LINEUP COMPOSITION EFFECTS ON EYEWITNESS IDENTIFICATION

A Thesis

Submitted to the Faculty of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Experimental and Applied Psychology

University of Regina

by

Ryan Joseph Fitzgerald

Regina, Saskatchewan

December 10, 2014

© 2014: R. J. Fitzgerald

UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

SUPERVISORY AND EXAMINING COMMITTEE

Ryan Joseph Fitzgerald, candidate for the degree of Doctor of Philosophy in Experimental & Applied Psychology**,** has presented a thesis titled, ***Lineup Composition Effects on Eyewitness Identification,*** in an oral examination held on December 10, 2014. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner:	*Dr. D. Stephen Lindsay, Univeristy of Victoria
Co-Supervisor:	Dr. Christopher Oriet, Department of Psychology
Co-Supervisor:	Dr. Heather Price, Department of Psychology
Committee Member:	Dr. Richard MacLennan, Department of Psychology
Committee Member:	Dr. Katherine Arbuthnott, Department of Psychology
Committee Member:	Dr. Rick Ruddell, Department of Justice Studies

Chair of Defense: Dr. Susan Johnston, Department of English

*via SKYPE

ABSTRACT

Whenever lineups are prepared for eyewitness identification, the investigator constructing the lineup must decide how to choose lineup members to appear in the lineup with the person under investigation. Accordingly, three research projects were conducted to inform lineup construction practices. The first project is a meta-analysis of lineup similarity effects. Results of the meta-analysis suggest that biased lineups yield the highest rate of suspect identifications, regardless of whether that person is guilty or innocent. When lineups with moderately similar members are compared with lineups containing highly similar members, increasing similarity reduces innocent suspect identifications and has only a small and nonsignificant effect on correct identifications. Two experimental studies were then conducted using morphing software to systematically manipulate the degree of similarity between lineup members who are known be innocent (fillers) and the lineup member who is under investigation (the suspect). In the first experiment, lineups with moderately high similarity fillers yielded a higher correct identification rate than did lineups with very high similarity fillers. When comparable procedures were used in the second experiment, fillers of low and moderately low similarity to the culprit yielded nearly identical correct identification rates. In both experiments, increasing suspect-filler similarity led to a decrease in innocent suspect misidentifications. The accumulation of evidence from the three research projects suggests lineups would be best constructed with fillers of moderate similarity to the suspect.

ACKNOWLEDGEMENTS

I thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding this project. These funds were provided through the Faculty of Graduate Studies and Research. I express my deepest gratitude to my supervisors, Drs. Chris Oriet and Heather Price, who made substantial contributions to the design of this project. The methodology in the experimental studies was the result of hours of discussion between the three of us. I also received some extremely helpful comments and suggestions from my committee members (Drs. Katherine Arbuthnott, Richard MacLennan, and Rick Ruddell). Dr. Steve Charman also deserves recognition for his contribution to the meta-analysis in Chapter 2. I also thank Chet Hembroff, Dallas Novokowski, and Andrew Templeton for their assistance with data collection.

POST-DEFENCE ACKNOWLEDGEMENT

I would like to thank my external examiner, Dr. Steve Lindsay, for his insightful questions and comments. It was an honour to have my work reviewed by someone so knowledgeable.

TABLE OF CONTENTS

ABSTRACT..... ii

ACKNOWLEDGEMENTS..... iii

POST-DEFENCE ACKNOWLEDGEMENT..... iii

LIST OF TABLES..... iv

LIST OF FIGURES..... vii

LIST OF APPENDICES..... vii

CHAPTER 1: INTRODUCTION..... 1

1.1 The Contribution of Eyewitness Misidentification to Wrongful Convictions..... 4

1.2 System Variables..... 6

1.3 The Technical Working Group for Eyewitness Evidence..... 7

1.4 Filler Selection Procedures..... 9

1.4.1 The match-to-appearance procedure..... 10

1.4.2 The match-to-description procedure..... 12

1.5 Defining Lineup Member Similarity..... 15

1.6 How Similar is “Too Similar”?..... 16

1.6.1 Filler search procedure: Match-to-description vs. match-to-appearance..... 17

1.6.2 Subjective similarity ratings..... 19

1.6.3 Tredoux’s E’..... 21

1.6.4 Euclidean distance..... 25

1.6.5 Computer-generated faces..... 26

1.6.6 Summary of empirical findings..... 26

1.7 Overview of Dissertation Research..... 28

CHAPTER 2: A META-ANALYSIS OF LINEUP MEMBER SIMILARITY EFFECTS... 29

2.1 Method..... 31

2.1.1 Literature search..... 31

2.1.2 Inclusion criteria..... 31

2.1.3 Exclusion criteria 32

2.1.4. Assignment of study weights..... 33

2.1.5 Coding of lineups..... 33

2.1.6 Effect size..... 33

2.1.7 Diagnosticity..... 34

2.1.8 Analysis of culprit-present and culprit-absent lineups..... 35

2.1.9 Analysis of lineup presentation subgroups..... 36

2.2 Results..... 37

2.2.1 High similarity versus biased..... 37

2.2.2 Moderate similarity versus biased..... 39

2.2.3 High versus moderate similarity..... 40

2.2.4 Heterogeneity..... 40

2.2.5 Subgroup analysis of simultaneous and sequential lineups..... 41

2.3 Discussion..... 44

2.3.1 Suspect and filler identifications..... 46

2.3.2 Lineup rejections..... 47

2.3.3 Limitations of the meta-analysis..... 50

2.3.4 Lineup construction recommendations vs. meta-analytic findings..... 51

CHAPTER 3: A NOVEL MANIPULATION OF LINEUP MEMBER SIMILARITY..... 53

3.1 Method..... 54
3.1.1 Participants..... 54
3.1.2 Design..... 54
3.1.3 Video event..... 54
3.1.4 Lineups..... 55
3.1.4.1 Culprit present..... 55
3.1.4.2 Culprit absent..... 60
3.1.5 Pilot tests..... 61
3.1.5.1 Relative judgement task..... 61
3.1.5.2 Find the nonmorph task..... 62
3.1.5.3 Suspect-filler similarity ratings..... 63
3.1.6 Procedure..... 63
3.2 Results..... 66
3.2.1 Effects of similarity on identification responses..... 66
3.2.2 Dispersion..... 69
3.2.3 Diagnosticity..... 70
3.2.4 Signal detection analyses..... 71
3.2.5 Confidence..... 72
3.3 Discussion..... 74
3.3.1 Diagnosticity..... 76
3.3.2 Confidence..... 79
3.3.3 Summary and conclusions..... 79
CHAPTER 4: EFFECTS OF FILLER HOMOGENEITY, SUSPECT-FILLER

SIMILARITY, AND LINEUP PRESENTATION..... 81

4.1 Potential Influences of Lineup Presentation on Filler Homogeneity Effects..... 83

4.2 Method..... 85

4.2.1 Participants..... 85

4.2.2 Excluded data..... 86

4.2.3 Design..... 88

4.2.4 Video event..... 89

4.2.5 Lineups..... 89

4.2.5.1 Low similarity homogeneous fillers..... 89

4.2.5.2 Low similarity heterogeneous fillers..... 89

4.2.5.3 Moderate similarity homogeneous fillers..... 94

4.2.5.4 Moderate similarity heterogeneous fillers..... 94

4.2.5.5 Innocent suspect..... 94

4.2.6 Similarity ratings..... 96

4.2.7 Procedure..... 102

4.3 Results..... 104

4.3.1 Diagnosticity ratios and signal detection theory analyses..... 108

4.3.2 Analysis of lab-only data..... 110

4.4 Discussion..... 114

4.4.1 A relative judgement strategy for sequential lineups?..... 116

4.4.2 Summary and conclusion..... 117

CHAPTER 5: GENERAL DISCUSSION..... 119

5.1 Reconceptualizing Suspect-Filler Similarity Categories..... 119

5.2 Summary of Experimental Findings on Suspect-Filler Similarity..... 120

5.3 Policy Considerations..... 124

5.3.1 Matching to the culprit's description..... 125

5.3.2 Matching to the suspect's appearance..... 126

5.3.3 A combination of methods..... 129

5.4 Concluding Remarks..... 132

REFERENCES..... 134

LIST OF TABLES

Table 1. Correct and false identification (ID) rates for Description-Matched (DM) and Appearance-Matched (AM) lineups..... 18

Table 2. Meta-analysis of suspect-filler similarity on identification choices..... 38

Table 3. Subgroup analysis: Lineup presentation influences on suspect-filler similarity effects..... 42

Table 4. Mean (*SE*) ratings of suspect-filler similarity in the Chapter 3 experiment..... 64

Table 5. Identification response probabilities (P) and standard errors (*SE*) in the Chapter 3 experiment..... 67

Table 6. Ethnic backgrounds of participants in the Chapter 4 experiment..... 87

Table 7. Mean (*SD*) ratings of similarity between the culprit and the fillers in the Chapter 4 experiment..... 97

Table 8. Mean (SD) ratings of culprit-filler similarity for heterogeneous lineups across Experiments 3 and 4..... 99

Table 9. Chapter 4 mean (SD) similarity ratings between the innocent suspect and the fillers..... 101

Table 10. Dispersion index (DI) values for the Chapter 4 experiment..... 105

Table 11. Identification response probabilities (P) for the Chapter 4 experiment ($N = 803$). 106

Table 12. Diagnosticity ratios and signal detection theory analyses (N = 803)..... 109

Table 13. Diagnosticity ratios and signal detection analyses for lab participants in Chapter 4 (n = 224)..... 111

Table 14. Identification response rates for lab participants in Chapter 4 (n = 224) 112

LIST OF FIGURES

Figure 1. Faces of the culprit and the fillers (prior to morphing)..... 56

Figure 2. Lineups for moderately high and very high similarity fillers in the Chapter 3 experiment..... 57

Figure 3. Standard morphing procedures used to create fillers in moderately high and very high similarity conditions in the Chapter 3 experiment..... 58

Figure 4. Proportion of suspect choices from among only those who picked someone from the lineup (choosers) in the Chapter 3 experiment..... 68

Figure 5. Post-identification confidence ratings in the Chapter 3 experiment..... 73

Figure 6. Example of a lineup from Chapter 4 with homogeneous fillers of low similarity to the culprit..... 90

Figure 7. Example of a lineup from Chapter 4 with heterogeneous fillers of low similarity to the culprit..... 91

Figure 8. Example of a lineup from Chapter 4 with homogeneous fillers of moderate similarity to the culprit..... 92

Figure 9. Example of a lineup from Chapter 4 with heterogeneous fillers of moderate similarity to the culprit..... 93

Figure 10. The lineup suspects for the Chapter 4 experiment..... 95

Figure 11. Identification responses across Chapter 3 and Chapter 4 experiments for lineups containing the culprit..... 122

Figure 12. Identification responses across Chapter 3 and Chapter 4 experiments for lineups that did not contain the culprit..... 123

LIST OF APPENDICES

Appendix 1. Lineup Coding Guide for the Meta-Analysis..... 150

Appendix 2. Research Ethics Board Approval Memo..... 152

Appendix 3. Full Contingency Table for Lab Participants in Chapter 4 (n = 224)..... 153

CHAPTER 1: INTRODUCTION

Police investigators are responsible for tracking down the perpetrators of crime and securing enough evidence to establish their guilt beyond a reasonable doubt. When investigators are successful in this pursuit, justice is served. Although the harm imposed by the criminal cannot be undone, future crime may be prevented and victims may get some satisfaction in knowing that perpetrators have been held accountable for their crimes. Unfortunately, the criminal justice system does not always serve its intended function. Police investigations can be led astray and innocent people are sometimes suspected of crimes they did not commit. Although innocent suspects may eventually be vindicated, the protections that have been crafted into the criminal justice system are not always sufficient. Given that more than 2000 people have been convicted and then exonerated since 1989 in the U.S. alone (Gross, 2013), it would appear that errors in the criminal justice system are relatively commonplace.

Miscarriages of justice have far-reaching consequences. Of course, the most direct consequence of a wrongful conviction is that an innocent person is unjustifiably penalized. The case of Ronald Cotton is a good example. Cotton spent more than 10 years in prison after his conviction for the rape of Jennifer Thompson. DNA evidence was ultimately used to exonerate Cotton and implicate Bobby Poole, who confessed when presented with the DNA evidence.

An often overlooked consequence is that whenever a person is wrongfully convicted, the true culprit does not face justice and is free to continue committing crimes. An unfortunate example comes from the wrongful conviction of Levon Brooks, who was accused of sexually assaulting and murdering a 3-year-old girl (Gross, 2012). Four

months after his conviction, a strikingly similar crime was committed. DNA testing eventually led investigators to interview Justin Johnson, who was in the home of the 3-year-old girl on the day of her death. Upon questioning, Johnson confessed to both crimes.

Wrongful convictions can also have financial implications. Substantial sums of money have been awarded as compensation to those who were convicted and later exonerated. In one such case, Réjean Hinse was convicted of armed robbery and served 5 years in prison. Hinse consistently asserted that at the time of the robbery he was not even in the city where the crime had taken place. After a long legal battle, the Supreme Court of Canada eventually ruled in Hinse's favour and issued an acquittal (*R. v. Hinse*, 1997). As compensation, the federal Government of Canada and provincial Government of Québec were ordered to pay Hinse a combined 13.1 million dollars (Weagant, 2011).

Over the past century, a list of confirmed false convictions has been continually accumulating. In the first systematic review, Borchard (1932) described 65 wrongful conviction cases. Some of these cases came to light following the discovery of new evidence that led to the true culprit. In others, no crime was committed at all. For example, in several cases the defendant was convicted of murdering a person who was later found to be alive and well. Over the years, an increasing number of wrongful conviction cases have been revealed in similar reviews (Brandon & Davies, 1973; Brooks, 1983; Du Cann, 1960; Devlin, 1976; Langdon & Wilson, 2005; Radin, 1964; Rattner, 1988).

False conviction case reviews are important for demonstrating the justice system's vulnerability to error. However, they typically do not address the question of how often

these errors occur. Estimating the prevalence of wrongful convictions is always a dubious endeavor, but that has not stopped legal scholars from trying. One approach to estimating wrongful convictions has been to compare the level of agreement between judges and juries. For example, after a jury conviction the presiding judge can be surveyed about whether the jury reached the correct verdict or not. Spencer (2007) applied probability models to judge-jury agreement rates and estimated that 10% of convictions are made in error. Another approach is to survey the major players in the legal system (police officers, judges, juries) for an estimate about the prevalence of false convictions. In one survey, U.S. officials estimated that between 1% and 3% of all convictions are wrongful (Zalman, Smith, & Kiger, 2008). Given that hundreds of thousands of people are convicted of crimes each year in the United States, these estimates are by no means trivial.

Gross (2013) took yet another approach and estimated the rate of U.S. defendants who have been wrongfully convicted and sentenced to death since 1973. Gross developed his estimates using data from two studies, one that examined exoneration rates in death penalty cases between 1973 and 1984 (Gross & O'Brien, 2008) and another that examined death penalty cases that were subsequently revealed to have been wrongful by DNA evidence (Risinger, 2007). These data led Gross (2013) to estimate that between 2% and 3% of the death penalty convictions eventually lead to exoneration.

Having established that wrongful convictions occur rather frequently, a logical next step is to pinpoint the factors that cause them. Substantial work has been carried out in this regard. Borchard (1932) drew attention to environmental factors that contribute to false convictions, such as public pressure to secure a conviction. Similarly, Radin (1964)

made note of how the justice system is played out like a game with prosecutors trying to "win" a trial rather than establishing the facts of the case and leading the court to the correct verdict.

Another commonly cited, systemic cause of wrongful convictions is unreliable evidence. False confessions are a prime example. A confession is often the primary objective of an interrogation because it provides a powerful piece of incriminating evidence. However, exoneration cases show that people sometimes admit to committing horrendous acts of which they have been falsely accused (Innocence Project, 2014a). Furthermore, psychological studies have illustrated how the situational aspects of an interrogation can make people with certain personality characteristics particularly susceptible to false confessions (e.g., Kassin & Kiechel, 1996).

My dissertation focuses on another type of evidence that has been associated with wrongful convictions: eyewitness identification. Eyewitness misidentification was a contributing factor in the wrongful convictions of Cotton, Brooks, Hinse, and countless others. The specific focus of my research is on the procedures used to select lineup members, which is a decision that must be made whenever a lineup is constructed. It has been more than 30 years since the first experimental demonstration that a lineup's composition can influence identification responses (Lindsay & Wells, 1980). Since then, however, the issue has received only modest attention in empirical studies. Accordingly, my objective is to first consolidate the existing knowledge of lineup composition effects and then to make new empirical and theoretical contributions.

1.1 The Contribution of Eyewitness Misidentification to Wrongful Convictions

The assertion that eyewitness misidentification contributes to wrongful convictions is widely agreed upon. In Borchard's (1932) review, false identification was cited as the single determining factor in almost half of the 65 wrongful conviction cases he reviewed. In numerous subsequent reviews, eyewitness misidentification has been cited as the most common contributing factor in false conviction cases (e.g., Radin, 1964; Rattner, 1988; Wells et al., 1998). Some of the strongest evidence linking eyewitness misidentification and wrongful convictions comes from DNA exonerations.

A variety of circumstances can lead to an exoneration. For example, a person may be exonerated because police misconduct was revealed during a review of the case. In these types of cases, the exoneration is not necessarily indicative of the person's innocence. It merely indicates that the evidence used to secure the conviction should not have been considered sufficient to establish the person's guilt beyond a reasonable doubt. Conversely, DNA-based exonerations typically prove the convicted person could not have committed the crime and in many cases lead investigators to the true culprit.

The Innocence Project (www.innocenceproject.org) is an organization dedicated to helping exonerate those who have been wrongfully convicted. The organization deals strictly with cases in which DNA evidence can be tested. In 1998, a subcommittee of the American Psychology-Law Society published a white paper on eyewitness identification that included an in-depth analysis of the first 40 DNA exoneration cases (e.g., Wells et al., 1998). The analysis showed that eyewitness misidentification contributed to 90% of the cases, supporting what many in the eyewitness identification research community had been suggesting for years; namely that even an honest, well-intentioned eyewitness is capable of making an identification error.

With the hopes of preventing future miscarriages of justice, the authors of the white paper sent a draft version to Janet Reno, who was the U.S. Attorney General at the time. After reading the white paper, Reno instructed the National Institute of Justice (NIJ) to commission a task force on eyewitness identification. The task force agreed on a set of key eyewitness identification reforms and developed a set of national guidelines for the administration of eyewitness lineups (Technical Working Group for Eyewitness Evidence, 1999, 2003). Although these guidelines were not binding, a recent survey suggests that many police agencies adopted the practices outlined in the guide (Police Executive Research Forum, 2013).

Some researchers in the eyewitness identification community have suggested that lineup reforms have not gone far enough (e.g., Levi & Lindsay, 2001). The most recent numbers suggest that eyewitness misidentifications contributed to 72% of the more than 300 DNA exonerations to date (Innocence Project, 2014b). Given that 90% of the first 40 DNA exoneration cases involved eyewitness misidentification, the reduced proportion of cases involving misidentification could be interpreted as evidence that the reformed have helped. However, 72% is still a substantive proportion and eyewitness misidentification remains the most common contributing factor to convictions overturned by DNA evidence.

1.2 System Variables

Experimental research has revealed numerous contributing factors that can lead eyewitnesses to identify people who are innocent. Some of these factors are beyond the control of the criminal justice system. For example, a plethora of studies have demonstrated that recognizing a member of one's own ethnicity is easier than recognizing

a member of another ethnicity (e.g., Malpass & Kravitz, 1969). However, if the ethnicity of the only witness happens to be different from the ethnicity of the perpetrator, the justice system cannot change that. Other factors that contribute to false identification can be controlled by the justice system. Wells (1978) called these 'system variables'.

The set of instructions that are given to an eyewitness before the lineup is presented is one example of a system variable. An early study showed that when witnesses were warned that the culprit might not be in the lineup, they were less likely to misidentify an innocent lineup member (Malpass & Devine, 1981). Subsequent research has shown that such instructions can also reduce the rate of correct identifications (Clark, 2005). Although the costs and the benefits of a pre-lineup admonition can be debated (Clark, 2012), the assertion that lineup instructions are within the justice system's control is unequivocal.

The research reported in this dissertation is focused on lineup composition, which is another system variable. Although lineup construction will likely always require some discretion on the part of those who are choosing the lineup members, investigative policies can be implemented to guide these practices. A key research question is how similar the lineup members should be. This is an issue with tremendous complexity. How much should innocent lineup members resemble the suspect? Should the similarity amongst the innocent lineup members be taken into account? How does the resemblance between the culprit and an innocent suspect contribute to lineup composition effects? These are the types of questions an investigator should be considering when constructing a lineup. At present, the Technical Working Group's training manual on eyewitness evidence is the best resource available to an investigator looking for answers.

1.3 The Technical Working Group for Eyewitness Evidence

Several recommendations for lineup construction are provided in *Eyewitness evidence: A trainer's manual for law enforcement* (Technical Working Group for Eyewitness Evidence, 2003). One recommendation advises having no more than one suspect in a lineup. Imagine a lineup that had no fillers and was composed entirely of suspects. This is precisely what happened after Crystal Mangum claimed to have been raped by members of the Duke lacrosse team (Wells, Cutler, & Hasel, 2009). The case investigators showed Mangum several lineups comprised entirely of Duke lacrosse team members. Procedures such as these are not really tests of memory because any choice translates into a suspect identification and could reflect familiarity with the lineup members in a context outside of the crime that is misattributed to the crime. To enable a test of memory that can be proven false, lineup should have only one person who is under investigation and several other lineups members – fillers – who are known to be innocent.

The training manual also recommends alternating the suspect's position within the lineup. This is important because if the public were to learn that the suspect was always placed in the same position, identifications could be made without relying on memory. Moreover, certain positions could be more likely than others to elicit identifications.

The training manual also provides guidance on selecting fillers. The section on lineup construction begins with the following policy statement: "The investigator should compose the lineup in such a manner that the suspect does not unduly stand out" (p. 32). In other words, dissimilar fillers should be avoided. This principle is further reinforced by advising the selection of fillers who "fit the witness description of the perpetrator" (p. 33) and, in the absence of an adequate description, "who resemble the suspect in significant

features" (p. 33). This guidance would be relatively straightforward were it not for the somewhat contradictory statements that follow. Specifically, investigators are cautioned that "[t]his does not mean that fillers must closely resemble the suspect" (p. 33) and that "making the fillers closely resemble the suspect is not advised because a lineup in which all the people look very similar to one another actually reduces the chances of an accurate identification" (p. 33). In sum, the report emphasizes that fillers should neither be very dissimilar nor very similar to the suspect.

Recommendations concerning suspect-filler similarity are inherently subject to interpretation. This is particularly true, however, when the recommendation is to select fillers who lie somewhere in the middle of the similarity spectrum. What constitutes too similar? What constitutes too dissimilar? The equivocal nature of these questions allows the Technical Working Group's (2003) recommendations to be exploited to one's advantage. For instance, in an effort to avoid making the lineup task too difficult, a police investigator could construct a biased lineup. Conversely, in an effort to defend their clients, defense attorneys could criticize lineups for having fillers who are not similar enough. Clearly, more specific recommendations are needed.

1.4 Filler Selection Procedures

Policy on lineup composition must provide guidance on how fillers should be selected. The literature on this topic has focused on two methods. The traditional procedure is to match fillers to the appearance of the suspect. Note that this procedure may not lead to the selection of fillers who look like culprit. This would only be the case when the police are investigating the right person. If they are investigating the wrong person (i.e., an innocent suspect), the match-to-suspect strategy might lead to the

selection of fillers who look a lot like the suspect, but nothing like the culprit. The alternative strategy is to choose fillers who match a description of the culprit that was given by an eyewitness. When a match-to-description strategy is employed, the guilt of the suspect should have no bearing on how much the fillers resemble the culprit. Rather, the similarity between the fillers and the culprit should be dependent upon the quality of the witness description.

The guide on constructing lineups suggests a preference for the match-to-description procedure (Technical Working Group for Eyewitness Evidence, 2003). The overarching policy in the Technical Working Group's training manual is that the suspect should not stand out from the fillers. The first recommended approach to achieving this objective is to choose fillers who match the witness description of the culprit. The training manual explicitly recommends against selecting fillers who strongly resemble the culprit and matching fillers to the suspect on significant features is only recommended if "there is a limited/inadequate description of the perpetrator provided by the witness, or when the description of the perpetrator differs significantly from the appearance of the suspect" (p. 33, Technical Working Group for Eyewitness Evidence, 2003). However, researchers have recently started to question the basis of this preference (Clark, Rush, & Moreland, 2013). Thus, the remainder of this section examines the relative merits of these two procedures.

1.4.1 The match-to-appearance procedure. The most commonly used method of selecting fillers is to match them to the appearance of the suspect (Wogalter, Malpass, & McQuiston, 2004; Police Executive Research Forum, 2013). When this method is used, the critical concern is determining the appropriate level of similarity between the

fillers and the suspect. Luus and Wells (1991) briefly mentioned the possibility of an optimal-similarity function in which the relation between suspect-filler similarity and lineup diagnosticity – that is, the likelihood that a suspect identification is an identification of the culprit rather than an innocent suspect – would be characterized by an inverted U-shape. They suggested low similarity lineups would be expected to have low diagnosticity because they would induce false identifications and high similarity lineups would also be expected to have low diagnosticity because having lineup members that look too similar to the suspect would impede correct identifications. Accordingly, an optimal lineup would include fillers who lie somewhere in the middle of the similarity spectrum.

Although fillers matched to the suspect were initially thought to protect innocent suspects from false identification (Lindsay & Wells, 1980), researchers have speculated that matching fillers to the suspect's appearance could "backfire" and actually increase the likelihood of innocent suspect misidentification. The first to note this possibility was Navon (1992), who pointed out that suspects and fillers are placed in appearance-matched lineups for different reasons. Unlike the fillers, who are selected because of their match to the suspect, innocent suspects are often in a lineup because of their match to a description of the culprit. Consequently, the innocent suspect would be the lineup member who is most similar to the perpetrator and thus would also be the most likely to be misidentified. Wogalter, Marwitz, and Leonard (1992) proposed an alternative mechanism by which appearance-matched lineups could lead to innocent suspect misidentifications. Wogalter and colleagues' reasoning rests on the notion that because suspects are the origin of suspect-matched lineups, they will be more similar to the fillers

than any of the fillers are to each other. Therefore, innocent suspects might be chosen because they represent the central tendency of the lineup. Consistent with these predictions, an innocent suspect was chosen from one appearance-matched lineup at a higher rate than all the fillers combined (Clark & Tunnicliff, 2001).

1.4.2 The match-to-description procedure. Luus and Wells (1991) recommended using fillers who fit the eyewitness description of the culprit, but who also possess additional features that differ from those of the culprit. They hypothesized that matching on features in the description would protect innocent suspects from false identification and that allowing features not mentioned in the description to vary would facilitate recognition. Wells, Rydell, and Seelau (1993) provided convincing evidence in support of this claim. Compared with lineups composed of fillers who were low in similarity to the culprit, matching fillers to the witness description resulted in a 30% reduction in false identifications of the innocent suspect and had no effect on correct identifications of the culprit. Moreover, compared with lineups composed of fillers high in similarity to the culprit, matching fillers to the description resulted in a 45% increase in correct identifications and an equivalent rate of false identifications. Thus, lineups with description-matched fillers offered protection to innocent suspects without impeding culprit identifications.

The match-to-description advantage observed by Wells et al. (1993) provided empirical support for the theoretical framework proposed by Luus and Wells (1991). When eyewitnesses describe a culprit, they draw on recall memory. The function of a lineup is to give the witness an opportunity to provide new, recognition-based information. When the suspect matches the description and the fillers do not, witnesses

need not rely on recognition memory because the fillers can be discounted based on the incongruence between the appearance of the fillers and what was recalled of the culprit's appearance. If the suspect happens to be innocent, the witness will be prone to false identification because the suspect will stand out from the fillers. Conversely, when fillers match the description, witnesses are forced to rely on recognition because all lineup members correspond with their recall. This explains the match-to-description procedure's protection of innocent suspects from false identification.

The advantage of matching to the witness description over matching to the suspect's appearance is explained by the complementary concepts of propitious heterogeneity and gratuitous similarity (Wells, 1993). Lineups with fillers who are matched only to the features in the description promote propitious heterogeneity, which is the idea that variations in the features not mentioned in the eyewitness description aid the process of recognition. Lineups with fillers who are matched to the appearance of the suspect promote gratuitous similarity, which is the idea that matching fillers to features of the suspect that are above and beyond those provided in the witness description eliminates important differences among the lineup members that are needed for recognition to operate effectively.

Although the match-to-description procedure showed early promise, it evinced little or no advantage over the match-to-appearance procedure in subsequent research. In two experiments, relative to appearance-matched lineups, description-matched lineups produced only nonsignificant increases in culprit identifications (Juslin, Olsson, & Winman, 1996; Lindsay, Martin, & Webber, 1994). Moreover, in one of those experiments (Lindsay et al., 1994), the innocent suspect misidentification rate was

significantly higher in description-matched lineups than in appearance-matched lineups, an effect the authors attributed to witness descriptions that were too vague. In more recent experiments, filler selection strategy had no effect on culprit or innocent suspect choice rates (Darling, Valentine, & Memon, 2008; Tunnicliff & Clark, 2000). To explain the absence of a match-to-description benefit in culprit identifications, Tunnicliff and Clark suggested the fillers in their appearance-matched lineups might not have resembled the suspect to the same degree as those in the experiment by Wells et al. (1993). These findings suggest matching to the witness description is not always the most advantageous filler selection procedure.

Furthermore, the match-to-description procedure is not always a viable option. Luus and Wells (1991) outlined three situations in which the fillers should not be matched to the eyewitness description: (1) When the description does not correspond with the appearance of the suspect; (2) When the description is so specific that it would be impossible to find fillers who match it; and (3) When multiple eyewitnesses to the same event report descriptions that contradict one another.

Another problem with the match-to-description procedure is that eyewitnesses often provide generic face descriptions, which can result in a lineup with fillers who fit the description yet look nothing like the suspect (Koehnken, Malpass, & Wogalter, 1996). Even more problematic is the situation in which eyewitnesses provide no information about the culprit's face. In one study (Lindsay et al., 1994) fewer than 10% of witnesses mentioned facial features in their description of a confederate to whom they had spoken for 3 minutes. Another problem with the match-to-description procedure is that the eyewitness description could be inaccurate (Meissner, Sporer, & Schooler, 2007; Wells et

al., 1998). These issues could explain why only 9% of respondents in a survey of police investigators indicated using witness descriptions as a basis for selecting fillers (Wogalter et al., 2004).

1.5 Defining Lineup Member Similarity

Regardless of whether a match-to-suspect or a match-to-description policy is in place, police investigators will inevitably need advice on how similar the lineup members should be. The match-to-description procedure might work well for cases with detailed witness descriptions, but these cases appear to be the exception rather than the rule and a match-to-suspect strategy is generally recommended when the match-to-description option is not practical. Moreover, the default option for the vast majority of police agencies is to match fillers to the suspect's appearance (Home Office, 2011; Police Executive Research Forum, 2013). Therefore, there is a need to establish the optimal level of lineup member similarity.

A first step in this direction is to settle upon a definition of just what is meant by lineup member similarity. The resemblance between two persons is best conceptualized within the framework of a similarity continuum. However, eyewitness researchers have traditionally conceptualized similarity in categorical terms. Police typically do not have the resources to generate fillers on a continuously defined scale of similarity, so the categorical approach is necessary for formulating lineup construction recommendations that can be practically implemented.

In this introductory chapter, I distinguish between three categories of lineups: biased, moderate similarity, and high similarity. Biased lineups are those which contain fillers who are highly dissimilar to the suspect or fillers who, for some reason, make the

suspect stand out. Moderate similarity lineups are those which contain fillers who match a general description of the target, but have not been closely matched to the suspect's appearance. High similarity lineups contain fillers who have been closely matched to the appearance of the suspect/target. Note that the distinction between categories should be considered in relative terms. For instance, although researchers have demonstrated greater similarity in the lineups categorized as 'high similarity' than in lineups categorized as 'moderate similarity', the lineups with high similarity fillers have typically yielded similarity ratings near the midpoint of any given scale. Thus, for a lineup to be categorized as high similarity, it need not score high on an absolute scale of similarity.

1.6 How Similar is "Too Similar"?

The upper bound of lineup similarity is a critical issue that has yet to be resolved. Investigators have been advised to ensure fillers do not bear too strong a resemblance to the lineup member suspected of the crime. The concern is that fillers who closely match the suspect's appearance will essentially result in a lineup of "clones" and make correct identifications too difficult (Wells & Luus, 1990). A multitude of sources have endorsed the not-too-similar recommendation (e.g., Brewer & Palmer, 2010; Malpass, Tredoux, & McQuiston-Surrett, 2007; Wells et al., 1998), so one might expect a relatively large database of rigorous empirical research demonstrating a negative effect of fillers who too strongly resemble the suspect. However, in the literature I review below I find only sparse empirical support for this recommendation.

At some level, less similar fillers make correct identifications easier. Numerous empirical investigations reveal that correct identification rates are higher for lineups with very dissimilar fillers relative to lineups with similar fillers (e.g., Carlson, Gronlund, &

Clark, 2008; Gronlund, Carlson, Dailey, & Goodsell, 2009; Lindsay et al., 1994). In other words, witnesses are better able to identify a target from a biased lineup than from a fair lineup; however, this would likely also be true of nonwitnesses who are given nothing more than the target's description (Doob & Kirshenbaum, 1973). Given that the extant literature has consistently shown that correct identifications are less likely to occur in fair relative to biased lineups, I focus my review on the more contentious distinction between lineups containing fillers of moderate and high similarity to the target.

Although a relatively substantial literature on lineup member similarity has been established, the wide variety of methods used to manipulate similarity has made a parsimonious interpretation of the findings elusive. Most commonly, researchers have manipulated similarity through filler selection strategies or by using similarity ratings to guide lineup construction; however, alternative methods can also be found in the literature. Given the potential influence of the type of similarity manipulation on the pattern of identification responses, I have organized the literature review according to the method by which similarity was manipulated. In addition to my primary interest in the effect of high similarity fillers on correct identification of a guilty suspect, I also review their effect on false identification of an innocent suspect.

1.6.1 Filler search procedure: Match-to-description vs. match-to-appearance. Wells et al. (1993) conducted the first empirical comparison between the match-to-

Table 1

Correct and false identification (ID) rates for Description-Matched (DM) and Appearance-Matched (AM) lineups

Experiment	Correct ID Rate		*z*	*p*	Odds Ratio, 95% CIs		
	DM	AM			OR	Lower	Upper
Wells et al., 1993	.67 (.07)	.21 (.06)	4.73	.001	7.33	2.76	19.48
Lindsay et al., 1994	.79 (.08)	.66 (.09)	1.10	.271	2.01	0.62	6.57
Juslin et al., 1996	.52 (.05)	.44 (.05)	1.10	.268	1.40	0.79	2.47
Tunnicliff & Clark, 2000, 1	.53 (.09)	.53 (.09)	0.00	1.000	1.00	0.38	2.67
Tunnicliff & Clark, 2000, 2[1]	.31 (.07)	.33 (.07)	-	-	-	-	-
Darling et al., 2008	.45 (.07)	.49 (.07)	0.40	.692	0.86	0.39	1.90
	False ID Rate						
	DM	AM					
Wells et al., 1993	.12 (.05)	.12 (.05)	0.00	1.000	1.00	0.27	3.75
Lindsay et al., 1994	.15 (.04)	.04 (.02)	2.11	.035	3.92	1.03	14.93
Juslin et al., 1996	.09 (.05)	.09 (.05)	0.00	1.000	1.00	0.19	5.37
Tunnicliff & Clark, 2000, 1	.13 (.06)	.03 (.03)	1.40	.161	4.43	0.47	42.02
Tunnicliff & Clark, 2000, 2	.08 (.04)	.19 (.06)	-	-	-	-	-
Darling et al., 2008	.05 (.03)	.04 (.02)	0.29	.769	1.34	0.18	9.92

Note: Standard errors for ID rates are in parentheses.

[1] In Tunnicliff and Clark's second experiment, a within-subjects design was employed. This represents a violation of the independence assumption for the two-proportions *z* test. Thus, *z* and *p* were not computed.

description and match-to-appearance procedures. As mentioned in a preceding section, the two procedures did not differ in false identifications and the correct identification rate was substantially reduced in appearance-matched lineups. However, subsequent comparisons of the two procedures revealed no significant differences in correct identifications (see Table 1) and in one experiment description-matched fillers increased false identifications (Lindsay et al., 1994). The overview of all comparisons between the match-to-description and match-to-appearance procedures in Table 1 shows the effect on correct identifications has diminished since the time the effect was first reported in the literature (i.e., a decline effect; Schooler, 2011). Furthermore, meta-analytic comparisons of the two procedures suggest that whether fillers are selected by matching to a description or matching to the suspect's appearance has no effect on the extent to which suspect identifications are diagnostic of guilt or of innocence (Clark & Godfrey, 2009; Clark, Howell, & Davey, 2008).

1.6.2 Subjective similarity ratings. In many of the comparisons between appearance- and description-matched lineups, researchers obtained ratings to measure the similarity between the suspect and the fillers. When such ratings have been obtained, suspect-filler similarity has been consistently higher for appearance-matched lineups than for description-matched lineups. Thus, the differences associated with similarity ratings in these studies can be inferred by examining Table 1 and substituting "moderate similarity" with description-matched and "high similarity" with appearance-matched.

In contrast to indirectly manipulating similarity through filler selection procedures, Brewer and Wells (2006) used subjective ratings to manipulate suspect-filler similarity. Brewer and Wells administered two lineup identification tasks, one for a thief

and one for a waiter. In both lineups, some fillers from a high similarity lineup were replaced with fillers who matched a description of the target, but were rated to be of lower similarity to the target. This similarity manipulation yielded mixed results. For the thief lineup, the correct identification rate was higher for the high similarity lineup, .40 [CI^{95} = .34, .46], than for the moderate similarity lineup, .34 [CI^{95} = .29, .40]. The opposite was true for the waiter lineup, which yielded a higher correct identification rate for the moderate similarity lineup, .66 [CI^{95} = .61, .72], than for the high similarity lineup, .57 [CI^{95} = .51, .63]. False identifications were comparable across similarity conditions. Given the small and contradictory effects on correct identifications and null effects on false identifications, this manipulation of similarity seems to have had minimal impact on identifications.

Sauer, Brewer, and Weber (2008) used subjective ratings to manipulate similarity in lineups for male and female targets. All lineup members were consistent with a general description of the target person, but similarity ratings were higher for the high similarity lineup members. In one condition, participants were instructed to respond to the lineup with a traditional identification response (binary condition). For these participants, increasing similarity led to small, nonsignificant reductions in correct identification rates. These trends were also observed for participants in another condition who were instructed to rate their confidence that a lineup member was the target rather than actually picking a lineup member. When collapsed across conditions, increasing similarity yielded a small but significant decrease in accuracy for target-present lineups and had virtually no effect on accuracy for target-absent lineups.

More recently, three colleagues and I used subjective ratings to manipulate similarity in lineups for child witnesses (Fitzgerald, Whiting, Therrien, & Price, 2014). When lineup member similarity was increased in Experiment 1, children were significantly less likely to identify the culprit (moderate = .23 [CI^{95} = .09, .37] vs. high = .07 [CI^{95} = -.02, .16]) and also significantly less likely to identify an innocent suspect (moderate = .30 [CI^{95} = .16, .44] vs. high = .04 [CI^{95} = -.03, .11]). In Experiment 2, similarity was manipulated in lineups for child and adult witnesses. For children, the pattern observed in Experiment 1 was replicated in Experiment 2. That is, relative to children in the moderate similarity condition, children in the high similarity condition were significantly less likely to identify the culprit (moderate = .74 [CI^{95} = .59, .89] vs. high = .48 [CI^{95} = .32, .64]) or misidentify the innocent suspect (moderate = .33 [CI^{95} = .17, .49] vs. high = .14 [CI^{95} = .02, .26]). Although the high similarity fillers also significantly reduced the adults' innocent suspect misidentification rate (moderate = .28 [CI^{95} = .14, .42] vs. high = .15 [CI^{95} = .04, .26]), the similarity manipulation had no effect on the adults' correct identification rate (moderate = .76 [CI^{95} = .62, .90] vs. high = .74 [CI^{95} = .59, .89]). Thus, although child witnesses seem to have difficulty with high similarity lineups, the high similarity fillers only had positive effects on the adult witnesses.

1.6.3 Tredoux's *E'*. Effective size represents the number of lineup members who are plausible alternatives to the suspect (Malpass, 1981). Effective size calculations require data from mock witnesses who are provided a description of the culprit and asked to choose the lineup member who best fits the description (Doob & Kirshenbaum, 1973). Effective size is estimated as the number of lineup members chosen at a rate that differs

from chance expectancy. For a 6-member lineup, an effective size of 3 has been proposed to be fair (Brigham, Ready, & Spier, 1990). Tredoux (1998) describes a similar measure, *E'*, which retains the favourable properties of effective size and also utilizes a known sampling distribution. Both effective size and *E'* are positively associated with suspect-filler similarity (Brigham & Brandt, 1992; Brigham et al., 1990; Tredoux, 2002).

Carlson et al. (2008) used photos from a mugshot database to construct "intermediate" (*E'* = 2.56) and "fair" (*E'* = 4.05) lineups. Intermediate lineups contained a mix of fillers who did and did not match the culprit's description. All fillers matched the description in fair lineups. The fair lineups led to a consistently lower rate of innocent suspect misidentifications and although the lineup manipulation did not affect overall correct identification rates (intermediate = .33 [CI^{95} = .23, .43]; fair = .36 [CI^{95} = .26, .46]), an interesting pattern emerged when the procedure used to present the lineup members was taken into account.

Lineup members can be presented to eyewitnesses all at once (simultaneously) or one at a time (sequentially). Simultaneous lineups have been criticized for allowing witnesses to adopt a relative judgment strategy (Wells, 1984). That is, witnesses viewing simultaneous lineups might be tempted to choose the person who looks most similar to the perpetrator in comparison to the other lineup members. Although using relative judgments should tend to lead to correct identifications from culprit-present lineups, this strategy is clearly problematic when the culprit is absent. Lindsay and Wells (1985) consequently developed the sequential lineup procedure to encourage witnesses to compare each individual lineup member to their memory of the culprit (i.e., an absolute judgment) rather than to other lineup members.

In the sequential lineup, each photo is shown individually and the witness is required to decide whether the photo is or is not the culprit before proceeding to the next one. Lindsay and Wells (1985) recommend showing each photo only once and not allowing witnesses to go back and look at previously viewed photos. In addition, to prevent the tendency to choose someone near the end of the array, they suggest withholding the number of photos that will be viewed from the witness (commonly referred to as "backloading"). When Lindsay and Wells directly compared the two lineup presentation formats, they observed a lower false identification rate on the sequential lineup (17% [CI^{95} = .07, .27]) than on the simultaneous lineup (.43 [CI^{95} = .30, .56]). Recent meta-analyses (Steblay, Dysart, Fulero, & Lindsay, 2001; Steblay, Dysart, & Wells, 2011) have supported the view that sequential lineups offer an effective safeguard against false identification. However, correct identification rates also tend to be lower in sequential compared to simultaneous lineups, leading some researchers to suggest the sequential procedure encourages witness to adopt a more conservative decision criterion (Flowe & Bessemer, 2011; Flowe & Ebbeson, 2007; Meissner, Tredoux, Parker, & MacLin, 2005). In other words, witnesses might be less willing to choose from sequential lineups than from simultaneous lineups.

Carlson et al. (2008) hypothesized that suspect-filler similarity could moderate lineup presentation effects. Specifically, they predicted that a sequential advantage would only emerge when an innocent suspect resembles the culprit and stands out from other lineup members who do not (i.e., when the lineup is biased), arguing the innocent suspect is much less likely to stand out in a sequential lineup because comparisons among lineup members are more difficult. Carlson et al. conducted two experiments to test their

hypothesis. In the first experiment, simultaneous and sequential lineups that only contained fillers highly similar to the innocent suspect were compared. In support of their hypothesis that the sequential advantage would only emerge when lineups were biased, the lineups with highly similar fillers led to equivalent rates of innocent suspect misidentifications between the simultaneous and sequential conditions. Furthermore, a simultaneous advantage was found for correct identifications from culprit-present lineups; however, Carlson et al. noted that this finding was likely a consequence of matching the fillers to the innocent suspect's appearance – even in culprit-present lineups – which would be expected to make the culprit stand out when the lineup members are presented simultaneously.

In their second experiment, Carlson et al. (2008) directly manipulated lineup fairness by constructing lineups containing fillers of low, moderate, and high similarity to the suspect. Again, compared to sequential lineups, simultaneous lineups produced a higher rate of correct identification from culprit-present lineups, albeit only for those containing low similarity fillers (i.e., biased lineups). In addition, in culprit-absent lineups the false identification rate was lower when presented sequentially than when presented simultaneously; however, similar to the simultaneous advantage for culprit-present lineups, the sequential advantage was only found in the case of biased lineups.

Although Carlson and colleagues' (2008) findings suggest a potential interaction between *E'* manipulations and lineup presentation format, the pattern did not replicate in subsequent research. Gronlund et al. (2009) used a combination of description-matching and appearance-matching to construct "medium" and "fair" lineups that differed in *E'*. Large pools of potential fillers for the two lineup conditions were obtained by instructing

research assistants to search an online mugshot database using different criteria. In the medium condition research assistants searched for fillers who matched the culprit on five descriptors, and then Gronlund selected five fillers from this pool who were neither very dissimilar nor very similar to the culprit's appearance, which produced *E'* values ranging from 2.33 to 3.15. In the fair condition different research assistants searched for fillers who matched the culprit on seven descriptors, and then Gronlund selected the fillers who best resembled the culprit, which produced *E'* values ranging from 3.75 to 4.51. This manipulation led to comparable innocent suspect misidentification rates and, contrary to the premise underlying the not-too-similar recommendation, the correct identification rate for fair lineups (.36 [CI^{95} = .31, .41]) was higher than for medium lineups (.30 [CI^{95} = .26, .35]), a trend that was consistent for simultaneous and sequential lineups.

1.6.4 Euclidean distance. Tredoux (2002) used a technique to manipulate similarity that is radically different from any of the previously described methods. Building on Valentine's (1991) euclidean multidimensional 'face space' framework, Tredoux used principal component analysis to systematically identify the euclidean distance (degree of similarity) between two faces. After establishing that this approach was comparable to perceptual similarity ratings, Tredoux (2002) conducted a lineup task comparing fillers who were very close in euclidean distance with fillers who were only moderately close. Tredoux did not specify separate accuracy rates for culprit-present and culprit-absent lineups; however, the similarity manipulation did not yield a significant difference in overall accuracy (moderate = .44 [CI^{95} = .32, .57]; high = .40 [CI^{95} = .28, .52]), and similarity did not interact with target-presence or lineup presentation.

1.6.5 Computer-generated faces. In contrast to each of previously-described experiments, which used photographs of real faces, Flowe and Ebbeson (2007) used a software program (FACES) to generate simulated faces that differed in similarity to a simulated target face. In the "random similarity" condition, fillers were pseudo-randomly selected from a database of 1000 faces (eye colour was required to match; all other facial features were required to mismatch). In the "matched similarity" condition, the lineup faces were matched on one facial feature. In two experiments, this manipulation yielded no significant differences in correct identifications. Consistent with the trend observed by Carlson et al. (2008), in Experiment 1 the correct identification rate for simultaneous presentation was higher for random similarity lineups (.52 [CI^{95} = .37, .67]) than for matched similarity lineups (.45 [CI^{95} = .30, .60]), but the correct identification rate for sequential presentation was lower for random similarity lineups (.33 [CI^{95} = .19, .47]) than for matched similarity lineups (.37 [CI^{95} = .23, .51]). The false identification rate for a "look-a-like" in target-absent lineups was approximately 10% higher for random than matched similarity lineups, an effect that was uninfluenced by lineup presentation. Flowe and Ebbeson did not report the correct identification rates separately for the similarity conditions in Experiment 2, only noting the absence of a significant difference.

1.6.6 Summary of empirical findings. Although concerns that highly similar fillers would make correct identifications too onerous to appear valid for child witnesses, similarity's effect on adults' ability to correctly identify a culprit was much less clear. In one experiment, highly similar fillers were associated with a dramatic reduction in correct identifications (Wells et al., 1993). In the other studies, however, the effect of highly similar fillers on correct identifications has been equivocal. In some cases highly similar

fillers were associated with a small decrease in correct identifications, whereas in other cases highly similar fillers were associated with a small increase in correct identifications.

Given the paucity of empirical research demonstrating the utility of the not-too-similar recommendation, one might ask why eyewitness scientists have accepted it so readily. On other lineup identification issues, researchers have required substantial evidence before uniformly endorsing a procedure. For example, the question of whether lineup members should be presented simultaneously or sequentially has been the subject of lively debate (e.g., Clark, 2012; Lindsay, Mansour, Beaudry, Leach, & Bertrand, 2009; McQuiston-Surrett, Malpass, & Tredoux, 2006; Steblay, Dysart, & Wells, 2011).

I suspect the not-too-similar recommendation has not been subjected to the same degree of rigor as other proposed lineup reforms because the similarity-difficulty relation is both intuitive and well-established in other domains within cognitive psychology. An eyewitness lineup is ultimately a multiple-choice recognition test and cognitive psychologists have used strong language to describe similarity's effect on such tests, noting that "we can make any recognition test as difficult as we want simply by making distractors similar to the correct alternative" (Glass, Holyoak, & Santa, 1979, p. 65). Psychometricians have also emphasized the association between similar distractors and item difficulty (Guttman & Schlesinger, 1967; Smith & Smith, 1988). Even in the eyewitness identification domain, correct identifications are more likely to occur when fillers are very dissimilar than when they are highly similar (e.g., Gronlund et al., 2009). If the relation between similarity and difficulty is linear, the idea that witnesses would have a better chance of identifying a target accompanied by moderate relative to high

similarity fillers seems a logical extension of this principle. However, the evidence from empirical investigations is remarkably unconvincing.

1.7 Overview of Dissertation Research

The body of the dissertation is divided into three chapters. In Chapter 2, I report a meta-analysis of experimental manipulations of suspect-filler similarity that were published between 1980 and 2012. The meta-analysis, which was performed in 2012, summarizes the identification decisions of more than 6500 participant-witnesses who made identifications from lineups categorized as biased, moderate similarity, and high similarity. In the following two chapters, I report the results of two experimental studies. In the Chapter 3 experiment, morphing software was used to systematically manipulate the degree to which fillers resemble the suspect. The key advantage of this approach was that the fillers in all of the conditions had a common origin. That is, rather than having faces of the fillers in lower similarity lineups belong to different people than the faces of the fillers in the higher similarity lineups, I used the same individuals to construct the lineups in all the similarity conditions and achieved the manipulation by morphing the individuals with the suspect to varying degrees. This novel method of manipulating similarity yielded fillers who were very similar to the target person, which was found to have a substantial influence on how participants responded to the identification task. In Chapter 4, I report a second experiment that utilized the morphing technique to explore two potential sources of bias: suspect-filler similarity and filler homogeneity similarity. Chapter 4 takes a more theoretical approach than in Chapters 2 and 3, and provides evidence of a complex interplay between lineup composition and lineup presentation. The final chapter provides a general discussion of the body of the dissertation.

CHAPTER 2

A Meta-Analysis of Lineup Member Similarity Effects

The meta-analysis reported in this chapter was conducted to provide a comprehensive overview of the existing research on suspect-filler similarity. A meta-analytic approach is helpful because the results of one study may, for example, be contingent upon the materials employed and might not generalize to all similarity manipulations (for a discussion of the need for stimulus sampling, see Wells & Windschitl, 1999). Although the number of studies with direct manipulations of suspect-filler similarity is relatively small, there have been several instances in which similarity has been indirectly manipulated (e.g., by adopting different filler selection strategies). This meta-analytic review summarizes studies that often differ in their stated intentions yet each contain a common element – a comparison between lineups that differ in suspect-filler similarity. Synthesizing data collected from different sources allows for a better understanding of how suspect-filler similarity affects lineup choices across different identification conditions. A comprehensive understanding is especially desirable in the case of suspect-filler similarity manipulations because researchers in this domain have yet to implement a standard method of objectively determining the similarity between two persons (Tredoux, 2002).

When examining suspect-filler similarity effects, one could contrast the rate at which culprits and innocent suspects are chosen between lineups with fillers of high versus low resemblance to the suspect. Essentially, this was the approach used by Clark and Godfrey (2009) in their broad review of the eyewitness literature. Based on a summary of seven studies, Clark and Godfrey concluded that both culprit and innocent suspect choices were more prevalent when suspect-filler similarity was low than when it

was high. The increase in innocent suspect misidentifications was greater than the increase in culprit identifications, suggesting that a low similarity lineup is more likely to lead to innocent suspect misidentifications than it is to facilitate culprit identifications. Clark and Godfrey noted, however, that three of the seven studies summarized contained a manipulation of clothing bias rather than a manipulation of facial appearance.

Clark and Godfrey (2009) also did not comment on how suspect-filler similarity affected filler selections and lineup rejections, focusing instead on suspect identifications. Filler selections are known errors, so they are less concerning than innocent suspect misidentifications (Wells & Lindsay, 1980). However, witnesses may be required to attempt lineup identifications on multiple occasions (Behrman & Davey, 2001); if a filler had been misidentified on a previous occasion, the credibility of that witness could be jeopardized (Tunnicliff & Clark, 2000). Furthermore, filler identifications have been shown to have diagnostic value; in fact, when the *a priori* likelihood of the suspect's guilt is relatively high, filler identifications may be more informative of a suspect's innocence than suspect identifications are of a suspect's guilt (Wells & Olson, 2002).

In this chapter, I use meta-analytic techniques to investigate the effect of suspect-filler similarity on all identification responses (suspect identifications, filler identifications, and lineup rejections). In addition, rather than strictly comparing high and low similarity lineups, both of which were recommended against by the Technical Working Group for Eyewitness Evidence (2003), I compared three levels of suspect-filler similarity: biased, moderate, and high. Note that these labels correspond to relative differences in similarity, rather than to objectively defined categories. Compared to moderate similarity lineups, biased similarity lineups were expected to increase both

correct identifications and false identifications. Conversely, high similarity lineups were expected to decrease correct identifications and false identifications.

2.1 Method

2.1.1 Literature search. To begin, I searched the PsycInfo and Web of Science databases for articles containing various combinations of the following search terms: eyewitness, identification, lineup, foil, filler, distractor, similarity, match, appearance, description, construction, fairness, composition. When a relevant article was identified, its reference list as well as the list of articles in which it had been cited were examined. Finally, *Google* and *Google Scholar* were searched to account for any articles that were not in the aforementioned databases. The search ended in April, 2012, with 17 independent studies that met the inclusion criteria, providing data from 6650 participants. Publication dates for the articles ranged between 1980 and 2011.

2.1.2 Inclusion criteria. To be included in the meta-analysis, the study needed to be an investigation of event memory (live events or video events) that included a comparison between two or more lineups that differed in suspect-filler similarity. Several methods of checking similarity manipulations were present in the literature. One method is to obtain similarity ratings between the suspect and each of the fillers. That is, raters observed two faces (the suspect and a filler) and indicated their judgment of similarity using a Likert scale. The number of points on the scale varied widely from study to study, with some as low as 4 points and others as high as 1001 points. Similarity judgments were typically, but not always (e.g., Juslin et al., 1996), conducted using participants who did not provide data for the study itself. Another method of checking similarity involves conducting mock witness tests in which a set of judges view a lineup and identify the

person who best fits an eyewitness description of the suspect (Doob & Kirshenbaum, 1973). The data obtained from mock witness tests can then be used to calculate effective size scores (Malpass, 1981; Tredoux, 1998), which are estimates of the number of lineup members who fit the description sufficiently to draw choices away from the suspect. Lineups with high effective size scores are judged to have higher suspect-filler similarity than lineups with low effective size scores (Brigham & Brandt, 1992; Brigham et al., 1990).

2.1.3 Exclusion criteria. Studies that manipulated lineup similarity within the context of face recognition paradigms (e.g., Flowe & Ebbeson, 2007) were excluded because they were not considered to adequately correspond with the experience of an eyewitness. Clothing bias manipulations were also excluded because they were considered fundamentally different from facial similarity manipulations. Furthermore, within-subject designs were excluded because I could not be certain that a repeated-measures design would be assessing the same effect as an independent-groups design (Morris & DeShon, 2002). In some studies, identifications were made from two lineups. For example, in one study a target-absent lineup was followed by a target-present lineup (Read, Tollestrup, Hammersley, McFadzen, & Christensen, 1990). These studies were dealt with by only including data from the first lineup that was shown. Following the procedure employed by previous meta-analysts in the psychology and law domain (Deffenbacher, Bornstein, Penrod, & McGorty, 2004), unpublished studies were excluded to accommodate the legal system's preference for published research (e.g., *Daubert v. Merrell Dow Pharmaceuticals*, 1993).

2.1.4. Assignment of study weights. Due to the relative nature of similarity judgments, variability in true effect sizes was assumed and study weights were assigned using the random-effects model (Hedges, 1992). In contrast to the fixed-effect model, which only takes the within-study variance into account (i.e., sampling error) and assigns substantially greater weight to larger studies than to smaller studies, in the random-effects model the weights assigned to smaller and larger studies are more evenly distributed because both the within-study variance and the between-study variance are taken into account (Borenstein, Hedges, Higgins, & Rothstein, 2010).

2.1.5 Coding of lineups. To account for the varying degrees of suspect-filler similarity, lineups were categorized as biased, moderate similarity, or high similarity. I developed a coding guide that provided a general description of the fillers, a range of mean similarity ratings, and a range of effective size scores for each lineup similarity category (descriptions of the categories as well as details about the coding guide are provided in Appendix A). The guide was provided to two coders, who independently categorized the lineups using information provided in the Methods section of each article. Cohen's kappa indicated the initial level of interrater reliability was acceptable, $\kappa = .87$. All coding discrepancies were resolved to consensus through discussion between the two coders.

2.1.6 Effect size. Typically, one of three choices is possible in lineup identification tasks: suspect identifications, filler identifications, or lineup rejections[2]. For the purpose of the present research, the outcomes of each of these possibilities were treated as binary (e.g., the suspect was identified or the suspect was not identified) and

[2] Researchers occasionally provided a "not sure" option. For ease of comparison among studies that did or did not provide this option, all "not sure" outcomes were treated as lineup rejections.

analyzed in separate meta-analyses. This approach involved a relatively high number of tests, and the increased likelihood of Type I errors should be noted. However, by analyzing each outcome, I was able to determine whether changes in the rates of suspect identifications corresponded with changes in filler identifications or changes in lineup rejections.

When dealing with binary data, meta-analysts have the option of computing an odds ratio, a risk ratio, or a risk difference. Of the three measures, odds ratio has the best mathematical properties. For example, risk ratio and risk difference are not typically able to assume their full range of values, but odds ratio is capable of assuming its full range of values (Cooper, Hedges, & Valentine, 2009). However, odds ratio is also the least intuitive of the three measures (Deeks, 2002). Risk difference is based on raw units and can be easily interpreted by both researchers and professionals who are unfamiliar with statistical techniques. For example, if rates of false identification were 25% for biased lineups and 35% for moderate similarity lineups, then the risk difference would be 10%. A meta-analysis of previous meta-analyses indicated similar conclusions were reached regardless of whether odds ratio or risk difference was calculated (Engels, Schmid, Terrin, Olkin, & Lau, 2000). Nevertheless, I calculated both measures to ensure the results would be both accurate and accessible. The statistical significance of the differences were assessed with *z* tests.

2.1.7 Diagnosticity. Diagnosticity ratios provide a measure of the probative value of a lineup procedure by indicating how much more likely a suspect identification is to correspond to the perpetrator, as opposed to an innocent person. Diagnosticity is calculated as the ratio of culprit identifications from culprit-present lineups to innocent

suspect identifications from culprit-absent lineups (Wells & Lindsay, 1980). For example, if a guilty suspect (culprit) is chosen by 60% of witnesses from a culprit-present lineup and an innocent suspect is chosen by 20% of witnesses from a culprit-absent lineup, the diagnosticity ratio would be 3.0 (60 ÷ 20), indicating that a guilty suspect is 3 times more likely to be chosen than an innocent suspect. I compared these ratios among lineups that differed in suspect-filler similarity to evaluate the extent to which lineup fillers affected diagnosticity. For diagnosticity to be compared, both similarity and culprit presence needed to be manipulated within the study. Because not all studies that met the inclusion criteria had both culprit-present and culprit-absent conditions, the diagnosticity statistics were based on a subset of the studies included in the meta-analysis. Furthermore, not all studies included biased, moderate, and high similarity lineups. Consequently, if a study included only a comparison between biased and moderate similarity lineups, data from the moderate similarity lineup from that study would not affect the diagnosticity ratios that were calculated for the comparison between moderate and high similarity lineups. This approach was important because it allowed for causal conclusions regarding any differences in diagnosticity between two lineup types.

2.1.8 Analysis of culprit-present and culprit-absent lineups. Suspect-filler similarity has been manipulated in both culprit-present and culprit-absent lineups. Whether the culprit is present or absent in a lineup necessarily determines the outcome that is considered accurate. For culprit-present lineups, the correct decision is to identify the suspect. For culprit-absent lineups, the correct decision is to reject the lineup. Given the fundamental difference between a lineup that contains a culprit and one that does not (Wells & Penrod, 2011), separate meta-analyses were conducted for culprit-present and

culprit-absent lineups instead of including culprit-presence as a subgroup variable within a larger meta-analysis.

In the case of culprit-present lineups, suspect-filler similarity always refers to the same thing: the degree to which the fillers resemble the culprit. For culprit-absent lineups, suspect-filler similarity could again refer to the degree of similarity between the culprit and the fillers or it could refer to the degree of similarity between the innocent suspect and the fillers. This discrepancy is a consequence of some researchers striving for experimental control by using the same fillers in culprit-present and culprit-absent lineups and other researchers striving for ecological validity by following the procedures that would be used when lineups are constructed by law enforcement personnel (Clark & Tunnicliff, 2001). For the present purposes, I made no distinction between culprit-absent lineups with fillers who were matched to the perpetrator and culprit-absent lineups with fillers who were matched to the innocent suspect, instead focusing on the similarity between lineups irrespective of the method used to manipulate it. Although it would be ideal to analyze the two variations of suspect-filler similarity on culprit-absent lineups separately, the limited number of studies examining suspect-filler similarity made this approach undesirable.

2.1.9 Analysis of lineup presentation subgroups. Given the different strategies that are hypothesized to play a role in evaluating simultaneous (relative judgment) and sequential (absolute judgment) lineups, lineup presentation could have an influence on suspect-filler similarity effects. Carlson et al. (2008) observed an influence of suspect-filler similarity on the effect of simultaneous versus sequential lineup presentation. In the present research, I took a different approach and tested whether or not lineup presentation

influences the effect of manipulating suspect-filler similarity by including lineup presentation as a subgroup variable.

2.2 Results

Table 2 presents descriptive and inferential statistics for the main effects of similarity. All analyses have been divided by whether the culprit was present or absent from the lineup. For each of the three comparisons (high similarity versus biased lineups, moderate similarity versus biased lineups, and moderate versus high similarity lineups), the proportions of suspect identifications, filler identifications, and lineup rejections are provided as well as the difference between those proportions (risk difference). The risk difference was computed such that a positive value would indicate that as similarity increased, so did the likelihood of a given outcome. Conversely, a negative risk difference would indicate that as similarity increased, the likelihood of a given outcome decreased. Table 2 also includes 95% confidence intervals, null hypothesis significance tests, and heterogeneity tests associated with the risk differences. Finally, Table 2 also includes the number of studies (K) and the number of participants (N) associated with each of the three comparisons.

2.2.1 High similarity versus biased. For culprit-present lineups, the correct identification rate was significantly lower for high similarity lineups compared with biased lineups. Correspondingly, the filler identification rate was significantly higher for high similarity lineups compared with biased lineups. The difference in incorrect rejections between high similarity and biased lineups was not reliable. Thus, it appears that including fillers who are highly similar to the suspect has the effect of

Table 2

Meta-analysis of suspect-filler similarity on identification choices

						Effect Size & CI^{95}			Null Test		Hetero-geneity	
Culprit	Choice	*k*	*n*	Similarity		RD	LL	UL	*z*	*p*	*Q*	*p*
				High	Biased							
Present	Suspect	5	1338	.44	.65	-.21	-.34	.08	3.2	.01	17.8	.01
	Filler	3	1138	.27	.06	.21	.14	.28	6.0	.01	3.7	.16
	Rejection	3	1138	.37	.34	.03	-.03	.08	1.0	.31	1.8	.41
Absent	Suspect	7	1775	.19	.37	-.18	-.27	-.10	4.2	.01	23.2	.01
	Filler	9	1965	.36	.18	.18	.11	.25	5.1	.01	26.1	.01
	Rejection	9	1965	.47	.49	-.02	-.09	.05	0.7	.51	15.7	.05
				Mod.	Biased							
Present	Suspect	6	1449	.47	.63	-.16	-.26	-.05	3.0	.01	15.0	.01
	Filler	5	1389	.24	.08	.16	.03	.29	2.4	.02	34.8	.01
	Rejection	5	1389	.32	.34	-.02	-.07	.03	0.7	.50	1.8	.77
Absent	Suspect	7	1551	.24	.40	-.16	-.23	-.09	4.4	.01	11.6	.07
	Filler	8	1583	.22	.10	.12	.06	.18	4.1	.01	13.9	.05
	Rejection	8	1583	.50	.48	.02	-.03	.07	0.8	.42	3.8	.80
				High	Mod.							
Present	Suspect	8	1996	.45	.47	-.02	-.12	.08	0.4	.66	29.2	.01
	Filler	7	1938	.19	.17	.02	-.06	.10	0.5	.64	29.7	.01
	Rejection	7	1938	.37	.38	-.01	-.06	.04	0.3	.76	7.1	.31
Absent	Suspect	9	1798	.12	.20	-.08	-.14	-.03	2.8	.01	19.7	.01
	Filler	11	2425	.29	.22	.07	.01	.13	2.2	.03	25.9	.01
	Rejection	11	2425	.57	.58	-.01	-.06	.05	0.3	.79	15.7	.11

Note. RD = Risk Difference. *Q* = Cochran's *Q*. Degrees of freedom for *Q* tests = *k* -1. Mean identification response rates are listed under the "Similarity" column.

drawing choices away from that suspect and toward the highly similar fillers, rather than toward rejection of the lineup.

When the culprit was absent, the rate of innocent suspect misidentifications was significantly lower for high similarity lineups compared with biased lineups. The rate of filler identifications was significantly higher for high similarity lineups compared with biased lineups. There was no reliable difference in correct rejections between high similarity and biased lineups. This pattern of results closely mirrors that found when the culprit was present. Thus, regardless of whether the culprit was present or absent, replacing biased fillers with high similarity fillers resulted in a decrease in suspect identifications and an increase in filler identifications.

The ratio of culprit identifications from culprit-present lineups to innocent suspect misidentifications from culprit-absent lineups was calculated for both high similarity and biased lineups to assess their diagnosticity. This analysis showed that the diagnostic value of suspect choices from high similarity lineups (2.42, 95% CI: 1.09, 5.37) was greater than the diagnostic value of suspect choices from biased lineups (1.51, 95% CI: 1.16, 1.95). Thus, compared to biased lineups, suspect identifications from high similarity lineups were more likely to be accurate choices.

2.2.2 Moderate similarity versus biased. Comparisons between moderate similarity and biased lineups produced results similar to those found when high similarity and biased lineups were compared. Specifically, compared to biased lineups, moderate similarity lineups produced a lower rate of suspect identifications, a higher rate of filler identifications, and had no reliable effect on lineup rejections. Again, this pattern was obtained both when the culprit was present and when the culprit was absent. The

diagnosticity ratio was greater for moderate similarity lineups (2.32, 95% CI: 1.21, 4.41) than for biased lineups (1.51, 95% CI: 1.16, 1.97).

2.2.3 High similarity versus moderate similarity. In contrast to the previous two comparisons, the effect of manipulating whether fillers were highly similar or moderately similar to the suspect was dependent upon whether the culprit was present or absent. For culprit-present lineups, there were no significant differences between high and moderate similarity lineups (for culprit identifications, filler identifications, and lineup rejections). In contrast, for culprit-absent lineups, the innocent suspect misidentification rate was significantly lower when fillers were highly similar compared with when they were moderately similar. A concomitant increase in the filler identification rate on high similarity lineups compared with moderate similarity lineups was also observed. Consistent with the high-biased and the moderate-biased comparisons, the difference in correct rejections between high and moderate similarity lineups was not reliable. The diagnosticity ratio for high similarity lineups (4.71, 95% CI: 2.03, 10.94) was greater than the diagnostic value for moderate similarity lineups (2.51, 95% CI: 1.47, 4.29).

2.2.4 Heterogeneity. The heterogeneity of effect sizes among studies was assessed by Cochran's *Q*. Cochran's *Q* tests the null hypothesis that the true effect size does not vary from study to study (Cochran, 1954). A significant *Q* value that is greater than expected by chance (that is, greater than the degrees of freedom) indicates the absence of a common effect size. Table 2 provides ample evidence of heterogeneity in effect sizes, indicating the assignment of weights using the random-effects model was appropriate. Many of the *Q* tests were significant, indicating the obtained summary

effects were often based on individual effect sizes that were subject to dispersion. This was particularly true for suspect and filler identifications, which were almost always significant. In contrast, only one *Q* test for lineup rejections reached significance.

2.2.5 Subgroup analysis of simultaneous and sequential lineups. Table 3 presents descriptive and inferential statistics for the moderator effect of lineup presentation, including the rate of each lineup outcome as a function of whether the lineup was presented simultaneously or sequentially, their associated risk differences, and a significance test (*Q*) of the moderating effect.

For culprit-present lineups, three significant moderator effects of lineup presentation were observed. First, in the comparison between moderate similarity and biased lineups, culprit identifications were influenced by whether the lineup was presented simultaneously or sequentially. For both simultaneous and sequential lineups, the culprit was more likely to be identified from a biased lineup than from a moderate similarity lineup; however the increase in culprit identifications was larger for sequential lineups than for simultaneous lineups. Second, when presented simultaneously the culprit identification rate was higher in moderate similarity lineups than in high similarity lineups. In contrast, when presented sequentially culprit identifications were less likely to occur in moderate similarity lineups than in high similarity lineups. Third, for simultaneous lineups the filler identification rate for moderate similarity lineups was slightly lower than the filler identification rate for high similarity lineups. Conversely, for sequential lineups the filler identification rate for moderate similarity lineups was higher than the filler identification rate for high similarity lineups.

Table 3

Subgroup analysis: Lineup presentation influences on suspect-filler similarity effects

Culprit	Lineup Choice	Presentation	*K*	Similarity		Risk Difference	Subgroup Test *Q*	*df*	*p*
				High	Biased				
Present	Suspect	Simultaneous	5	.45	.69	-.24	0.5	1	.495
		Sequential	2	.35	.51	-.16			
	Filler	Simultaneous	3	.31	.07	.24	1.4	1	.237
		Sequential	2	.24	.06	.18			
	Rejection	Simultaneous	3	.36	.26	.10	0.8	1	.369
		Sequential	2	.42	.44	-.02			
Absent	Suspect	Simultaneous	7	.21	.44	-.23	2.6	1	.104
		Sequential	5	.16	.27	-.11			
	Filler	Simultaneous	9	.43	.20	.23	2.1	1	.149
		Sequential	5	.20	.07	.13			
	Rejection	Simultaneous	9	.38	.42	-.04	0.1	1	.759
		Sequential	5	.64	.66	-.02			
				Mod.	Biased				
Present	Suspect	Simultaneous	6	.51	.66	-.15	5.8	1	.016
		Sequential	2	.22	.52	-.30			
	Filler	Simultaneous	5	.24	.08	.16	2.2	1	.137
		Sequential	2	.34	.06	.28			
	Rejection	Simultaneous	5	.26	.29	-.03	0.6	1	.427
		Sequential	2	.42	.41	.01			
Absent	Suspect	Simultaneous	7	.26	.45	-.19	0.7	1	.393
		Sequential	4	.21	.33	-.12			
	Filler	Simultaneous	8	.24	.11	.13	0.5	1	.477
		Sequential	4	.17	.07	.10			
	Rejection	Simultaneous	8	.47	.43	.04	0.1	1	.748
		Sequential	4	.61	.59	.02			

				High	Mod.				
Present	Suspect	Simultaneous	7	.42	.49	-.07	5.0	1	.022
		Sequential	2	.33	.23	.10			
	Filler	Simultaneous	6	.23	.20	.03	5.1	1	.024
		Sequential	2	.24	.35	-.11			
	Rejection	Simultaneous	6	.36	.35	.01	0.1	1	.743
		Sequential	2	.42	.45	-.03			
Absent	Suspect	Simultaneous	8	.13	.21	-.08	1.3	1	.258
		Sequential	4	.16	.19	-.03			
	Filler	Simultaneous	10	.37	.25	.12	3.7	1	.056
		Sequential	4	.21	.19	.02			
	Rejection	Simultaneous	10	.52	.55	-.03	0.6	1	.454
		Sequential	4	.62	.61	.01			

Note. Mean identification response rates are listed under the “Similarity” column.

For culprit-absent lineups, none of the moderating effects of lineup presentation reached significance. However, given the limited data available, it is worth considering some notable trends. Table 2 shows that decreases in suspect-filler similarity were associated with greater increases in innocent suspect misidentifications from simultaneous lineups than from sequential lineups. For example, compared with moderate similarity lineups, biased lineups increased innocent suspect misidentifications by 19% when presented simultaneously compared with 12% when presented sequentially. Similarly, in the comparison with high similarity lineups, biased lineups increased innocent suspect misidentifications by 23% when presented simultaneously compared with 11% when presented sequentially. Regardless of how the lineup was presented, innocent suspect misidentifications were consistently more likely with biased lineups than with moderate or high similarity lineups; however, it appears this effect may be partly mitigated by the use of sequential lineups.

2.3 Discussion

The meta-analysis revealed several key findings. First, suspect identifications were more common from biased lineups than from moderate or high similarity lineups. This was true both for culprit identifications and for innocent suspect misidentifications. Second, filler identifications were more common from moderate and high similarity lineups than from biased lineups. This finding was unaffected by whether the culprit was present or absent. Third, suspect-filler similarity had no reliable effects on lineup rejections, regardless of whether the culprit was present or absent. Fourth, whether the lineup contained moderate or high similarity fillers had no reliable effect on culprit identifications; however, innocent suspects were significantly more likely to be

misidentified from moderate similarity lineups than from high similarity lineups. Fifth, increases in suspect-filler similarity corresponded with increases in the degree to which suspect identifications were diagnostic of the suspect's guilt.

As predicted, the presence of biased fillers was associated with an increased likelihood of suspect identifications. This pattern was essentially uninfluenced by whether or not the culprit was present. Moreover, whether biased lineups were compared to moderate or to high similarity lineups also had little consequence. Such a robust finding emphasizes the value of ensuring the suspect does not stand out in a lineup. When filler similarity increased, there was a shift from suspect identifications to filler identifications rather than to lineup rejections. In other words, similarity manipulations had no reliable effect on whether a lineup member was chosen or not. Rather, the similarity of fillers only seemed to influence which lineup member was chosen. If fillers were dissimilar, the suspect was more likely to be chosen. If fillers were similar, a filler was more likely to be chosen.

Given that increasing similarity resulted in a shift from suspect to filler identifications regardless of whether the culprit was present or absent, the meta-analytic findings are consistent with Clark's (2012) assertion that policies designed to prevent innocent suspect misidentifications come at the cost of reducing correct identifications of the culprit. Clearly, culprits are more easily identified when fillers are dissimilar-looking than when they are similar-looking. However, the diagnosticity ratios indicated that any reduction in culprit identifications associated with increased filler similarity was outweighed by a more pronounced reduction in innocent suspect misidentifications. As similarity between the suspect and fillers increased, the diagnosticity of suspect

identifications also consistently increased. Thus, although it would be misleading to suggest increasing suspect-filler similarity had no cost, the reduction in culprit identifications was lower in magnitude than the reduction in innocent suspect misidentifications.

2.3.1 Suspect and filler identifications. Wells (1984) theorized that simultaneous lineups encourage witnesses to adopt a relative judgment strategy in which lineup members are compared to one another and the person who best resembles the culprit is chosen. Wells (1993) provided compelling evidence in support of this claim by comparing identification responses on a culprit-present lineup to a lineup that was identical except the culprit had been removed without replacement. From the culprit-present lineup, the culprit was chosen by approximately half of the witnesses and fillers were chosen by one quarter of the witnesses. If witnesses were using an absolute judgment strategy, removing the culprit would be expected to facilitate a shift from culprit identifications to lineup rejections and the filler identification rate should have been unchanged; however, that was not the case. On the contrary, removing the culprit resulted in a shift from culprit identifications to filler identifications, which more than doubled. More recently, using the removal-without-replacement procedure, Clark and Davey (2005) replicated the shift from culprit to filler identifications in simultaneous lineups. Interestingly, a similar shift was present when lineup members were presented sequentially, leading Clark and Davey to suggest relative decisions might also occur with sequential lineups.

The meta-analytic results provide further support for the notion that witnesses engage in a relative judgment strategy when making lineup decisions. In culprit-absent

lineups, the innocent suspects who were chosen by researchers typically either fit the culprit's description (Clark & Tunnicliff, 2001; Juslin et al., 1996; Lindsay & Wells, 1980; Tredoux, Parker, & Nunez, 2007; Wells et al., 1993) or were highly similar to the culprit's appearance (Carlson et al., 2008; Darling et al., 2008; Gronlund, Carlson, Dailey, & Goodsell, 2009; Lindsay et al., 1991). Thus, when lineups were composed of fillers who did not resemble the culprit, the innocent suspect would have been the lineup member who best matched the culprit's appearance. If participants were using relative judgments, this strategy should have increased the number of innocent suspect choices from biased lineups. Conversely, increasing the similarity of the fillers to the suspect should have increased the number of filler identifications, as it would have increased the likelihood that one of the fillers would have best resembled the culprit. This is precisely the pattern of results that was revealed in the meta-analysis.

2.3.2 Lineup rejections. In previous studies, manipulations of suspect-filler similarity have produced conflicting effects on lineup rejections. Increasing similarity between the suspect and fillers has been associated with increases in lineup rejections (e.g., Carlson et al., 2008; culprit-present, simultaneous lineups), decreases in lineup rejections (e.g., Lindsay et al., 1991, Experiment 3; culprit-present, simultaneous lineups), as well as having no effect on lineup rejections (Brewer & Wells, 2006; Charman, Wells, & Joy, 2011; Clark & Tunnicliff, 2001; Darling et al., 2008; Juslin et al., 1996; Lindsay et al., 1994; Lindsay & Wells, 1980; Tredoux et al., 2007; Wells et al., 1993). Tunnicliff and Clark (2000) explored one situation in which similarity seems likely to affect lineup rejections: when the lineup has been matched to the appearance of an innocent suspect who does not resemble the culprit. In two experiments, they found

that lineup rejections were commonplace when none of the lineup members resembled the culprit. Tunnicliff and Clark further discussed what might happen when the innocent suspect and the culprit are similar in appearance. If a similar-looking innocent suspect were placed into a lineup with dissimilar fillers, the lineup would be biased because the innocent suspect would stand out. In this scenario, a false identification seems more likely than a lineup rejection. Nevertheless, that has not always been the case. For example, when Lindsay et al. (1991, Experiment 3) biased a culprit-absent lineup towards an innocent suspect, identification responses were split almost evenly between false identifications of the innocent suspect and correct rejections of the lineup (fillers were never chosen). Moreover, the correct rejection rate for the biased lineup (46%) was twice the correct rejection rate for another lineup containing fillers who did resemble the culprit (23%). These data, combined with those reported by Tunnicliff and Clark, indicate that biased lineups are more likely to be rejected than lineups of moderate or high suspect-filler similarity.

It seems reasonable to hypothesize that lineup rejections would be inversely related to suspect-filler similarity. A lineup composed of fillers who are highly similar to the culprit should draw more choices than a lineup composed of fillers who bear little resemblance to the culprit. As intuitive as this idea may be, data suggesting the opposite have been reported. For example, Carlson et al. (2008) found a higher correct rejection rate for moderate similarity lineups (47%) than for biased lineups (24%), although this pattern was only found for simultaneous lineups. For sequential lineups, a nonsignificant trend in the opposite direction was observed. Although the effect observed by Lindsay et al. (1991) was consistent for simultaneous and sequential lineups, the results reported by

Carlson et al. suggest lineup presentation might influence whether or not suspect-filler similarity influences lineup rejections.

The meta-analysis suggests that similarity rarely had an effect on lineup rejections. The absence of an effect of similarity on rejections was perhaps the most consistent finding in the meta-analysis. In all comparisons, similarity effects on rejections were both small in magnitude and nonsignificant. Why was the rate of lineup rejections unchanged by manipulations of suspect-filler similarity? One possibility is that increasing the similarity of fillers produces contradictory effects. In his WITNESS model, Clark (2003) hypothesized that two factors contribute to suspect and filler identifications: (1) the extent to which a given lineup member matches the witness's memory of the culprit (i.e., an absolute judgment) and (2) the difference in strength of the recognition experience between the lineup member who best matches the witness's memory of the culprit and the next-best alternative. Therefore, increasing suspect-filler similarity could increase the likelihood that a lineup member will match the witness's memory of the culprit and thus exceed the criterion for a choice to be made while simultaneously decreasing the difference between the best match and the next-best match, in turn decreasing the witness’s confidence that the best match is in fact the culprit. Were this to be the case, these competing effects could, as observed in the present research, result in no net change in rejection rates. Of course, the effects would only negate each other if they are similar in strength. An effect of similarity on rejections could be expected if one of these competing effects was stronger than the other, which would explain why similarity has, at times, influenced rejections.

2.3.3 Limitations of the meta-analysis. There are, of course, limitations of the meta-analysis that should be noted. First, suspect-filler similarity was operationalized as the average similarity of the fillers to the suspect. Thus a moderate similarity lineup could consist entirely of fillers who moderately resemble the suspect or it could consist of some combination of biased, moderate similarity, and high similarity fillers. Furthermore, similarity relations are not limited to the resemblance between the suspect and the fillers. Other similarity relations that could affect eyewitness accuracy include the similarity between the culprit's photo and the culprit's physical appearance, the similarity amongst the fillers, and the similarity between the culprit and the innocent suspect. For instance, the extent to which the innocent suspect resembles the culprit would almost certainly influence suspect-filler similarity effects and ideally would have been included as a moderator variable. Unfortunately, this was not an option because ratings of the similarity between the culprit and the innocent suspect were rarely reported (but see Clark & Tunnicliff, 2001).

Second, I excluded unpublished studies from the analysis to accommodate the preference for published research in the legal system. Significant effects are generally more likely to be published than nonsignificant effects, so including unpublished studies into the meta-analysis might have resulted in smaller effects.

Third, despite more than 30 years having passed since the first exploration of suspect-filler similarity effects, the literature in this domain is relatively small. Consequently, the number of studies comprising appropriate tests of similarity limited the scope of the meta-analysis. For example, researchers varied in how they manipulated similarity and a larger database would have been needed to effectively examine whether

or not the type of manipulation influenced similarity effects. On a related note, because not all studies included a manipulation of lineup presentation, my conclusions about the moderating effect of this variable are tentative.

2.3.4 Lineup construction recommendations vs. meta-analytic findings. In the report developed by the Technical Working Group for Eyewitness Evidence (2003), police investigators are advised that the suspected culprit should not stand out from the lineup members who are known to be innocent. The meta-analysis results provide support for this recommendation. Compared with lineups that had fillers of moderate or high suspect-filler similarity, the rate of innocent suspect misidentifications nearly doubled when lineups contained biased fillers. The group's report further advises police investigators to ensure that fillers and the suspect are not too similar. The concern is that using extremely similar fillers will essentially result in a lineup of "clones" that would greatly diminish the likelihood that a culprit will be correctly identified. However, my synthesis of the existing literature did not support this assertion. Although the not-too-similar rule has a solid theoretical foundation (Luus & Wells, 1991) that was soon after supported by empirical research (Wells et al., 1993), I found no reliable difference in correct identifications between lineups within the categories of high and moderate suspect-filler similarity.

Wells (1993) reported concern among some eyewitness researchers that choosing fillers with features that vary from those of the suspect could result in lineups with an unintended bias towards innocent suspects. The present research suggests their concern may have been justified. Innocent suspects were significantly more likely to be misidentified from lineups of moderate suspect-filler similarity compared with lineups of

high suspect-filler similarity. This increase in innocent suspect misidentifications, taken together with the null effect in culprit identifications, suggests that either a) the rule of ensuring lineup members are not too similar to the suspect does not improve performance on culprit-present lineups and may actually contribute to wrongful convictions or b) the inability to obtain fillers who are truly of high resemblance to the suspect has led to an incongruity between theory and practice. Although the rule to avoid highly similar fillers may be theoretically sound, finding such fillers in practice may be more difficult than had been anticipated.

Lest it appear that I am advocating the dismissal of a rule that has been deemed best practice in lineup identification procedures (Turtle, Lindsay, & Wells, 2003), it is critical to emphasize that the similarity categories were developed in relation to one another and that the "high" similarity lineups might not have had the degree of similarity that has been cautioned against. Inspection of the similarity ratings provided by the researchers suggests this might very well have been the case. Although the lineups that were categorized as "high" had ratings higher than those categorized as "moderate", and researchers sought to create very high similarity lineups in many cases, the high similarity lineups rarely had mean similarity ratings that were above the midpoint of the scales that were used. The lineups were certainly not comprised of clones, but the relatively modest similarity ratings may also indicate a reluctance of those judging similarity to use the upper end of the scale (see Flowe & Ebbeson, 2007). In the following chapter, I report an experimental investigating using fillers that were unquestionably high in suspect-filler similarity.

CHAPTER 3

A Novel Manipulation of Lineup Member Similarity

The meta-analysis in Chapter 2 seemed to indicate that high similarity lineups provide the best balance in terms of protecting innocent suspects and facilitating culprit identifications. Biased lineups yielded the highest rates of correct identification, which should come as no surprise given the blatant biases that were characteristic of these lineups. For instance, some biased lineups contained fillers who were not even the same race as the suspect. Biased lineups also led to a very high rate of innocent suspect misidentifications, suggesting the biased fillers led to an inherent bias toward identification of the suspect, regardless of whether that suspect was guilty or innocent. Accordingly, the construction of biased lineups for eyewitness identification is not advised.

The question, then, is whether fillers should be of moderate or high similarity to the suspect. The meta-analysis suggested that, relative to moderate similarity fillers, high similarity fillers reduced misidentification of innocent suspects without impeding correct identifications of the culprit. This would seemingly suggest that lineup constructors would be best advised to select fillers who strongly resemble the suspect. One thing to note, however, is that the lineup categories in the meta-analysis were defined in relation to one another. Although suspect-filler similarity was greater in the high similarity category than in the moderate or biased categories, these were not objectively defined categories. Thus, in previous research, the lineups may not have been constructed with the degree of similarity that has been cautioned against.

One thing to consider is that the ability to create high similarity lineups in previous research has been limited by the availability of fillers who strongly resemble the

suspect. To circumvent this issue, I manipulated suspect-filler similarity in the present experiment using morphing software. The software is capable of creating lineups with an extremely high degree of similarity, which is critical for identifying the point at which fillers resemble the suspect to such a degree that correct identifications become too difficult. I carefully pilot-tested the materials to ensure that morphing per se did not influence responding and that participants actually perceived greater similarity in the lineup manipulated to be higher in similarity. In the experiment, witnesses viewed a video containing a target person and were subsequently asked to attempt his identification from lineups containing (a) the target or an innocent suspect and (b) moderately high or very high similarity fillers. I predicted that relative to the moderately high similarity fillers, the very high similarity fillers would reduce identifications of both the target and the innocent suspect.

3.1 Method

3.1.1 Participants. For the main experiment, I recruited 137 university undergraduate students (M_{age} = 20.70 years, SD = 3.76; 109 women) from a psychology department participant pool. Participants were compensated with partial course credit. The study was approved by the University of Regina Research Ethics Board (Appendix B).

3.1.2 Design. This experiment utilized a 2 (target: present vs. absent) × 2 (suspect-filler similarity: moderately high vs. very high) between-subjects design.

3.1.3 Video event. The 6-minute silent video began with a man and woman preparing to eat breakfast at a restaurant. They ordered food and received beverages, but got into an argument and left the restaurant before their food arrived. Outside the

restaurant, the argument resumed. Although the man (culprit) managed to sneak in a kiss, the woman ultimately pushed him away, got into a car, and fled the scene.

3.1.4 Lineups. The five faces in Figure 1 were altered using *Fantamorph* software to create fillers for four lineups that varied in target-presence and suspect-filler similarity (Figure 2). The original faces were selected because of their match to the culprit's appearance and their suitability for morphing.

3.1.4.1 Culprit present. The culprit-present lineups contained faces of the culprit and five fillers. A graphical representation of the morphing procedures used in the moderately high and very high similarity lineups is provided in Figure 3.

In the moderately high similarity lineup, the fillers in Figure 1 were morphed with the culprit to create a new face that was 40% culprit and 60% filler. Although this procedure only changed the appearance of the fillers slightly from their unmorphed photograph, morphing was performed nonetheless to avoid having a very high similarity lineup that contained morphed fillers and a moderately high similarity lineup that contained unmorphed fillers.

A simple method of producing fillers who resemble the culprit more than the fillers in the moderately high similarity lineup would be to increase the degree of morph with the culprit. For example, I could have created faces that were 70% culprit and 30% filler. However, increasing the morph to this degree would produce fillers that are indistinguishable from both the culprit and each other. Even though I am well-acquainted with the person acting as the culprit, I was not even able to identify him from a lineup with fillers morphed with him by 70%.

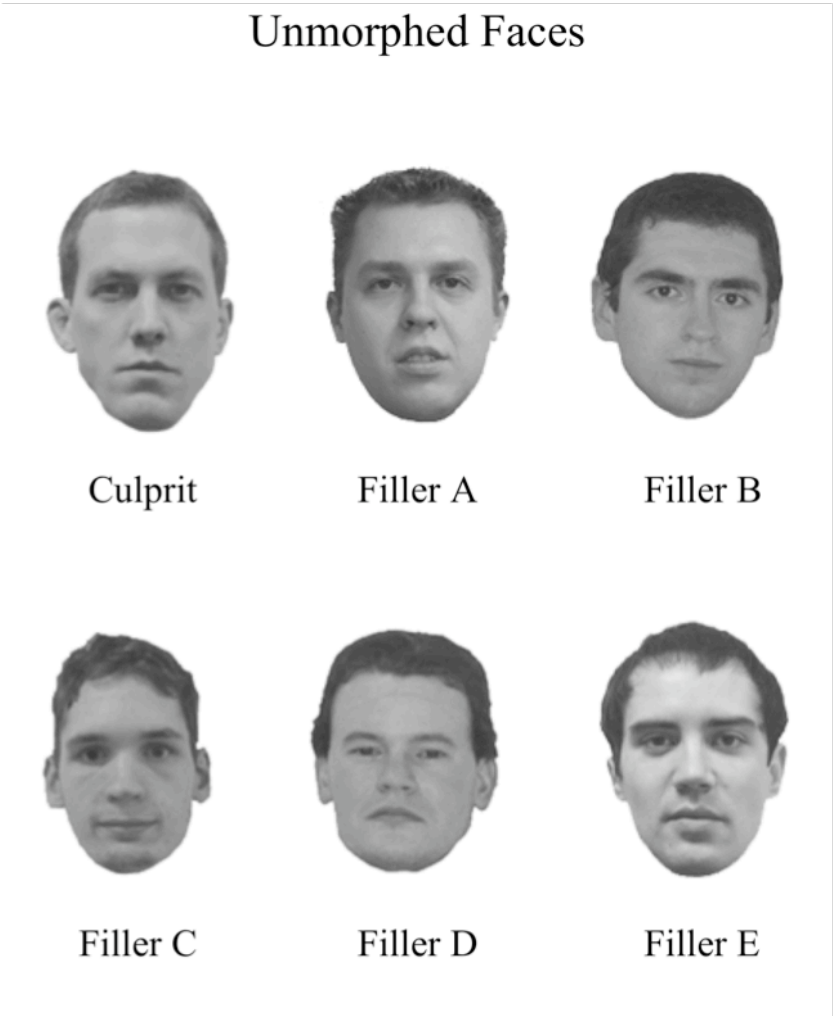

Figure 1. Faces of the culprit and the fillers (prior to morphing).

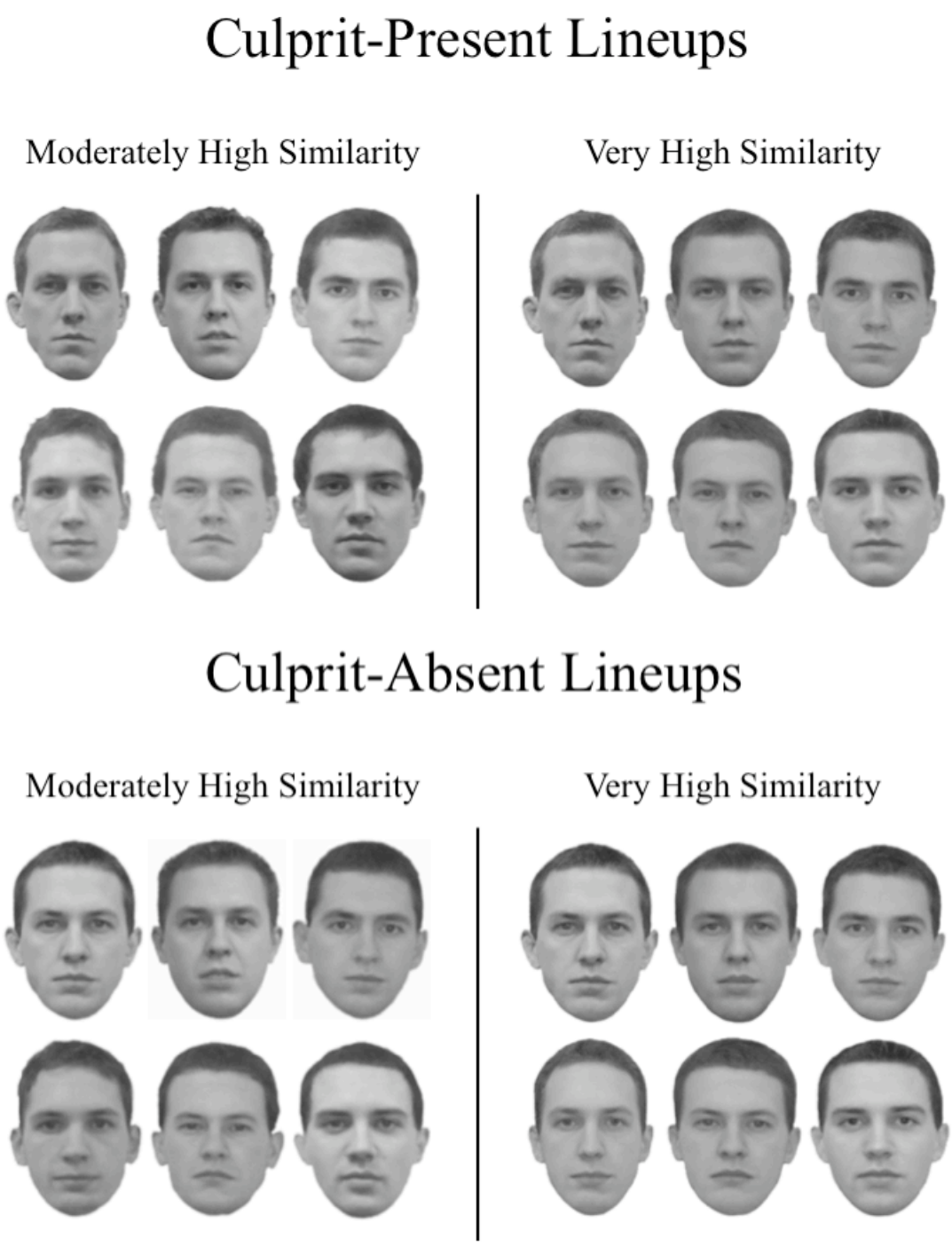

Figure 2. Lineups for moderately high and very high similarity fillers in the Chapter 3 experiment. In these lineups, the suspect is always positioned in the top-left corner. In the experiment, suspect position was counterbalanced.

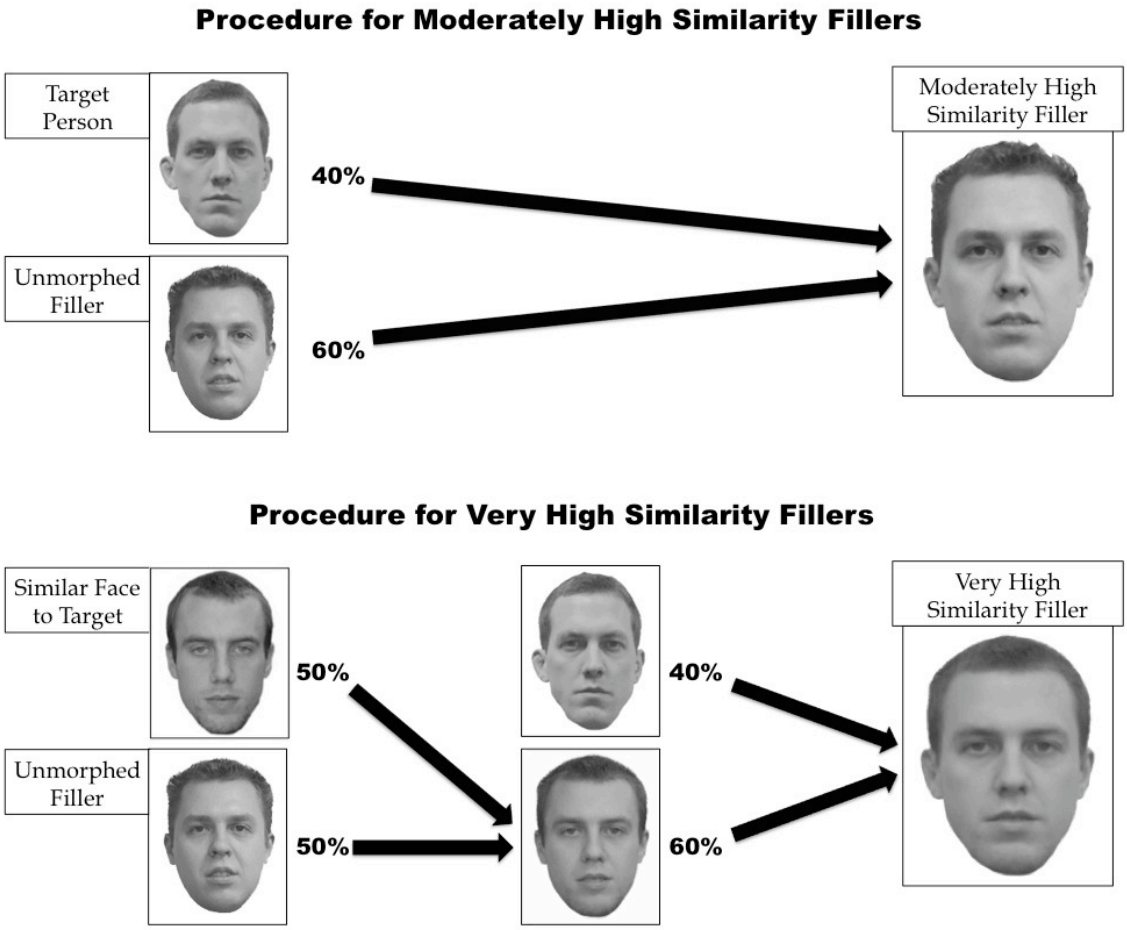

Figure 3. Standard morphing procedures used to create fillers in moderately high and very high similarity conditions in the Chapter 3 experiment. The target and filler images were photographed locally. The "Similar Face to Target" was obtained from the Glasglow Unfamiliar Face Database (Burton, White, & McNeill, 2010).

To avoid this problem, fillers in the very high similarity lineup were morphed twice. First, I obtained similarity ratings to identify faces of high similarity to the culprit. Participants (n = 5) compared the culprit to 277 other faces and assigned a number from 0 (not at all similar) to 10 (highly similar). These data were used to select five faces of high similarity to the culprit, which were then morphed 50% with the faces in Figure 1. Each face was morphed with a different person of high similarity to the culprit to avoid having a particularly homogenous set of fillers, which could make the culprit stand out. Then, to ensure that the very high similarity fillers resembled the culprit more than the moderately high similarity fillers, the morphed faces were morphed again, with the culprit (40% culprit and 60% filler). This procedure produced fillers who were very similar to the culprit, but not so similar that the author who was acquainted with the culprit could not distinguish between the culprit and the fillers.

Although pilot testing (the relative judgement task, described below in the Pilot Tests section) indicated this procedure produced an effective difference in culprit-filler similarity between the moderately high and very high similarity conditions for three of the fillers, the difference in culprit-filler similarity between the moderately high and very high similarity conditions was equivocal for Fillers C and E. This was addressed by adjusting the morph for Fillers C and E in the very high similarity condition to create faces that were 50% culprit and 50% filler instead of 40% culprit and 60% filler. Follow-up pilot tests indicated that this procedure created Fillers C and E who were more similar to the culprit in the very high similarity condition than in the moderately high similarity condition.

3.1.4.2 Culprit absent. Culprit-absent lineups comprised an innocent suspect and five fillers. Fillers were created by morphing the faces in Figure 1 with the innocent suspect's face. In previous studies, researchers have often used the same fillers in the culprit-present and culprit-absent lineups. That is, they simply replaced the culprit with an innocent suspect. By only changing one element of the lineup, this design has the advantage of high experimental control; however, this approach does not correspond with how the match-to-appearance procedure would be implemented by legal investigators constructing lineups for an innocent suspect because the appearance of the culprit would not be known (Clark & Tunnicliff, 2001). To simulate the lineup construction procedures that would occur in the field, I matched fillers in the culprit-absent lineups to the innocent suspect's appearance. When fillers are matched to an innocent suspect's appearance, the innocent suspect is hypothesized to resemble the culprit more than any of the fillers (cf., Navon, 1992; Wogalter et al., 1992). To create this effect in the lineups, the innocent suspect's face was morphed with the culprit's face (50%) and the filler faces were not morphed with the culprit's face.

The morphing procedure for fillers in the culprit-absent lineup was similar to that employed for fillers in the culprit-present lineup. In the moderately high similarity lineup, faces in Figure 1 were morphed to create a face that was 40% innocent suspect and 60% filler. In the very high similarity lineup, faces were morphed twice. First, each of the faces in Figure 1 was morphed 50% with a unique face that was judged to bear a strong resemblance with the innocent suspect in pilot testing (n = 4; same rating procedure as was used with the culprit). The resulting faces were then morphed with the innocent suspect (40% innocent suspect and 60% filler).

Pilot testing (the relative judgement task) indicated this procedure did not produce an effective difference in suspect-filler similarity between three pairs of moderately high and very high similarity fillers (Fillers B, C, and E). This was addressed by morphing these three very high similarity fillers with the innocent suspect 50% instead of 40%. For each of the three pairs, follow-up pilot tests indicated this procedure created very high similarity fillers who resembled the innocent suspect more than did the moderately high similarity fillers.

3.1.5 Pilot tests. The effectiveness of the morphing software in producing a manipulation of similarity was evaluated in a series of pilot tests.

3.1.5.1 Relative judgement task. In the first set of pilot studies, participants made relative judgements about which of two lineup members (a moderately high similarity filler vs. a very high similarity filler) was more similar to the suspect. In the first pilot study, judges (n = 12) completed multiple trials in which three faces were presented in a row. The suspect (i.e., the culprit or the innocent suspect) was always in the middle position. Fillers were always positioned on the left and right sides. The number "1" was always displayed over the face on the left side and the number "0" was always displayed over the face on the right side. Judges were instructed to determine whether Person 0 or Person 1 looked more similar to the person in the middle. For example, the culprit would be in the middle, Filler A from the moderately high similarity culprit-present lineup would be on the left, and Filler A from the very high similarity culprit-present lineup would be on the right. If the very high similarity filler resembles the culprit more than does the moderately high similarity filler, participants should choose the very high similarity filler more frequently.

For the similarity manipulation to be acceptable, I set an arbitrary criterion stating that the very high similarity filler needed to garner at least twice as many choices as the moderately high similarity filler. This criterion (i.e., 8/12) was met in 5 of the 10 comparisons. In the other five comparisons, the number of choices for the moderately high and very high similarity fillers was approximately evenly split. This was addressed by making a slight increase in the extent to which very high similarity fillers were morphed with the suspect and then conducting a follow-up pilot study with new participants (n = 12) to confirm the new faces met the previously mentioned criterion. The modifications to these fillers are detailed in the Materials section.

3.1.5.2 Find the nonmorph task. In the culprit-present lineups, the photograph of the culprit is the only nonmorphed photograph. Thus, witnesses could potentially choose the culprit not because they remember him from the video, but rather because they can tell the other faces have been altered in some way and that the culprit's photograph is the only one that has not been digitally altered. To assuage such concerns, pilot studies were conducted with the culprit-present lineups to ensure that the culprit's face could not be distinguished from the fillers in the absence of recognition.

In the pilot studies, participants (n = 84) who did not see the crime video were informed they would view a simultaneous lineup containing five digitally manipulated faces and one unaltered face (location of culprit was counterbalanced). Their task was to choose the face that had not been manipulated. For the 43 participants who viewed the moderately high similarity lineup, the proportion (P) who chose the culprit (P = .16, SE = .06) did not differ from chance (.17). For the 41 participants who viewed the very high

similarity lineup, the proportion who chose the culprit (P = .20, *SE* = .06) also did not differ from chance.

3.1.5.3 Suspect-filler similarity ratings. The similarity between the suspects and the morphed fillers was further assessed through subjective similarity ratings. A new set of judges (n = 17) viewed suspect-filler face pairs and assigned a number from 0 (not at all similar) to 10 (highly similar). These judges also rated the similarity between the culprit and the innocent suspect (M = 6.53, SE = 0.61). Paired-samples t-tests were used to assess the difference between ratings for a filler in the moderately high similarity condition and for the corresponding filler in the very high similarity condition. Table 4 shows reliable differences in the predicted direction for Fillers A, B, C, and D, but no reliable differences for Filler E. Additional paired-samples t-tests were used to contrast average ratings for the five fillers in the moderately high and very high similarity conditions. For culprit-present and culprit-absent conditions, the moderately high similarity fillers were rated to be significantly less similar than the very high similarity fillers.

3.1.6 Procedure. The study followed the ethics guidelines put forth in the Tri Council of Canada Policy Statement and was approved by the University of Regina Research Ethics Board (the approval memo is provided in Appendix B). To prevent knowledge of an upcoming memory test, the study was advertised as an investigation of media influences on gender roles. When participants first arrived at the lab, they completed an informed consent sheet that described their rights as participants. Initially, participants were led to believe that they would watch a video and then answer questions about gender roles. Participants viewed the video described in the Materials section on a

Table 4

Mean (SE) ratings of suspect-filler similarity in the Chapter 3 experiment

		Similarity Condition				Cohen's *d* and 95% CIs		
		Moderately	Very				Lower	Upper
Suspect	Filler	High	High	*t*	*p*	*d*	Limit	Limit
Culprit	A	5.37 (0.56)	7.63 (0.47)	3.65	.002	0.91	0.25	1.57
	B	3.71 (0.45)	6.00 (0.56)	3.43	.003	0.82	0.15	1.48
	C	4.88 (0.54)	7.50 (0.43)	6.15	.001	1.58	0.88	2.26
	D	3.56 (0.57)	6.31 (0.58)	3.32	.005	0.83	0.08	1.57
	E	4.47 (0.59)	5.00 (0.53)	1.28	.217	0.32	-0.04	0.67
	Average	4.39 (0.36)	6.50 (0.40)	5.63	.001	1.41	0.72	2.09
Innocent	A	5.31 (0.78)	8.06 (0.44)	3.51	.003	0.90	0.24	1.55
	B	6.06 (0.63)	7.81 (0.53)	2.43	.028	0.60	-0.02	1.22
	C	4.94 (0.60)	7.18 (0.55)	3.53	.003	0.86	0.25	1.41
	D	5.12 (0.55)	6.88 (0.49)	3.85	.001	0.93	0.42	1.45
	E	6.25 (0.54)	6.06 (0.57)	0.31	.764	-0.08	-0.44	0.60
	Average	5.39 (0.47)	7.08 (0.42)	5.64	.001	1.46	0.83	2.08

Note: Scale ranged from 0 (Not at all similar) to 10 (highly similar). All comparisons had 16 degrees of freedom.

21-inch computer screen and then provided a verbal report about what happened in the video to the experimenter. No follow-up questions were asked and no feedback was given. The experimenter subsequently informed participants that the experiment was actually about eyewitness identification, not gender roles. After learning of the true purpose of the experiment, participants asked for consent to proceed with the identification task. None of the participants opted to withdraw.

After obtaining consent to proceed with the identification task, the experimenter instructed participants to imagine that the man from the video committed a crime and that they were the only person in a position to identify him. The experimenter verbally instructed the participants that the culprit may or may not be in the lineup and that they could reject the lineup if they did not think the culprit was present. To facilitate double-blind lineup administration, the lineup task was completed on a computer. The experimenter who opened the computer program did not know whether the culprit was in the lineup or not, nor did he know which fillers the lineup contained.

Lineup members were presented simultaneously in a 2 × 3 array. Each face was associated with a number (1-6). The spatial location of the suspect's photograph was counterbalanced. The computer program instructed participants that if the man from the video was present, they were to press the number associated with his image. If absent, participants were instructed to press "0". After the identification task, participants provided a confidence assessment about their identification decision using a 6-point scale, ranging from 1 ("not all at confident") to 6 ("highly confident"). In an exit interview, participants were asked to report their demographic information, previous

experience participating in eyewitness experiments, and whether or not they knew that the experiment involved an identification test before viewing the video.

3.2 Results

The exit interview revealed seven participants who claimed awareness that the experiment involved an identification test before watching the video. These seven participants were omitted from all data analyses, which had no impact on whether any of the differences were significant or not.

3.2.1 Effects of similarity on identification responses. The significance of associations between categorical variables was assessed with *z*-tests for the difference between two proportions. For effect size measures, odds ratios (OR) with 95% confidence intervals are reported with all *z*-tests. An odds ratio of 1.00 indicates perfect unity between two groups in the odds of an identification response between conditions (Bland & Altman, 2000). For the analyses that follow, an odds ratio above 1.00 indicates a greater likelihood of a response in the moderately high similarity condition than in the very high similarity condition.

On culprit-present lineups, similarity was associated with suspect and filler choices (Table 5). A greater proportion of culprit identifications were made when similarity was moderately high relative to very high, $z = 2.48$, $p = .01$, OR = 4.04, 95% CI = 1.24, 13.23. Correspondingly, a smaller proportion of filler selections were made in the moderately high similarity condition relative to the very high similarity condition, $z = 2.77$, $p = .01$, OR = 0.25, 95% CI = 0.09, 0.71. Similarity had no effect on incorrect lineup rejections, $z = 0.49$, $p = .62$, OR = 1.34, 95% CI = 0.43, 4.20.

Table 5

Identification response probabilities (P) and standard errors (SE) in the Chapter 3 experiment

Culprit	**Similarity**	***n***	**Identification Response**					
			Suspect		**Filler**		**Reject**	
			P	*SE*	P	*SE*	P	*SE*
Present	Moderately High	32	.44	.09	.28	.08	.28	.08
	Very High	31	.16	.07	.61	.09	.23	.08
Absent	Moderately High	33	.30	.08	.12	.06	.58	.09
	Very High	34	.24	.07	.38	.08	.38	.08

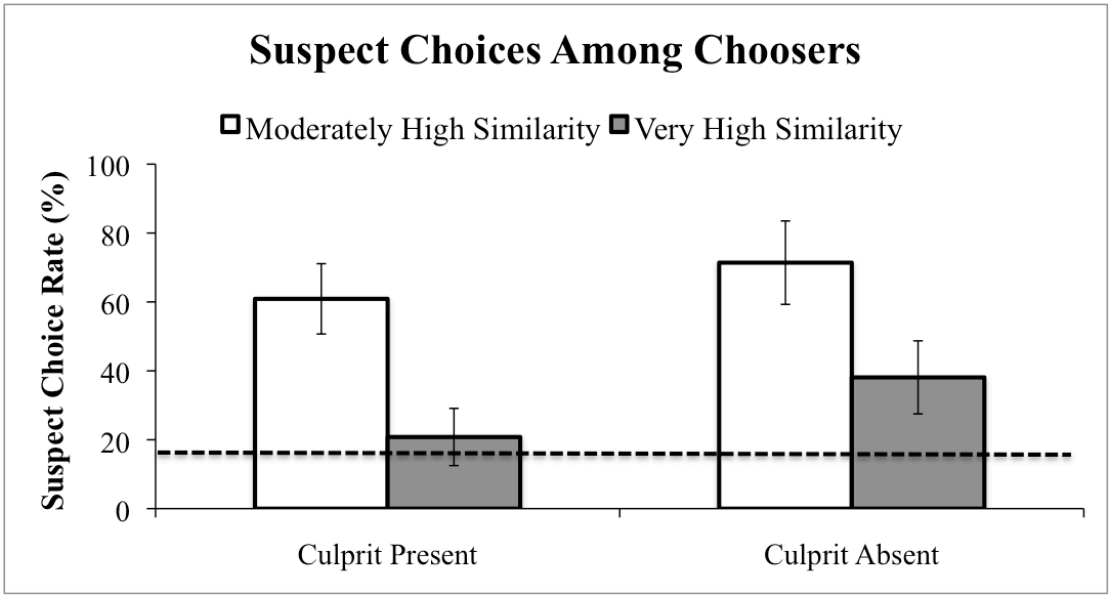

Figure 4. Proportion of suspect choices from among only those who picked someone from the lineup (choosers) in the Chapter 3 experiment. The dotted line indicates the suspect choice rate expected by chance. Error bars represent +/- 1 standard error.

On culprit-absent lineups, the innocent suspect was misidentified more frequently from the moderately high similarity lineup than from the very high similarity lineup (Table 5); however, the difference was small and nonsignificant, $z = 0.62$, $p = .54$, OR = 1.41, 95% CI = 0.48, 4.19. The only significant effect in the culprit-absent condition was for filler selection rates, which were lower for moderately high relative to very high similarity lineups, $z = 2.54$, $p = .01$, OR = 0.22, 95% CI = 0.06, 0.78. The correct rejection rate was higher for the moderately high similarity lineup than for the very high similarity lineup; however, the difference did not reach significance, $z = 1.59$, $p = .11$, OR = 2.19, 95% CI = 0.83, 5.83.

To explore similarity effects on only those who chose a lineup member (choosers), the analyses were repeated after excluding those who rejected the lineup. Among choosers, similarity had comparable effects on culprit-present and culprit-absent lineups (Figure 4). In particular, both the culprit ($z = 2.99$, $p = .003$, OR = 5.91, 95% CI = 1.62, 21.54) and the innocent suspect ($z = 2.01$, $p = .04$, OR = 4.06, 95% CI = 0.95, 17.43) were more likely to be chosen from moderately high relative to very high similarity lineups. Thus, regardless of whether the culprit was present or absent, increasing similarity for choosers resulted in a shift from identification of the suspect to identification of a filler.

3.2.2 Dispersion. When presented with a 6-person lineup and an option to reject, an eyewitness has seven options to choose from. In the preceding analyses, choices were divided into three categories: suspect, filler, reject. Although collapsing the filler choices into a single category was advantageous for conducting inferential statistics with enough power to detect significant differences between the two similarity conditions, this

approach does not fully inform about how the lineup choices were dispersed among the lineup members.

MacLennan (2002) developed a measure for quantifying the dispersion of incorrect choices in multiple-choice tests. The dispersion index (DI) measures the randomness of responses to distractor items and produces values ranging from 0 (complete bias) to 1 (complete randomness). The dispersion index is calculated using the total frequency count (N), the number of categories (k), and frequency count for each category (n_i). The equation for the dispersion index is provided below:

$$DI = \frac{N \times Log(N) - \sum_{i=1}^{k} n_i \times Log(n_i)}{N \times Log(k)}$$

The dispersion index provides a useful tool for assessing the extent to which identification responses are spread among the lineup members. The dispersion index is designed to assess dispersion among the distractor items (e.g., fillers) and, thus, is uninfluenced by the frequency count for the correct item (e.g., the culprit). For culprit-present lineups, the measure provides a summary statistic to indicate how choices were dispersed among the five fillers. Regardless of whether the culprit was present or absent, the filler selections were more randomly distributed for very high similarity lineups (culprit-present: $DI = 0.95$; culprit-absent: 0.95) than for moderately high similarity lineups (culprit-present: $DI = 0.78$; culprit-absent: 0.42).

3.2.3 Diagnosticity. Diagnosticity ratios, which are the ratio of an identification response probability in culprit-present and culprit-absent lineups, can be calculated to assess the extent to which a response is indicative of the suspect's guilt or innocence (Wells & Lindsay, 1980). Diagnosticity ratios for suspect identifications are typically

used to provide information about guilt, whereas diagnosticity ratios for filler identifications and lineup rejections are typically used to provide information about innocence (Wells & Olson, 2002).

Diagnosticity ratios for suspect selections were calculated by dividing the culprit selection rate by the innocent suspect selection rate. A ratio above unity (i.e., 1.00) would suggest that suspect identifications from a given lineup are indicative of guilt. A larger diagnosticity ratio was found for the moderately high similarity lineups (1.45, 95% CI = 0.75, 2.75) than for very high similarity lineups (0.69, 95% CI = 0.25, 1.87). Thus, a suspect identified from a moderately high similarity lineup was indicative of guilt, whereas a suspect identified from a very high similarity lineup was not indicative of guilt.

Diagnosticity ratios for filler selections and lineup rejections were calculated by dividing the response rate for the culprit-absent lineup by the response rate for the culprit-present lineup. A ratio above unity (i.e., 1.00) would suggest that filler and rejection responses from a given lineup are indicative of innocence. For filler selections, the ratios in both similarity conditions were less than 1.00 (moderately high = 0.43, 95% CI = 0.14, 1.26; very high = 0.62, 95% CI = 0.37, 1.04). For lineup rejections, the ratios in both similarity conditions were greater than 1.00 (moderately high = 2.05, 95% CI = 1.09, 3.83; very high = 1.69, 95% CI = 0.78, 3.69). Thus, rejections were indicative of innocence, but filler selections were not.

3.2.4 Signal detection analyses. Although the diagnosticity ratio has traditionally been used to provide an overall assessment of identification performance across culprit-present and culprit-absent lineups, this technique has been criticized for its susceptibility to influences on response criterion (Wixted & Mickes, 2012). As a procedure becomes

increasingly conservative, the diagnosticity ratio will generally also tend to increase. Mickes, Moreland, Clark, and Wixted (2014) described a measure of discriminability (*d'*) derived from signal detection theory as an alternative. Although *d'* has traditionally been applied to studies where lists of items are studied and tested, Mickes et al. (2014) report that *d'* can be applied to eyewitness identification experiments to accurately approximate receiver operator characteristic curves.

In a memory test for a list of items, a *d'* value can be calculated for each participant; however, eyewitness identifications typically only involve one identification decision. Thus, *d'* can only be computed at the group level. This is accomplished by treating the suspect identification in the culprit-present lineups as the hit rate and treating the innocent suspect identification rate in the culprit-absent lineups as the false alarm rate. This produced a higher *d'* score for the moderately high similarity lineups (0.37) than for the very high similarity lineups (-0.28). Given that only a single *d'* score is available for each group, the traditional significance tests that would normally be used to compare discriminability across conditions cannot be applied. However, Gourevitch and Galanter (1967) describe a method that can be used to estimate the variance of a single *d'* score, which in turn allows for a statistical comparison of two *d'* scores. This method indicated the difference between *d'* scores was nonsignificant, $G = 1.38$, $p = .17$.

3.2.5 Confidence. Influences on post-identification confidence were assessed with two 2 (lineup similarity) × 2 (identification accuracy) analyses of variance (ANOVA): one for culprit-present lineups and one for culprit-absent lineups (Figure 5). The ANOVA for culprit-present lineups revealed a main effect of similarity, $F(1,59) =$

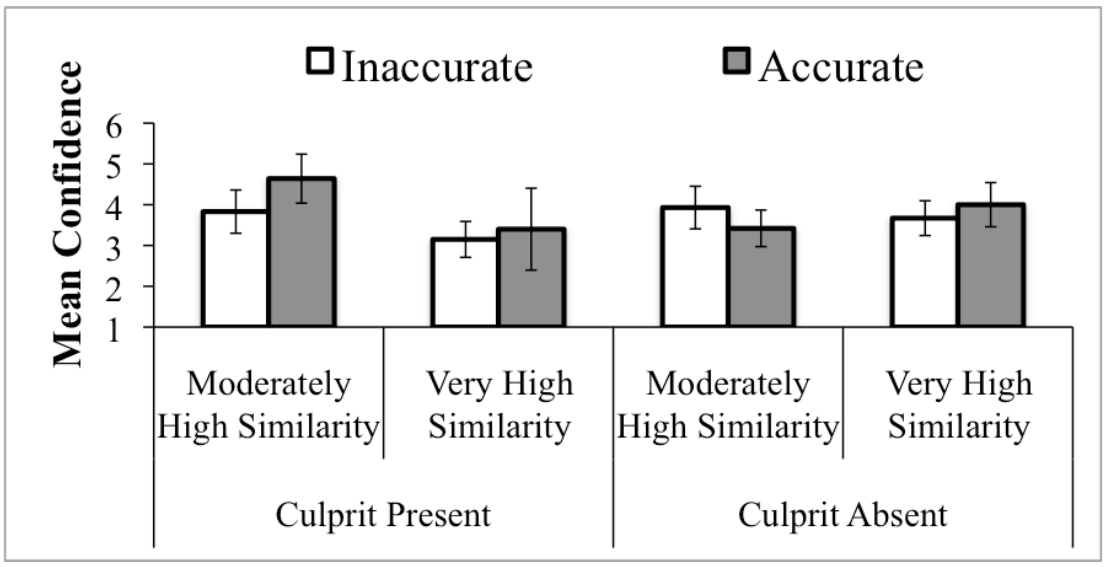

Figure 5. Post-identification confidence ratings in the Chapter 3 experiment. The scale ranged from 1 (not at all confident) to 6 (very confident). Error bars represent 95% confidence intervals.

8.03, $p = .006$, $d = 0.89$, 95% CI = 0.60, 1.17, indicating greater confidence for the moderately high similarity lineups relative to the very high similarity lineups. There was no main effect of accuracy and no interaction. The ANOVA for culprit-absent lineups revealed no main effects and no interaction.

Consistent with previous research (Sporer, Penrod, Read, & Cutler, 1995), a stronger confidence-accuracy correlation was observed for choosers ($r = .29$, $p = .007$) relative to nonchoosers ($r = -.01$, $p = .93$); however, the difference between the correlations (.30; 95% CI = -.06, .64) did not reach significance, $z = 1.65$, $p = .10$. When calculated only for moderately high similarity lineups, the confidence-accuracy relation was much stronger for choosers ($r = .38$, $p = .02$) than for nonchoosers ($r = -.24$, $p = .21$) and the difference between correlations (.62; 95% CI = .20, .97) was significant, $z = 2.46$, $p = .01$. Conversely, when similarity was very high, the relation was stronger for nonchoosers ($r = .30$, $p = .20$) than for choosers ($r = .00$, $p = 1.00$), and the difference between correlations (.30; 95% CI = -.25, .76) was nonsignificant, $z = 1.08$, $p = .28$.

3.3 Discussion

This experiment was designed to establish the upper bounds of suspect-filler similarity. Consistent with my prediction, participants were much more adept at identifying the culprit from the moderately high similarity lineup than from the very high similarity lineup. The correct identification rate for the very high similarity lineup was quite low, which suggests the morphing software was successful at establishing the degree of similarity required to impede correct identifications. The decrease in correct identifications corresponded with an increase in filler identifications, suggesting the very

high similarity fillers drew choices away from the culprit. These data suggest suspect-filler similarity can be "too high", at least with the use of morphing software.

Only partial support was found for the hypotheses concerning similarity's effect on culprit-absent lineups. Although the increase in suspect-filler similarity led to a concomitant increase in filler selections, it yielded only a small and nonsignificant reduction in the overall rate of innocent suspect misidentifications. However, these rates were skewed by different choosing rates across similarity conditions. When only choosers were considered, both the innocent suspect and the culprit were significantly more likely to be identified from the moderately high similarity lineup than from the very high similarity lineup. Thus, costs and benefits were associated with very high similarity fillers.

The absence of an effect of similarity on rejection rates was one of the most consistent findings in the Chapter 2 meta-analysis, so the difference in correct rejection rates between the moderately high (58%) and very high (38%) similarity lineups in the present research was not anticipated. In Chapter 2, I used Clark's (2003) WITNESS model to interpret the null effect of similarity on choosing/rejection rates. According to WITNESS, lineup choices may occur because one lineup member is a strong match with the witness's memory of the culprit or because one lineup member is a much better match than any of the other lineup members. I noted that as suspect-filler similarity increases, two effects on choosing were possible: (1) choosing could increase because of the higher likelihood that one of the lineup members would exceed a threshold for choosing, or (2) choosing could decrease because of the reduced difference between best and next-best matches.

I suggested that if these two competing effects are of equal strength, similarity would have no effect on choosing. However, in the present research one effect may have been stronger than the other. In particular, morphing the very high similarity fillers with the innocent suspect may have increased the likelihood that one of the lineup members exceeded the decision criterion for making a positive selection. Although the increase in suspect-filler similarity should also have decreased the difference between the best and next-best matches, this might not have had any effect because the resemblance between the suspect and the fillers was relatively high in both the moderately high and very high similarity conditions. In other words, the difference between the best and next-best matches may not have been sufficiently large in either of the similarity conditions to warrant a positive identification, thus forcing witnesses to base their decision on whether the recognition experience elicited by one of the lineup members exceeded their decision threshold or not. Given the increased chance of a lineup member in the very high similarity lineup exceeding this threshold, this interpretation is consistent with the increased rate of choosing in the very high similarity condition.

3.3.1 Diagnosticity. The diagnosticity ratios indicated that suspect identifications from the moderately high similarity lineup were indicative of guilt, but suspect identifications from the very high similarity lineup were not. This is interesting because the Chapter 2 meta-analysis suggests that diagnosticity typically increases as fillers become more similar to the suspect. I suspect that the "high similarity" fillers in previous research were not as similar as the fillers that I created through morphing software, which could explain the atypical effect of similarity on diagnosticity. Consistent with this interpretation, similarity had a minimal effect on innocent suspect misidentifications and

the lower diagnosticity ratio for the very high similarity lineups was primarily reflective of a drop in correct identifications. In other words, reducing similarity to only moderately high made it easier to identify the target.

Reducing the similarity between fillers and the innocent suspect might be expected to cause an increase in innocent suspect selections, so it is noteworthy that this did not happen. For this issue, two points are worthy of note. First, the match between the suspect's photograph and the memory of the culprit should typically be greater for the culprit than for an innocent suspect. Therefore, even if suspect-filler similarity were identical across target-present and target-absent lineups, the difference between the best match and the next best match would be greater for target-present than target-absent lineups. Second, the similarity ratings in Table 4 suggest that suspect-filler similarity in this experiment was not identical across target-present and target-absent lineups. Although I used comparable procedures for constructing target-present and target-absent lineups, suspect-filler similarity for the moderately high similarity condition was rated higher for the target-absent lineup ($M = 5.39$) in comparison to the target-present lineup ($M = 4.39$). The combination of these two factors can be expected to have created a larger difference between the best and the next-best match in the target-present condition than in the target-absent condition, which would explain the asymmetrical effect of similarity on suspect identifications in moderately high similarity lineups.

It should also be noted that diagnosticity ratios for suspect identifications in both similarity conditions (0.69-1.45) were substantially smaller than the average diagnosticity ratio of around 10.00 for high similarity lineups in the Chapter 2 meta-analysis. Note, however, that the size of a diagnosticity ratio is heavily influenced by the innocent

suspect misidentification rate. Therefore, if a lineup or procedure leads to low innocent suspect misidentification rates, larger diagnosticity ratios can be expected (Wixted & Mickes, 2012). Given the high innocent suspect misidentification rates in this experiment, the small diagnosticity ratios should come as no surprise.

Diagnosticity ratios for filler and rejection responses were relatively unaffected by lineup similarity. In both conditions, rejections provided information about the suspect's innocence, whereas filler selections were actually indicative of the suspect's guilt. These results can be explained as a product of the lineup construction procedures.

The diagnosticity values for rejections were above unity, indicating that a rejection was more likely to occur for a culprit-absent lineup than for a culprit-present lineup. Fillers in the culprit-absent condition were matched to the innocent suspect's appearance, as opposed to the common procedure of using the same fillers in culprit-present and culprit-absent lineups. This procedure should have the effect of making fillers in the culprit-present lineup match the representation of the culprit in memory better than the fillers in the culprit-absent lineup. As a consequence, eyewitnesses will be more likely to reject the culprit-absent lineups than the culprit-present lineups.

The diagnostic values for filler selections were below unity, indicating that a filler selection is more likely to occur if the culprit is present than if the culprit is absent. Given that culprit-present fillers were matched to the culprit's appearance, they can be expected to match the memorial representation of the culprit to a greater extent than fillers who were matched to the innocent suspect's appearance. This finding is consistent with previous research showing that filler selections are indicative of the suspect's innocence

for the match-to-description procedure, but not for the match-to-appearance procedure (Clark & Wells, 2008).

3.3.2 Confidence. According to models of information accumulation (Van Zandt, 2000; Vickers, 1979), confidence in a multiple choice task is a product of the difference between the recognition experience elicited by the chosen and non-chosen options. In the very high similarity condition, the extreme homogeneity of the lineup members could be expected to produce only a small difference in the feeling of familiarity between the culprit and the fillers, which would lead to uncertainty in the witnesses. In the moderately high similarity condition, the culprit could be expected to appear notably more familiar than the fillers and lead to greater certainty than in the very high similarity condition. Consistent with the model of information accumulation, participants in the culprit-present condition were more confident in their identification responses for moderately high similarity lineups than for very high similarity lineups. Similarity had no effect on confidence in the culprit-absent condition, which could indicate that the difference in familiarity between the innocent suspect and the fillers was not sufficiently greater in the moderately high similarity condition than in the very high similarity condition to affect confidence.

3.3.3 Summary and conclusions. The present findings suggest a strategy of matching fillers as closely to the suspect as possible could be problematic. I found that if fillers resemble the suspect too much, the correct identification rate is substantially reduced and the overall innocent suspect misidentification rate is only slightly reduced. Of course, whether investigators in the field could create lineups akin to the very high similarity lineups tested here or not is debatable. Whenever similarity is manipulated

using a methodology that deviates from how police construct lineups, the ecological validity of the findings will be questionable. Nevertheless, the experiment in this chapter was important for validating the morphing methodology and showing that the similarity manipulation works in the same way that more natural similarity manipulations have worked. In the following chapter, I use the morphing methodology to address issues of a more theoretical nature.

CHAPTER 4

Effects of Filler Homogeneity, Suspect-Filler Similarity, and Lineup Presentation

Eyewitnesses have been noted for their susceptibility to demand characteristics. They tend to look for cues in the lineup that provide an indication of which lineup member the administrator wants them to choose (Wells, Lieppe, & Ostrom, 1979). When the suspect appears distinct from the fillers, witnesses may be able to discern the identity of the suspect in the absence of recognition. Accordingly, a long-standing recommendation in the eyewitness identification literature is to "compose the lineup in such a manner that the suspect does not unduly stand out" (p. 32, Technical Working Group for Eyewitness Identification, 2003). In previous chapters, I have focused on the biasing effect of fillers who are low in similarity to the suspect. However, additional factors could also cause a suspect to stand out, such as how the suspect is clothed (Lindsay, Wallbridge, & Drennan, 1987) or how the suspect is photographed (Buckhout & Friere, 1975), relative to other lineup members. The suspect might also stand out if the fillers have commonalities that are not shared by the suspect. Thus, this chapter explores the potentially biasing effect of filler homogeneity.

Filler homogeneity could manifest in a variety of ways. Telephone interviews with police agency representatives from across the United States revealed that police officers are commonly called upon to serve as the lineup fillers (Police Executive Research Forum, 2013). Unless the suspect is also a police officer, this practice could conceivably provide cues about the identity of the suspect because of characteristics that are common among police officers. These commonalities may derive from a variety of circumstances: people with a certain type of upbringing might be more or less interested

in a career in policing; police agencies might be more or less inclined to select certain types of people when hiring officers; or police agencies might have codes of conduct that dictate how their officers should appear. Although these commonalities may not be immediately recognizable to the person conducting the lineup, the implication is that when police officers are used as fillers the suspect does not belong to the same group as the rest of the lineup members.

The experiment in this chapter provides an empirical examination of filler homogeneity bias. I manipulated filler homogeneity bias using a similar methodology as was used in the previous chapter. However, this time I created a homogeneous set of fillers by morphing each of them with the same face. As the only lineup member who was not morphed with that face, the suspect was anticipated to stand out from the fillers. In addition to their degree of homogeneity, fillers were also manipulated according to their similarity with the suspect. In particular, fillers were designed to be of "low" and "moderately low" similarity to the culprit. In the Chapter 3 experiment, participants made more correct identifications from lineups with moderately high similarity fillers than from lineups with very high similarity fillers; however, even in the moderately high similarity condition the correct identification rate was below 50%. Thus, in addition to allowing for an exploration of interactions between suspect-filler similarity and filler homogeneity, this experiment allows for an examination of how lower levels of suspect-filler similarity affect identification responses. Finally, given recent suggestions that a suspect would stand out more in a simultaneous lineup than in a sequential lineup (Carlson et al., 2008), lineup presentation was included as an additional manipulation.

4.1 Potential Influences of Lineup Presentation on Filler Homogeneity Effects

The homogeneous lineups were composed of five fillers who were all morphed with the same face and one suspect who was not morphed with that face. This was designed to make the suspect stand out from the fillers. Carlson et al. (2008) suggested that lineup members who stand out should be more susceptible to false identification when presented simultaneously than when presented sequentially because of the increased ease with which relative judgements can be made for simultaneous lineups. This will be referred to as the simultaneous stand-out hypothesis. Data supporting the simultaneous stand-out hypothesis would support previous assertions that sequential presentation reduces the effect of biasing procedures (c.f., Lindsay et al., 1991). However, an alternative theoretical account, referred to as the sequential contrast hypothesis, predicts the opposite effect.

The sequential contrast hypothesis draws upon knowledge from the psychophysical and social psychological literatures to predict that homogeneous fillers will produce higher innocent suspect misidentification rates for lineups presented sequentially than for lineups presented simultaneously. This alternative hypothesis relies on a mechanism known as the contrast effect. The contrast effect is a perceptual error that occurs when items that are presented in sequence appear more different than they actually are (Holland & Lockhead, 1968). For example, lifting a 20 lb weight prior to lifting a 10 lb weight makes the 10 lb weight feel lighter than if it had not been preceded by the heavier weight (Cialdini, 2009). In the social psychology literature, contrast effects have been implicated as underlying mechanisms in a variety of sequential compliance-gaining

techniques (e.g., Miller, Seligman, Clark, & Bush, 1976). Contrast effects have also been used to explain perceptual judgements of faces (e.g., Hsu & Yang, 2013).

Research on context effects in physical attractiveness judgements is particularly relevant to the present research. Geiselman, Haight, and Kimata (1984) instructed participants to judge the attractiveness of faces in several experiments. In the first three experiments, faces were rated less attractive when presented with unattractive faces and were rated more attractive when presented with attractive faces. Thus, rather than a contrast effect, the faces were perceived as more similar to the faces with which they appeared. This is referred to as assimilation. Note, however, that the faces in these experiments were presented simultaneously. In a fourth experiment, Geiselman and colleagues found assimilation for faces presented simultaneously and found contrast effects for faces presented sequentially. Thus, contrary to the effects found with simultaneous presentation, sequentially-presented faces were perceived to be more attractive when presented after unattractive faces and were perceived to be less attractive when presented after attractive faces (see also Wedell, Parducci, & Geiselman, 1987).

The sequential contrast hypothesis predicts that assimilation will occur in the simultaneous homogeneous lineups and that contrast will occur in the sequential homogeneous lineups. Of course, to make this prediction it is necessary to assume that relative judgements have an influence on sequential lineup decisions. Given that the sequential lineup was designed to curb relative judgement strategies (Lindsay & Wells, 1985; Wells, 1984), this assumption is somewhat controversial.

Although previous researchers have noted that systematically comparing all lineup members is more difficult for sequential than simultaneous lineups, sequential

presentation might promote another type of relative judgement. Specifically, when lineup members are presented sequentially, relative judgments could be expected to focus on comparisons between the lineup member currently in view and the lineup member most recently in view. Such comparisons could be particularly sensitive to changes or similarities in the appearance of one lineup member to the next. For example, if one sequentially-presented lineup member is judged to be a poor match to the stored representation of the culprit and that lineup member is followed by a lineup member with even a moderate match to the stored representation of the culprit, a contrast effect could occur and the difference between the two matches could be perceived as greater than they actually are. If the difference between these two lineup members is perceived to be sufficiently large, this could facilitate the identification of the lineup member with the greater match.

4.2 Method

4.2.1 Participants. The final sample consisted of 814 participants. Of these participants, 468 were online community members and 346 were University of Regina students. The internet community members were recruited from *Cint*'s online sampling service and compensated according to a marketplace points system, which rewards participants based on the time required to complete the task. The points can be redeemed for either cash or merchandise. The University of Regina students were recruited from a psychology department participant pool and compensated with partial course credit. The University of Regina students either completed the task in the lab (n = 224) or online (n = 122).

The demographic characteristics of the two samples had similarities and differences. For both samples, the majority of participants were women (University of Regina = 76%; online community = 71%) and White (University of Regina = 69%; online community = 69%). Table 6 provides a detailed breakdown of ethnicity in the two samples. The online community members consisted of younger participants (M_{age} = 29.18, SD = 5.39, range = 18-37 years, n = 262) and older participants (range = 37 years and older, specific ages not reported, n = 187). On average, the University of Regina students (M_{age} = 21.21, SD = 4.44, range = 17-53 years) were younger than both groups of the online community members.

4.2.2 Excluded data. Empirical research suggests that good quality data can be collected from online communities, even with relatively inexpensive compensation rates (e.g., Buhrmester, Kwang, & Gosling, 2011). Nevertheless, participants who complete a task online are not subject to the same controls that are generally in place during laboratory participation. Accordingly, safeguards were developed to flag individuals to exclude for not completing the experiment as instructed. In total, 1182 participants completed the identification task. However, 368 participants appear not to have completed the online version of the task properly and were excluded from the analyses.

A total of 300 participants (298 online community members and 2 University of Regina students) were excluded because they completed the experiment too quickly. The experimental procedure comprised several tasks: reading the study description, providing informed consent, watching the target video, reporting what happened in the video, completing the identification task, answering questions about video content, and providing demographic information. Given that the target video alone was 6 minutes

Table 6

Ethnic backgrounds of participants in the Chapter 4 experiment

	Online Community Members (n = 468)	University of Regina Students (n = 346)
Aboriginal	0.9 %	3.1 %
Asian	5.7 %	9.7 %
Black	13.0 %	5.9 %
Hispanic	7.1 %	1.4 %
Middle Eastern	0.4 %	2.3 %
South Asian	0.0 %	4.8 %
Southeast Asian	0.0 %	0.3 %
White	69.1 %	69.4 %
Mixed	2.4 %	1.1 %
Other	0.9 %	0.0 %
Prefer not to answer	0.5 %	2.0 %

long, the entire experiment was expected to require approximately 15 minutes. Accordingly, all participants who completed the experiment in 10 minutes or less were excluded from the analyses.

As an additional safeguard, the online version of the experiment included questions about the video content. One question was open-ended and three questions were multiple-choice. The open-ended question requested a narrative report of what happened in the video. The multiple-choice questions asked about three specific details of the video: one question for each of the three two-minute segments. The multiple choice questions were used to identify participants who potentially did not watch the entire video. If more than one multiple choice question was answered incorrectly, the data for that participant were excluded from all data analyses. Taking into consideration the possibility that a multiple choice error could reflect a memory error rather than a lack of adherence to task instructions, the responses to the open-ended question of all participants who answered a single multiple-choice question incorrectly were scrutinized. If the response to the open-ended question showed evidence that the entire video was viewed, the participant's data were retained for analysis. If the response to the open-ended question showed any indication that the participant did not attend to the entire video, the participant's data were excluded from all analyses. A total of 68 participants (all online community members) were excluded because their responses to the video content questions suggested they did not sufficiently attend to the video.

4.2.3 Design. This experiment was a 2 (target: present vs. absent) × 2 (suspect-filler similarity: low vs. moderate) × 2 (filler homogeneity: homogeneous vs.

heterogeneous) × 2 (lineup presentation: simultaneous vs. sequential) between-subjects design.

4.2.4 Video event. The video was the same as in the Chapter 3 experiment.

4.2.5 Lineups. The five faces in Figure 1 were altered using *Fantamorph* software to create fillers for lineups that varied in target-presence, suspect-filler similarity, and filler homogeneity. In addition to five fillers, culprit-present lineups contained the man from the video and culprit-absent lineups contained an innocent suspect. The lineup fillers belonged to one of four categories: low similarity homogeneous (Figure 6); low similarity heterogeneous (Figure 7); moderate similarity homogeneous (Figure 8); and moderate similarity heterogeneous (Figure 9).

4.2.5.1 Low similarity homogeneous fillers. These fillers were designed to be moderately low in similarity to the culprit, but high in similarity to each other. A total of five lineups were constructed for this condition using the five faces in Figure 1 and five faces that were judged in Chapter 3 pilot-testing to have low similarity to the culprit (ratings ranged from 0.8-1.2 out of 10). For each lineup, the fillers were created by morphing the Figure 1 faces with a single low similarity face. All morphs were 50% Figure 1 face and 50% low similarity face. Given that these fillers were a mix of faces of low and moderate similarity to the culprit, their precise similarity to the culprit would be best described as "moderately low". However, for brevity, I refer to them as "low similarity" for the remainder of the chapter.

4.2.5.2 Low similarity heterogeneous fillers. As with the previously described condition, low similarity heterogeneous fillers were designed to have moderately low similarity to the culprit; however, contrary to the previously described condition, these

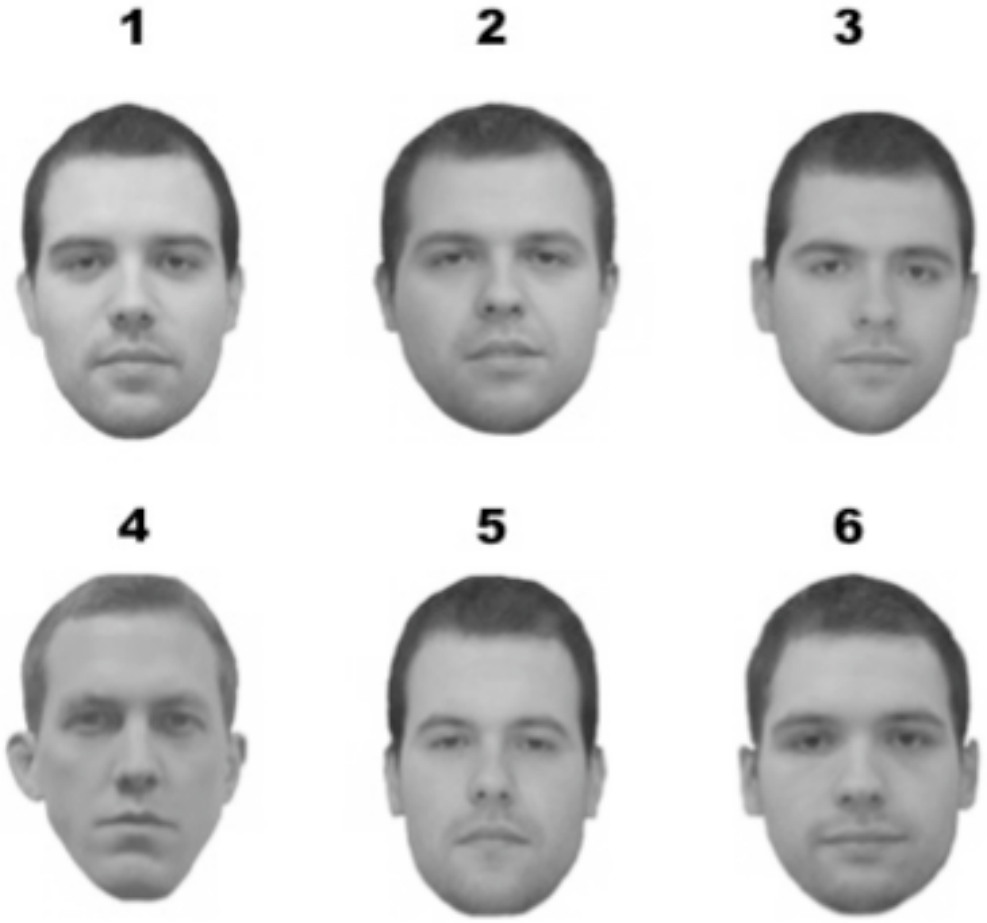

Figure 6. Example of a lineup from Chapter 4 with homogeneous fillers of low similarity to the culprit. In this lineup, the target is in Position 4. In the experiment, suspect position was counterbalanced.

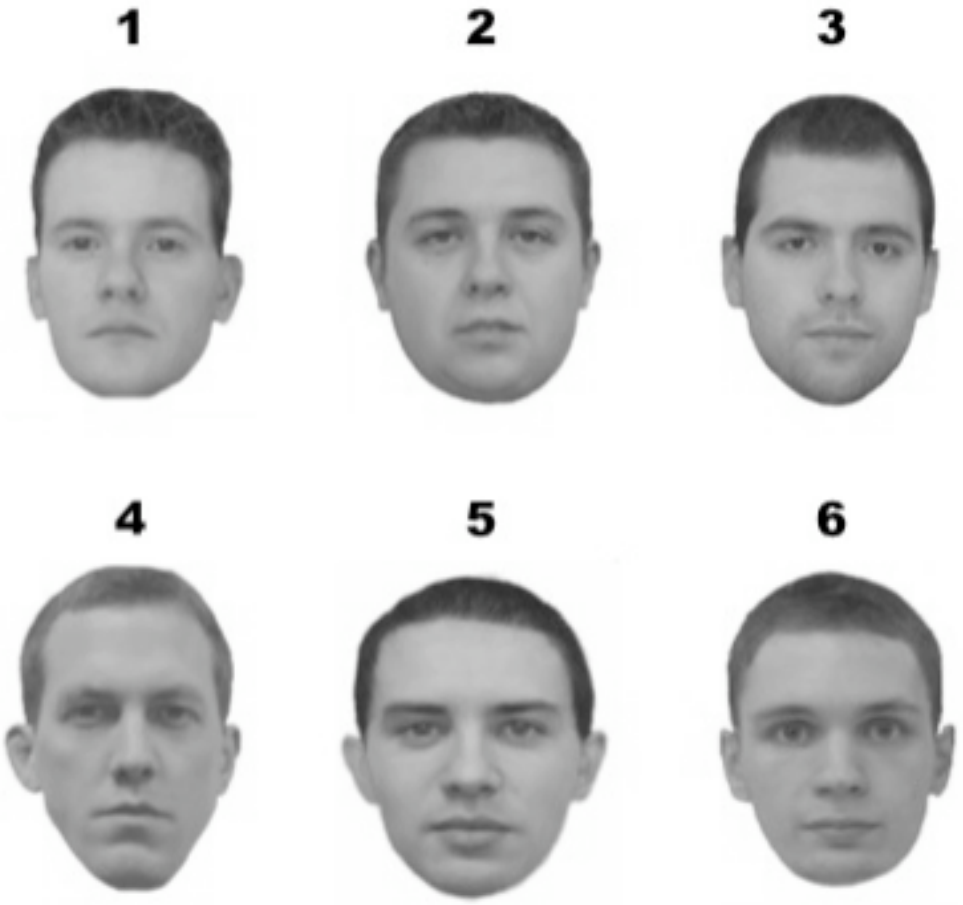

Figure 7. Example of a lineup from Chapter 4 with heterogeneous fillers of low similarity to the culprit. In this lineup, the target is in Position 4. In the experiment, suspect position was counterbalanced.

Figure 8. Example of a lineup from Chapter 4 with homogeneous fillers of moderate similarity to the culprit. In this lineup, the target is in Position 4. In the experiment, suspect position was counterbalanced.

Figure 9. Example of a lineup from Chapter 4 with heterogeneous fillers of moderate similarity to the culprit. In this lineup, the target is in Position 4. In the experiment, suspect position was counterbalanced.

fillers were designed have a relatively moderate degree of variation from one another. This was achieved by constructing a single lineup that comprised one filler from each of the five lineups in the low similarity homogeneous condition.

4.2.5.3 Moderate similarity homogeneous fillers. The fillers in this condition were designed to have moderate similarity to the culprit and high similarity to each other. This was achieved by repeating the process used for the low similarity homogeneous condition with the exception that the Figure 1 faces were morphed with a face of moderate similarity to the culprit (ratings ranged from 3.0-3.2 out of 10). Thus, five lineups were created and each comprised the Figure 1 faces morphed with the same moderate similarity face. Again, all morphs were 50% Figure 1 face and 50% low similarity face.

4.2.5.4 Moderate similarity heterogeneous fillers. These fillers were designed to have moderately high similarity to the culprit and to vary moderately from one another. This was achieved by constructing a single lineup that comprised one filler from each of the five lineups in the moderate similarity homogeneous condition.

4.2.5.5 Innocent suspect. In the Chapter 3 experiment, the innocent suspect was selected quite frequently even when placed in a lineup with fillers of very high similarity. The fact that the innocent suspect's face was morphed with the culprit's face likely contributed to this effect. Given the plan to construct lineups with low and moderate similarity fillers, an innocent suspect who resembles the culprit so strongly could be expected to yield even higher innocent suspect misidentification rates. Accordingly, a new innocent suspect was created for this experiment.

Figure 10. The lineup suspects for the Chapter 4 experiment. The innocent suspect was the suspect in culprit-absent lineups and the target person was the suspect in culprit-present lineups.

The objective for the new innocent suspect was to create a face that was more similar to the culprit than the fillers who would be in the lineup, but not as similar as the innocent suspect used in the Chapter 3 experiment. This was achieved by morphing the faces of two individuals (50% each) who were judged to be of high similarity to the culprit (Figure 10). The high similarity faces were selected using the pairwise similarity ratings obtained for the culprit in the Chapter 3 experiment. One face was rated as a 4.40 (out of 10) and the other face was rated as a 5.40 (out of 10). Although these ratings would be considered moderate on an absolute scale, they correspond with ratings assigned to high similarity fillers in the lineups that were meta-analysed in Chapter 2.

4.2.6 Similarity ratings. I obtained subjective similarity ratings to verify that the similarity manipulations produced effective differences in similarity. Judges (n = 8) provided pairwise ratings of similarity between the culprit and (a) the innocent suspect, (b) the 50 fillers used in this experiment, and (c) the fillers and the innocent suspect from the Chapter 3 experiment. All ratings were made on a scale from 0 (not at all similar) to 10 (highly similar).

The similarity ratings suggest the morphing manipulations worked as intended. Table 7 presents the average culprit-filler similarity ratings for stimuli used in this experiment. Relative to fillers in the low similarity condition, the fillers in the moderate similarity condition received higher ratings of similarity to the culprit. This was true for fillers used in the heterogeneous condition and for fillers used in the homogeneous condition. According to Cohen's (1977) guidelines, all effect sizes are large.

Similarity ratings for the Chapter 3 stimuli were collected to provide context for the ratings of the new stimuli. Given that the ratings for the Chapter 3 experiment are

Table 7

Mean (SD) ratings of similarity between the culprit and the fillers in the Chapter 4 experiment

		Similarity Condition				Cohen's *d* and 95% CIs		
Condition	Filler	Low	Moderate	*t*	*p*	*d*	Lower Limit	Upper Limit
Heterogen.	A	2.50 (1.60)	3.38 (1.51)	1.99	.087	0.88	0.22	2.37
	B	2.25 (1.98)	3.50 (1.60)	1.42	.199	1.56	0.45	2.93
	C	1.75 (0.88)	4.25 (2.38)	2.89	.023	3.43	1.78	4.04
	D	2.13 (1.25)	3.25 (1.39)	1.93	.094	1.90	0.93	2.76
	E	2.25 (1.28)	3.88 (2.17)	2.60	.035	2.11	0.61	3.00
	Average	2.18 (1.26)	3.65 (1.28)	3.36	.012	2.59	1.70	3.46
Homogen.	A	2.57 (0.86)	3.88 (1.57)	2.50	.041	2.41	1.32	3.01
	B	2.92 (1.61)	4.38 (1.17)	4.56	.003	2.35	1.54	3.46
	C	2.52 (0.80)	3.83 (1.40)	2.52	.039	2.66	1.69	3.22
	D	2.43 (0.77)	3.80 (1.28)	4.72	.002	2.99	2.10	3.52
	E	2.65 (0.88)	3.80 (1.20)	2.69	.031	2.47	1.64	3.08
	Average	2.62 (0.87)	3.94 (1.13)	3.14	.004	2.95	2.17	3.55

Note: Scale ranged from 0 (Not at all similar) to 10 (highly similar). All comparisons had 7 degrees of freedom.

most comparable to the heterogeneous lineup fillers used in the present research, only the heterogeneous lineups were used for comparison. The ratings in Table 8 suggest lineups in the present research have lower culprit-filler similarity than the lineups that were used in Chapter 3. Paired-sample t-tests indicate the average culprit-filler similarity ratings for the low similarity lineup were significantly lower than that ratings for the moderately high similarity lineup, $t(7) = 4.53$, $p = .003$, $d = 4.22$ [95% CI = 3.48, 5.10], and the very high similarity lineup, $t(7) = 11.47$, $p < .001$, $d = 11.01$ [95% CI = 10.09, 11.88]. The average culprit-filler similarity ratings for the moderate similarity lineup are also significantly lower than the ratings for the moderately high similarity lineup, $t(7) = 3.25$, $p = .014$, $d = 1.37$ [95% CI = 0.63, 2.26], and the very high similarity lineup, $t(7) = 12.90$, $p < .001$, $d = 8.40$ [95% CI = 7.47, 9.28]. Finally, the average similarity rating between the culprit and the new innocent suspect ($M = 4.13$, $SD = 2.30$) are significantly lower than the average rating for the Chapter 3 innocent suspect ($M = 8.25$, $SD = 1.68$), $t(7) = 3.78$, $p = .007$, $d = 4.63$ [95% CI = 3.47, 6.22]. Thus, relative to the Chapter 3 experiment, the fillers and the innocent suspect in the present research are less similar to the culprit.

Recall that in Chapter 3, fillers were created by morphing faces with the innocent suspect and, consequently, different fillers were used in culprit-present and culprit-absent lineups. This approach corresponds with how culprit-absent lineups would be constructed in the field if the match-to-suspect strategy were utilized. In this experiment, culprit-absent lineups were created simply by replacing the culprit with the innocent suspect. As a result, the same fillers were used for culprit-present and culprit-absent lineups. Although this approach would not correspond with how a match-to-suspect strategy for

Table 8

Mean (SD) ratings of culprit-filler similarity for heterogeneous lineups across Experiments 3 and 4

Filler	Chapter 4 Lineups		Chapter 3 Lineups	
	Low	Moderate	Moderately High	Very High
A	2.50 (1.60)	3.38 (1.51)	6.25 (3.01)	9.75 (0.46)
B	2.25 (1.98)	3.50 (1.60)	5.75 (2.25)	7.63 (2.77)
C	1.75 (0.88)	4.25 (2.38)	2.88 (1.64)	8.00 (2.07)
D	2.13 (1.25)	3.25 (1.39)	3.50 (1.30)	9.25 (2.12)
E	2.25 (1.28)	3.88 (2.17)	3.50 (2.82)	8.13 (1.45)
Average	2.18 (1.26)	3.65 (1.28)	4.38 (1.07)	8.55 (1.33)

Note: Scale ranged from 0 (Not at all similar) to 10 (highly similar).

an innocent suspect would be implemented in the field, there is greater experimental control when changing only one lineup member as opposed to changing all of them (Clark & Tunnicliff, 2001). Given the highly artificial manipulation of similarity, the culprit-absent lineups arguably would not have corresponded with how suspect-matched lineups are constructed in the field even if the fillers had been matched to the suspect.

When the culprit-absent fillers were created, their similarity to the innocent suspect was not taken into account. Nevertheless, I was interested in whether the manipulation of culprit-filler similarity affected the similarity between the fillers and the innocent-suspect or not. To my knowledge, this question has never been empirically tested. Clark and Tunnicliff (2001) conducted the only study that sheds some light on the issue. In that study, fillers who were matched to the innocent suspect were rated as more similar to the innocent suspect than were fillers who were matched to the culprit. However, these data do not specifically address the question of whether culprit-filler similarity manipulations have an influence on similarity between the fillers and the innocent-suspect or not.

Clark and Tunnicliff (2001) suggest that if the culprit and the innocent suspect are highly similar, the similarity between the fillers and the innocent suspect should be relatively unaffected by whether fillers are matched to the culprit or the innocent suspect. The culprit and the innocent suspect in this experiment were judged to be fairly similar, so it is possible that the culprit-filler manipulation would have the same effect on innocent suspect-filler similarity. To examine this possibility, I recruited another set of judges (n = 7) to make pairwise ratings of similarity between the innocent suspect and the fillers. Table 9 suggests that some (but not all) of the fillers in the moderate similarity

Table 9

Chapter 4 mean (SE) similarity ratings between the innocent suspect and the fillers

		Similarity Condition				Cohen's *d* and 95% CIs		
Condition	Filler	Low	Moderate	*t*	*p*	*d*	Lower Limit	Upper Limit
Heterogen.	A	3.00 (2.08)	5.00 (1.29)	2.23	.068	2.65	1.70	4.19
	B	4.86 (3.34)	5.29 (2.21)	0.34	.747	0.35	-1.29	2.82
	C	3.71 (2.43)	3.00 (2.00)	0.67	.526	-0.72	-2.20	1.08
	D	3.71 (1.50)	4.00 (1.83)	0.55	.604	0.39	-0.97	1.50
	E	3.43 (1.99)	4.86 (1.86)	2.50	.047	1.66	0.28	3.14
	Average	3.74 (1.36)	4.43 (1.30)	2.20	.070	1.16	0.73	1.61
Homogen.	A	3.34 (1.41)	4.46 (1.23)	2.72	.034	1.90	0.99	2.94
	B	3.51 (1.66)	4.82 (1.21)	1.90	.106	2.04	1.14	3.27
	C	3.94 (1.43)	3.91 (1.23)	0.09	.934	-0.05	-0.96	1.01
	D	4.22 (1.31)	3.91 (1.42)	0.71	.502	-0.51	-1.56	0.46
	E	3.82 (2.02)	6.11 (0.83)	4.51	.004	3.59	2.98	5.09
	Average	3.77 (1.44)	4.64 (1.07)	2.92	.027	1.55	0.76	2.62

Note: Scale ranged from 0 (Not at all similar) to 10 (highly similar). All comparisons had 6 degrees of freedom.

condition were judged to be more similar to the innocent suspect relative to the fillers in the low similarity condition. Although the increase in similarity to the innocent suspect for fillers in the low and moderate conditions was not nearly as large as the increase in similarity to culprit, the ratings averaged across Fillers A-E suggest that similarity to the innocent suspect was generally higher in the moderate similarity condition than in the low similarity condition. For consistency, this variable is referred to as suspect-filler similarity.

4.2.7 Procedure. As with the Chapter 3 experiment, all participants were recruited for a study about gender roles. The basic procedure involved watching a 6-minute video, recalling what happened in the video, learning about the true purpose of the experiment, and completing a lineup identification task. Participants were randomly assigned to the simultaneous and sequential conditions and the computer program that administered the lineups randomly assigned participants to the various lineup composition conditions.

The procedure for lab participants in the simultaneous condition was very similar to the procedure described in Chapter 3. The only exception was that in this experiment, the suspect never appeared in the top-left position of the lineup. That is, the suspect's face was rotated between five positions rather than six. When presenting lineups sequentially, administrators typically avoid putting the suspect in the first position (e.g., Memon, Havard, Clifford, Gabbert, & Watt, 2011). This may be done to avoid criticism by a defence lawyer for presenting their client first (Wells, Steblay, & Dysart, 2014). Although the suspect's position might not be as critical in simultaneous lineups, this

change was implemented to make the simultaneous condition comparable with the sequential condition.

The procedure for the lab participants in the sequential condition was the same as for the lab participants in the simultaneous condition until it was time to explain the lineup task. The experimenter instructed participants in the sequential condition that they would be viewing faces one at a time and that for each face they would be asked for a yes/no response to the question, "Is this the man from the video?". The experimenter explained that each face would only be shown once and that the decision for each face would be considered a final decision. The experimenter further clarified that it would not be possible to move back and forth between faces. Participants were informed that if they responded "yes" to a face, the remaining faces would continue to be shown but the experimenter warned that only the first "yes" would count. The final instruction for participants in the sequential condition (and in the simultaneous condition) was that the man from the video may or may not be in the lineup. Participants in the sequential condition were not informed about how many lineup members they would be viewing. After each yes/no decision, participants in the sequential condition were asked for a confidence assessment (scale of 1 to 6). After decisions had been made for all six faces, participants in the sequential condition were asked for an overall confidence assessment for all of their identification decisions.

The procedure for the online participants was made as comparable as possible to the lab procedure. However, some potentially important differences should be noted. For example, after viewing the video, participants in the lab conditions were asked to orally report what happened in the video. The decision to have participants report this

information orally was intended to keep the procedures in this experiment as similar as possible to the procedures in the Chapter 3 experiment. In the online version, participants were asked to "Report what happened in the video. Provide as much detail as necessary. Point form is acceptable." Because of limitations in how online responses could be collected, online community members typed their response. The only other major difference between the online and lab tasks is that online participants were asked three multiple-choice questions about the video after the lineup task was completed.

4.3 Results

Identification responses were categorized as suspect identifications, filler identifications, or lineup rejections. A suspect identification was either an identification of the target person from a culprit-present lineup (correct identification) or an identification of an innocent suspect from a culprit-absent lineup. Filler identifications were erroneous selections of any lineup member other than the suspect. If none of the lineup members was selected, the lineup was rejected. A lineup rejection is the correct decision for culprit-absent lineups and an incorrect decision on culprit-present lineups. In the sequential condition, participants made decisions for all six lineup members and in some cases more than one lineup member was selected. If two lineup members were selected, the first identification was categorized as the final decision. A small number of participants (n = 11) were excluded from the analyses because they selected more than two lineup members. As a consequence, 803 participants were included for analysis.

As a check of the filler homogeneity manipulation, the dispersion index (MacLennan, 2002) was computed for each of the 16 experimental conditions (Table 10). If homogeneity was indeed greater in the homogeneous filler condition than in the

Table 10

Dispersion index (DI) values for the Chapter 4 experiment

Filler Homogeneity	Target	Suspect-Filler Similarity	Lineup Presentation	DI
Heterogeneous	Present	Low	Simultaneous	.39
			Sequential	.85
		Mod	Simultaneous	.68
			Sequential	.79
	Absent	Low	Simultaneous	.67
			Sequential	.42
		Mod	Simultaneous	.78
			Sequential	.81
			Mean	.67
Homogeneous	Present	Low	Simultaneous	.94
			Sequential	.84
		Mod	Simultaneous	.83
			Sequential	.83
	Absent	Low	Simultaneous	.77
			Sequential	.94
		Mod	Simultaneous	.85
			Sequential	.85
			Mean	.86

Table 11

Identification response probabilities (P) for the Chapter 4 experiment (N = 803)

Target	LP	Heterogeneous Lineups													
		Low Similarity							Moderate Similarity						
		Suspect		Filler		Reject			Suspect		Filler		Reject		
		P	*SE*	P	*SE*	P	*SE*	*n*	P	*SE*	P	*SE*	P	*SE*	*n*
Present	Sim	.39	.08	.31	.07	.31	.07	39	.40	.07	.19	.05	.42	.07	53
	Seq	.33	.07	.29	.06	.37	.07	51	.19	.06	.36	.07	.45	.07	42
Absent	Sim	.44	.07	.22	.06	.35	.07	46	.23	.06	.23	.06	.54	.08	43
	Seq	.37	.07	.08	.04	.55	.07	51	.20	.05	.39	.07	.41	.07	54
		Homogeneous Lineups													
Present	Sim	.22	.06	.35	.06	.44	.07	55	.26	.06	.26	.06	.49	.07	47
	Seq	.36	.06	.30	.06	.34	.06	61	.38	.08	.38	.08	.24	.07	42
Absent	Sim	.26	.06	.26	.06	.49	.07	51	.31	.06	.33	.06	.36	.07	55
	Seq	.38	.06	.29	.06	.33	.06	58	.33	.06	.35	.06	.33	.06	55

Notes: LP = Lineup Presentation; Sim = Simultaneous; Seq = Sequential.

heterogeneous filler condition, filler identifications should be more randomly distributed in the homogeneous condition. Consistent with this prediction, dispersion index values were typically greater for lineups with homogeneous fillers (M = .86) than for lineups with heterogeneous fillers (M = .67), $t(7) = 2.77$, $p = .02$, $d = 1.48$ [95% CI = 1.44, 1.60].

A contingency table containing the identification response data is provided in Table 11. To identify any significant associations, I performed a 2 (culprit: present vs. absent) × 2 (filler homogeneity: homogeneous vs. heterogeneous) × 2 (suspect-filler similarity: low similarity vs. moderate similarity) × 2 (lineup presentation: simultaneous vs. sequential) × 3 (identification response: suspect vs. filler vs. rejection) hierarchical log-linear (HILOG) analysis. HILOG analysis produces likelihood ratios that test for global effects at each level of the hierarchy as well as partial associations that test each possible interaction for the lower levels of the hierarchy. The global test for the 5-way interaction (i.e., the highest order effect) was nonsignificant, $\chi^2(2) = 4.68$, $p = .10$, which suggests the model that retained all factors did not provide an adequate fit for the data. The global test for a 4-way interaction was also nonsignificant, $\chi^2(11) = 11.92$, $p = .37$. The highest order effect that reached significance was a 3-way interaction, $\chi^2(27) = 41.50$, $p = .04$. This suggests at least one of the 3-way interactions was significant. Examination of the partial associations revealed two significant 3-way interactions.

One of the significant interactions involved suspect-filler similarity, lineup presentation, and identification response, $\chi^2(2) = 6.25$, $p = .04$. For this interaction, the effect of suspect-filler similarity on filler selection rates depended on how the lineups were presented. For sequentially-presented lineups, the filler selection rate was

significantly higher for moderate similarity lineups (37%) than for low similarity lineups (24%), $z = 2.73$, $p = .006$, OR = 1.80 [1.18, 2.75]. By contrast, suspect-filler similarity had no effect on filler selections for lineups presented simultaneously (moderate similarity = 25%; low similarity = 28%).

The other significant interaction involved filler homogeneity, lineup presentation, and identification response, $\chi^2 (2) = 10.57$, $p = .005$. This effect was primarily driven by opposing effects of lineup presentation on suspect identifications in heterogeneous and homogeneous lineups. In the heterogeneous condition, the suspect identification rate was higher for simultaneous presentation (37%) than for sequential presentation (28%), $z = 1.81$, $p = .07$, OR = 0.17 [0.05, 0.60]. Conversely, in the homogeneous condition, the suspect identification rate was lower for simultaneous presentation (26%) than for sequential presentation (36%), $z = 2.26$, $p = .02$, OR = 0.62 [0.41, 0.94]. Note that this interaction did not involve the culprit-presence variable, which suggests the effect occurred regardless of whether the suspect was the target person or the innocent suspect.

4.3.1 Diagnosticity ratios and signal detection theory analyses. To obtain an overall sense of how the manipulations affected discrimination between the target person and the innocent suspect, signal detection theory was used to calculate *d'*. Diagnosticity ratios were also calculated for comparative purposes. Two general observations can be gleaned from these calculations (see Table 12). The first observation is that the two indices led to similar results. The signal detection measure has been recommended as an alternative to the diagnosticity ratio because the diagnosticity ratio can become inflated if responses are overly conservative (Mickes et al., 2014). One explanation for the consistency between the two measures in this experiment is that the innocent suspect was

Table 12

Diagnosticity ratios and signal detection theory analyses (N = 803)

	Heterogeneous Lineups											
	Low Similarity						Moderate Similarity					
	Diagnosticity			Signal Detection			Diagnosticity			Signal Detection		
LP	DR	LL	UL	*d′*	LL	UL	DR	LL	UL	*d′*	LL	UL
Sim	0.88	0.53	1.48	-0.12	-0.19	-0.06	1.70	0.90	3.22	0.47	0.41	0.52
Seq	0.89	0.53	1.51	-0.11	-0.16	-0.05	0.94	0.41	2.12	-0.05	-0.11	0.01
	Homogeneous Lineups											
Sim	0.86	0.43	1.70	-0.12	-0.18	-0.06	0.83	0.44	1.55	-0.16	-0.22	-0.10
Seq	0.95	0.59	1.52	-0.05	-0.09	-0.01	1.16	0.68	2.00	0.15	0.09	0.20

Notes: LP = Lineup Presentation; Sim = Simultaneous; Seq = Sequential;

DR = Diagnosticity Ratio; LL = 95% CI lower limit; UL = 95% CI Upper Limit

selected quite frequently. The second observation is that none of the discrimination scores were particularly high. Most of the *d'* scores were below zero, indicating that the innocent suspect misidentification rate was generally higher than the correct identification rate.

The signal detection measures indicated that the innocent suspect was just as likely as the target person to be identified from the lineup. This could explain why none of the interactions revealed by the HILOG analysis involved the culprit-presence variable. That is, if participants were unable to discriminate between a suspect that was guilty and a suspect who was innocent, the manipulated variables would be expected to have similar effects on culprit-present and culprit-absent lineups.

One possible explanation for the low discrimination accuracy scores is that a majority of the participants completed the experiment online. The online participants were only included in the analysis if they took a reasonable amount of time to complete the experiment and performed well on the video content questions. Nevertheless, how the experiment was experienced by the online participants is largely unknown. The controlled environment of a laboratory is designed to produce less noise relative to online data collection, where participants would likely be more susceptible to distractions, interruptions, and other external influences. To explore the possibility that collecting data from online participants contributed to the low *d'* scores, the data from only those who participated in the lab were analysed separately.

4.3.2 Analysis of lab-only data. Tables 13 presents the diagnosticity ratios and discrimination scores for only those who participated in the laboratory version of the present experiment. Contrary to the *d'* scores for the entire sample, which were mostly

Table 13

Diagnosticity ratios and signal detection analyses for lab participants in Chapter 4 (n = 224)

	Heterogeneous Lineups											
	Low Similarity						Moderate Similarity					
	Diagnosticity			Signal Detection			Diagnosticity			Signal Detection		
LP	DR	LL	UL	*d′*	LL	UL	DR	LL	UL	*d′*	LL	UL
Si	1.16	0.54	2.46	0.18	-0.02	0.37	4.57	1.19	17.60	1.33	1.16	1.50
Se	3.80	1.20	12.06	1.26	1.05	.147	1.18	0.30	4.72	0.06	-0.17	0.30
	Homogeneous Lineups											
Si	5.55[3]	0.32	96.97	1.01	0.70	1.33	5.38	0.78	36.96	1.38	1.09	1.66
Se	0.72	0.41	1.45	-0.38	-0.55	-0.20	0.97	0.39	2.40	-0.03	-0.24	0.18

Notes: LP = Lineup Presentation; Si = Simultaneous; Se = Sequential;

DR = Diagnosticity Ratio; LL = 95% CI lower limit; UL = 95% CI Upper Limit

[3] The data for the simultaneously-presented, low similarity, homogeneous lineup were adjusted for computation of the diagnositicity ratio and *d'* score because the innocent suspect misidentification rate was zero. Following Snodgrass and Corwin (1988), 1.0 was added to the denominator and 0.5 was added to the numerator that were used for calculating the correct identification and innocent suspect misidentification rates in this condition.

Table 14

Identification response rates for lab participants in Chapter 4 (n = 224)

Fillers	Target	Lineup Presentation	Suspect Identifications		*n*
			Rate	*SE*	
Heterogeneous	Present	Simultaneous	.56	.08	36
		Sequential	.41	.11	22
		Total	.50	.07	58
	Absent	Simultaneous	.28	.08	29
		Sequential	.19	.07	32
		Total	.23	.05	61
Homogeneous	Present	Simultaneous	.37	.09	21
		Sequential	.46	.08	31
		Total	.42	.06	52
	Absent	Simultaneous	.05	.04	27
		Sequential	.55	.10	26
		Total	.35	.07	53

Note: Data are collapsed across the suspect-filler similarity variable.

negative, the *d'* scores for the lab-only sample are mostly positive and generally indicative that the participants were substantially more likely to select the culprit than they were to select the innocent suspect.

Given the sample size (n = 224), some conditions were collapsed for the analyses involving lab-only participants to increase statistical power. The sequential contrast hypothesis is most concerned with differences in lineup presentation effects on suspect identifications from homogeneous lineups. Accordingly, data were collapsed across the two levels of suspect-filler similarity and the two types of non-suspect identification responses (i.e., filler selections and rejections). The suspect identification rates for the reduced design are provided in Table 14. The complete contingency table can be found in Appendix C.

A 2 (culprit: present vs. absent) × 2 (filler homogeneity: homogeneous vs. heterogeneous) × 2 (lineup presentation: simultaneous vs. sequential) × 2 (identification response: suspect vs. non-suspect) hierarchical loglinear (HILOG) analysis revealed a significant 4-way interaction, χ^2 (2) = 3.94, p = .047. Thus, the model that retained all factors was an adequate fit to the data. Given that the dependent variable was treated as a binary response, chi-square tests were used to explore the nature of the interaction. All expected cell counts exceeded 5, so all assumptions of the chi-square test were met.

In the heterogeneous condition, culprit-presence was the major influence on suspect identifications. In particular, the target person was significantly more likely to be selected from a culprit-present lineup (50%) than was the innocent suspect from a culprit-absent lineup (23%), χ^2 (1) = 9.43, p = .002, OR = 3.36 [1.52, 7.38]. The effect was consistent across the lineup presentation manipulation. The effect for simultaneous

lineups was significant (56% vs. 28%), χ^2 (1) = 5.12, p = .02, OR = 3.28 [1.15, 9.34]; however, the effect for sequential lineups only approached significance (41% vs. 19%), χ^2 (1) = 3.19, p = .07, OR = 3.00 [0.88, 10.25]. Signal detection measures for simultaneous (d' = 0.74) and sequential lineups (d' = 0.66) did not significantly differ.

In the homogeneous condition, lineup presentation was the major influence on suspect identifications. The suspect identification rate was significantly lower for simultaneous presentation (23%) relative to sequential presentation (51%), χ^2 (1) = 8.64, p = .003, OR = 0.16 [0.07, 0.38]. Although the pattern of results was consistent across culprit-present and culprit-absent lineups, the effect was primarily driven by differences in innocent suspect identifications. In the culprit-present lineup, the difference in correct identifications for simultaneous (37%) relative to sequential (46%) presentation was small and nonsignificant, χ^2 (1) = 0.45, p = .50, OR = 0.68 [0.23, 2.06]. By contrast, in the culprit-absent lineups, the difference in innocent suspect misidentifications between the simultaneous (5%) and sequential (55%) conditions was large and significant, χ^2 (1) = 13.87, p < .001, OR = 0.04 [0.01, 0.34]. The discrimination score for simultaneous presentation (d' = 1.33) was significantly greater than the score for sequential presentation (d' = -0.22), G = 2.48, p = .01.

4.4 Discussion

Two competing hypotheses were developed to predict differences in innocent suspect misidentifications in lineups containing homogeneous fillers. Given the increased ease with which comparisons can be made for simultaneous lineups, the simultaneous stand-out hypothesis predicted that an innocent suspect would be at greater risk of false identification when presented simultaneously than when presented sequentially.

Conversely, the central prediction of the sequential contrast hypothesis is that an innocent suspect presented *after* homogeneous fillers would be more likely to be misidentified than a suspect presented *with* homogeneous fillers.

Consistent with the sequential contrast hypothesis, sequential presentation of lineups containing homogeneous fillers was associated with an increased rate of suspect identifications. In the full sample (lab and online participants), this effect was uninfluenced by whether the lineup contained the target person or the innocent suspect. When the online data were included, the participants were just as likely to misidentify the innocent suspect as they were to correctly identify the target person, which could explain why the effect of lineup presentation on suspect identifications was consistent across the culprit-present and culprit-absent conditions.

When the analyses were restricted to only those who participated in the lab, the correct identification rate was typically substantially higher than the innocent suspect misidentification rate. This was particularly true for lineups containing heterogeneous fillers. Regardless of how the heterogeneous lineups were presented, lab participants were more likely to identify the target person than the innocent suspect. However, in the homogeneous condition, lab participants only demonstrated the ability to discriminate between the target and the innocent suspect when lineups were presented simultaneously.

When the homogeneous lineups were presented sequentially, even the lab participants were unable to discriminate between the target and the innocent suspect. More than half of the lab participants falsely identified the innocent suspect from a sequentially-presented lineup containing homogeneous fillers (17/31). By contrast, lab participants rarely chose the innocent suspect from a simultaneous lineup containing

homogeneous fillers (1/21). These data suggest that sequential presentation increased the difference between the innocent suspect and the fillers (contrast effect) and that simultaneous presentation increased the similarity between the innocent suspect and the fillers (assimilation).

4.4.1 A relative judgement strategy for sequential lineups? The data supporting the sequential contrast hypothesis are consistent with the suggestion that sequential presentation allows for relative judgements. However, further research will be needed to explore the nature of relative judgment strategies for sequential presentation, which could be qualitatively different than the kind of relative judgements that occur for simultaneous lineups. Although eye tracking research shows that simultaneous presentation typically does not lead to an exhaustive comparison of all lineup members (Mansour, Lindsay, Brewer, & Munhall, 2009), simultaneous presentation allows witnesses to pick and choose lineup members for comparison.

Sequential presentation imposes limitations on how lineup members can be compared. If, for example, a 9-person lineup were presented sequentially, comparing the 2nd lineup member to the 7th lineup member would require the use of working memory and the ability to suppress the interference created by the intervening lineup members. The ability to make comparisons is particularly restricted when lineup members are only shown once and each decision is considered final. If this strict form of the sequential lineup is implemented, the decision for the lineup member currently in view can only be influenced by the lineup members who have already been viewed. Given these factors, it seems likely that relative judgements for sequential lineups would be focused on the lineup member currently in view and the immediately preceding lineup member.

Although the suggestion that relative judgements play a role in decision-making for sequential lineups is at odds with the majority of the literature on lineup presentation, I am not the first to make such a proposition. Goodsell, Gronlund, and Carlson (2010) initially assumed that simultaneous lineups would promote a relative judgement strategy and that sequential lineups would promote an absolute decision strategy; however, in order to fit Clark's (2003) WITNESS model to existing data, the assumed association between lineup presentation and decision strategy had to be modified. In particular, after rejecting a certain number of sequentially-presented lineups members, witnesses were assumed to switch from an absolute to a relative judgement strategy. Consistent with my interpretation of the present results, Goodsell and colleagues proposed that a sequentially-presented lineup member who is a considerably better match than the immediately preceding lineup member could create a "pop-out effect" and facilitate an identification.

4.4.2 Summary and conclusion. The results suggest that an innocent suspect who stands out because of homogeneous fillers is at an increased risk of false identification if the lineup is presented sequentially. As with the experiment in the previous chapter, it will be important to replicate these findings using nonmorphed faces. Police typically select fillers from mugshots or driver's license photographs (Police Executive Research Forum, 2013), which could cause filler homogeneity biases. For example, given recent findings associated with criminal face biases (Flowe & Humphries, 2011), suspect identification rates might be influenced by placing a suspect (who is likely to have a criminal past) in a lineup with fillers chosen from a driver's licence photograph database (who are less likely to have a criminal past). This is a potentially promising avenue for future research. It would also be worthwhile to examine whether or not mock

witnesses are able to discern the identity of the suspect when police officers are used as fillers. In future research, these and other manifestations of filler homogeneity bias should be explored.

CHAPTER 5

General Discussion

The body of this final chapter is organized into three sections. The first section revisits the definitions of the similarity categories that were used in the Introduction and articulates a new conceptualization of suspect-filler similarity. The second section applies the new conceptualization of similarity to interpret the dissertation findings as a whole. The third section considers policy recommendations. To conclude the chapter, I provide some final considerations.

5.1 Reconceptualizing Suspect-Filler Similarity Categories

As mentioned in the Introduction, a continuous conceptualization of suspect-filler similarity would be most desirable; however, practical constraints suggest a categorical approach will have to suffice for now. In my literature review of suspect-filler similarity research, I referred to three categories of lineups: biased, moderate similarity, and high similarity. These categories were also used for the meta-analysis. However, over the course of this dissertation, it became clear that three categories of similarity are by no means sufficient.

Going forward, I recommend conceptualizing similarity with five categories: biased, low similarity, moderately low similarity, moderately high similarity, and very high similarity. Biased lineups would be those with fillers who are either particularly low in similarity to the suspect or lineups with fillers who do not match an eyewitness description of the culprit. To be categorized in one of the three middle categories (low, moderately low, moderately high), fillers would at the very minimum need to be consistent with an eyewitness description of the culprit. Which of the three categories the fillers would fall into would depend on either the quality of the description, subjective

ratings of suspect-filler similarity, or both. Low similarity fillers would match a relatively vague description of the culprit and/or would be rated to have low similarity to the suspect. Moderately low similarity fillers would match a more detailed description of the culprit and/or would be rated to have moderate similarity to the suspect. Moderately high similarity fillers would match a very detailed description of the culprit and/or would be rated to have high similarity to the suspect. Very high similarity fillers would have the level of similarity that was obtained in the Chapter 3 experiment. At this point, it is not clear whether lineups with very high similarity could be constructed without using artificial means (e.g., morphing), but such fillers would likely need to match an exceptionally detailed description and/or would be rated to have particularly high similarity to the suspect.

5.2 Summary of Experimental Findings on Suspect-Filler Similarity

Given the comparable methodologies in the Chapter 3 and 4 experiments, between-experiment comparisons were examined. Participants were not randomly assigned to these experiments, so the ability to make causal inferences is somewhat limited. However, given that no previous comparison has involved four levels of similarity, examining similarity effects across these two experiments could provide some guidance for future research that is capable of showing causal effects.

The conditions in the Chapter 4 experiment that were most similar to the Chapter 3 experiment involved the heterogeneous lineups that were presented simultaneously. Thus, any lineups that involved homogeneous fillers or sequential presentation were excluded from the cross-experiment comparison. In addition, given that all participants in the Chapter 3 experiment participated in the lab, all online participants from the Chapter

4 experiment were excluded. The lineups from Chapter 4 that were included in the comparison are referred to here as “low similarity” and “moderately low similarity”. The lineups from Chapter 3 are referred to as “moderately high similarity” and “very high similarity.” The ratings in Table 8 suggest these labels are appropriate and that the differences in similarity between conditions are reliable.

Figure 11 presents the identification responses across the four levels of similarity for lineups containing the culprit. The correct identification rates for the low and moderately low similarity lineups are the highest. When similarity is increased to moderately high, the correct identification rate is slightly reduced and the filler identification rate is slightly increased. However, when similarity is increased to very high, the impact is substantially more pronounced. The correct identification rate plummets and the filler identification rate skyrockets. This suggest that when suspect-filler similarity reaches particularly high levels, correct identifications become overly difficult and choices shift toward the very high similarity fillers.

Figure 12 presents the identification responses across the four levels of similarity for culprit-absent lineups containing an innocent suspect. The comparisons across experiments for the culprit-absent lineups are not as clean as for the culprit-present lineups because different innocent suspects were used across experiments, whereas the culprit remained constant across experiments. In the Chapter 4 experiment, what could be considered a very similar innocent suspect was created by morphing a similar face with the culprit’s face. In the Chapter 3 experiment, what could be considered a moderately

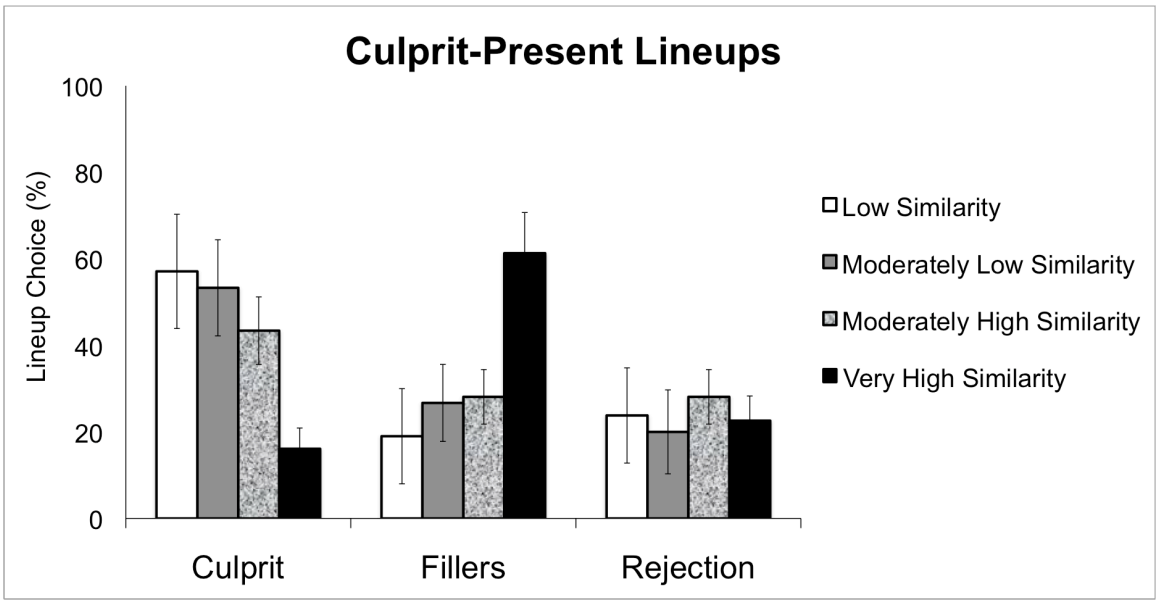

Figure 11. Identification responses across Chapter 3 and Chapter 4 experiments for lineups containing the culprit. All lineups contained the same culprit.

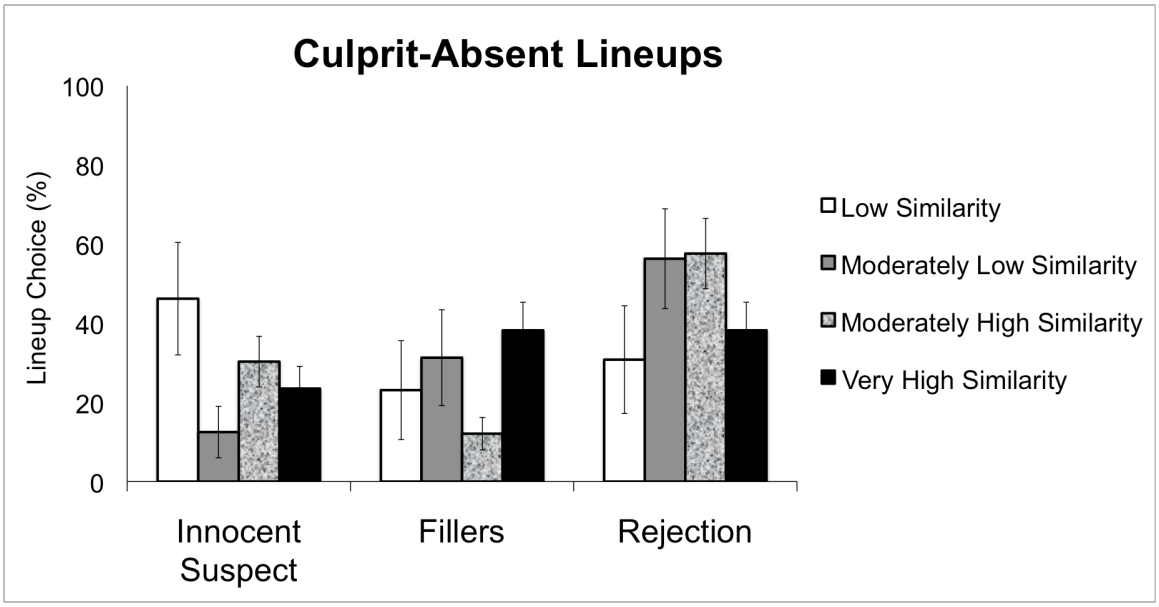

Figure 12. Identification responses across Chapter 3 and Chapter 4 experiments for lineups that did not contain the culprit. The low and moderately low similarity lineups contained an innocent suspect whose similarity to the culprit was moderately high and the moderately high and very high similarity lineups contained an innocent suspect whose similarity to the culprit was very high.

high similarity innocent suspect was created by morphing two faces that were rated as moderately high in similarity to the culprit.

Given that the culprit-absent fillers and the innocent suspect were lower in similarity in the Chapter 4 experiment than in the Chapter 3 experiment, comparing the pattern of results across experiments is more appropriate than comparing responses across the four conditions. From this perspective, the decrease in innocent suspect misidentifications when similarity is increased from low to moderately low is greater than when similarity is increased from moderately high to very high. Taken together, the culprit-present and culprit-absent lineups suggest that low similarity lineups create too much bias towards innocent suspects and very high similarity lineups create too much difficulty for witnesses trying to identify the culprit.

5.3 Policy Considerations

The meta-analysis in Chapter 2 suggests that an innocent suspect is at increased risk of false identification if the lineup contains fillers who do not sufficiently resemble the suspect. However, the Chapter 3 experiment suggests that if fillers resemble the suspect too much, correct identifications will be substantially reduced. Taken together, these findings present a paradox: fillers should be neither very dissimilar nor very similar to the suspect. The Technical Working Group for Eyewitness Evidence (2003) demonstrated awareness of this paradox in their training manual, which first states that the suspect should not stand out and then cautions that fillers should not resemble the suspect too strongly. The objective, then, is to determine the best method of selecting fillers who resemble the suspect, but not overly so. In this section, I consider three

options for achieving this end: (1) matching to the culprit's description; (2) matching to the suspect's appearance; or (3) a combination of the first two options.

5.3.1 Matching to the culprit's description. Luus and Wells (1991) hypothesized that matching fillers only on the features in the witness description of the culprit would lead to fillers who are similar to the suspect, but not so similar that they would impede correct identifications. The match-to-description procedure requires that all lineup members match on what the witness can recall of the culprit's appearance, which forces the witness to rely on recognition memory. The procedure also permits variation among lineup members, which should aid recognition of the culprit.

Although Wells et al. (1993) found a strong advantage of matching to a witness description over matching to the culprit's appearance, these results did not replicate in subsequent investigations (Darling et al., 2008; Juslin et al., 1996; Lindsay et al., 1994; Tunnicliff & Clark, 2000). Even if the encouraging results found by Wells et al. (1993) had replicated, some form of matching to the suspect's appearance would nonetheless be required in many instances because matching to a description is not always practical. If witnesses routinely provided a good description of the culprit, the match-to-description procedure could be a viable option. However, witnesses are notorious for providing vague descriptions, particularly for facial descriptors (Lindsay et al., 1994). If the description is not specific enough, the match-to-description procedure could lead to the selection of fillers who bear little resemblance to the suspect, yet fit the description (Koehnken et al., 1996). Another problem can occur if the eyewitness description does not correspond with the appearance of the suspect (Luus & Wells, 1991; Meissner et al.,

2007). In both instances, matching fillers to the witness description would likely make the suspect stand out.

The Technical Working Group (1999, 2003) indicated a preference for matching to the culprit's description. However, Clark et al. (2013) recently questioned the basis of this preference. They reported meta-analytic results indicating that, relative to matching to the suspect's appearance, matching to the culprit's description typically increases innocent suspect misidentifications and reduces the probative value of suspect identifications. Although I believe matching to the culprit's description is important, I also do not support a recommendation to exclusively match fillers to a description because (a) the empirical findings on its effectiveness have been inconsistent and (b) it is not always practical.

5.3.2 Matching to the suspect's appearance. Surveys consistently show that the majority of investigators in the field (81-83%) select fillers by matching to the suspect's appearance (Police Executive Research Forum, 2013; Wogalter et al., 2004). Although their rationale for this preference was not probed, I suspect the frequency of insufficient descriptions is a contributing factor. In the more recent survey, respondents indicated using the match-to-appearance procedure in different ways. Of the 81% of agencies that reported using a match-to-appearance procedure, 50% indicated using fillers who were matched on "the general characteristics of the suspect" (p. 59) and 31% indicated using "fillers who look as much like the suspect as possible" (p. 59). Presumably, police agencies would be keen to learn which of these two methods is best supported by empirical research.

Arguments could be made in favour of either of these two methods. The meta-analysis in Chapter 2 indicated that relative to moderate similarity lineups, high similarity lineups reduced innocent suspect misidentifications without reducing correct identifications. This finding could be interpreted as support for the construction of lineups with the most similar fillers available; however, this would only be true if the high similarity lineups in the meta-analysis were constructed with the most similar fillers available. It is not clear whether this was the case or not. Moreover, the fact that highly similar lineup members have been shown to reduce correct identifications in some studies (e.g., Chapter 3; Wells et al., 1993) could justify the construction of lineups with fillers who only match on the suspect's general characteristics.

Researchers have also presented arguments in favour of abandoning the match-to-appearance procedure altogether. Luus and Wells (1991) reasoned that if the procedure were taken to its extreme, the lineup would consist of clones and correct identifications would be impossible. In addition to making the identification task too difficult, the match-to-appearance procedure has also been criticized for its potential to increase innocent suspect misidentifications. For example, Wogalter et al. (1992) suggested matching to the suspect's appearance leads to a prototype effect. They proposed that appearance-matching encourages the selection of fillers who resemble the suspect more than they resemble any of the other fillers, which makes the suspect appear distinctive. In support of their claim, participants who had not previously observed a target person were able to identify the suspect from appearance-matched lineups more often than expected by chance. Wogalter et al. concluded that witnesses chose the suspect because he represented the central tendency of the lineup.

The prototype theory is not the only path by which matching to the suspect's appearance may increase innocent suspect misidentifications. Navon (1992) theorized that if an innocent suspect were in the lineup for his match to the culprit's description and fillers were in the lineup for their match to the suspect's appearance, the suspect would resemble the culprit to a greater extent than any of the fillers. Thus, rather than protecting an innocent suspect, the match-to-appearance procedure would "backfire" and bias the lineup towards the innocent suspect. In the Chapter 3 experiment, I simulated this backfire effect by creating lineups with an innocent suspect who was morphed with the culprit and fillers who were morphed with the innocent suspect (and not with the culprit). This manipulation led to quite high innocent suspect misidentification rates, which is consistent with the backfire effect. However, the high innocent suspect misidentification rates might also have been a function of the similarity between the innocent suspect and culprit. Given that high innocent suspect misidentification rates were also observed in the Chapter 4 experiment (when culprit-absent fillers were matched to the culprit's appearance), the strong resemblance of the innocent suspects to the culprit seems to provide the most plausible explanation for the high innocent suspect misidentification rates.

In Chapters 2-4, I consistently found that innocent suspect misidentifications were reduced when the similarity between fillers and the suspect was increased. Given the consistency of the findings for culprit-absent lineups and the finding that correct identifications only seem to be reduced when fillers very strongly resemble the culprit, my dissertation on the whole suggests that matching fillers to the suspect is an effective strategy for protecting innocent suspects. Nevertheless, the principles underlying the

match-to-description procedure are hard to disagree with. If the witness reports features of the culprit's appearance, it seems prudent to ensure that all the lineup members have those features. Depending on how it is specified, the match-to-appearance procedure might result in fillers who match on the features of the witness description. However, 50% of U.S. police officers apply the match-to-appearance procedure by selecting fillers who match the *general* appearance of the suspect. Thus, a pure match-to-appearance procedure might not always produce fillers who match the witness description of the culprit. Given that an eyewitness can instantly discount any filler who does not match their recall of the culprit, I do not support a recommendation to only match fillers to the suspect's appearance.

5.3.3 A combination of methods. In a recent review of lineup construction techniques, Clark et al. (2013) recommend a combination of the match-to-description and match-to-appearance procedures. Although they did not specify how this should be done, the literature provides some options. For example, Gronlund et al. (2009) obtained a pool of fillers who matched the culprit's description and then selected fillers from this pool based on their match to the suspect's appearance. I find the notion of combining the two procedures more compelling than purely matching to a description or purely matching to the suspect's appearance. In particular, I believe that a well-designed combination approach has the most potential for protecting innocent suspects from misidentification, and that this protection of innocent suspects need not come at the expense of low correct identification rates.

A major drawback of the pure description-matching procedure is that witnesses do not always provide a good description. For instance, Lindsay et al. (1994) implicated

poor descriptions as the cause of increases in innocent suspect misidentifications in description-matched lineups relative to appearance-matched lineups. However, if the witness provides a poor description, a combination approach can take into account the aspects of the description that are useful while also incorporating aspects of the suspect's appearance.

One frequently cited problem with the pure appearance-matching procedure is that the suspect and the fillers may be in the lineup for different reasons. That is, the innocent suspect may be in the lineup because he matches the culprit's description and the fillers may be in the lineup because they match the suspect's appearance, rendering a lineup with an innocent suspect who resembles the culprit more than do any of the fillers (Navon, 1992). However, this would not occur with a combination approach because the innocent suspect and the fillers would both be matched to the culprit's description.

Another problem associated with the pure appearance-matching procedure is that the fillers resemble the suspect more than they resemble the other fillers (i.e., the prototype effect; Wogalter et al., 1992). To solve this problem, Wogalter et al. proposed taking the appearance of some fillers into account when selecting other fillers. Turtle et al. (2003) describe a combination approach capable of achieving that very objective: (1) Compile a pool of potential fillers by searching for photographs of individuals who match the witness description of the culprit; (2) Choose the person who most closely resembles the suspect to be Filler 1 and remove the suspect's photograph from view; (3) Choose the person who most closely resembles Filler 1 to be Filler 2 and remove Filler 1 from view; (4) Repeat this procedure until one more than the desired number of lineup fillers have been obtained; (5) Construct a lineup containing all fillers except Filler 1; (6) Compare

each lineup member to the culprit's description and replace any that do not match. This procedure would ensure that each filler resembles the suspect. One potential drawback to this approach is that the fillers might resemble each other too much and create the type of filler bias that I explored in Chapter 4. The iterative nature of this procedure might be sufficient for keeping the fillers relatively heterogeneous (e.g., Filler 2 should not be particularly similar to Filler 8); nevertheless, evaluating the lineup as a whole and ensuring the fillers are not overly homogeneous might be considered as a final step to add to the protocol.

Perhaps the most commonly cited problem associated with the pure appearance-matching procedure is that fillers would resemble the suspect too much. Consistent with this concern, very high similarity fillers in Chapter 3 had an undesirable effect on correct identifications. Three considerations are relevant to this issue. First, I developed the very high similarity lineups using morphing software. Researchers using a match-to-appearance procedure with nonmorphed images have generally not observed correct identifications rates as low as in our very high similarity condition (but see Wells et al., 1993). Second, the meta-analysis of "moderate" and "high" similarity lineups in Chapter 2 and the analysis of lineups ranging from "low" to "moderately high" similarity all revealed identification rates of 45% plus or minus 2%. Obviously, a hit rate of greater than 45% is desirable. However, the consistency of these rates suggests that as long as the lineup is fair, the similarity of the fillers is unlikely to influence whether or not an eyewitness with a good memory of the culprit will make a correct identification. Only with biased lineups is the correct identification rate greater than 50%, suggesting some of these identifications are lucky guesses. Third, and related to the first point, whether

investigators in the field could create lineups akin to the very high similarity lineups from Chapter 3 or not is debatable.

Police investigators typically obtain fillers from databases containing mug shots or drivers license photographs (Police Executive Research Forum, 2013). Although I am not aware of any previous researchers using a database containing drivers' license photographs, researchers have used mugshot databases on occasion. For instance, Gronlund et al. (2009) used a combination of culprit description-matching and suspect appearance-matching to select fillers from the Florida Supervised Offenders database (http://www.dc.state.fl.us/activeoffenders/search.asp), which contains more than 100,000 mug shots. As described in detail in the Introduction, Gronlund et al.'s combination approach used quite rigorous procedures to create "fair" lineups comprised of fillers who resembled the suspect and correct identifications were still relatively commonplace. Of the 409 witness administered a culprit-present lineup in the fair condition, 149 (36%) correctly identified the culprit. This provides empirical evidence that a combined approach does not make the lineup task too difficult.

5.4 Concluding Remarks

Based on the accumulation of knowledge that was produced in this dissertation, I recommend constructing lineups that have a moderately high degree of similarity. This might be achieved by first obtaining a pool of fillers who match a description of the culprit and then choosing fillers who match the appearance of the suspect. If the search for fillers yields a dead-ringer for the suspect, the suggestion by Turtle and colleagues (2003) to omit that person from the lineup is sensible. I suspect the most similar fillers

available will be appropriate in most instances; however, this would almost certainly depend on the size of the filler database.

An important caveat to the conclusions from this work is that my dissertation focused on adult witnesses. The research I have conducted with children suggests they are not as able as adults to identify a target from among similar fillers (Fitzgerald et al., 2014). Further, I am not aware of any published manipulation of lineup member similarity with older adults, but I have started to collect data on similarity effects in older adults and preliminary results suggest they too are less able than young adults to identify a target from lineups with similar fillers.

In terms of future research, a logical direction is to focus on replicating the results reported here using nonmorphed faces. The morphing methodology is undeniably quite different from the typical lineup construction procedures utilized in applied settings, and it could have led to fillers who were fundamentally different than the fillers that would be available to police. Similarly, most researchers using more traditional lineup construction strategies could be critiqued for selecting photographs from relatively small face databases. Police typically have access to hundreds of thousands of mugshots or driver's license photographs (Police Executive Research Forum, 2013). Thus, it is critical to replicate the lab findings using the databases that are available to investigators constructing lineups in the field. Such research would speak to the external validity of experimental findings reported here and provide further evidence that proper lineup construction is fundamental to preventing miscarriages of justice.

REFERENCES

Borchard, E. M. (1932). *Convicting the innocent: Errors of criminal justice.* New Haven, CT: Yale University Press.

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

Bland, J. M., & Altman, D. G. (2000). The odds ratio. *BMJ, 320*, 1468. DOI:10.1136/bmj.320.7247.1468

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97-111. DOI: 10.1002/jrsm.12

Buckhout, R., & Friere, V. (1975). *Suggesting in lineups and photospreads: A casebook for lawyers.* New York, Center for Responsive Psychology. Brooklyn College, Brooklyn, CUNY.

Brandon, R., & Davies, C. (1973). *Wrongful imprisonment: Mistaken convictions and their consequences.* London: George Allen & Unwin.

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*, 77-96. DOI:10.1348/135532509X414765

Brewer, N., & Wells, G. L. (2006). The confidence–accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30. DOI:10.1037/1076-898X.12.1.11

Brigham, J. C., & Brandt, C. C. (1992). Measuring lineup fairness: Mock witness responses versus direct evaluations of lineups. *Law and Human Behavior, 16,* 475-489.

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

Brooks, N. (1983). *Pretrial Eyewitness Identification Procedures: Police Guidelines: a Study Paper*. Law Reform Commission of Canada.

Buhrmester, M. D., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3-5. DOI:10.1177/1745691610393980

Burton, A. M., White, D., & McNeill, A. (2010). 'The Glasgow Face Matching Test'. *Behavior Research Methods*, *42*, 286-291. DOI: 10.3758/BRM.42.1.286

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 14,* 118-128. DOI: 10.1037/1076-898X.14.2.118

Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The dud effect: Adding highly dissimilar fillers increases confidence in lineup identifications. *Law and Human Behavior, 35,* 479-500. DOI: 10.1007/s10979-010-9261-1

Cialdini, R. B. (2009). *Influence: Science and practice* (5th Ed.). Boston: Allyn & Bacon.

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17,* 629-654. DOI: 10.1002.acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29,* 395-424.

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7,* 238-259. DOI: 10.1177/1745691612439584

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29,* 151-172. DOI:10.1007/s10979-005-2418-7

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review, 16,* 22-42. DOI:10.3758/PBR.16.1.22

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 32,* 187-218. DOI:10.1007/s10979-006-9082-4

Clark, S. E., Rush, R. A., & Moreland, M. B. (2013). Constructing the lineup: Law, reform, theory, and data. In B. L. Cutler (Ed.), *Reform of eyewitness identification procedures* (pp. 87-112). Washington, DC: American Psychological Association.

Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behavior, 25,* 199-216.

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32,* 406-422. DOI:10.1007/s10979-007-9115-7

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10,* 101-129.

Cohen, J. (1977). *Statistical power analysis for behavioral sciences* (revised ed.). New York: Academic Press.

Cooper, H. M., Hedges, L. V., & Valentine, J. (2009). *The Handbook of Research Synthesis and Meta-Analysis* (2nd ed.). New York: The Russell Sage Foundation.

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

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

Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S.Ct. 2786 (1993).

Deeks, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine, 21,* 1575-1600. DOI:10.1002/sim.1188

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

Devlin, Lord P. (1976). *Report to the Secretary of State for the Home Department of the Departmental Committee on Evidence of Identification in Criminal Cases.* London: Her Majesty's Stationery Office.

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

Du Cann, C. G. L. (1960). *Miscarriages of justice.* London: Frederick Muller.

Engels, E. A., Schmid, C. H., Terrin, N., Olkin, I., & Lau, J. (2000). Heterogeneity and statistical significance in meta-analysis: An empirical study of 125 meta-analyses. *Statistics in Medicine, 19,* 1707-1728.

Fitzgerald, R. J., Whiting, B. F., Therrien, N. M., & Price, H. L. (2014). Lineup member similarity effects on children's eyewitness identification. *Applied Cognitive Psychology, 28,* 409-418. DOI: 10.1002/acp.3012

Flowe, H., & Bessemer, A. (2011). The effect of target discriminability and criterion placement on accuracy rates in sequential and simultaneous target-present lineups. *Psychology, Crime, & Law, 17,* 587-610. DOI:10.1080/10683160903397540

Flowe, H. D., & Ebbesen, E. B. (2007). The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law and Human Behavior, 31,* 33-52. DOI:10.1007/s10979-006-9045-9

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

Geiselman, R. E., Haight, N. A., & Kimata, L. G. (1984). Context effects on the perceived attractiveness of faces. *Journal of Experimental Social Psychology, 20,* 409–424.

Glass, A. L., Holyoak, K. J., & Santa, J. L. (1979). *Cognition*. Reading, MA: Addison-Wesley.

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior, 34,* 445-459. DOI: 10.1007/s10979-009-9215-7

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika, 32,* 25-33.

Gross, A. (2012). Levon Brooks. *The National Registry of Exonerations*. Retrieved from http://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3058

Gross, S. R. (2013). How many false convictions are there? How many exonerations are there? In C. R. Huff & M. Killias (Eds.), *Wrongful convictions and miscarriages of justice: Causes and remedies in North American and European criminal justice* (pp. 45-59). New York: Routledge.

Gross, S. R., & O'Brien, B. (2007). Frequency and predictors of false conviction: Why we know so little, and new data on capital cases. *Journal of Empirical Legal Studies, 5,* 927-962.

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15,* 140-152. DOI:10.1037/a0015082

Guttman, L., & Schlesinger, E. M. (1967). Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement, 27,* 569-580.

Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics, 17,* 279–296. DOI:10.2307/1165125

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics, 3,* 409–414.

Home Office (2011). Police and Criminal Evidence Act. (1984). Codes of Practice, Code D. Retrieved on September 24, 2014 from

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253831/pace-code-d-2011.pdf

Hsu, S. M., & Yang, L. X. (2013). Sequential effects in facial expression categorization. *Emotion, 13,* 573-586. DOI: 10.1037/a0027285.

Innocence Project. (2014a). Understand the causes: False confessions. Retrieved September 21, 2014, from http://www.innocenceproject.org/understand/False-Confessions.php

Innocence Project. (2014b). Understand the causes: Eyewitness misidentification. Retrieved September 21, 2014, from http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php

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.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological Science, 7,* 125-128.

Koehnken, G., Malpass, R. S., & Wogalter, M. S. (1996). Forensic applications of lineup research. In S. L. Sporer, R. S. Malpass & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 205-231). Mahwah, NJ: Erlbaum.

Langdon, J., & Wilson, P. (2005). When justice fails: A follow-up examination of serious criminal cases since 1985. *Current Issues In Criminal Justice, 17,* 179-202.

Levi, A. M., & Lindsay, R. C. L. (2001). Lineup and photo spread procedures: Issues concerning policy recommendations. *Psychology, Public Policy, and Law, 7,* 776-790.

Lindsay, R. C. L., James, A. L., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76,* 796-802.

Lindsay, R. C. L., Mansour, J. K., Beaudry, J. L., Leach, A. M., & Bertrand, M. I. (2009). Sequential lineup presentation: Patterns and policy. *Legal and Criminological Psychology, 14,* 13-24. DOI:10.1348/135532508X382708

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

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do 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. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4,* 303-313.

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.

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

MacLennan, R. N. (2002). Two statistical approaches to distractor analysis for multiple-choice test items. Unpublished manuscript.

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

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

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*, 330-334.

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). Lineup construction and lineup fairness. In R. Lindsay, D. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology, Volume 2: Memory for people* (155-178). Mahwah, NJ: Erlbaum.

Mansour, J., Lindsay, R. C. L., Brewer, N., & Munhall, K. G. (2009). Characterizing visual behavior in a lineup task. *Applied Cognitive Psychology, 23,* 1012–1026.

McQuiston-Surrett, D., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy, and Law, 12,* 137-169. DOI:10.1037/1076-8971.12.2.137

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

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. E. Ross, J. D. Read, & M. P. Toglia

(Eds.), *Handbook of eyewitness psychology: Memory for people* (Vol. 2), Mahwah, NJ: Erlbaum.

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

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute *d'*, not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition, 3,* 58–62. DOI: 10.1016/j.jarmac.2014.04.007

Miller, R. L., Seligman, C., Clark, N. T., & Bush, M. (1976). Perceptual contrast versus reciprocal concession as mediators of induced compliance. *Canadian Journal of Behavioral Science, 8,* 401-409.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105-125. DOI:10.1037//1082-989X.7.1.105

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

Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identification. *Law and Human Behavior, 17,* 11-26.

Police Executive Research Forum. (2013). *A National Survey of Eyewitness Identification Processes in Law Enforcement Agencies*. Washington, DC: Author. Retrieved from http://policeforum.org/library/eyewitness-identification/NIJEyewitnessReport.pdf

R. v. Hinse. (1997) 1 S.C.R. 3

Radin, E. D. (1964). *The innocents*. New York: William Morrow & Company.

Read, J. D., Tollestrup, P., Hammersley, R., McFadzen, E., & Christensen, A. (1990). The unconscious transference effect: Are bystanders ever misidentified? *Applied Cognitive Psychology, 4,* 3-31.

Risinger, D. M. (2007). Innocents convicted: An empirically justified factual wrongful conviction rate. *Journal of Criminal Law and Criminology, 97,* 761–807.

Rubin, A. (2007). *Statistics for evidence-based practice and evaluation* (1st ed.). Belmont, CA: Thompson Brooks/Cole.

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

Schooler, J. W. (2011). Unpublished results hide the decline effect. *Nature, 470,* 437. DOI:10.1038/470437a

Smith, R. L., & Smith, J. K. (1988). Differential use of item information by judges using Angoff and Nedelsky procedures. *Journal of Educational Measurement, 25,* 259–274.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50. doi:10.1037/0096-3445.117.1.34

Spencer, B. D. (2007). Estimating the accuracy of jury verdicts. *Journal of Empirical Legal Studies, 4,* 305-329.

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

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

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17,* 99-139. DOI:10.1037/a0021650

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: National Institute of Justice.

Technical Working Group for Eyewitness Evidence. (2003). *Eyewitness evidence: A trainer's manual for law enforcement.* Washington, DC: National Institute of Justice.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22,* 217–237.

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied, 8,* 180-193. DOI:10.1037//1076-898X.8.3.180

Tredoux, C. G., Parker, J. F., & Nunez, D.T. (2007). Predicting eyewitness identification accuracy with mock witness measures of lineup fairness: Quality of encoding interacts with lineup format. *South African Journal of Psychology, 37,* 207-222.

Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? *Law and Human Behavior, 24,* 231-258.

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

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology, 43,* 161-204. DOI:10.1080/14640749108400966

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 26,* 582-600. DOI: 10.1037//0278-7393.26.3.582

Vickers, D. (1979). *Decision processes in visual perception.* New York: Academic Press.

Weagant, T. (2011, April 14). Nearly 50 years later, Réjean Hinse obtains compensation for wrongful conviction. *CNW*. Retrieved from http://www.newswire.ca/en/story/769733/nearly-50-years-later-rejean-hinse-obtains-compensation-for-wrongful-conviction

Wedell, D. H., Parducci, A., & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness — Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology, 23,* 230-249.

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

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89-103. DOI: 10.1111

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

Wells, G. L., Cutler, B. L., & Hasel, L. E. (2009). The Duke-lacrosse rape investigation: How not to do eyewitness identification procedures. In M. L. Siegel (Ed.), *Race to injustice: Lessons learned from the Duke University lacrosse rape case* (pp. 307-321). Durham, NC: Carolina Academic Press.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776-784.

Wells, G. L., & Luus, C. A. E. (1990). The diagnosticity of a lineup should not be confused with the diagnostic value of non-lineup evidence. *Journal of Applied Psychology, 75,* 511-516.

Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior, 3,* 285-293.

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviours. *Journal of Experimental Psychology: Applied, 8,* 155-167. DOI:10.1037//1076-898X.8.3.155

Wells, G. L., & Penrod, S. D. (2011). Eyewitness identification research: Strengths and weaknesses of alternative methods. In B. Rosenfeld, & S. D. Penrod (Eds.), *Research methods in forensic psychology* (pp. 237-256). John Wiley and Sons, Hoboken, NJ.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78,* 835-844.

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,* 1-38.

Wells, G. L., Stebley, N. K., & Dysart, J. E. (2014). Double-blind photo-lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*. Advance Online Publication. DOI:10.1037/lhb0000096

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115-1125.

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon "probative value" and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science, 7,* 275-278. DOI:10.1177/1745691612442906

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of U. S. police on preparation and conduct of identification lineups. *Psychology, Crime & Law, 10,* 69-82. DOI:10.1080/106831604100016418

Wogalter, M. S., Marwitz, D. B., & Leonard, D. C. (1992). Suggestiveness in photospread lineups: Similarity induces distinctiveness. *Applied Cognitive Psychology, 6,* 443-453.

Zalman, M., Smith, B., & Kiger, A. (2008). Officials' estimates of the incidence of "actual innocence" convictions. *Justice Quarterly, 25,* 72-100. DOI:10.1080/07418820801954563

Appendix A

Lineup Coding Guide for the Meta-Analysis

Before the lineups were coded, guidelines were developed to facilitate a reliable method of categorizing lineups as having biased, moderate similarity, or high similarity. I developed the guidelines by reading the Methods sections of relevant articles and taking note of how labels researchers assigned to lineups (e.g., low similarity, high similarity) corresponded with similarity ratings and effective size values. Two coders were instructed to attend to more than just the similarity ratings and the effective size scores because these values could be influenced by the scale that was used (e.g., 7-point scale vs. 100-point scale) as well as the instructions that the researchers provided. Moreover, quantitative measures of similarity were not reported in some articles, further necessitating a more comprehensive approach to categorizing the lineups. To encourage the coders to consider more than the quantitative ratings and adopt a more holistic evaluation of the lineups, the guidelines included a range of mean similarity ratings for each of the categories that overlapped with one another. This provided coders with the flexibility to use information in addition to the similarity ratings and effective size scores when assigning a code to a lineup.

The range of similarity ratings and effective size scores for each category is provided below. All similarity ratings were converted to a 100-point scale for ease of comparison between studies. The lower end of the range of similarity ratings for the high similarity lineups may seem low on an absolute scale; however, there is good reason to suspect judges tend to be conservative when assessing the similarity between two faces. For instance, when Flowe and Ebbeson (2007) collected judgments of similarity between two computer-generated faces that apart from one feature were identical, those faces were

assigned a similarity rating of 70 (on a 101-point scale). Furthermore, of all the lineups included in the meta-analysis, not a single one exceeded a similarity rating of 60 (on a 101-point scale). Thus, in relative terms, a lineup with similarity ratings near the midpoint of a scale can be considered quite high. With regard to the effective size scores, our guidelines correspond well with Brigham (1990), who suggested a lineup with an effective size of 3 (for a 6-member lineup) should be considered "fair".

Biased

Fillers bear little resemblance to the suspect. Similarity ratings range between 0 and 35 (on a 101-point scale). Effective size scores around 1-2 (for a 6-member lineup).

Moderate Similarity

Fillers resemble the suspect to some degree, but not as much as other potential fillers. Similarity ratings range between 25 and 50 (on a 101-point scale). Effective size scores around 2-3 (for a 6-member lineup).

High Similarity

Fillers closely resemble the suspect. Similarity ratings range between 40 and 100 (on a 101-point scale). Effective size scores around 4-5 (for a 6-member lineup).

Appendix B

Research Ethics Board Approval Memo

University of Regina

OFFICE OF RESEARCH SERVICES
Regina, Saskatchewan, Canada S4S 0A2
Phone: 306.585.4775 Fax: 306.585.4893
www.uregina.ca/research

DATE: December 22, 2011

TO: Ryan Fitzgerald
4723 Singer Crescent
Regina, SK S4X 4T6

FROM: Dr. Bruce Plouffe
Chair, Research Ethics Board

Re: **Determining Optimal Lineup Member Similarity (File # 35S1112)**

Please be advised that the University of Regina Research Ethics Board has reviewed your proposal and found it to be:

☑ 1. APPROVED AS SUBMITTED. Only applicants with this designation have ethical approval to proceed with their research as described in their applications. For research lasting more than one year (Section 1F). **ETHICAL APPROVAL MUST BE RENEWED BY SUBMITTING A BRIEF STATUS REPORT EVERY TWELVE MONTHS.** Approval will be revoked unless a satisfactory status report is received. Any substantive changes in methodology or instrumentation must also be approved prior to their implementation.

☐ 2. ACCEPTABLE SUBJECT TO MINOR CHANGES AND PRECAUTIONS (SEE ATTACHED). Changes must be submitted to the REB and approved prior to beginning research. Please submit a supplementary memo addressing the concerns to the Chair of the REB.** Do not submit a new application. Once changes are deemed acceptable, ethical approval will be granted.

☐ 3. ACCEPTABLE SUBJECT TO CHANGES AND PRECAUTIONS (SEE ATTACHED). Changes must be submitted to the REB and approved prior to beginning research. Please submit a supplementary memo addressing the concerns to the Chair of the REB.** Do not submit a new application. Once changes are deemed acceptable, ethical approval will be granted.

☐ 4. UNACCEPTABLE AS SUBMITTED. The proposal requires substantial additions or redesign. Please contact the Chair of the REB for advice on how the project proposal might be revised.

D. Bruce Plouffe
Dr. Bruce Plouffe

cc: Dr. Chris Oriet and Dr. Heather Price – Psychology

** supplementary memo should be forwarded to the Chair of the Research Ethics Board at the Office of Research Services (Research and Innovation Centre, Room 523) or by e-mail to research.ethics@uregina.ca

Phone: (306) 585-4775
Fax: (306) 585-4893
www.uregina.ca/research

Appendix C

Full contingency table for lab participants in Chapter 4 (n = 224)

Culprit	LP	Heterogeneous Lineups													
		Low Similarity							Moderate Similarity						
		Suspect		Filler		Reject			Suspect		Filler		Reject		
		P	*SE*	P	*SE*	P	*SE*	*n*	P	*SE*	P	*SE*	P	*SE*	*n*
Present	Sim	.53	.13	.27	.12	.20	.10	15	.57	.11	.19	.09	.24	.09	21
	Seq	.60	.16	.10	.10	.30	.15	10	.25	.13	.33	.14	.42	.14	12
Absent	Sim	.46	.14	.23	.12	.31	.13	13	.13	.08	.31	.12	.56	.12	16
	Seq	.16	.08	.00	.00	.84	.08	19	.23	.12	.23	.12	.54	.14	13
		Homogeneous Lineups													
Present	Sim	.21	.11	.29	.12	.50	.13	14	.54	.14	.15	.10	.31	.13	13
	Seq	.50	.14	.14	.09	.36	.13	14	.42	.14	.17	.11	.42	.14	12
Absent	Sim	.00	.00	.27	.13	.73	.13	11	.10	.10	.50	.16	.40	.16	10
	Seq	.65	.12	.06	.06	.29	.11	17	.43	.13	.21	.11	.36	.13	14

Notes: LP = Lineup Presentation; Sim = Simultaneous; Seq = Sequential.