of exposure angle, participants watched the videos of the four targets described earlier. The order in which the targets appeared, as well as which angle they were facing, were counterbalanced across participants so that position and angle occurred an equal number of times per target. Immediately after each video, participants answered questions regarding the description of the person they just viewed. Participants were presented with set questions – rather than providing open-ended responses – to ensure that participants would give roughly the same amount of detail and time when providing the descriptions. After seeing the videos and answering the questions for all targets, participants were given an anagram task for 30 seconds. After the anagram task, participants were asked to recall the angle each target was facing in the video. A brief, relevant description of each target was presented along with the 5-image lineup (Figure 1) to cue participants to the target they should recall. Once participants stated which angle each of the four targets was facing, they were debriefed and the study ended.

Results

As in Experiment 1, all analyses were run with all data, as well as excluding lineups where members were recognized. Again, the latter analyses will be noted only in the text when they differ substantially from analyses including all data, using the same thresholds. Information included in any Table below is of analyses using all of the data. The targets were named using the same convention as in Experiment 1 (e.g., F = female, M = male); the numbers do not overlap as an entirely different set of targets was used for Experiment 2. A check for a lineup order effect showed the presentation order of lineups (e.g., first, second, third) had no effect on accuracy either as a main effect ($p = .24$) or as an interaction with target presence ($p = .80$). The threshold used to determine significance for inferential analyses was $\alpha = .05$.

Percent Guilty

The first hypothesis was examined by computing the frequency of participants making 0, 1, 2, or 3 suspect identifications, for TP and TA lineups, both overall (see Table 4) and by-target (see Appendix E). Chi-square analyses of frequencies were not run, because the overall data were not independent and because not all by-target analyses met the criterion minimum expected cell frequency of 5.

When examined overall, there was a steady increase in percent guilty as the number of suspect identifications increased, with the lowest percent guilty seen with zero identifications (28%) and the highest seen with three identifications (83%). Overall, 58% of participants made multiple identifications of the guilty suspect, with 34% identifying the suspect from all three lineups.

As in Experiment 1 however, the patterns of percent guilty did vary when broken down by-target. The expected pattern of a continuously increasing percent guilty was found with only six of the eight targets (see Appendix E). Examining the percentages of targets who made multiple identifications by-target, over half of participants made two or more target identifications with six of the targets, with percentages ranging from 54% to 83%. Close to half of the participants were able to identify the guilty suspect from all three lineups for two targets.

Table 4

Frequencies, Percent Guilty, and Percentage of Participants Making Identifications Collapsed Across Target

Target		Number of Identifications Made			
		0	1	2	3
Overall	TP	95 (24.74%)	68 (17.71%)	92 (23.96%)	129 (33.59%)
	TA	241 (62.76%)	77 (20.05%)	40 (10.42%)	26 (6.77%)
	Percent Guilty	28.27%	46.90%	69.70%	83.22%

Note. TP = Target-Present lineup data; TA = Target-absent lineup data. Percentage of participants identifying the target given each number of identifications is provided in brackets.

The number of suspect selections was examined as a function of Target presence (present, absent) and Exposure angle (front, ¾, profile, ¼; see Table 5, Appendix F). The greater the number of identifications engendered by an exposure view, the closer to ‘3’ the value for that cell would be – with higher values good for TP lineups and bad for TA lineups. As the sphericity assumption was not met for the exposure factor, the Greenhouse-Geisser adjustment was used for this analysis. The main effect for target presence was significant, $F(1, 95) = 226.90$, $p < .01$, (significantly more suspect identifications were made with TP lineups than with TA lineups) as was the main effect for exposure angle, $F(2.69, 225.37) = 7.57$, $p < .01$, and the interaction between target and exposure angle, $F(3, 285) = 5.38$, $p < .01$.

For exposure angle, simple effects analyses were examined between face-front (1.03), face-¾ (1.38), face-profile (1.19), and face-¼ exposures (0.96). The average number of suspect identifications with the face-¾ exposure was significantly greater than the average number of identifications of the face-front, $t(95) = -4.26$, $p < .01$, face-profile, $t(95) = 2.32$, $p = .02$, and face-¼ exposures, $t(95) = 3.80$, $p < .01$. The difference between the profile and ¼ views was also significant, $t(95) = 2.26$, $p = .03$. In examining the interaction, in the TP lineups, the face-¾ exposure (2.08) again garnered significantly more suspect identifications than the face-front (1.54), $t(95) = -3.61$, $p < .01$, face-profile (1.79), $t(95) = 1.98$, $p = .05$, and the face-¼ exposures (1.24), $t(95) = 5.13$, $p < .01$. The face-profile exposure (1.29) engendered significantly more TP suspect identifications than the face-¼ exposure $t(95) = 3.20$, $p < .01$. In contrast, none of the TA paired-samples *t*-tests were significant (*p*'s = .14 to 1.00). Thus, in contrast to Experiment 1, the face-¾ view does appear to be beneficial, at least when presented as an exposure angle. Percent guilty calculations are between 75% and 76% for the face-front, face-¾, and face-profile exposures, but much lower for the face-¼ exposure (65%).

The number of correct lineup decisions was examined as a function of Target presence (present, absent) and Exposure angle (front, ¾, profile, ¼; see Table 5 and Appendix F). This second analysis is subtly different than the first as a correct decision involves either identifying the target in TP lineups, or rejecting the lineup (e.g., stating the target is not present) in TA lineups. The greater the number of correct decisions a condition engenders, the closer to ‘3’ the value for that cell will be. Significant results were found for the main effect of target presence, $F(1, 95) = 33.66$, $p < .01$, (significantly more correct decisions were made in TP lineups than in TA lineups), the main effect of exposure view, $F(3, 285) = 13.98$, $p < .01$, and the interaction, $F(3, 285) = 2.66$, $p = .05$.

Table 5

Average Number of Suspect Identifications by Exposure Angle and Target Presence or Absence

Dependent Variable		*N*	Exposure Angle			
			Face-Front	Face-¾	Face-Profile	Face-¼
Suspect Identifications	TP	96	1.54 (1.23)	2.08 (1.03)	1.79 (1.10)	1.24 (1.19)
	TA	96	0.51 (0.83)	0.68 (0.92)	0.58 (0.92)	0.68 (1.01)
	Percent Guilty		75.12%	75.36%	75.53%	64.58%
Correct Decisions	TP	96	1.54 (1.23)	2.08 (1.03)	1.79 (1.10)	1.24 (1.19)
	TA	96	1.26 (1.06)	1.25 (0.95)	1.13 (1.03)	0.75 (0.98)

Note. TP = Target-present lineup data; TA = Target-absent lineup data. Values above are the number of correct responses averaged across participants. Standard deviations are bracketed.

For the main effect of exposure view, simple effects analyses were run on the face-front (1.40), face-¾ (1.67), face-profile (1.46), and face-¼ exposures (1.00). Once again, the face-¾ view had significantly more correct decisions than the face-front, $t(95) = -2.75$, $p < .01$, face-profile, $t(95) = 2.04$, $p = .04$, and face- ¼ exposures $t(95) = 6.66$, $p < .01$, while the face-¼ exposure view had significantly fewer correct decisions than the face-front, $t(95) = 3.66$, $p < .01$, and face-profile exposures $t(95) = 3.84$, $p < .01$. In examining the interaction, the significant differences in the TP lineups are the same as in the prior analysis, as suspect and correct identifications are synonymous when the target is present in the lineup. Again, the face-¾ exposure had significantly more correct identifications than the face-front, face-profile, and the face-¼ exposures, and the face-profile exposure had significantly more correct identifications than the face-¼ exposure. In the TA lineups, the face-¼ exposure (0.75) garnered significantly fewer correct responses than did the face-front (1.26), $t(95) = 3.81$, $p < .01$, face-¾, $t(95) = 3.98$, $p < .01$, and face-profile (1.12) exposures, $t(95) = 2.64$, $p = .01$.

Again, the face-¾ was beneficial in the current Experiment (as an exposure angle), as correct responses were significantly higher than all other exposure angles when the suspect was present in the lineup. The face-¼ exposure view was detrimental, as it often led to significantly fewer identifications of the suspect (when present in the lineup), and significantly fewer correct decisions (i.e., lineup rejections) when the suspect was absent from the lineup. Percent guilty was not calculated for this analysis, as the dependent variable did not concern suspect identifications.

Lineup Angle

The proportion of participants making a suspect identification was examined as a function of Target presence (present, absent), Lineup angle (front, ¾, profile), and Exposure angle (front, ¾, profile, ¼; see Table 6 and Appendix G). This analysis examined the second hypothesis, whether accuracy is greatest when lineups correspond to the angle of exposure. Sphericity was significant for the exposure angle main effect ($p < .01$) and the target by lineup angle interaction ($p = .04$), so the Greenhouse-Geisser adjustment was used for these. Significant main effects were found for all of: target presence, $F(1, 95) = 226.90$, $p < .01$ (a significantly greater proportion of participants made suspect identifications in TP lineups than in TA lineups); lineup angle, $F(2, 190) =$

3.40, $p = .04$; and exposure angle, $F(2.69, 255.37) = 7.46$, $p < .01$. The target presence by exposure angle interaction was significant, $F(3, 285) = 5.38$, $p < .01$, as was the target presence by lineup angle by exposure angle interaction, $F(6, 570) = 4.07$, $p < .01$. When participants who recognized lineup members were excluded listwise, the only change was that lineup angle was no longer significant ($p = .20$).

As with the previous analyses, simple effects analyses were run to see where the significant effects lay. For the main effect of lineup, the only significant difference was between the face-¾ (.36) and face-profile lineups (.41), wherein a significantly greater proportion of participants made suspect identifications in the face-profile lineups, $t(95) = -2.80$, $p < .01$. For the main effect of exposure angle, a significantly larger proportion of participants made suspect identifications when they saw the face-¾ angle (.46) in the exposure video than those who saw the face-front (.34), $t(95) = -4.26$, $p < .01$, face-profile (.40), $t(95) = 2.32$, $p = .02$, and face-¼ (.32) exposures $t(95) = 3.80$, $p < .01$. A significantly greater proportion of participants made suspect identifications with the face-profile exposure as compared to the face-¼ exposure, $t(95) = 2.26$, $p = .03$.

Examining the target by exposure interaction, for TP lineups the proportion of participants making suspect identifications was significantly greater with the face-¾ (.69) exposure than with the face-front (.51), $t(95) = -3.61$, $p < .01$, face-profile (.60), $t(95) = 1.98$, $p = .05$, and face-¼ (.41) exposures, $t(95) = 5.13$, $p < .01$. Also, the face-profile exposure (.60) garnered a significantly larger proportion of suspect identifications than the face-¼ exposure, $t(95) = 3.20$, $p < .01$. With TA lineups, none of the paired-samples *t*-tests were significant. In short, seeing the target in a face-¾ view led more participants to make more identifications, while seeing the target in a face-¼ view led them to make fewer identifications. In contrast to Experiment 1, this enhancement was beneficial as it was found with guilty suspect identifications, not with innocent suspect identifications.

Table 6

Proportions of Participants Making Suspect Identifications in Each Condition

Exposure View			Lineup Type		
		N	Face-Front	Face-¾	Face-Profile
Face-Front	TP	96	**.63 (.49)**	.44 (.50)	.48 (.50)
	TA	96	**.14 (.34)**	.18 (.38)	.20 (.40)
	Percent Guilty		81.82%	70.97%	70.59%
Face-¾	TP	96	.70 (.46)	**.70 (.46)**	.69 (.47)
	TA	96	.20 (.40)	**.19 (.39)**	.29 (.46)
	Percent Guilty		77.78%	78.65%	70.41%
Face-Profile	TP	96	.51 (.50)	.62 (.49)	**.66 (.48)**
	TA	96	.21 (.41)	.18 (.38)	**.20 (.40)**
	Percent Guilty		70.83%	77.50%	76.74%
Face-¼	TP	96	.38 (.49)	.39 (.49)	.48 (.50)
	TA	96	.25 (.44)	.17 (.38)	.26 (.44)
	Percent Guilty		60.32%	69.64%	64.86%

Note. TP = Target-present lineup data; TA = Target-absent lineup data.

In terms of the 3-way interaction, this interaction indicates that the target by lineup angle 2-way interaction differs depending on what exposure angle was seen. As a way to limit the number of analyses run, the proportion of participants making suspect identifications was examined as a function of Target presence (present, absent) and Lineup angle (front, ¾, profile) for each of the exposure angles (front, ¾, profile, ¼; see Appendix H). Of the exposure angles, only the face-front showed a significant interaction between these factors, $F(2, 190) = 7.40$, $p < .01$. Further simple effects analyses showed that in the TP lineups for the face-front exposure angle, the face-front lineup (.62) had a significantly higher proportion of suspect identifications than either the face-¾ (.44), $t(95) = 3.76$, $p < .01$, or face-profile (.48) lineups, $t(95) = 3.12$, $p < .01$. None of the simple effects analyses in the TA lineups showed significant differences (p's = .16 to .62). In general, the only situation where a lineup provided significantly larger proportions of suspect identifications (particularly of the target) when the same angle was seen in exposure was with the face-front exposure.

Examining the percent guilty for each lineup broken down by lineup exposure, the highest is 82%, which is still slightly below the overall percent guilty for three identifications (83%). Most percent guilty were in the 70’s and were considerably lower than that for three identifications. The lowest percent guilty were for the ¼ exposure, which reflects the generally poorer performance seen with this exposure angle. Notably, even though performance was generally poorer with the ¼ exposure angle, the percent guilty were not always substantially lower than the percent guilty found with the other exposures. In contrast to Experiment 1, and the earlier analyses in Experiment 2, the face-

¾ lineup was beneficial, as it provided the highest percent guilty for all exposure angles, except for face-front. However, the face-¾ exposure did not provide substantially higher percent guilty (thus differing from the earlier analyses), and the correspondence between lineup and exposure did not provide a substantially higher percent guilty, which indicates that the ¾ advantage may be limited.

Confidence

Originally, the third hypothesis – examining the confidence-accuracy relationship – was to be examined by analyzing weighted confidence as a function of Target presence (present, absent) and Exposure angle (front, ¾, profile, ¼). However, this analysis had only n = 2. Because of low n, the confidence-accuracy relationship was instead examined by comparing confidence scores of decisions from lineups matching the exposure angle against confidence scores of decisions from lineups differing from the exposure angle. As the face-¼ view was only included in exposure and not in a lineup, it was excluded from this analysis. For the dependent variable, the confidence score from the lineup that matched the exposure was compared against the average confidence score from the two lineups that were different from the exposure angle (see Table 7; see Appendix I for confidence data by exposure and lineup angles).

Weighted confidence for suspect identifications was examined as a function of Target presence (present, absent) and Angles (match, different; see Appendix J). The main effect for target presence was significant, $F(1, 31) = 11.77$, $p < .01$, meaning that confidence ratings were significantly higher when the target was present (75.42%) in the lineup than when the target was absent (65.29%). The main effect for angles was also significant, $F(1, 31) = 4.29$, $p = .05$, indicating that confidence ratings were significantly higher when the lineups matched the exposure (72.44%) than when they differed from the exposure (68.27%).

Table 7

Confidence Ratings for Lineups that Match Exposure View Versus are Different

Type of Lineup Decision			Exposure and Lineup Angles	
		N	Match	Different
Suspect Identifications	TP	32	78.58 (17.40)	72.25 (17.40)
	TA	32	66.30 (19.43)	64.28 (14.59)
Correct Decision	TP	63	73.94 (18.60)	68.93 (19.88)
	TA	63	68.24 (20.28)	64.96 (18.94)

Note. TP = Target-present lineup data; TA = Target-absent lineup data. Standard deviations are bracketed. For suspect identifications, $n = 32$ participants made identifications from all four cells (match/different, TP/TA). For correct decisions, $n = 63$ participants made correct decisions for all four cells (match/different, TP/TA).

Weighted confidence for correct identifications was also examined as a function of Target presence (present, absent) and Angles (match, different; see Appendix J). Examining confidence in correct decisions, the only significant finding was for angles, $F(1,62) = 9.58$, $p < .01$, wherein confidence ratings for lineup angles that matched the exposure angles (71.09%) were significantly higher than when the two angles were different (67.44%). In general, a correspondence between exposure and lineup angles led to significantly greater confidence in both correct decisions and suspect identifications. As well, identifications of guilty suspects were made with significantly more confidence than identifications of innocent suspects.

As in Experiment 1, the diagnosticity of weighted confidence was examined. The confidence for each suspect identification (by-target) was summed to provide the weighted confidence score. Weighted confidence again could vary from 0 (indicating no suspect identifications, or 0% confidence in any suspect identifications) to 300 (indicating suspect identified all three times with 100% confidence). Weighted confidence was divided into the following ranges: 1-60%; 61-120%; 121-180%; 181-240%; and 241-300% (0% was excluded for the same reason stated in Experiment 1). Frequencies for the number of participants with weighted confidence scores falling in each range were then calculated for both TP and TA lineups (see Table 8 for overall; Appendix K for by-target). Inferential statistics (chi-square analyses of frequencies) were not performed, as the overall data were not independent and the by-target data produced expected cell values of 5 per cell for all cells.

Table 8

Frequencies for Weighted Confidence Scores Collapsed Across Targets

Target		Weighted Confidence Score				
		1 – 60%	61 – 120%	121 – 180%	181 – 240%	241 – 300%
Overall	TP	48	61	65	54	61
	TA	43	56	27	12	5
	Percent Guilty	52.75%	52.14%	70.65%	81.82%	92.42%

Note. TP = Target-present lineup data; TA = Target-absent lineup data; F = female target; M = male target.

Overall, weighted confidence was relatively low for identifications from TA lineups, as most scores were in the 1 – 120% range. However, data from TA lineups did fall into all of the categories. For TP lineups, data also fell into all of the categories, with relatively fewer scores in the lowest category (1 – 60%). Percent guilty was near chance at the lowest two confidence ranges (1 – 60% and 61 – 120%) and increased steadily over the remaining confidence levels. The highest percent guilty (92%) was found for the highest weighted confidence range (241 – 300%). The pattern of percent guilty increasing steadily as weighted confidence increased was not present when the targets were examined individually (see Table 8). When only weighted confidence between 121 – 300% was examined, only three targets had percent guilty that increased. An additional target produced percent guilty of 100% for all three of the highest confidence ranges.

Recall of Exposure Angle

For the fourth hypothesis, regarding how well exposure angle is recalled, proportions of correct responses for each exposure view were calculated and compared descriptively between the four exposure angles (see Table 9). The greatest proportion of correct recall came with the face-front exposure – 92% of participants correctly recalled seeing the target in this position. When participants seeing this exposure were inaccurate, they were most likely to recall the target as facing the ¾ position (6%). The next highest proportion of correct recall was with the face-¼ exposure (81%), where the most likely error was to recall seeing the target in profile (10%), although 2% recalled seeing the target face-front. Proportion of correct recall was lower with the face-profile exposure (75%), with a sizable percentage of participants incorrectly recalling the target facing ¾ (15%). Lastly, proportion correct was lowest with the face-¾ (71%) exposure, with the largest error being recall of the target as facing-front (20%). The poorer performance of the face-¾ exposure angle in this circumstance indicates that this exposure angle is not beneficial when recalling the angle seen, though it can be beneficial in terms of recognizing the target.

Table 9

Percentage of Participants Making Each Response for Angle Recalled

Angle Recalled	N	Exposure Angle Face-front	Face-¾	Face-profile	Face-¼
Front	96	**91.67% (*n* = 88)**	19.79% (*n* = 19)	5.21% (*n* = 5)	2.08% (*n* = 2)
¾	96	6.25% (*n* = 6)	**70.83% (*n* = 68)**	14.58% (*n* = 14)	6.25% (*n* = 6)
Profile	96	1.04% (*n* = 1)	6.25% (*n* = 6)	**75.00% (*n* = 72)**	10.42% (*n* = 10)
¼	96	1.04% (*n* = 1)	-	4.17% (*n* = 4)	**81.25% (*n* = 78)**
Back	96	-	3.12% (*n* = 3)	1.04% (*n* = 1)	-

Note. Front = Face-front; 3/4 = Face-¾; Profile = Face-profile; 1/4 = Face-¼; Back = Face-Back. For discrepancy means, standard deviations are bracketed. Dashes indicate no data for the cell. Percentages reflect percent of participants providing a response in that cell. Bold font indicates cells with correct recall.

Discussion

In terms of the first hypothesis, that percent guilty would increase with number of identifications, results from the overall data, and five of eight targets, supported this assertion. In contrast to expectations however, percent guilty were not lower than those in Experiment 1. With overall data, the percent guilty were comparable (and higher in the case of two identifications). For individual targets, many percent guilty for three identifications were higher than the percent guilty in Experiment 1. Of the targets, percent guilty increased to 93% or above for four targets, and increased to 86% and 83% for two others. The results for F4 were notable however, as the percent guilty remained at or around 50%, regardless of the number of identifications made. In short, the performance of the multiple lineup method again varied depending on the target – which again may reflect some combination of target, lineup, and subject factors.

Comparing the by-target patterns of percent guilty found in Experiment 2 to Experiment 1, percent guilty was much more likely to steadily increase in Experiment 2 and was often also higher. In particular, in Experiment 1, the highest percent guilty for three targets was between 82-85% and for one it was 100%. In Experiment 2 however, for two targets the highest percent guilty was 83-86%, while for four targets it was 93-96%. Yet, in terms of the percentage of participants making multiple identifications (by-target), fewer participants were able to do so in Experiment 2 – though this does not mean that performance was absolutely poor in the second study. In fact, over half of the participants were still able to make multiple identifications for six of the targets, and almost half of the participants identified the guilty suspect from all three lineups for two of the targets.

When examining the average number of suspect identifications elicited by each exposure angle, significantly more identifications were made when the ¾ exposure angle was seen, as opposed to all other exposures. In contrast to Experiment 1, this larger number of identifications seems to be driven by more guilty suspect identifications from TP lineups, rather than more innocent suspect identifications from TA lineups. When examining the average number of correct decisions (i.e., target identifications or correct rejections) by exposure angle, significantly more correct decisions were made with the ¾ exposure than with the front, profile, and ¼ exposures. There was also an interaction with correct decisions: in TP lineups, the ¾ exposure led to significantly more correct decisions than all other exposures, while in TA lineups, the ¼ exposure led to significantly fewer correct decisions than all other exposures. In short, when the target was present, participants were significantly more likely to identify the target in the lineup with the ¾ exposure, but when the target was absent, participants were significantly less likely to state that the target was not in the lineup.

In terms of the second hypothesis, that accuracy would be greatest for the lineups that correspond to the view seen in the exposure video, very limited support was found. For these results, increased accuracy would mean high proportions of participants make TP suspect identifications, but small proportions of participants make TA suspect identifications. In the analyses, the expected pattern was found for only one of the exposures: face-front. With this exposure, the proportion of participants identifying the innocent suspect did not vary significantly across the three lineup angles, although a significantly greater proportion of participants identified the guilty suspect when shown the face-front lineup compared to the other two lineups. In short, the same-angle advantage was only found when a face-front exposure and lineup corresponded.

Regarding the third hypothesis, that confidence would increase as accuracy increased, some support was found. Although the planned analysis (which would have directly assessed this question) could not be run, the analyses that were run addressed the question in part. When suspect identifications were examined, identifications of suspects in TP lineups (i.e., guilty suspects) garnered significantly higher confidence ratings than incorrect identifications (i.e., identifications of innocent suspects). For correct decisions, there was no significant difference in confidence between TP and TA lineups – participants correctly identified the target or correctly rejected the lineup with equivalent amounts of confidence. In other words, correct decisions were made with similar confidence regardless of whether it was identifying the target or rejecting the lineup. Correct identifications (of guilty suspects) also were made with significantly more confidence than were misidentifications (of innocent suspects). The other question assessed by the confidence analyses, how confidence ratings are impacted when exposure and lineup angles do or do not match, showed that confidence in correct decisions and suspect identifications are significantly higher when the two angles match than when they do not.

When percent guilty were calculated for weighted confidence, percent guilty increased as confidence increased (at least for the higher levels of weighted confidence). As with many of the analyses, this pattern was clear when data were considered overall, and much more variable when examined by-target. In all, identifications of guilty suspects were generally made with more confidence than were identifications of innocent suspects.

As there was no hypothesis connected to the fourth question, the hypothesis could be neither supported or unsupported. Yet, the percent of participants correctly recalling the angle was higher for some exposure angles (e.g., face-front) and much lower for others (e.g., ¾). When errors were made, participants were more likely to recall a face-front exposure as facing more towards the target's right (i.e., as a ¾ exposure), while other exposures were more likely to be recalled facing more towards the target's left (i.e., closer to the face-front position than what the exposure actually was). In short, if participants saw the target facing the front, they were more likely to state (when incorrect) that the target was facing more towards the side – which was also the only possible way to be incorrect for this exposure. If the target was facing one of the other positions (i.e., facing more towards the side), participants were more likely to make errors wherein they recalled seeing more of the target's face than they actually did.

One limitation to Experiment 2 regards the examination of how exposure views not presented in the lineups can impact accuracy. In particular, this question was assessed using only one such exposure view – the ¼ view. Although this view does provide some indication of how the multiple lineup method can perform when exposure view does not correspond to the lineup views, it is also confounded by the fact that very little face information is available in this angle. So, while the ¼ exposure seemed to lead to a poorer performance than the other exposure angles, this could also be explained by the fact that recognition may be generally poor for this exposure view, especially as there is relatively little distinguishing information available.

A second exposure-related issue is that the exposure videos were flipped, which creates potential problems as no person is perfectly symmetrical. Thus, participants would be exposed to information that differed slightly from the information presented at test. If this method was a disadvantage to correct identifications, participants should perform more poorly in Experiment 2 compared to Experiment 1. In particular, the proportion of participants making identifications of guilty suspects should decrease, as the information about the guilty suspect differed between exposure and test. However, proportion of participants identifying the guilty suspect (and percents guilty) generally differed minimally between Experiments 1 and 2. However, if participants in Experiment 2 performed more poorly, a difficulty in interpretation would arise as the poorer performance could be explained in multiple ways. For example, poorer performance could be due to the materials used or because the exposure provided to participants was limited to only one angle (which was originally hypothesized).

A third limitation is that although it is good to assess the utility of the multiple lineup procedure using as many different targets as possible, using different targets in Experiments 1 and 2 limits the extent to which results from the two experiments can be compared. Any differences in the results could be explained as being due to the different targets and lineups used, either of which could engender different proportions of correct responses and different proportions of misidentifications. Particularly, the effectiveness of the procedure may disappear if foils are overly similar to the target (as is a notable possibility with F4 in Experiment 2). In order to perform a more stringent comparison of the results, the same targets (with the same respective lineups) should be used in both designs.

Chapter 4: General Discussion

Overall, while some worrisome results were found (particularly in by-target analyses), the multiple lineup procedure held up well, even though the lineups were not independent. By using only face lineups, it was possible that any diagnosticity found with the multiple lineup method in prior studies would be destroyed if participants identified a lineup member and recognized that same person from subsequent lineups (regardless of whether or not it was the target). Although the method did not perform perfectly for all targets and analyses, there was generally no massive decline in percent guilty, particularly when suspects were identified from all three lineups. As well, the low percent guilty with one identification (just under 50%) may indicate that witnesses who provide a single identification (when presented with multiple face lineups) are a poorer source of evidence than witnesses who make multiple identifications. Thus, although percent guilty for two or three identifications was not a substantially higher than the percent guilty for each individual lineup, the use of multiple lineups may help in terms of differentiating between witnesses who provide poorer versus better identification evidence.

While the perfect diagnosticity (percent guilty = 100%) found in both Pryke et al.'s (2004) Experiment 2 and Boyce et al.'s (2007) Experiment 2 occurred only once in the current studies, the percent guilty for 3 suspect identifications in the current analyses are generally on par with those found in Pryke et al. (2004) Experiment 1 (91%) and Boyce et al. (2007) Experiment 1 (85%), both of which included independent lineups. As well, even though percent guilty can only be computed for 2 suspect identifications in Sauerland and Sporer (2008), the percent guilty for 3 identifications in the current studies are generally similar to the 74% and 89% that they found. Looking at the current studies' overall analyses, percent guilty were 77% and 85% (Experiment 1) and 83% (Experiment 2), with individual targets often having percent guilty above 90% (particularly in Experiment 2). Even the lowest overall percent guilty (77%; for uninformed data in Experiment 1) was still similar to some of the prior findings. In all, the multiple lineup procedure worked well, even when using three face lineups.

Given that the multiple lineup method appears to maintain its utility (or loses very little of it), the next goal to assess is whether including lineups that are not independent helps increase the percentage of people making multiple identifications. In Experiment 1, an average of 64% of participants made two or more target identifications when informed and 43% made three target identifications. This finding is comparable to (albeit slightly higher than) Experiment 2, where an average of 58% of participants made two or more target identifications and 34% made three identifications. For the uninformed data in Experiment 1, the percentages were slightly lower: 50% from two or more lineups, 28% from all three. As in the other analyses, the percentages by-target varied dramatically.

In comparison to the results in Pryke et al. (2004) and Boyce et al. (2007), percentages for two or more identifications are similar, while percentages for three or more identifications are much higher in the current analyses. For example, Pryke et al. (2004) found only 13% of participants made three target identifications in Experiment 1, and only 17% made three or more identifications in Experiment 2. Boyce et al. (2007) found slightly higher percentages (e.g., 29%, 22%, and 40% making at least three target identifications), though they used both face-front and face-profile lineups, which may have increased the percentage of participants able to make multiple target identifications. Even though percentages for Sauerland and Sporer (2008) could only be calculated for two identifications, their percentages (e.g., 2%, 12%) were much lower than the ones found in the current experiments. Overall, using only face-lineups appears to have both maintained the utility of the multiple lineup procedure, as well as increased the frequency of people making the multiple identifications.

When it comes to the ¾ view, some evidence was found for a ¾ advantage, although this advantage was not strong and was mainly limited to the exposure angle. The main advantage for the ¾ lineup was in Experiment 2, where this lineup provided the highest percent guilty for the ¾, profile, and ¼ exposures. In terms of the relatively poorer performance of the ¾ lineup in Experiment 1, it is difficult to assess whether the lower percent guilty was due to the issue of lineup order, the issue of how exposure angles were manipulated, or a combination of these issues. At the same time, there was no disadvantage in using this lineup, in terms of both specific analyses and the generally high diagnosticity of multiple identifications. Only partially in line with the face recognition literature however, was the limited evidence for increased accuracy when exposure and test angles correspond, which only occurred with the face-front view. Additional evidence for correspondence was found with the ¼ exposure, which matched no lineup, as accuracy was generally poorer under this condition. However, this disadvantage may also be due to the very small amount of information available for recognition in that view.

Given such findings, it would be interesting to see how the face-only multiple lineup method functions when the exposure angle is a mirror image to the lineup angle (e.g., exposure to right profile and test with left profile). While the symmetrical exposure would provide similar information about facial features (albeit about the other side of the face), research in the facial recognition field is somewhat divided, with some research indicating that exposure views that are symmetrical to test views aid recognition in limited ways (Hill et al., 1997; Jeffery et al., 2006) and other research indicating that symmetry strongly aids recognition (Busey & Zaki, 2004).

In assessing the confidence-accuracy relationship in the two studies, one potential criticism that could be levied is that the relationship was not assessed using a point-biserial correlation (due to the data involved), and that the analyses performed are less valid when examining how these two factors are related. However, though the analyses differed from the usual test, debate exists about the best way to capture this relationship. In particular, some argue that a better indicator of this relationship is to look at calibration, where the percent of participants making correct identifications is plotted against confidence levels (from 0 to 100%; Brewer, Keast, & Rishworth, 2002; Leippe & Eisenstadt, 2007), as correlations are not indicative of diagnosticity (Juslin, Olsson, & Winman, 1996). Perfect calibration would occur if 100% of participants who have confidence levels of 100% are accurate, if 80% of participants who have confidence levels of 80% are correct, and so on (Brewer et al., 2002; Leippe & Eisenstadt, 2007). Even when confidence-accuracy correlations are relatively low, good (or perfect) calibration can occur (Juslin et al., 1996). However, the utility of calibration is limited to confidence scores from 50-100%, is better when participants have used absolute judgment strategies, and is better for identifications as opposed to lineup rejections (Weber & Brewer, 2003).

In fact, the current analyses provide some indication of calibration. The percent guilty calculation – or percentage of identifications that are accurate (i.e., of the guilty suspect) – generally increased gradually as weighted confidence increased. Although the calculations in the current Experiments were not exact replications of the calibration discussed by the above authors, they are a form of calibration.

Another potential confidence-related limitation of the experiments conducted above is that the exposure to the targets in the videos was very short. It is possible that clearer relationships between confidence and accuracy would be found if participants were provided with a longer exposure to each target, as there is some indication that being able to view targets for longer periods of time does strengthen the relation between confidence and accuracy (Leippe & Eisenstadt, 2007). Yet, as with many of the factors discussed here, evidence for this view is not conclusive. Still, it may be useful to increase the amount of time participants are able to see the target in future studies, in order to see whether clearer relationships between confidence and accuracy are found with this method.

An additional limitation of the current studies involves the length of the delays between exposure and lineup (and between lineups). Although similar to other studies (e.g., Sauerland & Sporer, 2008), the 30 second delays used in the current research could be too short and may have artificially inflated the diagnosticity of the results. Increasing the amount of delay would allow for an assessment of the face-only multiple lineup procedure under more realistic conditions.

A further limitation of these studies involves the fact that they did not account for the cross-race effect. Boyce et al. (2007) did include an examination of this effect with the multiple lineup procedure, however their results may not be entirely applicable to the current studies. While Boyce et al. (2007) did include more than one face lineup in their procedure, they had multiple arrays which provided far fewer cues as to the lineup member's ethnicity. With the current procedure, faces were included in all of the lineups, meaning that cues to ethnicity were prevalent. As the current study was not designed to assess the cross-race effect (or additional biases, such as the own-sex bias), examination of such effects is hampered by the relatively small number of participants in certain subgroups (i.e., non-European; males). Thus, assessing the extent to which the cross-race effect influences the results for the face-only multiple lineup method is another avenue to explore with future research.

Lastly, further research can examine what factors cause the variability seen in the by-target analyses of the face-only multiple lineup procedure. In particular, whether the variability is due to aspects of the target (e.g., distinctiveness), lineup (e.g., lineup fairness), witness (sex, ethnicity, age), interactions these factors (e.g., cross-race effects), or other as yet unidentified factors. For example, the poor diagnosticity found with F4 in Experiment 2 may be because the target and innocent suspect were highly similar to each other. While the iterative procedure used in the current experiments was based on that of Turtle et al. (2003), there are two reasons that the procedure may not have worked in this research. The first reason why the iterative procedure may not have worked is that the first selected foil (the one with the greatest similarity to the target) was kept in the lineup, thus ensuring that at least one lineup member was chosen based on a direct similarity to the suspect. A foil chosen due to similarity to the target would likely be mistakenly identified more often than foils who appear to be less similar to the target.

A second reason that may have undermined the effect of the iterative procedure is the use of the worst case scenario for selecting the innocent suspect. This method ensures that any accidental match to the appearance of the target inflates the false identification rate because that lineup member will be treated as the suspect, while in real-world situations it would be a random event whether the most similar member of a target-absent lineup was the suspect. Addressing what causes the by-target variability is important as it could help mitigate the poor performance seen with some of the targets. If the multiple lineup method could perform similarly across all targets, this would help its applicability in forensic settings, as its utility would be more reliable for single targets, particularly as police forces would use the method on specific targets (i.e., criminal suspects), and not use it as an aggregate tool.

However, it should also be noted that the variation across targets is not necessarily ruinous to the procedure – the effectiveness of other procedures already in use by police forces, including sequential lineups, likely varies across targets. As well, the effectiveness of a procedure for any one target will likely change depending on how target, lineup, subject, or other factors interact with each other. In all, a more pertinent issue is the extent to which the procedure is an improvement over other methods. Creating an identification procedure which always and only provides accurate identifications may be unrealistic, due to the nature of human memory. Creating a procedure which provides even slight increases in diagnosticity however, will be helpful. Taken in the context of actual cases and the absolute number of identifications made, slight increases will have large impacts on the number of innocent suspects being misidentified (Lindsay, Mansour, Beaudry, & Bertrand, 2009).

In all, solid support was found for the face-only multiple lineup procedure, particularly in terms of its diagnosticity (in the form of percent guilty) and in terms of the percentage of participants who could make multiple target identifications. In terms of the specific hypotheses, support was more mixed and sometimes varied between Experiments 1 and 2 (e.g., ¾ advantage). Although some limitations to the studies exist, there is enough support and advantages to warrant further investigation of this procedure.

References

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.

Boyce, M. A., Lindsay, R. C. L., Dupuis, P., & Beaudry, J. L. (2007). Multiple lineup identifications: Optimizing a procedure to calibrate eyewitness accuracy. Unpublished manuscript, Queen's University.

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*, 44-56.

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

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.

Bruce, V., Valentine, T., & Baddeley, A. (1987). The basis of the ¾ view advantage in face recognition. *Applied Cognitive Psychology, 1*, 109-120.

Burton, A. M., & Jenkins, R. (in press). Unfamiliar face perception. In A. Calder, G. Rhodes, J. Haxby, & M. Johnson (Eds), *The Handbook of Face Perception*, Oxford University Press.

Busey, T. A., & Zaki, S. R. (2004). The contribution of symmetry and motion to the recognition of faces at novel orientations. *Memory & Cognition, 32*, 916-931.

Carbon, C. C., & Leder, H. (2006). When faces are heads: View-dependent recognition of faces altered relationally or componentially. *Swiss Journal of Psychology, 65*, 245-252.

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 Behavior, 30*, 287-307.

Dupuis, P. R., & Lindsay, R. C. L. (2007). Radical alternatives to traditional lineups. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 179-200). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Haw, R. M., Dickinson, J. J., & Meissner, C. A. (2007). The phenomenology of carryover effects between show-up and line-up identification. *Memory, 15*, 117-127.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*, 201-222.

Innocence Project. (n.d.). *Eyewitness Identification Reform*. Retrieved June 28, 2009, from http://www.innocenceproject.org/Content/165.php.

Jeffery, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. *Psychological Science, 17*, 501-505.

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.

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

Laeng, B., & Rouw, R. (2001). Canonical views of faces and the cerebral hemispheres. *Laterality, 6*(3), 193-224.

Laughery, K. R., Alexander, J. F., & Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology, 55*, 477-483.

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 377-426). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Levi, A. M. (2006). An analysis of multiple choices in MSL lineups, and a comparison with simultaneous and sequential ones. *Psychology, Crime, and Law, 12*, 273-285.

Levi, A. M. (2007). Research note: Evidence for moving to an 84-person photo lineup. *Journal of Experimental Criminology, 3*, 377-391.

Lindsay, R. C. L. (1994). Expectations of eyewitness performance: Jurors' verdicts do not follow from their beliefs. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult Eyewitness Testimony: Current Trends and Developments*. New York: Cambridge University Press.

Lindsay, R. C. L., Mansour, J. K., Beaudry, J. L., & Bertrand, M. I. (2009). Sequential lineup presentation: Patterns and policy. *Legal and Criminological Psychology, 14*, 13-24.

Lindsay, R. C. L., Semmler, C., Weber, N., Brewer, N., & Lindsay, M. R. (2008). How variations in distance affect eyewitness reports and identification accuracy. *Law and Human Behavior, 32*, 526-535.

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,* 463-478.

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

Lindsay, R. C. L., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses. *Law and Human Behavior, 13,* 333-339.

Liu, C. H., & Chaudhuri, A. (2002). Reassessing the ¾ view effect in face recognition. *Cognition, 83*, 31-48.

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. *Applied Cognitive Psychology, 1*, 53-69.

Melara, R. D., DeWitt-Rickards, T. S., & O'Brien, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology, 74*, 706-713.

Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law & Human Behavior, 22*, 549-570.

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.

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

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.

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

Steblay, N., 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.

Stephan, B. C. M., & Caine, D. (2007). What is in a view? The role of featural information in the recognition of unfamiliar faces across viewpoint transformation. *Perception, 36*, 189-198.

Turtle, J., Lindsay, R. C. L., & Wells, G. L. (2003). Best practice recommendations for eyewitness evidence procedures: New ideas for the oldest way to solve a case. The Canadian Journal of Police and Security Services, 1, 5-18.

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition., 490-499.

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

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14(2)*, 89-103.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22*, 603-647.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law, 7*, 119-133.

Appendix A

Face-front lineup

Face-profile lineup

Face-¾ lineup

Appendix B

Frequencies, Percent Guilty, and Percentage of Participants Making Identifications By-Target

Target		Number of Identifications Made			
		0	1	2	3
F1	TP	12 (40.00%)	9 (30.00%)	4 (13.33%)	5 (16.67%)
	TA	22 (73.33%)	3 (10.00%)	3 (10.00%)	2 (6.67%)
	Percent Guilty	35.29%	75.00%	57.14%	71.43%
F2	TP	4 (13.33%)	4 (13.33%)	6 (20.00%)	16 (53.33%)
	TA	17 (56.67%)	4 (13.33%)	6 (20.00%)	3 (10.00%)
	Percent Guilty	19.05%	50.00%	50.00%	84.21%
F3	TP	1 (3.33%)	0 (0.00%)	7 (23.33%)	22 (73.33%)
	TA	9 (30.00%)	8 (26.67%)	9 (30.00%)	4 (13.33%)
	Percent Guilty	10.00%	0.00%	43.75%	84.62%
M1	TP	10 (33.33%)	7 (23.33%)	7 (23.33%)	6 (20.00%)
	TA	19 (63.33%)	4 (13.33%)	4 (13.33%)	3 (10.00%)
	Percent Guilty	34.48%	63.64%	63.64%	66.67%
M2	TP	8 (26.67%)	6 (20.00%)	7 (23.33%)	9 (30.00%)
	TA	14 (46.67%)	10 (33.33%)	4 (13.33%)	2 (6.67%)
	Percent Guilty	36.36%	37.50%	63.64%	81.82%
M3	TP	3 (10.00%)	1 (3.33%)	6 (20.00%)	20 (66.67%)
	TA	23 (76.67%)	3 (10.00%)	4 (13.33%)	0 (0.00%)
	Percent Guilty	11.54%	25.00%	60.00%	100.00%

Note. TP = Target-Present lineup data; TA = Target-Absent lineup data; F = Female target; M = Male target. Percent of participants making that number of suspect identifications are bracketed.

Appendix C

Analysis of Variance (ANOVA) Results for Uninformed and Informed Analyses

Uninformed				
Source	*df*	SS	*F*	*P*
		Between subjects		
Intercept	1	26.04	67.08	< .01
Target (T)	1	4.45	11.46	< .01
Between-group error	70	27.176	(0.39)	
		Within subjects		
Lineup angle (L)	2	0.11	0.47	.63
L x T	2	0.59	2.49	.09
Within-group error	140	16.63	(0.12)	

Informed				
Source	*df*	SS	*F*	*P*
		Within subjects		
L	2 (142)	0.60 (6.29)	6.76	< .01
T	1 (71)	14.46 (9.65)	106.39	< .01
L x T	2 (142)	0.64 (5.36)	8.42	< .01

Note. Error for sum of squares and degrees of freedom in the informed repeated-measures ANOVA are in brackets.

Appendix D

Frequencies for Weighted Confidence Scores By-Target

Target		Weighted Confidence Score				
		1 – 60%	61 – 120%	121 – 180%	181 – 240%	241 – 300%
F1	TP	5	6	3	1	3
	TA	1	4	2	1	0
	Percent Guilty	83.33%	60.00%	60.00%	50.00%	100.00%
F2	TP	1	4	6	4	11
	TA	2	5	4	2	0
	Percent Guilty	33.33%	44.44%	60.00%	66.67%	100.00%
F3	TP	0	4	4	9	12
	TA	7	6	5	2	1
	Percent Guilty	0.00%	40.00%	44.44%	81.82%	92.31%
M1	TP	3	7	5	4	1
	TA	2	5	1	2	1
	Percent Guilty	60.00%	58.33%	83.33%	66.67%	50.00%
M2	TP	3	4	5	5	5
	TA	7	5	1	2	1
	Percent Guilty	30.00%	44.44%	83.33%	71.43%	83.33%
M3	TP	1	2	4	12	8
	TA	3	3	1	0	0
	Percent Guilty	25.00%	40.00%	80.00%	100.00%	100.00%

Note. TP = Target-present lineup data; TA = Target-absent lineup data; F = female target; M = male target.

Appendix E

Frequencies, Percent Guilty, and Percentage of Participants Making Identifications By-Target

Target		Number of Identifications Made			
		0	1	2	3
F4	TP	19 (39.58%)	10 (20.83%)	11 (22.92%)	8 (16.67%)
	TA	21 (43.75%)	10 (20.83%)	8 (16.67%)	9 (18.75%)
	Percent Guilty	47.50%	50.00%	57.89%	47.06%
F5	TP	12 (25.00%)	5 (10.42%)	8 (16.67%)	23 (47.92%)
	TA	42 (87.50%)	4 (8.33%)	1 (2.08%)	1 (2.08%)
	Percent Guilty	22.22%	55.56%	88.89%	95.83%
F6	TP	3 (6.25%)	5 (10.42%)	16 (33.33%)	24 (50.00%)
	TA	29 (60.42%)	12 (25.00%)	6 (12.50%)	1 (2.08%)
	Percent Guilty	9.38%	29.41%	72.73%	96.00%
F7	TP	13 (27.08%)	9 (18.75%)	8 (16.67%)	18 (37.50%)
	TA	34 (70.83%)	7 (14.58%)	4 (8.33%)	3 (6.25%)
	Percent Guilty	27.66%	56.25%	66.67%	85.71%
M4	TP	11 (22.92%)	8 (16.67%)	12 (25.00%)	17 (35.42%)
	TA	25 (52.08%)	10 (20.83%)	5 (10.42%)	8 (16.67%)
	Percent Guilty	30.56%	44.44%	70.59%	68.00%
M5	TP	13 (27.08%)	6 (12.5%)	15 (31.25%)	14 (29.17%)
	TA	32 (66.67%)	9 (18.75%)	6 (12.50%)	1 (2.08%)
	Percent Guilty	28.89%	40.00%	71.43%	93.33%
M6	TP	13 (27.08%)	16 (33.33%)	9 (18.75%)	10 (20.83%)
	TA	19 (39.58%)	18 (37.50%)	9 (18.75%)	2 (4.17%)
	Percent Guilty	40.62%	47.06%	50.00%	83.33%
M7	TP	11 (11.46%)	9 (9.38%)	13 (13.54%)	15 (15.62%)
	TA	39 (81.25%)	7 (14.58%)	1 (2.08%)	1 (2.08%)
	Percent Guilty	22.00%	56.25%	92.86%	93.75%

Note. TP = Target-Present lineup data; TA = Target-absent lineup data; F = female target; M = male target. Percentage of participants identifying the target given each number of identifications is provided in brackets.

Appendix F

ANOVA Results for Number of Suspect Identifications and Correct Decisions, by Exposure View

Suspect Identifications				
Source	*df*	SS	*F*	*p*
		Within subjects		
Exposure angle (E)	2.69 (225.37)	20.34 (259.16)	7.57	< .01
Target (T)	1 (95)	212.52 (88.98)	226.9	< .01
E x T	3 (285)	18.72 (330.78)	5.38	< .01
Correct Decisions				
Source	*df*	SS	*F*	*p*
		Within subjects		
E	3 (285)	45.53 (309.47)	13.98	< .01
T	1 (95)	61.88 (174.62)	33.66	< .01
E x T	3 (285)	8.09 (289.41)	2.66	0.05

Note. Error for sum of squares and degrees of freedom are in brackets.

Appendix G

3-Way ANOVA Results for Proportion of Participants Making Identifications

Source	*df*	SS	*F*	*p*
		Within subjects		
Target (T)	1 (95)	70.84 (29.66)	226.9	< .01
Lineup angle (L)	2 (190)	0.96 (26.87)	3.4	0.04
Exposure angle (E)	2.69 (255.37)	6.78 (86.39)	7.46	< .01
T x L	1.87 (178.00)	0.04 (18.46)	0.23	0.78
T x E	3 (285)	6.24 (110.26)	5.38	< .01
L x E	6 (570)	1.11 (66.72)	1.58	0.15
T x L x E	6 (570)	2.90 (67.60)	4.07	< .01

Note. Error for sum of squares and degrees of freedom are in brackets. For Exposure angle main effect and Target by Lineup angle interaction, values provided are the Greenhouse-Geisser corrections.

Appendix H

ANOVA Results for Target by Lineup Angle Interaction, by Exposure Angle

Source		*df*	SS	*F*	*p*
		Within subjects			
Face-front	Target (T)	1 (95)	17.02 (38.15)	42.37	< .01
	Lineup angle (L)	2 (190)	0.51 (20.82)	2.35	.10
	T x L	2 (190)	1.54 (19.79)	7.4	< .01
Face-¾	T	1 (95)	31.64 (31.53)	95.35	< .01
	L	2 (190)	0.25 (25.08)	0.96	.38
	T x L	2 (190)	0.38 (24.95)	1.47	.23
Face-profile	T	1 (95)	23.36 (34.97)	63.46	< .01
	L	2 (190)	0.45 (23.55)	1.81	.17
	T x L	1.86 (176.91)	0.73 (23.93)	2.91	.06
Face-¼	T	1 (95)	5.06 (35.27)	13.64	< .01
	L	2 (190)	0.86 (24.14)	3.38	.04
	T x L	2 (190)	0.28 (17.38)	1.54	.22

Note. Error for sum of squares and degrees of freedom are in brackets. For the face-profile Target by Lineup angle interaction, values provided are the Greenhouse-Geisser corrections.

Appendix I

Average Confidence Judgments for Suspect Identifications, by Exposure and Lineup Angles

Exposure View		Lineup Type		
		Face-Front	Face-¾	Face-Profile
Face-Front	TP	73.82%	72.26%	67.07%
	TA	63.85%	56.18%	66.53%
Face-¾	TP	73.78%	74.84%	72.85%
	TA	56.00%	71.89%	62.82%
Face-Profile	TP	72.86%	74.07%	74.86%
	TA	61.45%	66.12%	66.21%
Face-¼	TP	60.53%	65.38%	69.50%
	TA	57.00%	64.38%	54.84%

Note. TP = Target-present lineup data; TA = Target-absent lineup data. Cell *n*'s range from 13 to 67.

Appendix J

Statistics for Confidence ANOVAs

Source		*df*	SS	*F*	*p*
		Within subjects			
Suspect Identifications	Target (T)	1 (31)	3283.37 (8647.62)	11.77	< .01
	Angles (A)	1 (31)	557.15 (4026.66)	4.29	.05
	T x A	1 (31)	148.89 (3526.73)	1.31	.26
Correct Decisions	T	1 (62)	1185.60 (21713.99)	3.38	.07
	A	1 (62)	837.13 (5420.84)	9.58	< .01
	T x A	1 (62)	116.58 (4739.33)	1.52	.22

Note. Error for sum of squares and degrees of freedom are in brackets.

Appendix K

Frequencies for Weighted Confidence Scores By-target

Target		Weighted Confidence Score				
		1 – 60%	61 – 120%	121 – 180%	181 – 240%	241 – 300%
F4	TP	4	12	5	7	1
	TA	4	9	10	3	1
	Percent Guilty	50.00%	57.14%	33.33%	70.00%	50.00%
F5	TP	4	4	10	4	14
	TA	3	3	0	0	0
	Percent Guilty	57.14%	57.14%	100.00%	100.00%	100.00%
F6	TP	5	4	12	9	15
	TA	6	8	5	0	0
	Percent Guilty	45.45%	33.33%	70.59%	100.00%	100.00%
F7	TP	8	6	4	5	12
	TA	6	2	5	1	0
	Percent Guilty	57.14%	75.00%	44.44%	83.33%	100.00%
M4	TP	5	9	12	9	2
	TA	5	8	3	4	3
	Percent Guilty	50.00%	52.94%	80.00%	69.23%	40.00%
M5	TP	5	6	10	9	5
	TA	4	9	2	1	0
	Percent Guilty	55.56%	40.00%	83.33%	90.00%	100.00%
M6	TP	12	9	5	4	5
	TA	12	12	2	2	1
	Percent Guilty	50.00%	42.86%	71.43%	66.67%	83.33%
M7	TP	5	11	7	7	7
	TA	3	5	0	1	0
	Percent Guilty	62.50%	68.75%	100.00%	87.50%	100.00%

Note. TP = Target-present lineup data; TA = Target-absent lineup data; F = female target; M = male target.