Eyewitness identification for multiple perpetrator crimes: Examining underlying issues in memory and decision-making

Nina Tupper

September 2017

This thesis is submitted in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy of The University of Portsmouth
Abstract

The over-arching aim of the research conducted for this thesis was to examine underlying issues in memory and decision-making that impact eyewitness identification procedures in the context of multiple perpetrator crimes. In one survey and five experiments, we (i) explored key concerns in multiple perpetrator identifications in police practice in three EU countries (Police Survey), (ii) tested the independence of multiple identification decisions made successively (Experiments 1, 2 and 3) and (iii) examined the purported utility of using other faces as contextual cues for recognizing the faces of multiple perpetrators (Experiments 4 and 5). In the survey we asked police officers (from Sweden, Belgium, and the Netherlands) to describe how agencies in various countries conduct and regulate identification procedures with multiple perpetrators. Results demonstrated sizeable differences in police practice between countries and highlighted the importance of determining whether there are consequences of testing memory on multiple lineups presented in succession. In Experiments 1 and 2, participants watched a mock-crime film involving three perpetrators and later made three showup identification decisions, one showup for each perpetrator. Experiments 1 and 2 used similar procedures, with the exception of varied patterns of target-presence. Across both experiments, evidence for sequential dependencies for choosing behavior was inconsistent. In Experiment 1, responses on the second, target-present showup assimilated towards previous choosing. However, in Experiment 2, responses on the second showup contrasted previous choosing regardless of target-presence. Experiment 3 examined whether methodological differences between the recognition and eyewitness paradigms used in previous research on sequential dependencies might account for the inconsistent findings in Experiments 1 and 2. Participants studied pairs of words, landscapes, or
faces, and were later tested for recognition. Sequential dependencies were detected in recognition decisions over many trials, including recognition for faces: the probability of a yes response on the current trial increased if the previous response was also yes (vs. no). However, choosing behavior on previous trials did not predict individual recognition decisions on the current trial. In Experiments 4 and 5, we sought replicate facilitative effects in cued face recognition, to (i) investigate the mechanisms underlying those effects, and (ii) determine whether such effects would extend to more than two faces. Participants encoded sets of individual, paired, or groups of four faces and were tested with no cues, correct cues (a face previously studied with the target test face), or incorrect cues (a never-before-seen face). Hit rates were not affected by either cue type or face encoding condition, but cuing of any kind (correct or incorrect) appeared to provide a protective buffer to reduce false-alarm rates in the two- and four-face conditions through increased sensitivity, but mostly reduced response bias. The present research on sequential dependencies for identification decisions suggests that the integrity of identification and recognition decisions is not likely to be impacted by making multiple decisions in a row. However, our findings suggest that cued face recognition may be a useful technique to use for reducing false recognition rates in contexts with multiple faces. Throughout the thesis, we argue for the systematic examination of influential factors that are both unique and inherent to practice, memory, and decision-making for multiple perpetrator identification and recognition.
Table of Contents

Chapter 1: Introduction to the thesis ................................................................. 14
Identification in practice .................................................................................. 17

Police practice in Sweden, Belgium, and the Netherlands ...................... 20
Participants and recruitment ................................................................. 21
Method ............................................................................................................ 22
Results ............................................................................................................. 24
Discussion ....................................................................................................... 29

Overview of the current thesis ...................................................................... 13

Chapter 2: Showup identification decisions for multiple perpetrator crimes:
Testing for sequential dependencies ............................................................... 18

Introduction ..................................................................................................... 19
Experiments 1 and 2 ....................................................................................... 28
Participants and Design .................................................................................. 28
General Method for Experiments 1 and 2 .................................................... 29

Materials .......................................................................................................... 29
Procedure ......................................................................................................... 30
Results .............................................................................................................. 31

Descriptive statistics for choosing on showups ......................................... 32
Experiment 1: Testing for sequential dependencies ..................................... 32
Experiment 2: Testing for sequential dependencies ..................................... 35
Discussion ....................................................................................................... 35

Experiment 3 .................................................................................................... 39
Participants and design .................................................................................. 39
Method ............................................................................................................. 40

Materials .......................................................................................................... 40
Procedure ......................................................................................................... 41
Results .............................................................................................................. 42

Preparation of data and calculation of contingency rates .......................... 43
Sequential effects as a function of stimulus type ........................................ 43
Sequential effects as a function of section .................................................... 44
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Sequential effects in choosing.</td>
<td>46</td>
</tr>
<tr>
<td>4. Predicting choosing on individual trials.</td>
<td>47</td>
</tr>
<tr>
<td>Discussion</td>
<td>48</td>
</tr>
<tr>
<td>General Discussion</td>
<td>49</td>
</tr>
<tr>
<td>Chapter 3: Face value: Testing the utility of contextual face cues for face recognition</td>
<td>55</td>
</tr>
<tr>
<td>Introduction</td>
<td>56</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>67</td>
</tr>
<tr>
<td>Method</td>
<td>67</td>
</tr>
<tr>
<td>Participants</td>
<td>67</td>
</tr>
<tr>
<td>Design</td>
<td>67</td>
</tr>
<tr>
<td>Materials</td>
<td>68</td>
</tr>
<tr>
<td>Procedure</td>
<td>69</td>
</tr>
<tr>
<td>Results</td>
<td>72</td>
</tr>
<tr>
<td>Orienting the effect of cuing.</td>
<td>73</td>
</tr>
<tr>
<td>Replicating and extending effects of cued face recognition.</td>
<td>76</td>
</tr>
<tr>
<td>Discussion</td>
<td>82</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>83</td>
</tr>
<tr>
<td>Method</td>
<td>85</td>
</tr>
<tr>
<td>Participants</td>
<td>85</td>
</tr>
<tr>
<td>Materials</td>
<td>85</td>
</tr>
<tr>
<td>Procedure</td>
<td>85</td>
</tr>
<tr>
<td>Results</td>
<td>86</td>
</tr>
<tr>
<td>Orienting the effect of cuing.</td>
<td>86</td>
</tr>
<tr>
<td>Replicating and extending effects of cued face recognition.</td>
<td>88</td>
</tr>
<tr>
<td>Sensitivity and response bias.</td>
<td>89</td>
</tr>
<tr>
<td>Discussion</td>
<td>90</td>
</tr>
<tr>
<td>General Discussion</td>
<td>90</td>
</tr>
<tr>
<td>Chapter 4: Thesis General Discussion</td>
<td>99</td>
</tr>
<tr>
<td>Summary of findings</td>
<td>99</td>
</tr>
<tr>
<td>Theoretical implications</td>
<td>102</td>
</tr>
<tr>
<td>Practical implications</td>
<td>105</td>
</tr>
</tbody>
</table>
Methodological considerations and future directions..............................................112
Conclusion..............................................................................................................117
References..............................................................................................................119
Abbreviations

ANOVA: Analysis of variance

$BF_{01}$: Bayes factor, expressed as evidence in support of null hypothesis

$BF_{10}$: Bayes factor, expressed as evidence in support of alternative hypothesis

CR: Correct-rejection

FA: False-alarm

FR: False-rejection (miss)

PACE: Police and Criminal Evidence Act (U.K.)

TA: Target-absent

TP: Target-present
Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Word count: 46,351
Table of Figures and Tables

Figures
Figure 1.1 Illustration of multiple suspects for single perpetrators vs. multiple perpetrators

Figure 1.2 Police Survey: Frequency of responses by country regarding multiple suspect lineups for single vs. multiple perpetrator crimes

Figure 2.1 Experiment 3: Example stimuli pairs for faces, places, and words

Figure 2.2 Experiment 3: Hit rate and false-alarm rate contingencies

Figure 3.1 Example encoding stimuli using images from AR Face database

Figure 3.2 Experiments 4 and 5: Procedure for encoding and testing

Figure 3.3 Experiment 5: False-alarm rates as a function of cue type and face encoding condition

Tables
Table 1.1 Police Survey: Range of Age and Job Experience (Mean) and Status of Certification Requirement for Police Respondents on Multiple Perpetrator Identification Survey

Table 1.2 Questions for Police on Multiple Perpetrator Crimes and Eyewitness Identification Procedures

Table 1.3 Procedure and results for eyewitness identification experiments that test identification procedures created or adapted for multiple perpetrator crime

Table 2.1 Experiments 1 and 2: Proportion (Frequency) of Choosing Across Showups and Overall
Table 2.2 Experiments 1 and 2: Final Models of Logistic Regressions Predicting Choosing on Second and Third Showups Based on Previous Choosing and Target-Presence ..........................................................70

Table 2.3 Experiment 3: Results for ANOVAs on Current Hit Rates, False-Alarm Rates, and Choosing Rates Given Previous Responses and Condition.........81

Table 2.4 Experiment 3: Results for ANOVAs on Current Hit Rates, False-Alarm Rates, and Choosing Rates Given Previous Responses and Testing Section……81

Table 2.5 Experiment 3: Results of Logistic Regression Predicting Choosing on Second and Third Recognition Test Trials Based on Previous Choosing and Target-Presence..........................................................83

Table 3.1 Experiments 4 and 5: Mean Hit Rates and False-Alarm Rates (95% CI) by Cue Type and Face Encoding Condition ........................................110

Table 3.2 Experiments 4 and 5: Mean Sensitivity and Response Bias Rates (95% CI) by Cue Type and Face Encoding Condition ...............................111

Table 3.3. Experiments 4 and 5: Results for ANOVAs and t-tests Comparing Hit Rates and False-Alarm Rates Across Cue Types and Face Encoding Condition.................................................................114

Table 3.4 Experiments 4 and 5: Results for Bayesian ANOVAs Comparing Hit Rates (HR) and False-Alarm Rates (FAR) Across Cue Types and Face Encoding Condition .................................................................115

Table 3.5 Experiments 4 and 5: Results for t-tests and ANOVA’s Comparing Sensitivity and Response Bias as a Factor of Face Encoding Condition or Cue Type ...........................................................................117
Acknowledgments

I would like to extend my deepest gratitude to the House of Legal Psychology for planting and growing this incredible international collaboration with its vast and varied network of candidates and established researchers from all over the world. I want to thank all of the supervisors, mentors, and visitors to the program for contributing their time and resources to our training, for imparting their knowledge, and for offering their shoulders on which our dwarfly selves can stand.

On that note, I would especially like to thank my direct supervisors: Lorraine Hope, Jim Sauer, and Melanie Sauerland for consistently pulling me (and my writing) out of the clouds and down to the ground, for providing generous support, critiques, advice, and laughs through it all.

Thank you to my colleagues, office-mates, and friends in Portsmouth and Maastricht for the many daily stats-chats, psych-chats life-chats, and soundboard sessions. Andy, Joanne, and Nathalie, thank you for stability, companionship and all the good times as we shared this nomadic, academic life.

I would like to thank my family for their constant support, particularly my parents. It is no easy job, but I am especially grateful for your continuous encouragement from across the ocean time and time again. You instilled in me a love for learning and the value of follow-through, which have brought me to where I am today.

I would lastly like to acknowledge and thank the European Union and the Erasmus Mundus Joint Doctorate Program in Legal Psychology for funding this PhD fellowship.
**Dissemination**

**Conference Presentations**

*Experiment 1 was presented at the European Association for Psychology and Law conference in July 2015:*


*Experiments 1 and 2 were presented at the American Psychology and Law Society conference in March 2016 and the 20th Anniversary Faculty of Psychology and Neuroscience Research Day in May, 2016:*


*Experiments 1, 2, and 3 (Chapter 2) as well as Experiment 4 were presented at the Society of Applied Research in Memory and Cognition conference in January 2017:*


**Publications**


*The Police Survey (Chapter 1) is being prepared for submission for publication*


*Experiments 1, 2, and 3 (Chapter 2) are being revised for submission for publication.*

*Experiments 4 and 5 (Chapter 3) are being prepared for submission for publication.*

Chapter 1: Introduction to the thesis

A case of identity: Misidentification and multiple perpetrator crimes

In July of 1981, three men broke into a Florida home, threatened the five victims with a shotgun, and robbed the residence. Three of the victims were tied up and left in the house. Two victims, 12-year-old Isabelle and 36-year-old Michelle, were forced into the trunk of the car, driven to a dark, wooded area, and were raped. They were left tied to trees as the perpetrators drove away. The victims in the house untied themselves in time to record the license plate number of the perpetrators’ car for police. Isabelle and Michelle also eventually managed to untie themselves from the trees and sought help from a nearby home. Soon after the crime, the police showed the victims photographic lineups for two suspects. Michelle identified both Douglas James and Alan Crotzer, while one of the other victims also identified Douglas. The next day, police showed the victims more photos, and three of them identified Corlenzo James, the brother of Douglas, as the third perpetrator. During their subsequent trials, Corlenzo pled guilty to robbery and assault, Douglas defended himself by claiming the rape was permitted by the adult victim (i.e., consent defense), but Alan maintained his innocence and claimed he had no knowledge of the crime. All three were convicted, with Alan Crotzer given a prison sentence of 130 years. (Innocence Project, 2017)

Unfortunately, the victims in this case were only correct about two of the three perpetrators. In 2003, new analyses of the DNA found on the rape victim’s clothing definitively excluded Alan Crotzer as a rapist. Despite the fact that all five eyewitnesses made in-court identifications of him, Alan was an innocent man convicted of a crime he did not commit. Douglas eventually revealed the third perpetrator.

1 Not their real names
perpetrator to be a childhood friend and admitted that they had never met Alan before he was accused of their crime. In 2006, after 24 years in jail, Alan was finally released.

While we know a lot about eyewitness memory in general—four decades of research provides a number of insights as to what might have produced a faulty memory of Alan Crotzer— it is fair to say that we do not know much specifically about eyewitness memory for multiple perpetrator crimes. We do know that multiple perpetrator crimes are prevalent around the world and that such crimes are among the most difficult to solve (Dauvergne & Li, 2006; Liem et al., 2013). We also know that nearly two-thirds of surveyed U.K. police report issues and confusion in conducting multiple suspect identification procedures, that eyewitness identification accuracy tends to decrease as the number of perpetrators increase, and that attempts to address this multiple perpetrator identification disadvantage through adapted lineup procedures have had limited success (Hobson, Wilcock, & Valentine, 2012; Horry, Halford, Brewer, Milne, & Bull, 2014).

However, research has done little to uncover or explore those factors that are unique to eyewitness memory in the context of multiple perpetrator crimes. This leaves us with an incomplete picture when trying to understand eyewitness memory and decision-making in multiple suspect lineups like for the crime depicted above. Furthermore, there is a paucity of information regarding how police construct identification procedures for a multiple perpetrator crime, to what extent those procedures are similar or different for a single perpetrator crime, and what, if any, issues police or eyewitnesses face in this applied context (cf. U.K. police; Hobson et al., 2012). Thus, while there have been some attempts to create novel identification procedures adapted specifically to the context of multiple perpetrator crimes, it is
unclear which theories in memory and decision-making should drive such attempts, whether the theories used for previous attempts are relevant to multiple perpetrator crime identification, and to what extent the traditional practice-as-usual comparison groups (i.e., control groups) reflect actual police methods.

This thesis specifically aims to address this gap in the literature concerned with eyewitness memory for multiple perpetrator crimes. Across five empirical studies and one exploratory survey, this thesis reviews police practice in three European countries in the context of multiple suspect identification (Chapter 1: Police Survey), tests the independence of multiple identification decisions made successively (Chapter 2: Experiments 1, 2 and 3), and examines the purported utility of face context cues for recognizing the faces of multiple perpetrators (Chapter 3: Experiments 4 and 5). This thesis seeks to extend a small, but growing, field in eyewitness memory for multiple faces and eyewitness identification for multiple perpetrators so that compiled evidence may someday be used to support identification procedures adapted specifically to the context of multiple perpetrator crimes.

The purpose of this chapter is to place this thesis in context by (1) establishing applied questions for research in multiple perpetrator identification, (2) reviewing the extant literature that has thus far adapted traditional lineups techniques for multiple perpetrator crimes, and (3) exploring the theoretical background for two factors in memory and decision-making that are both unique to multiple perpetrator crimes and have yet to be empirically tested. These two factors—the associations between multiple decisions and the associations between memories for multiple faces—are the focus of the thesis.
Identification in practice

The applied eyewitness experiment: Identification of multiple perpetrators. The case described at the outset of this thesis is just one example of the many crimes that are committed by multiple perpetrators—crimes that are often violent, premeditated and goal-driven. Gang violence (Juodis, Woodworth, Porter, & Ten Brinke, 2009), hate crimes (Sandholtz, Langton, & Planty, 2013), rapes (Horvath & Kelley, 2009), and assaults (Hobson, et al., 2012) are often committed by strangers working as a group. In fact, the rising rate of such crimes appears to be a global phenomenon. For example, in Finland, Sweden, and the Netherlands, 13-17% of homicides between 2003 and 2006 involved two or more perpetrators (Liem et al., 2013), while the proportion of homicides with multiple perpetrators in the U.S. reached 20% in 2008 (nearly double that reported in 1980; Cooper & Smith, 2011). In Canada, one third of recorded homicides are committed by two or more perpetrators, gang-related homicides tripled between 1995 and 2000 (Fedorowyckz, 2001), and doubled again between 2001 and 2008 to 0.42 gang-related homicides per 100,000 people (Statistics Canada, 2016).

Such crimes often involve victims or bystanders as eyewitnesses who may be asked to identify multiple suspects related to the multiple perpetrators. A lineup, with one suspect and at least five known-to-be-innocent look-alikes (called fillers), can be constructed using photographs or videos of the lineup members to show the eyewitness, or by having lineup members physically present (i.e., behind a one-way mirror). This lineup can be seen as an applied scientific experiment with tangible and immediate consequences for the livelihoods of the suspect, the victim(s), and the wider community in which they live. In this ‘experiment’, police test the hypothesis that the suspect is the actual perpetrator of the crime and the eyewitness can either
confirm the police hypothesis by choosing the suspect, or they can falsify the police hypothesis by rejecting the lineup (i.e., saying the perpetrator is not there) or by mistakenly identifying the known-innocent filler. It is critical to understand how factors within the control of the justice system (i.e., instructions to witnesses; lineup construction) and factors outside of the control of the justice system (i.e., encoding conditions of the witnessed event; mechanics of memory) might influence an eyewitness’s memory and decisions in eyewitness identification procedures because we can use this knowledge to improve identification procedures or to make informed decisions about the reliability of the evidence after the fact.

There is an extensive history of police practice inspiring empirical tests for eyewitness memory (e.g., witness instructions; Malpass & Devine, 1981; post-identification feedback effect; Wells & Bradfield, 1989), but only one such experiment specifically in the case of multiple perpetrator identification procedures (Hobson, & Wilcock, 2011). One source of information on current practice is surveys of police practitioners, such as the most recent National Institute of Justice survey of police practice across the United States (Police Executive Research Forum, 2013). However, there is only one police survey of law enforcement agencies that asked about police procedures specifically in the context of multiple perpetrator crimes, which is reviewed below (Hobson et al., 2012). Also presented are the recent results of a new survey conducted for this thesis, which extends the scope of Hobson and colleague’s survey for three additional countries within the European Union (EU).

**Police practice in the U.K.** Despite being the only region in the world that addresses lineups for multiple perpetrator crimes in legislation (PACE 1984), guidelines for law enforcement officials in the U.K. are sparse, non-specific, and not empirically vetted. On multiple suspect showings (i.e., there are multiple suspects for
a single perpetrator crime), PACE procedures instruct, “When all members of a similar group are possible suspects, separate identification parades shall be held for each,” (PACE Code D, 2011, p. 181). This might apply to a situation where a single member of a gang committed a crime, but the police do not know which one of the members is the perpetrator. In this case, the eyewitness would view one lineup for the first suspect and make a decision about that lineup before viewing the next one. This is a logical extension of the PACE rules for single suspect showings, where the established procedures are simply replicated separately for each one of the multiple suspects. For crimes with multiple perpetrators, an adaption was added, instructing, “Only one suspect shall be included in an identification parade unless there are two suspects of roughly similar appearance,” (PACE Code D 2011, Annex A, p. 47). In other words, while PACE rules forbid multiple suspects to appear in the same lineup for a single perpetrator crime, they make an exception for multiple perpetrator crimes.

According to the results of a survey of U.K. police by Hobson and colleagues (2012), all 29 law enforcement agencies from England, Wales, and Northern Ireland report creating individual lineups for each suspect and requiring the witness to make a decision about the first before moving on to the next. However, officers also report frequently running into difficulty implementing such lineups, including having to adapt instructions, being unsure of how to handle out-of-ordinary requests from eyewitnesses, and receiving complaints of “blindness” from witnesses having to view too many faces. For example, one officer reported that although they normally follow PACE procedure, they would allow the witness to see all lineups before making identification decisions if the witness specifically requested to do so. Officers also reported that witnesses would sometimes try to pick multiple suspects from the same lineup, even when there was only one suspect in each lineup. When this happens,
some officers insisted that the eyewitness could no longer see the following lineups, but most reported that they would show the other lineups anyway. In such a case, it is unclear how officers should proceed, and neither current police protocol nor psychological research have addressed this concern.

While the 2012 survey was an important first step to understand police practice for multiple perpetrator identifications, the results are limited to one region that specifically addresses multiple perpetrator identifications in mandated police procedures. Therefore we do not know what officers would do in the absence of such instructions. Furthermore, the original survey failed to address some of the concerns that are unique to multiple perpetrator identification procedures, including if and when multiple identifications are administered (i.e., witnesses view lineups for perpetrators as suspects become available, or all at the same time), who is responsible for constructing the multiple lineups (i.e., does the same person construct all lineups for a multiple perpetrator crime?), and how officers perceive witnesses that can identify some, but not all, of the presented suspects.

**Police practice in Sweden, Belgium, and the Netherlands.** To address this gap, a new survey was developed to extend this previous survey. The original questions from Hobson and colleagues (2012) were adapted and translated into two languages (Swedish and Dutch) along with new questions to address the concerns above. This new survey was distributed online to police agencies in three EU countries: Sweden, Belgium, and the Netherlands (Tupper, Sauerland, Sauer, & Hope, in prep). In this survey, we aimed to (a) understand the prevalence and characteristics of multiple perpetrator crimes from the perspective of law enforcement agencies, (b) discern how agencies in various countries conduct identification procedures (e.g., lineups, photo-arrays, showups) with multiple perpetrators, and (c) gain insight into
how law enforcement agents and eyewitnesses experience the identification process in the context of such crimes.

**Participants and recruitment.** In total, we received responses from 72 law enforcement officials from Sweden (n = 25), Belgium (n = 27), and the Netherlands (n = 20) who had experience administering identification procedures to eyewitnesses. However, attrition occurred throughout the study, such that the final questions received 39 participant responses. Participants from Sweden completed the Swedish-translated version of the survey and participants from Belgium and the Netherlands completed the Dutch-translated version of the survey. We used the snowball sampling method of recruitment, meaning that we used initial police contacts in each country to recruit colleagues and distribute the online survey link in a way that best suited the structure of the police in that particular country. Participants received the survey link via an e-mail, which also included a short explanation of the purpose and contents of the survey. Participants could access the survey through Qualtrics (Provo, Utah), where they read a more detailed information sheet, provided informed consent, and completed the survey questions (survey in appendices).

Table 1.1

<table>
<thead>
<tr>
<th>Country</th>
<th>Age</th>
<th>Job Experience</th>
<th>Certification Required?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>27-61 (M = 40.44)</td>
<td>0-25 (M = 7.67)</td>
<td>No</td>
</tr>
<tr>
<td>Belgium</td>
<td>27-55 (M = 43.15)</td>
<td>1-30 (M = 14.00)</td>
<td>No</td>
</tr>
<tr>
<td>Netherlands</td>
<td>33-68 (M = 47.40)</td>
<td>1-30 (M = 10.89)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Note. Age and job experience express range of years reported by all police respondents from each country.*

2 The survey was not translated to French. Therefore we did not target recruitment from Belgian regions that are primarily French speaking (i.e., Brussels).
**Method.** We received the original version of the police survey (Hobson, et al., 2012) with permission to adapt it. Ten of the 11 original survey questions were used, and 11 questions were added (see Table 1.2 for a list of questions). Given that the survey would be distributed to multiple countries with different procedure regulations and varying levels of institutional or legal obligation to those regulations, two of the original 10 questions were adapted to allow officers to indicate which procedures they used in their own job and one was dropped. Original questions, adapted questions, and novel questions are marked in Appendix B. The survey included multiple choice responses and open-ended questions throughout five main sections: (1) general information about the officer (e.g., police jurisdiction, years of experience; 5 questions), (2) estimation of the proportion of crimes and the type of crimes officers had handled that involved multiple perpetrators (3 questions), (3) current procedures for multiple identifications (13 questions), (4) problems experienced while administering multiple identification procedures (2 questions) and (5) officer’s experiences in their interactions with eyewitnesses viewing lineups for multiple perpetrator crimes (2 questions). There was also space at the end for their suggestions regarding if/how to adjust or change current procedures (1). For the purpose of this thesis, we focus only on sections 3-5. However, Table 1.1 provides descriptive statistics of general information, including age, job roles, job experience, and whether certification is required to administer identification lineups.

As part of the development of the survey, police contacts in each of the three countries to be surveyed received both the English and the appropriate translated version of the survey. They were asked to provide feedback on the survey, including ease of understanding, the accuracy of terminology used in translation, the appropriateness and relevance of the procedures described, and whether they felt there
Table 1.2 *Questions for Police on Multiple Perpetrator Crimes and Eyewitness Identification Procedures*

### 1. General Information*
1. Gender (male/female/other)
2. Age?
3. How many years of experience in conducting eyewitness identification procedures do you have?
4. What is your job role?
5. Jurisdiction?

### 2. The Criminal Offences*
1. Of the crimes you have dealt in the last 12 months, what proportion involved multiple suspect showings? (0-100%)
2. How many suspects are typically involved in the multiple perpetrator cases you have dealt with? (Please select the box for the category that applies most often) (2-10)
3. In the past 12 months, what types of crimes have you dealt with that typically involve multiple perpetrators? (robbery / burglary/ assault/ sexual assault/ homicide/ other)

### 3. Current Procedures

**Scenario 1**
4a. Choose the option that resembles what you would do in this case (A1 or A2/ A1 and A2 separately/ A1 and A2 together)
4b. In your work with multiple perpetrator crimes, Scenario 1 occurs: (never/ sometimes/ often/ always)

**Scenario 2**
5a. Choose the option that resembles what you would do in this case (A or B/ A and B separately/ A and B together)
5b. In your work with multiple perpetrator crimes, Scenario 2 occurs: (never/ sometimes/ often/ always)

6. **In what manner do you present the parades to witnesses in a multiple suspect identification?** Select all the options that apply: (Lineups: live/ photo/ video; Format: simultaneous/ sequential/ other; Show-ups: live/ photo/ video)

7. Are there any procedural requirements or guidelines in place for multiple suspect identifications?
8. How do you organize the identification presentations for eyewitnesses in the case of a multiple-perpetrator crime? (witness views when: all lineups available/ as lineups become available/ other)

9*. What instructions do you give to a witness for multiple perpetrator identifications?
10*. Do you ask the witness to look for a specific suspect?
11*. Do you ask the witness to describe the role of the suspect they are identifying?
12. Do you record all eyewitness identification decisions in a crime with multiple perpetrators?
13. Do you record confidence for all suspect identifications for multiple suspect identifications?
14. Who is responsible for constructing the lineups? Is the same person responsible for all suspect lineups in a given case involving multiple perpetrators?

### 4. Issues with Current Practice*
15. Do you, as someone who administers identifications, experience any problems with multiple suspect identifications?
16. Do you think witnesses experience any problems with multiple suspect identifications?

### 5. Perceptions of Eyewitnesses
17. How do you think eyewitnesses of a multiple perpetrator crime perform in identifications compared to eyewitnesses of a single perpetrator crime? Generally eyewitnesses to crimes committed by multiple perpetrators are _____ compared with eyewitnesses to crimes committed by a single perpetrator: (worse/ as good as/ better)
18. In your opinion, how useful is a witness for you if they identify one, but not all of the suspects presented?

**Your suggestions**
19. Do you have any ideas about how multiple suspect identifications could be improved from the point of view of the police?

* indicates original survey question from Hobson, Wilcock, and Valentine (2012)
** see Figure 1.1 for graphic illustration of scenarios
was anything in particular missing from the survey. Adaptations were made for each country accordingly.

Officers answered in their native language, but their responses were translated to English by native Dutch and Swedish speakers for the purpose of analysis. The full English version of the survey can be found in Appendix B of this thesis, and recruitment materials (i.e., recruitment e-mail, information form, and debriefing form) can be found in Appendices B, C, and D.

**Results.** Given the exploratory nature of this study, we did not conduct statistical comparisons between countries or procedures, but rather sought to describe the pattern of practice across countries. Due to attrition, not all questions have the same number of respondents. The number of respondents still in the survey are reported following the percentage, and the percentage represents the number of respondents out of the remaining respondents (cf. all respondents). This thesis will first cover the procedures that are common to all three countries and then focus on reported practices specific to each country.

**Current procedures: Sweden, Belgium, and the Netherlands.** To begin, we focus on procedures that the three countries report having in common. Officers in all countries expressed knowledge of national and/or local guidelines for police protocol, none of which were legally binding. Thus, it was important for police to understand what the courts required to admit evidence. A majority of responses (96%; \( n = 47 \)) showed that police officers in all countries most often use photographic, sequential lineup procedures. When given an open question to provide instructions given to witnesses, 40% \( (n = 40) \) of officers reported providing unbiased witness instructions (i.e., telling the witness that the perpetrator may or may not be present in the lineup and that they should only choose someone if they are certain) and 15% provided hair-
specific appearance-change instructions (i.e., encouraging witnesses to focus on features like eyes and lips because hair may change). In the Netherlands, instructions are provided to eyewitnesses as written instructions sheets, and all but two officers either reported unbiased and appearance-change instructions, or directed us to the eyewitness information sheet on which the instructions are printed. A majority of all officers (77%; n = 39) reported that they record confidence only if it is spontaneously provided by the eyewitness.

One question that appeared to confuse officers was the question, “Do you ask the witness to look for a specific suspect (e.g., identify the one who was driving the car)?” This was intended to determine whether officers inform the eyewitness which lineup is for which perpetrator before any identification decision is made. However, most officers that reported “yes” subsequently explained that they only did this following a positive identification. Although it is clear that this question did not elicit the responses expected, qualitative responses suggest officers do not generally instruct eyewitnesses to look for specific suspects. At least 90% (n = 40) of officers in some form indicated that role in the crime was only elicited after the identification procedure.

Current procedures in Sweden. National guidelines are available in Sweden, but one officer complained that these guidelines are updated frequently and it becomes difficult to know which rules to follow. Officers conducting identification procedures in Sweden sometimes use video or live lineups and require eyewitnesses to view the entire sequential lineup first. Afterwards witnesses are allowed to make an identification decision in the second viewing. A majority of officers (92%; n = 12) reported that they record all identification decisions, whether those are positive or
negative responses, and one officer wrote that they would consider it wrong, if not
criminal, to *not* report all decisions.

*Current procedures in Belgium.* While most officers expressed that they
record *all* identification decisions (i.e., including if the eyewitness identifies a filler or
does not make an identification)82% (*n* = 17), others reported they would only record
positive decisions (i.e., eyewitness identifies the suspect only).

*Current procedures in the Netherlands.* Police procedures in The Netherlands,
unlike Sweden and Belgium, require officers to be trained and certified in order to
administer identification procedures. Therefore, responses from Dutch police officers
were the most consistent, always referring to the national guidelines (van Amelsvoort,
2013). These guidelines are considered by the courts when admitting eyewitness
evidence. Eyewitnesses are only allowed to view the lineup once and can make their
decision during or after the lineup is presented. Officers conducting identification
procedures in the Netherlands sometimes use video lineups, sometimes live lineups
(called Oslo confrontations), and sometimes a unique procedure called the *Chroma-
key technique*, which aims to reinstate the context of a witnessed event by
superimposing the typical static video lineup over the image of the environment in
which the perpetrator was seen (e.g., the crime scene). The identification procedures
are administered on a computer, meaning that each lineup conducted is automatically
recorded in a national database along with the eyewitness decisions. All instructions
are standardized and presented via the computer. The identifications must be
presented double-blind, so the officer who created the lineup must find another
colleague who does not know the suspect in order to administer to the eyewitness.
Figure 1.1 Illustration of multiple suspects for single perpetrators vs. multiple perpetrators.

Lineups for multiple suspects. In this survey, we were particularly keen to understand how police construct lineups when there are multiple suspects. However, it was important to distinguish between situations in which there are multiple suspects for a single perpetrator vs. situations in which there are multiple suspects related to multiple perpetrators. Therefore, the survey included two scenarios to distinguish between these two instances (see Appendix B). In both cases, officers were asked how they would construct lineups for multiple perpetrators with respect to three options:

Option A: *The eyewitness sees one line up, only for Suspect 1 or only for Suspect 2, not both.*
Option B: *The eyewitness sees two line ups, one for Suspect 1 and one for Suspect 2.*
Option C: *The eyewitness sees one lineup, with both Suspects 1 and 2 in the same lineup.*

As Figure 1.2 demonstrates, a majority of officers chose Option B for both scenarios (Sweden = 53-71%; Belgium = 65-78%; Netherlands = 46-38%), but many also chose C (Sweden = 18-29%; Belgium = 22-39% Netherlands = 54-62%)\(^3\). This demonstrates that when confronted with multiple suspects for different perpetrators, without specific recommendations, officers most often choose to conduct separate multiple lineups, but may also conduct a single lineup for both suspects.

\(^3\) n\text{Sweden} = 17; n\text{Belgium} = 18; n\text{Netherlands} = 13
Police perceptions of eyewitnesses of multiple perpetrator crimes. We also asked officers for their perceptions of how eyewitnesses of multiple perpetrator crimes perform on identification procedures compared to those of single perpetrator crimes. Most officers felt that eyewitnesses of multiple perpetrator crimes perform “worse than” (Sweden = 90%; Belgium = 39%; Netherlands = 45%) or “as good as” (Sweden = 10%; Belgium = 61%; Netherlands = 54%) eyewitnesses of single perpetrator crimes, but never “better”\textsuperscript{4}. They were also asked to comment on the investigative and/or probative value of an eyewitness that identified one suspect, but not the other(s). Most officers (89%; \(n = 38\)) reported that any identification is valuable, even if they do not identify all suspects; justifications for this included that one identification could provide more leads in the case, that one identification was better than none, and that this evidence was perfectly usable in court. Some officers in Sweden and Belgium noted that failure to identify all perpetrators did not undermine the probative value of the identification as long as there was a sufficient explanation (e.g., the eyewitness had longer exposure to the perpetrator they identified, or that

\textsuperscript{4}n_{\text{Sweden}} = 10; n_{\text{Belgium}} = 18; n_{\text{Netherlands}} = 11
perpetrator was more distinctive, etc.). In the Netherlands, one officer reported that they felt this made the eyewitness seem reliable because it demonstrated that they had followed instructions to only identify someone if with absolute certainty.

Eyewitness perceptions of lineups in multiple perpetrator lineups. Police were also asked to provide their subjective impression of the eyewitness experience. Swedish police (80%; n = 10) asserted that eyewitnesses of multiple perpetrator crimes often mixed up the roles and clothing of the perpetrators when linking suspects to perpetrators. Belgian police (50%, n = 12) generally asserted that lineups in general were difficult because they rely on memory; the other 50% reported no problems. Meanwhile Dutch police generally did not consider these lineups to be particularly problematic because of good organization (80%; n = 10). The two Dutch officers that did mention problems discussed general memory issues not specific to multiple perpetrator crimes.

Discussion. This survey demonstrates the differences in police practice between countries that have various degrees of regulation for identification procedures. For example, the Netherlands, where training and certification is mandatory to administer identification procedures, has little variance between the responses of law enforcement officials when discussing the details and order of those procedures. This is also reflected in their perceptions of the identification experience where they express little difficulty in administering lineups and little difficulty for eyewitnesses to understand the procedure and follow directions. Meanwhile, there is less regulation in Sweden and Belgium, which appears to be associated with greater difficulties for the officers and greater confusion for the eyewitnesses in the context of eyewitness identification. However, there are also many similarities shared by the countries, such as using primarily sequential, photographic lineups, and, for the most
part, concluding that it is better not to inform the eyewitness which lineup represents which perpetrator.

Perhaps the most interesting results of this survey were derived from the scenarios of creating a lineup for multiple suspects in the context of a single perpetrator vs. multiple perpetrator crime. Across the three countries in which the survey was conducted, officers generally treated the two cases the same; most often, officers reported separating the two suspects into separate lineups, but many alternatively chose to present them in the same lineup while increasing the number of fillers. In the absence of regulations specifically and separately targeting these two situations (cf. PACE rules), it is not unreasonable that the two scenarios should be treated the same. Nevertheless, there is clearly confusion over whether to keep the suspects together or apart for the purpose of formal identification procedures. The eyewitness identification literature has considered whether suspects should be presented in the same lineup in the context of a single perpetrator crime; the general consensus is that placing multiple suspects of a single perpetrator crime in the same lineup violates the capacity to falsify the police hypothesis that the suspect is a perpetrator by providing what is akin to a multiple choice test (e.g., Wells et al., 1998). This option increases the probability of identifying an innocent suspect and it is recommended to be avoided. However, no research to date addresses the applied format of multiple-perpetrator multiple-suspect lineups reported here. The reported police practice of placing multiple suspects in the same lineup serves to highlight the importance of understanding how memory for multiple perpetrators is stored and how the presence of the multiple suspects may enhance or undermine memory for the actual perpetrators. Furthermore, given that the alternative is to separate multiple

---

5 One experiment by Wells and Pozzulo (2006) has tested a novel lineup format including multiple suspects related to multiple perpetrators.
suspects into separate lineups, and that some officers specifically reported presenting these multiple identification procedures successively on the same day, it is likewise important to determine whether there are consequences of making multiple identification decisions on multiple lineups presented in succession.

The independence or non-independence of memory and decision-making for eyewitness identifications in the context of multiple perpetrator crimes is a critical topic for future research. The survey results reported here highlight the two overarching research questions that are the focus of the current thesis: (1) are multiple identification decisions independent (i.e., does a previous identification decision impact following ones)? and (2) are memories for multiple faces linked and, if so, can this link be used to enhance recognition performance? In order to contextualise these questions and the methods used to answer them in this thesis, it is first necessary to review the literature on multiple perpetrator identification, on the links between multiple decisions, and the associative nature of memory for multiple items (i.e., faces).

**Eyewitness identification for multiple perpetrators: Experiments in context.** Clifford and Hollin (1981) were the first to experimentally test for the difficulty of eyewitness identification in the context of multiple perpetrator crimes compared to single perpetrator crimes. In one condition, participants viewed a non-violent event with one, three, or five perpetrators. Despite only having to select the main perpetrator from a target-present lineup immediately following the crime, only 30% of participants in the three-perpetrator condition and 20% in the five-perpetrator condition made accurate identifications (compared to 40% in the one-perpetrator condition). Shepherd (1983) subsequently demonstrated the poor accuracy of eyewitnesses with as few as two perpetrators when there was a delay of at least one
month between the witnessed event and the attempted identifications. Out of 41 participants, only 20% could identify one of the two men shown in the video, and only one person accurately identified both.

Three published studies and one unpublished thesis to date have attempted to address the multiple perpetrator identification disadvantage demonstrated above, with limited and varying success, through adapted identification procedures (see Table 1.3 for results as a function of target presence and perpetrator). First, Wells and Pozzulo (2006) introduced what they called a two-person serial lineup for a crime involving two perpetrators. After watching a video of a theft, participants viewed a series of suspect photographs two at a time – one photo (filler or suspect) for the assailant lineup was always paired with one photo (filler or suspect) for the accomplice lineup. Participants viewed a series of these coupled photographs until they had seen six photos for the assailant and six for the accomplice. While the lineups for each are flashed side-by-side on the screen, the two suspects are never shown simultaneously; each suspect is always paired with a filler. In theory, the context of one face should aid our ability to recognize or reject the other face. While this novel procedure did not produce significantly more correct identifications when a target was present, it did produce significantly more correct rejections when the target was absent. Implications for these results, and the ones to follow, will be discussed later.

Another novel identification procedure which adapted current PACE procedures in the U.K. was proposed by Hobson and Wilcock (2011) for a three-perpetrator crime. Participants viewed sequential video lineups for the perpetrators, each one after the other (i.e., perpetrator 1-2-3), but were not allowed to make any identification decisions. Then, they viewed the sequential lineups in reverse order (i.e., perpetrator 3-2-1) and were asked to make an identification decision before
moving on to the next lineup. The authors hypothesized that this technique would allow the participants to anchor their memories of the event with the appropriate lineup before making any decisions, thus avoiding source monitoring errors in confusing which perpetrator performed which action during the crime. Meanwhile, the order in which the perpetrators were presented was intended to reduce the high cognitive load participants experience as a result of viewing all lineups. This novel presentation style produced an increase in correct identifications in target-present lineups for two out of the three perpetrators, but had no effect on correct rejections for target-absent lineups for any of the innocent suspects.

Next, Dempsey and Pozzulo introduced the elimination lineup for use with adults (2008) and children (2013) making identification decisions after witnessing a two-perpetrator crime. An elimination lineup is a two-step decision process whereby witnesses are initially asked to look at a simultaneous lineup with six photos and choose the person that looks most like the perpetrator. All other photos are then removed, and the witness is asked to compare his/her memory of the perpetrator to the chosen photo and determine if this person is the actual perpetrator. In this study, when the thief was absent from the lineup, the elimination lineup procedure produced significantly more correct rejections compared to a standard simultaneous lineup. When the accomplice was present in the lineup, the sequential lineup produced more correct identifications compared to the elimination lineup.

Lastly, Dempsey (2012) conducted two experiments for two-perpetrator crimes, attempting to use the face of one perpetrator to cue the identification of the other. In both experiments, participants viewed a two-perpetrator crime and 5-10 days later were asked to identify one of the perpetrators. As a cue, they were provided either with the correct face of the second perpetrator (correct cue), a second face that
they had never actually seen before (incorrect cue), or no face (no cue). In Experiment 1, accurate cues and no cues led to more correct identifications of the actual perpetrator compared to incorrect cues. However, there was no difference in performance between participants who received an accurate cue and those who received no cue at all. Correct rejection rates of innocent suspect lineups did not vary based on the cues. In Experiment 2, participants were given the same types of cues, but were also given either biased or neutral instructions. There were no differences between groups based on cue veracity.

In summary, the adapted sequential lineup (Hobson & Wilcock, 2011) increased correct identifications for two out of three perpetrators, the elimination lineup (Dempsey & Pozzulo, 2008; 2013) increased adult’s correct rejections for both innocent suspects of both perpetrators, but only increased children’s correct rejections for one perpetrator and decreased children’s accurate identifications for the other perpetrator. All other methods had no significant effect on correct identifications of actual perpetrators or correct rejections of innocent suspects. Although Dempsey’s (2014) cued lineup experiments demonstrated that correct cues were associated with higher rates of correct identification, there was no meaningful difference between providing a correct cue and providing no cue. These results suggest that incorrect cues may have undermined identification performance, but that providing a correct cue did not enhance identification performance and therefore does not represent a benefit to the novel procedure. More importantly, these findings are difficult to explain and provide limited insight to apply to future research on multiple perpetrator identifications. This is because when these new methods produce null or inconsistent findings, it is unclear if it is because (1) the proposed adaptations did not address the mechanisms they were intended to, or (2) the theories used to justify these adaptations
Table 1.3 Procedure and results for eyewitness identification experiments that test identification procedures created or adapted for multiple perpetrator crime

<table>
<thead>
<tr>
<th>Lineup</th>
<th>Perpetrator 1</th>
<th></th>
<th>Perpetrator 2</th>
<th></th>
<th>Perpetrator 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target-present</td>
<td>Target-absent</td>
<td>Target-present</td>
<td>Target-absent</td>
<td>Target-present</td>
<td>Target-absent</td>
</tr>
<tr>
<td></td>
<td>Hit</td>
<td>Foil</td>
<td>FR</td>
<td>CR</td>
<td>Foil</td>
<td>Hit</td>
</tr>
<tr>
<td>Wells &amp; Pozzulo, 2006</td>
<td>.12 (3)</td>
<td>.44 (11)</td>
<td>.44 (33)</td>
<td>.76 (19)</td>
<td>.24 (6)</td>
<td>.24 (6)</td>
</tr>
<tr>
<td>N = 150; Delay = 25 min</td>
<td>Sequential</td>
<td>.12 (3)</td>
<td>.36 (09)</td>
<td>.52 (13)</td>
<td>.44 (11)</td>
<td>.56 (14)</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
<td>.24 (6)</td>
<td>.40 (10)</td>
<td>.36 (9)</td>
<td>.56 (14)</td>
<td>.44 (11)</td>
</tr>
<tr>
<td>Hobson &amp; Wilcock, 2011</td>
<td>Adapted sequential</td>
<td>.72 (13)*</td>
<td>.06 (1)</td>
<td>.22 (4)</td>
<td>.33 (6)</td>
<td>.67 (12)</td>
</tr>
<tr>
<td>N = 72; Delay = 30 min</td>
<td>Sequential</td>
<td>.22 (4)</td>
<td>.50 (9)</td>
<td>.28 (5)</td>
<td>.33 (6)</td>
<td>.67 (12)</td>
</tr>
<tr>
<td>Dempsey &amp; Pozzulo, 2008</td>
<td>Elimination</td>
<td>.65 (20)</td>
<td>.03 (1)</td>
<td>.32 (10)</td>
<td>.80 (28)*</td>
<td>.20 (7)</td>
</tr>
<tr>
<td>N = 132; Delay N/A</td>
<td>Simultaneous</td>
<td>.56 (19)</td>
<td>.12 (4)</td>
<td>.32 (11)</td>
<td>.53 (17)</td>
<td>.47 (15)</td>
</tr>
<tr>
<td>Dempsey &amp; Pozzulo, 2013</td>
<td>Elimination</td>
<td>.42 (8)</td>
<td>.26 (5)</td>
<td>.32 (6)</td>
<td>.75 (18)*</td>
<td>.25 (6)</td>
</tr>
<tr>
<td>N = 90; Delay = 25 min</td>
<td>Simultaneous</td>
<td>.28 (8)</td>
<td>.38 (11)</td>
<td>.34 (10)</td>
<td>.44 (8)</td>
<td>.56 (10)</td>
</tr>
<tr>
<td>Dempsey, 2012a</td>
<td>Accurate cue</td>
<td>.47 (8)</td>
<td>.24 (4)</td>
<td>.29 (5)</td>
<td>.71 (12)</td>
<td>.29 (5)</td>
</tr>
<tr>
<td>N = 221; Delay = 5-10 days</td>
<td>Inaccurate cue</td>
<td>.19 (3)</td>
<td>.19 (3)</td>
<td>.63 (10)</td>
<td>.38 (6)</td>
<td>.63 (10)</td>
</tr>
<tr>
<td></td>
<td>No cue</td>
<td>.50 (8)</td>
<td>.44 (7)</td>
<td>.06 (1)</td>
<td>.33 (5)</td>
<td>.67 (10)</td>
</tr>
</tbody>
</table>

Note. This chapter references comparisons between tested lineup procedures as they affect correct answers only (i.e., hits for target-present lineups and correct rejections for target-absent lineups). No lineup procedure improved hits for all perpetrators, but the Dempsey and Pozzulo (2008) elimination lineup improved correct rejections for all perpetrators for adult witnesses, but not child witnesses. Note that, with the exception of Hobson and Wilcock (2011), all studies designated a central perpetrator and accomplice. In this table, the central perpetrator is under Perpetrator 1 and the accomplice is Perpetrator 2. FR stands for false-rejection and CR stands for correct-rejection.

+ comparisons where significant differences favored the adapted lineup procedure
* comparisons where significant differences favored the old procedure or control group

6 Experiment b is an extension of Experiment a, but with the addition of biased vs. neutral witness instructions. Experiment b is not included in this table because the data was reported collapsed across variables that do not allow for appropriate comparison here. However, results are similar to Experiment a.
are not relevant to the multiple perpetrator identification disadvantage. For example, in the two-person serial lineup (Wells & Pozzulo, 2006), the context of one face should theoretically aid our ability to recognize or reject the other face. But when this lineup does not improve identification accuracy, it is unclear if it is because contextual cuing is not useful for faces in a lineup context or because the suspects are never shown together, and thus are not cuing memory. Ultimately, it is difficult to interpret these results because research to date on identification lineup solutions reflects a premature attempt to fix problems that are still not well understood, meaning the adapted lineups amount to trial-by-error solutions. Sections 1.3 and 1.4 take a step back to consider underlying issues in memory and decision-making that may impact on the performance of these adapted lineup procedures.

**Decision-making in recognition and identification tasks**

An identification decision is composed of two overlapping, but distinct, processes for the eyewitness: memory and decision-making. Memory is the representation of the perpetrator, like a picture in the mind’s eye of the person to be remembered. Because memory is not a tape-recorder, it is likely for this representation to be degraded over time, and it is possible for it to be eventually forgotten, distorted by external information, or even replaced (Deffenbacher, Bornstein, McGorty, & Penrod, 2008; Loftus, 2005). Decision-making here refers to the choice made at the time of the identification. Ideally, the decision to identify a suspect or reject a lineup is based on the actual memory of the perpetrator, even though that memory may have been degraded or distorted over time. However, it is entirely possible to make a decision in the absence of memory; an eyewitness may reject a lineup because they do not remember the perpetrator, or they may choose a suspect in spite of the fact that they do not clearly remember the perpetrator (e.g.,
Horry, Halford, Brewer, Milne, & Bull, 2014; Innocence Project, 2017). It is also possible to make a decision based on a wrong memory; an eyewitness may, for example, use an inaccurate memory to misidentify the innocent suspect as the perpetrator.

To date, eyewitness researchers and practitioners have assumed that an eyewitness is making a decision based solely on the information related to that particular identification procedure, meaning that it is presumed that an eyewitness making multiple identification decisions is making a series of independent decisions. The dearth of research and discussion on the relationship between multiple identification decisions suggests that researchers and practitioners have not considered otherwise. However, research in perception and memory demonstrates that a current decision \((i)\) can be influenced by a previous one \((i - j)\), so that a current response may favor (assimilation) or disfavor (contrast) the preceding responses (Treisman & Williams, 1984). In other words, in a series of trials presented one-after-another, the responses, although separate, are not independent. These **sequential dependencies** appear in perception, classification, and recognition tasks where participants make multiple, sequential decisions (see Malmberg & Annis, 2012)—tasks which present a theoretical overlap with making multiple eyewitness identification decisions.

**Sequential dependencies: The link between multiple decisions.** Sequential dependency can be demonstrated in its simplest form in a traditional detection experiment. Howarth and Bulmer (1956) seated participants in a dark room with a flash-bulb set at a 50% detection rate at a given intensity, meaning that the light was bright enough to be detected, but dim enough that participants only reported seeing it half of the time. The momentary flashes were accompanied by the sound of a bell, so
that when participants heard the bell ring, they indicated whether or not they had seen the flash of light (yes vs. no). At 50% detection, participants made errors half of the time; errors that should theoretically display natural fluctuations and therefore appear randomly throughout the hundreds of trials. However, participants demonstrated a tendency to assimilate responses towards previous ones, meaning that a no response was more likely to be followed by another no response than a yes response. At some points, the light signal was omitted so that the bell rang without the accompanying light flash. When the experimenters forced a sequence of three of these blank trials (no-no-no), they found the same degree of assimilation for the subsequent fourth response as for three natural occurring negative responses (i.e. participants were more likely to respond no’s after a sequence of no’s). Such sequential dependencies are found in a variety of tasks, including absolute judgments of sound (Holland & Lockhead, 1968), and the perceptual classification of facial expressions (Hsu & Yang, 2013).

The mechanism underlying sequential dependencies remains an ongoing debate (see Malmberg & Annis, 2012), with attempts to model sequential dependencies favoring one of the two systems involved in a perception task: the cognitive system and the decisional processes. Some models consider sequential dependencies to arise from biased decision-making (e.g., Treisman & Williams, 1984). According to these models, assimilation results from the observer’s short-term assumption that the most recent stimulus is also the most likely to occur again. However, patterns of contrasting answers are the result of the observer attempting to correct decisional criteria to a desirable level in the long-term. Proponents of these models argue that fluctuations in response bias explain why judgments show assimilation immediately following trial i, but revert to contrast after a few trials.
(Treisman & Williams, 1984; Treisman, 1985). On the other side of the debate are models arguing that sequential dependencies arise either entirely, or at least in part, from the cognitive system (Brown, Marley, Donkin, & Heathcote, 2008; Matthews & Stewart, 2009; Stewart, Brown, & Chater, 2005). In these models, sequential dependencies arise as a result of inappropriate information being carried forward from the previous trial, affecting the perception of the current stimulus.

Malmberg and Annis (2012) were the first to demonstrate sequential dependencies in recognition memory. They presented a series of experiments using traditional recognition paradigms and judgments of frequency recognition tasks to approximate the perception and categorization tasks that routinely demonstrate sequential dependencies. For example, in one experiment, participants studied 40 word pairs and were later tested on their recognition for those words among never-studied words. As with Howarth and Bulmer’s (1956) light-detection experiment, participants were more likely to respond old if they had responded old (rather than new) on the previous trial, regardless of whether the previous response was correct (hit) or incorrect (false-alarm). The appearance of sequential dependencies was consistent across several replications with different stimuli, including landscape images, and picture-word pairs. Interestingly, although sequential dependencies arose as expected, the patterns for these effects across four follow-up experiments were different from those previously found in perception experiments. For example, assimilation was not impacted by varying time between trials (cf. the strength of assimilation decaying over time), assimilation was present in judgment of frequency tasks regardless of whether feedback was provided, and did not reverse to contrast with increased trials between responses (cf. assimilation after the first trial, but reversed to contrast at later trials when feedback is provided). For the purpose of this
dissertation, further details of differences are not relevant, but rather the fact that the differences exist. Because perception tasks and recognition tasks share equivalent decisional processes, these discrepant patterns suggest that sequential dependencies arise from the different cognitive systems supporting perceptual vs. mnemonic tasks (Malmberg & Annis, 2012). Therefore, their results reinforce perception models that posit response interference from the previous trials as the root of sequential dependencies.

**Sequential dependencies and eyewitness identification decisions.** Multiple perpetrator crimes present a framework in which relatively few sequential decisions are made, and in which these decisions have serious consequences. Sequential dependencies measured in the recognition paradigm have no substantial effect on overall recognition accuracy because the beneficial and detrimental sequences of dependencies will typically balance out over the many trials, reducing its impact on the overall accuracy for recognition (e.g., Malmberg & Annis, 2012). Given that identification paradigms lack the many trials needed to balance out recognition accuracy, the appearance of sequential dependencies in the multiple perpetrator identification context would be a matter of substantial impact and cause for concern. Despite the clear theoretical relevance of sequential dependencies to witnesses making identification decisions for multiple perpetrators (Hobson et al., 2012), it is surprising that the phenomenon remains untested in the eyewitness identification context.

In summary, when people are asked to make multiple decisions in a row, a previous decision can influence the decisions that follow. In other words, these multiple decisions are connected. Critically, this pattern of connection appears to hold true for successive tests of recognition—decisions which we otherwise assume
to rely exclusively on the memory for the previously-studied word, object, or landscape. Furthermore, these decisions are similar to the identification decisions that eyewitnesses are asked to make during an investigation involving multiple perpetrators, and are therefore a critical avenue of research within the field of multiple perpetrator identifications. One of the aims of this thesis is to test for the independence of making multiple identification decisions for the multiple suspects of a multiple perpetrator crime.

**Memory in recognition and identification tasks**

The multiple face recognition disadvantage. Memory is the other integral portion of the identification process, and it is desirable to preserve eyewitness memory as much as possible in order to use memory evidence within an investigation. In particular, researchers seek to identify and therefore protect eyewitnesses from external influences on memory (i.e., post-event information; Frenda & Loftus, 2011) and to provide tools to reliably elicit as much memorial information as possible from eyewitnesses (i.e., self-administered interviews; Hope, Gabbert & Fisher, 2011). This is not dissimilar from the same way that forensic scientists seek to safeguard DNA evidence from contamination, degradation over time, and develop new technologies to collect and test the available evidence.

However, the preservation of memory evidence is dependent upon the strength of that initial memory strength to begin with. The recognition and identification of an unfamiliar face from memory is a difficult task, and it becomes even more challenging as the number of unfamiliar faces to be remembered increases (Bindemann, Sanford, Gillatt, Avetisyan, & Megreya, 2012; Clifford & Hollin, 1981; Megreya & Bindemann, 2012; Nortje, Tredeoux, & Vredeveldt, 2017). Indeed, Megreya and Burton (2006) found that eyewitness accuracy dropped between one and
two unfamiliar faces to be encoded. Participants saw either one face or two faces simultaneously and were instructed to study them until they felt confident they could recognize the faces in a subsequent test. When participants were immediately tested with a simultaneous suspect lineup, accuracy for identifying the culprit dropped from 70% in the single face condition to 54% in the two face condition. Megreya and Bindemann (2012) extended the above findings by systematically varying the gender of the perpetrator and accomplice, a personal feature that should be easily distinguishable for eyewitnesses (Bruce & Young, 1986). Results again showed that participants who viewed a single perpetrator had more correct identifications in target-present lineups (48-60%) and more correct rejections in target-absent lineups (61-68%) compared to participants who viewed a double perpetrator crime (20-38% and 58-68%, respectively). Critically, results showed that the two-face disadvantage persisted across conditions of gender combinations, meaning that it did not matter if the two perpetrators were the same gender (highly similar) or different genders (highly dissimilar).

If the so-called two-face disadvantage had disappeared in conditions with the perpetrator and accomplice of different genders, the authors could have reasonably argued that the disadvantage was a result of transference of details between encoded faces during memory retrieval, because eyewitnesses would be mixing up details of faces only when those faces were of the same gender. However, the persistence of the two-face disadvantage indicates that the effect may be a result of divided attention during the event that limits the richness of details encoded in the first place. Some evidence suggests that people are only able to process one face at a time; specifically, it is proposed that we suppress the processing of features of one face in order to accurately encode the features of the other (Bindemann, Burton, & Jenkins, 2005;
Bindemann, Jenkins, & Burton, 2007; Boutet & Chaudhuri, 2001; Palermo & Rhodes, 2002). This impediment could explain the mixed results for which targets are most-accurately identified in the eyewitness context: in some cases, the central role (i.e., perpetrator) is best recognized (Dempsey & Pozzulo, 2008), while other times the peripheral role (i.e., accomplice) is best recognized (Wells and Pozzulo, 2006). It is possible that the circumstances of the mock-crimes used in different experiments somehow direct the attention towards one particular target over the other, but that conditions forcing participants to split these resources lead to poor accuracy for both. In any case, it appears that attempting to divide these facial-processing resources between perpetrators interferes with proper encoding of perpetrator features. This poor initial memory strength would make subsequent identifications much more difficult (Clifford & Hollin, 1981).

While it is clear that divided attention plays a role, it does not entirely account for the multiple-face recognition disadvantage. For example, Bindemann and colleagues (2012) asked participants to study either one face alone or two faces simultaneously and then immediately showed them a lineup for only one of those faces. Participants were much more likely to be accurate when they were shown only one face (67% hits), than if they were shown two (48% hits). Tracking participant eye-movements during encoding demonstrated that while participants tended to split their concentration evenly between the pair of face, the duration spent looking at the target while studying the faces correlated with identification accuracy (Bindemann et al., 2012). This again clearly demonstrates that this drop in accuracy is, in part, because the presence of multiple targets exacerbates the difficulty in encoding unfamiliar faces through divided attention (e.g., Bindemann, Jenkins, & Burton, 2007).
However, the participants in two out of three of Bindemann et al.’s experiments (2012, Experiments 1 and 2) had unlimited time to encode the presented faces: they were instructed to move on to the identification only once they felt confident they would remember the face(s) later. So while it was true that participants had divided attention, they were compensated with unlimited time in which to sufficiently encode each face. Furthermore, a follow-up experiment controlled for divided attention: participants either saw two faces presented sequentially for a fixed amount of time (e.g., 2250 ms) or were shown one face for the same fixed amount of time (2250 ms) followed by an equivalent delay. When participants were shown only one face, their subsequent identifications were significantly more accurate (55% hits) than when they were shown two faces (39% hits). The persistence of the two-face disadvantage led Bindemann and colleagues to conclude that the drop in accuracy also reflects an increase in memorial demand when people attempt to hold two faces in their memory for any period of time. This conclusion is supported by the finding that the only condition that produced similar accuracy rates to the single-face condition was when researchers told participants which of two faces would be tested, ostensibly allowing participants to hold only one face in memory instead of two.

In summary, it appears that the capacity to recognize multiple faces is impacted by factors at encoding, such as divided attention, but also factors at storage, such as memorial demand of holding more faces in memory. Interestingly, the factor that makes multiple-face recognition so challenging— the presence of multiple faces— may actually provide a solution to support memory for multiple faces, and thus provide a tool to better access the memory evidence.

**Enhanced recognition with contextual cues.** Memory researchers have long known that context matters for retrieval. Contextual cues are often implemented to
help individuals recall seemingly-forgotten details in episodic memory (encoding specificity principle; Thomson & Tulving 1970), including to facilitate eyewitness recall during investigative interviews (context reinstatement; Geiselman, Fisher, MacKinnon, & Holland, 1986). It has been hypothesized that these cues work because the retrieval of a memory is dependent upon the way it was stored, and an item of episodic memory is, by nature, nested within our experience of the relevant event (Tulving & Thomson, 1973; Polyn, Norman, & Kahana, 2009). Thus, episodic memory is not only tied to temporal markers (when an event occurred), but can also be integrated with the other memory traces of that event. For example, according to the Context Maintenance Retrieval Model, we naturally recall items in semantic clusters (i.e., book is semantically related to paperback, hardcover, nonfiction, bestseller) as a result of our long-standing associations between items in our experience. However, we also have a tendency to recall items in temporal clusters (i.e., the order in which the items appeared) and source clusters (i.e., information from experimenter vs. confederate vs. own knowledge) as a result of the associations we form during study phase between items (Polyn et al., 2009). These patterns represent associations between memories. A contextual cue is used at retrieval to take advantage of the associative nature of memory whereby, for example, peripheral details of the environment can be used to cue retrieval of the critical details.

Thomson and Tulving (1970) demonstrated this experimentally when they had participants memorize pairs of words: a target word and a word that was semantically unrelated (e.g., ocean and piano). During the testing phase, the semantically-unrelated pair word (weak cue) or a semantically-related, but previously-unstudied word (strong cue) was presented in order to elicit the target word. If the target word is “ocean”, a strong cue might be “land”, while the weak cue might be “piano”. Results showed that
participants were better able to recall the target word if it was presented along with the weak cue (i.e., the cue with which it was originally encoded), compared to the strong cue, suggesting that the context of encoding the target word mattered more than the strength of the semantic association.

Contextual cues also benefit recognition memory, including when recognizing pictures (Palmer, 1975) and faces (e.g., Watkins, et al., 1976). Just within the field of face recognition, a variety of external contexts have been tested as a means to enhance recognition for the target faces, including using backgrounds, descriptions, clothing, and other faces (see Davies, 1988 for a review). Winograd and Rivers-Bulkeley (1977) asked participants to memorize pairs of faces during the study-phase—one of which served as the target-face while the other served as the cue. During the test-phase, participants were presented with a target face alongside either a face they had previously studied (correct cue), a face that had not been previously studied (incorrect cue), or no cue at all. In this forced-choice paradigm, recognition performance for target faces was enhanced by the presentation of correct cue faces and impaired by incorrect cue faces, while performance with no cues fell in-between. Similarly, Watkins and colleagues (1976) demonstrated reduced hit rates for face recognition when the incorrect cues were previously-studied faces that had been paired with another face during study. In other words, simply swapping context, as opposed to introducing new context, also affected face recognition performance.

However, although the above experiments found an enhancing effect of using correct cues, other research has not (e.g. Bower & Karlin, 1974) and while Kan, Giovanello, Schnyer, Makris, and Verfaellie (2007) found that incorrect cuing undermined hit rates, they found no benefit for correct cues (cf. no cues). Issues in encoding and recognition likely underpin these discrepant results. For example, the
context must be strongly encoded in association with the target for recognition. In other words, if the participant did not pay attention to the contextual information, or did not link it with the target information, then the cue will not help to recognize the target (Peris, 1985, as cited in Davies, 1988). Additionally, it appears that context is useful as a recognition memory cue only when other, stronger cues are lacking (Davies, 1988). To explain this, it is first important to understand that some theories of recognition hold that recognition is comprised of two mechanisms (e.g., Mandler, 1980; Peris & Tiberghien, 1984): The first is the perceptual system that is automatically activated and produces fast answers that hinge on the feeling of familiarity and the second is the cognitive system, which is activated when the first does not immediately provide an answer. The second system is slower and searches for other, external information, like context, which will aid its response. Context cues would thus only be used when this second system is activated and there is not a stronger source of information. This process is also captured in the rationale of the outshining hypothesis, which contends that we use the most relevant cues available to recognize faces. When our memory trace is strong, that memory outshines the utility of environmental context (Smith & Vela, 2001). However, for weak memory traces, such as when there was suboptimal encoding or longer retention intervals, context may support memory to improve recognition performance (Mandler, Pearlstone, & Koopmans, 1969).

Cued recognition for faces. Cued recognition presents an interesting means of enhancing face recognition, a concept that may prove useful to the applied fields of eyewitness identification or wanted-persons recognition. The second face in a two-perpetrator crime provides a naturally-occurring contextual cue for the eyewitness. A second face is also a particularly relevant cue for humans, given that it is our natural
tendency to orient and focus attention towards other human faces, thus increasing the chance of incidental associative encoding (Di Giorgio, Turati, Altoè, & Simion, 2012; Langton, Law, Burton, & Schweinberger, 2008). Perhaps face cues, then, could be used to support eyewitnesses of a multiple perpetrator crime while viewing a suspect lineup. This is not an entirely novel idea. Indeed, at least two published experiments (Hobson & Wilcock, 2011; Wells & Pozzulo, 2006) and one unpublished thesis (Dempsey, 2014) have attempted to apply face cuing to support eyewitness memory and adapt identification procedures in the context of multiple-perpetrator crimes. In neither case did results provide convincing evidence that this method of cuing memory could aid lineup identification decisions, and it is of interest to understand why this might be.

Existing research also exposes questions regarding the limit of such an effect, specifically the number of additional faces to be encoded. Research on cued face recognition to date has held encoding conditions constant while manipulating the conditions at retrieval (e.g. Watkins, Ho, & Tulving, 1976; Winograd & Rivers-Bulkeley, 1977). However, pairing faces at encoding represents the minimum number of faces we might encode when attempting to implement cued face recognition. Limiting our consideration to pairs of faces fails to reflect the variability in group sizes that we encounter every day, and it is unclear if the benefits identified in previous work extend to conditions in which participants need to encode more stimuli. In other words, if these effects replicate, can they also be useful in situations with more than just two faces? Although encoding additional faces may provide contextual cues to aid subsequent recognition, cognitive load is likely to increase with each additional face requiring encoding. As this load increases, the resources needed to
successfully encode the target face and to develop the associations required to support subsequent cued-recognition may be reduced.

In summary, memory for multiple faces (and thus multiple perpetrators) is impacted by encoding (i.e., divided attention), but also separately by the increased difficulty of storing those multiple faces (i.e., memorial demand). One way to support memory under suboptimal conditions is to use context to enhance the memory trace that exists. In the context of multiple faces, some research suggests that memory for one face of a studied pair can be enhanced using the context of the other face to cue that memory. Such contextual cuing could be useful in the context of multiple perpetrator recognition and identification; however we do not fully understand these effects for pairs of faces, nor whether such effects are also present when more than two faces are present. This thesis aims to examine the effect of cued face recognition and tests this effect with more than two faces.

**Overview of the current thesis**

The current thesis comprises four chapters, including the survey discussed earlier in this chapter and, additionally, five experiments to examine underlying issues in memory and decision-making that impact eyewitness identification in the context of multiple perpetrator crimes. Chapter 2 will present research testing the independence of multiple identification decisions made successively (Experiments 1, 2 and 3). Chapter 3 will present experiments examining the purported utility of associative memory for recognizing the faces of multiple perpetrators (Experiments 4 and 5). In the discussion, an overview of the key findings are presented, followed by theoretical implications for memory and decision-making, practical implications for researchers and police involved with multiple perpetrator identification. These chapters are summarized in further detail below.
Police Survey on Multiple Perpetrator Identifications: Sweden, Belgium, and the Netherlands. A new, exploratory survey was developed to extend previous work by Hobson et al. (2012), which was translated into two languages (Swedish and Dutch) to be distribute to other European countries (Tupper, Sauerland, Sauer, & Hope, in prep). This survey aimed to (a) inform our understanding of the prevalence and characteristics of multiple perpetrator crimes from the perspective of law enforcement agencies, (b) discern how agencies in various countries conduct identification procedures (e.g., lineups, photo-arrays, showups) with multiple perpetrators, and (c) gain insight into how law enforcement agents and eyewitnesses experience the identification process in the context of such crimes. Results demonstrated the differences in police practice between countries that have various degrees of regulation for identification procedures. The reported police practice of placing multiple suspects in the same lineup serves to highlight the importance of understanding how memory for multiple perpetrators is stored and how the presence of the multiple suspects may enhance or undermine memory for the perpetrators. Furthermore, the more commonly-reported practice of separating multiple suspects into separate lineups similarly highlights the important to examine the consequence of testing memory on multiple lineups presented in succession.

Experiments 1 and 2: Testing for sequential dependencies in eyewitness showup identifications. Research in perception and recognition demonstrates that a current decision (i) can be influenced by previous ones (i - j), meaning that subsequent responses are not always independent. In Experiments 1 and 2, we examined the relation of previous identification decisions to subsequent choosing behavior in the context of the multiple showup identification decisions for a multiple perpetrator crime. That is, if it is possible to predict current choosing on a showup
identification decision from previous choosing, it would provide initial evidence that sequential effects may be present in multiple showup identification decisions. In both experiments, participants watched a mock-crime film involving three perpetrators and later made three showup identification decisions, one showup for each perpetrator. Given that research has previously demonstrated that sequential dependencies in recognition are a result of interference from previous trials (Malmberg & Annis, 2012), Experiments 1 and 2 considered both previous signal (target-presentation: present vs. absent) and previous response (Choosing: yes vs. no) as predictors of current choosing behavior (Matthews & Stewart, 2009). Experiments 1 and 2 were similar in procedure, with the exception that Experiment 1 used four trials conditions of target-presentation such that the first two trials were consistent in target-presentation of the stimulus (i.e., target-absent/ target-absent or target-present/ target-present) and the third was either consistent or similar. However, in retrospect, we could not observe the separate effects of previous target-presentation and previous response between the first two trials. Therefore, Experiment 2 extended Experiment 1 by using all patterns of target-presentation. We expected that initial showup responses would predict choosing for subsequent showup responses. In other words, choosing on a previous showup identification would lead to choosing on subsequent ones, and rejecting on a previous showup identification would lead to rejecting on subsequent ones. We also expected previous target-presentation to exert a separate influence on the current identification decision, such that when the previous target being present would predict current choosing and the previous target being absent would predict current rejecting (e.g., Matthews & Stewart, 2009).

**Experiment 3: Testing for sequential dependencies in faces as a function of number of trials.** While consecutive recognition decisions have been shown to
produce sequential dependencies, such effects were found to be inconsistent within the eyewitness identification context. Experiment 3 examined whether methodological differences between the recognition and eyewitness paradigms used in previous research on sequential dependencies could account for the inconsistent findings presented in Experiments 1 and 2. This experiment therefore sought to replicate previous recognition research in sequential dependencies using word and landscape stimuli, and extend these effects to face stimuli. This experiment also examined whether the strength of these sequential dependencies changed as a result of the number of test trials (i.e., beginning vs. middle of experiment). Participants studied pairs of words, landscapes, or faces, and were later tested for recognition. We expected that sequential dependencies would arise in all categories of stimuli, but that the strength of this effect would be weaker in the beginning compared to the middle of the experiment.

**Experiments 4 and 5: Cued face recognition: Testing a tool to enhance eyewitness performance.** In contrast to the previous experiments, Experiments 4 and 5 focused on the association between memories for multiple faces. The presence of multiple faces increases memorial demand, but also provides a naturally-occurring contextual cue that may promote recognition. Early experiments demonstrated the benefits of cues on face recognition when faces were studied as pairs (e.g., Watkins, Ho, & Tulving, 1976), and some more recent experiments have tried to apply such techniques to the context of eyewitness identification when multiple perpetrators are involved in the crime (e.g., Wells & Pozzulo, 2006). However, there is a paucity of contemporary research systematically examining cued face recognition effects in applied contexts. Experiments 4 and 5 sought replicate previously-reported enhancing effects in cued face recognition, to investigate the mechanisms underlying those
effects, and to determine whether such effects could include more than two faces, as many crimes involve more than two perpetrators. To do this, we included the traditional condition in which participants study paired faces, and added both a control condition in which participants studied single faces and another experimental condition in which participants studied groups of four faces. At test, participants in the single-face condition were tested only on those individual faces without cues. Participants in the two and four-face conditions were tested using no cues, correct cues (a face previously studied with the target test face), or incorrect cues (a never-before-seen face). We hypothesized that correct cuing would enhance recognition accuracy for target faces compared to no cue and incorrect cues, but that this effect would be stronger for participants studying two faces compared to participants studying four faces at a time.

**Overview of chapters.** Initial analyses from the Police Survey were presented in the Introduction (Chapter 1). Experiments 1, 2 and 3 are presented together in Chapter 2, while Experiments 4 and 5 are presented together in Chapter 3. In Chapter 4, the results of the survey and five experiments, theoretical implications for the field of eyewitness memory and identification, as well as implications for current and future police practice are discussed.
Chapter 2: Showup identification decisions for multiple perpetrator crimes:

Testing for sequential dependencies

Abstract

Research in perception and recognition demonstrates that a current decision \( (i) \) can be influenced by previous ones \( (i - j) \), meaning that subsequent responses are not always independent. Experiments 1 and 2 tested whether initial showup identification decisions impact choosing behavior for subsequent showup identification responses. Participants watched a mock-crime film involving three perpetrators and later made three showup identification decisions, one showup for each perpetrator. Across both experiments, evidence for sequential dependencies for choosing behavior was inconsistent. In Experiment 1, responses on the second, target-present showup assimilated towards previous choosing. Yet, in Experiment 2, responses on the second showup contrasted previous choosing regardless of target-presents. Experiment 3 examined whether differences in stimuli and number of test trials in the eyewitness (vs. basic recognition) paradigm could account for the absence of hypothesized patterns of sequential dependencies in Experiments 1 and 2. Sequential dependencies were detected in recognition decisions over many trials, including recognition for faces: the probability of a yes response on the current trial increased if the previous response was also yes (vs. no). However, choosing behavior on previous trials did not predict individual recognition decisions on the current trial. That sequential dependencies did not impact observed choosing behavior on identification decisions suggests that the integrity of identification and recognition decisions is not likely to be impacted by making the multiple decisions in a row.

*Experiments 1, 2, and 3 are presented together in Chapter 2 because these experiments are being prepared for publication together.*
Introduction

In October 2015, news outlets (e.g., “Frontière belge”, 2015) featured security footage of an unresolved case: the attempted abduction of a truck driver on the French-Belgian border. As the truck driver walked around the rear of his truck, two men appeared and attacked him. While the two perpetrators struggled to force the driver into the back of a waiting car, an elderly passerby intervened, pulling at the perpetrators’ jackets and trying to place himself between them and the truck driver. Following the failed abduction and a hurried, but equally unfruitful search for the truck driver’s keys, the two men fled the scene by car.

This case is just one example of the many violent crimes that are committed by multiple perpetrators. Gang violence (Juodis et al., 2009), hate crimes (Sandholtz, Langton, & Planty, 2013), rapes (Horvath & Kelley, 2009), and assaults (Hobson, et al., 2012) are often committed by perpetrators working as a group. In fact, the rising rate of such crimes appears to be a global phenomenon. For example, in Finland, Sweden, and the Netherlands, 13-17% of homicides between 2003 and 2006 involved two or more perpetrators (Liem et al., 2013), while the proportion of homicides with multiple perpetrators in the U.S. reached 20% in 2008 (nearly double that reported in 1980; Cooper & Smith, 2011). These crimes often involve victims or bystanders as eyewitnesses—like the driver and the passerby above—who may be asked to identify multiple suspects related to the multiple perpetrators. Yet, the decades of research focused on uncovering and understanding factors that affect accuracy for eyewitness identification procedures typically considers only identifications of a single perpetrator, providing little empirical evidence to support or oppose recommendations in protocols specific to the context of multiple perpetrator crime. Should police departments, for instance, follow the example of the U.K. and multiply “best practice”
by creating a new lineup for each suspect of a different perpetrator (Police and Criminal Evidence Act, 1984)? If so, does the order of presentation of identification tests affect the reliability of the evidence obtained? Or does the act of making multiple identification decisions affect the decisions themselves?

In this paper, we address this last question, examining the consequences of testing memory for multiple perpetrators (e.g., Malmberg, Lehman, Annis, Criss, & Shiffren, 2014). We present three experiments examining whether initial showup identification decisions affect witness choosing behavior and accuracy on subsequent showup decisions. We aimed to determine whether sequential dependencies (i.e., where choosing behavior on previous tests influences choosing on a current test) should be considered in cases when eyewitnesses are asked to make multiple identification decisions, specifically when those decisions pertain to the different suspects in a multiple perpetrator crime.

**Identification of multiple perpetrators**

Clifford and Hollin (1981) first revealed the difficulty of eyewitness identification in the context of multiple perpetrator crimes when they had participants view a non-violent event with one, three, or five perpetrators. Despite only having to select the main perpetrator from a target-present lineup immediately following the crime, only 30% of participants in the three-perpetrator condition and 20% in the five-perpetrator condition made accurate identifications (compared to 40% in the one-perpetrator condition). More recently, Megreya and Binnewa (2012) demonstrated a similar drop in accuracy with as few as two unfamiliar faces to be encoded. Participants viewed a mock crime with one perpetrator alone or with an accomplice and were subsequently asked to identify the perpetrator. The presence of a second person at encoding was associated with decreased identification accuracy in target-
present lineups (lower hit rates and higher miss rates). Approximately 54% of participants who saw the perpetrator alone were able to accurately identify him/her, compared to only 29% of participants who saw the perpetrator with an accomplice.

To date, three procedures have been proposed to address the applied issue of the multiple perpetrator identification disadvantage. The two-person serial lineup (Wells & Pozzulo, 2006), the elimination lineup (Dempsey & Pozzulo, 2008, 2013), and an adapted sequential identification procedure (Hobson & Wilcock, 2011) were each tested against traditional simultaneous lineups, sequential lineups, or both. The results were mixed, and any improvements associated with these methods depended upon which target identity was being presented (i.e., accomplice vs. perpetrator), the presence or absence of the target in the lineups, or both. Unfortunately, when these new methods fall short, we do not know if it is because the proposed adaptations did not address the mechanisms they intended to, or if the theories used to justify these adaptations are ultimately not relevant to the multiple perpetrator identification disadvantage. For example, the two-person serial lineup is intended to provide context to aid memory by presenting the sequential lineups of the culprit and of the accomplice at the same time (Wells & Pozzulo, 2006). While the lineups for each are flashed side-by-side on the screen, the two suspects are never shown simultaneously, but always paired with a filler. In theory, the context of one face should aid our ability to recognize or reject the other face. But when this lineup does not improve identification accuracy, is it because contextual cuing is not useful for faces in a lineup context? Or is it because the suspects are never shown together, and thus are not cuing memory? Perhaps it is difficult to interpret their results because they are premature attempts to fix problems that are still not well understood, meaning the adapted lineups amount to trial-by-error solutions.
Other research has aimed to test the theoretical causes of the multiple perpetrator identification disadvantage. Shallow encoding (Megreya & Burton, 2006) or increased memorial demand (Bindemann, Sanford, Gillatt, Avetisyan & Megreya, 2012) have more recently been explored as reasons for the decreased identification performance for multiple perpetrator crimes, and both appear to play a role. However, there is another independent factor that is unique to multiple perpetrator identification that has yet to be considered: the decisional structure of making multiple identifications. Below, we explore how the act of making multiple identifications may undermine the integrity of those decisions.

**Sequential dependencies in perception and recognition**

An individual police lineup has been likened by researchers to a real-world signal detection decision, but with the modification to include filler (i.e., non-suspect) misidentifications (Palmer, Brewer, & Weber, 2010, 2012). The signal detection model, however, mathematically assumes independence of trials, for which a decision is based solely on the evidence present in that trial. In contrast, research in perception and memory demonstrates that a current decision \(i\) can be influenced by a previous one \(i-j\), so that a current response may favor (assimilation) or disfavor (contrast) the preceding responses (Treisman & Williams, 1984). In other words, in a series of trials presented one-after-another, the responses, although separate, are not independent. These *sequential dependencies* appear in perception, classification, and recognition tasks where participants make multiple, sequential decisions—tasks which present a theoretical overlap with making multiple eyewitness identification decisions.

Sequential dependency can be demonstrated in its simplest form in a traditional detection experiment. Howarth and Bulmer (1956) seated participants in a
dark room with a flash-bulb set at a 50% detection rate at a given intensity, meaning that the light was bright enough to be detected, but dim enough that participants only reported seeing it half of the time. The momentary flashes were accompanied by the sound of a bell, so that when participants heard the bell ring, they indicated whether or not they had seen the flash of light (yes vs. no). At 50% detection, participants will make errors half of the time; errors that should theoretically display natural fluctuations and therefore appear randomly throughout the hundreds of trials. However, participants demonstrated a tendency to assimilate responses towards previous ones, meaning that a no response was more likely to be followed by another no response than a yes response. Further still, at some points, the light signal was omitted so that the bell rang without the accompanying light flash. When the experimenters forced a sequence of three of these blank trials (no-no-no), they found the same degree of assimilation for the subsequent fourth response as for three natural occurring negative responses. Such sequential dependencies are found in a variety of tasks, including absolute judgments of sound (Holland & Lockhead, 1968), and the perceptual classification of facial expressions (Hsu & Yang, 2013).

The mechanism underlying sequential dependencies remains an ongoing debate, with attempts to model sequential dependencies favoring one of the two systems involved in a perception task: the cognitive system and the decisional processes. Some models consider sequential dependencies to arise from biased decision-making (e.g., Treisman & Williams, 1984). According to these models, assimilation results from the observer’s short-term assumption that the most recent stimulus is also the most likely to occur again. However, patterns of contrasting answers are the result of the observer attempting to correct decisional criteria to a desirable level in the long-term. These fluctuations in response bias purport to explain
why judgments show assimilation immediately following trial \( i \), but revert to contrast after a few trials. On the other side of the debate are models arguing that sequential dependencies arise either entirely, or at least in part, from the cognitive system (Brown, Marley, Donkin, & Heathcote, 2008; Matthews & Stewart, 2009; Stewart, Brown, & Chater, 2005). In these models, sequential dependencies arise as a result of inappropriate information being carried forward from the previous trial, affecting the perception of the current stimulus.

Malmberg and Annis (2012) were the first to demonstrate sequential dependencies in recognition memory. They presented a series of experiments using traditional recognition paradigms and judgments of frequency recognition tasks to approximate the perception and categorization tasks that routinely demonstrate sequential dependencies. For example, in one experiment, participants studied 40 word pairs and were later tested on their recognition for those words among never-studied words. As with Howarth and Bulmer’s (1956) light-detection experiment, participants were more likely to respond old if they had responded old (rather than new) on the previous trial, regardless of whether the previous response was correct (hit) or incorrect (false-alarm). The appearance of sequential dependencies was consistent across several replications with different stimuli, including landscape images, and picture-word pairs. Interestingly, the patterns for sequential dependencies were different from those found in perception experiments. Because perception tasks and recognition tasks share equivalent decisional processes, these discrepant patterns suggest that sequential dependencies arise from the different cognitive systems supporting perceptual vs. mnemonic tasks (Malmberg & Annis, 2012). Therefore, their results reinforce perception models that posit response interference from the previous trials as the root of sequential dependencies.
The current research

Studies investigating the cause of the multi-face recognition disadvantage (e.g., Bindemann et al., 2012; Megreya & Burton, 2006) tend to focus on the encoding conditions: how factors that affect perception and attention interfere with encoding, and thus damage chances of identification from the outset. Consequently, studies adapting lineups that were originally designed for single-perpetrator crimes so far considered these encoding difficulties and adjusted methodology in attempts to compensate for the resulting impoverished memory (e.g., Hobson & Wilcock, 2011; Wells & Pozzulo, 2006). While this is a reasonable starting point to investigate multiple perpetrator identifications, it is also important to explore other factors that may affect identification decisions. In this vein, we investigated the possibility of sequential dependencies within the eyewitness paradigm. Specifically, how does the act of making multiple identification decisions for unique perpetrators affect the reliability of those decisions? Despite the clear theoretical relevance of sequential dependencies to witnesses making identification decisions for multiple perpetrators, the phenomenon remains untested in the eyewitness identification context.

Multiple perpetrator crimes present a framework in which relatively few sequential decisions are made, and in which these decisions have serious consequences. Sequential dependencies measured in the recognition paradigm have little substantial effect on overall recognition accuracy because the beneficial and detrimental sequences of dependencies will typically balance out over the many trials, reducing its impact on the overall accuracy for recognition (e.g., Malmberg & Annis, 2012). Considering identification paradigms lack the many trials needed to balance out recognition accuracy, the appearance of sequential dependencies in this context would be a matter of substantial impact and cause for concern. Therefore, we tested
for sequential dependency effects within the eyewitness identification context by having participants make multiple, consecutive showup decisions.

Show ups are particularly well-suited for an initial test for sequential dependencies within the eyewitness context for three reasons. First, showups (live or photographic) are a common identification procedure around the world (e.g., Davis, Valentine, Memon, & Roberts, 2015; Police Executive Research Forum, 2013). Second, forced-report showup decisions (Is this the perpetrator? Yes vs. no) emulate the binary-decision tasks in which sequential dependencies have been consistently observed. Third, showups permit a controlled investigation of sequential dependencies on identification decision-making free from the influence of lineup construction variables (e.g., filler similarity, lineup presentation method). If sequential dependencies are found to affect showup decision-making, subsequent investigations can determine how these effects interact with lineup composition and presentation variables.

Across two initial experiments, we examined the relation of previous identification decisions to subsequent choosing behavior in the context of the multiple showup identification decisions for a multiple perpetrator crime. That is, if it is possible to predict current choosing on a showup identification decision from previous choosing, it provides initial evidence that sequential effects may be present in multiple showup identification decisions. Given that research has previously demonstrated that sequential dependencies in recognition are a result of interference from previous trials (Malmberg & Annis, 2012), Experiments 1 and 2 consider both previous signal (target-presence: present vs. absent) and previous response (Choosing: yes vs. no) as predictors of current choosing behavior (Matthews & Stewart, 2009). We expected that initial showup responses would predict choosing for subsequent
showup responses. In other words, choosing on a previous showup identification would be associated with choosing on subsequent ones, and rejecting on a previous showup identification would be associated with rejecting on subsequent ones. We also expected previous target-presence to be separately associated with the current identification decision, such that when the previous target being present would predict current choosing and the previous target being absent would predict current rejecting (e.g., Matthews & Stewart, 2009).

We also consider the possibility of an interaction between current target-presence and previous choosing on the current choosing, such that participants confronted with a target-absent trial would be influenced more by the previous target-presence and choosing compared to a target-present trial. Non-memorial factors tend to exert stronger effects on recognition memory tasks when the target stimulus is absent and there is no opportunity for genuine recognition (Palmer, Sauer, & Holt, 2017). In other words, if memory is not able to provide the answer, people look for other cues to influence that decision. In this way, sequential dependencies might represent an attempt to use imperfect cues to guide decision-making under conditions of uncertainty (Gigerenzer, Hoffrage, & Kleinboelting, 1991). Although straightforward sequential dependencies should arise regardless of target presence, it is possible that the strength of the effect will vary depending on whether the target is present or not.

Although Experiments 1 and 2 were conducted separately, they used similar methodologies and analyses to answer the same question. Thus, although the data is not collapsed across experiments, the methods and results are presented together.
Experiments 1 and 2

Participants and Design

A total of 411 participants were tested, 404 of which were included in analyses. Participants either completed the experiment in the lab (Experiment 1, \( N = 120 \)) or online (Experiment 2, \( N = 291 \)). The average age of participants was 21 years \( (M = 20.77, SD = 3.64) \). They were compensated with a €5 gift voucher (Experiment 1) or participation credit (Experiments 1 and 2).

Participants viewed a three-person mock crime video and were subsequently presented with three photographic showups, one for each of the three perpetrators. This number of showups was chosen in order to create a minimum sequence of responses within which sequential dependencies could be tested. In Experiment 1, we aimed to provide an initial test of sequential dependency in facial identification. Four conditions were chosen to optimize conditions for sequential dependencies through an established pattern of target-present and target-absent showup photographs (Howarth & Bulmer, 1956). The first and second showups were always consistent in target-presence; they were either both target-absent (TA) or both target-present (TP), while the third showup was either consistent or different, leading to four conditions with targets: (1) TA/TA/TP, (2) TA/TA/TA, (3) TP/TP/TA, and (4) TP/TP/TP. In retrospect, we realized this also meant that we were not able to disentangle the effect of target-presence between showups 1 and 2 on showup 2. Thus, Experiment 2 implemented all combinations of target presence by adding four additional conditions with targets: (5) TA/TP/TA, (6) TA/TP/TA, (7) TP/TA/TP, and (8) TP/TA/TA. Presentation order of targets (i.e., 123, 132, 231, 213, 312, 321) was counterbalanced for both experiments.
General Method for Experiments 1 and 2

Materials.

Crime video. In the 2:45 min mock-crime video, the male victim arrives by bike and locks it against a railing with other bikes. Three target people, one woman and two men, are shown in the background gesturing towards the victim. When the victim walks into a nearby building, the thieves use a hand-saw to break the locks of two bikes, including the victim’s, and walk away with the bikes. Each target actor in the video has approximately 15-20 s of close-up shots in which their faces are clearly visible.

Targets. All three targets were Caucasian university students. The female target had an average build, long, straight, blond hair and was judged by pilot participants to be approximately 22 years old. The first male target had an average build with short, dark-blonde hair, and beard-scruff, and was judged to be approximately 25 years old. The second male target was comparatively shorter to the first target, with short, blond hair, and was judged to be approximately 23 years old.

Showups. Three target-absent and three target-present showups were constructed, one for each of the three perpetrators. The showups consisted of color photographs 4.39 x 5.89 cm in size. The targets were photographed on the same day as the stimulus event was filmed, but wore different clothing. One innocent suspect was selected as a replacement for each target in the target-absent showups. The replacements were chosen based on similarity to the actual target, as established by a pilot study with $N = 22$ participants (age: $M = 27.45$, $SD = 12.14$). Specifically, replacements were rated as statistically similar to the perpetrator with regard to memorability, distinctiveness, and typicality (cf. Wickham, Morris & Fritz, 2000; Vokey & Read, 1992). Participants were also asked to judge the similarity of the
target faces paired with each of their possible replacements. This comparison score was used to match for similarity across the three target-replacement pairings, so that each of the three target-absent showups would be equally difficult for participants to judge. Results of the pilot study are available in supplementary material and can be found in Appendix F.

**Procedure.** Participants arrived at the lab for individual testing sessions (Experiment 1) or received a Qualtrics (Provo, UT) link to complete the experiment online (Experiment 2). Participants were informed that the experiment would be administered using a self-paced computer task. After giving informed consent, participants were told that they would be shown a video and were instructed to pay close attention as they would be asked questions about it later. After watching the mock-crime video, participants completed a 20-30 min filler task by answering a series of questionnaires (Experiment 1 and 2) or by completing a combination of search tasks and word-generation games (Experiment 2). Next, participants were reminded that they had seen a film of three thieves stealing a bike, and were now considered eyewitnesses. They were instructed: *You will be shown a series of three photographs. Each photograph is one suspect for each of the three bike thieves. For each photograph, please decide whether or not the person shown was one of the perpetrators. Once you make a decision, you will move on to the next photo.* A subsequent screen displayed a one-time warning that the persons in the photographs may or may not be the actual perpetrators.

Participants were then shown a photo for one of the perpetrators (Suspect 1: present or absent). A forced-report question asked if the person shown was one of the perpetrators (*yes* or *no*), after which they were asked to indicate how certain they were in their decision (0-100%). The procedure was repeated for Suspects 2 and 3.
Although suspects are numbered here for convenience, presentation order of targets was counterbalanced; meaning Suspect 1 for the eyewitness could correspond to any of the three perpetrators. Following all identification decisions and confidence ratings, participants were shown the photos of those they had positively identified and asked to name the role each played in the crime. However, role assignment and confidence are outside of the scope of the current research and are therefore not addressed further. Finally, participants were thanked for their time and debriefed.

Experiment 2 differs from Experiment 1 in two additional ways. In order to determine whether participants had watched the entire video, a still image of a white arrow and the text “This is a white arrow, please remember this arrow as you will be asked about it later” was added for the last 7 s of the video (after the target event). Following the filler task, participants were asked to name the shape and color presented at the end of the video. This section of the computer task was timed so that the task advanced automatically after 2:52 min regardless of whether or not the video was paused. Therefore, participants that could not correctly name the shape and color \((n = 4)\) were assumed to have not completed the video and were removed from all analyses. A final question prompted participants to describe the environment in which they completed the experiment (e.g., time of day, location, presence of others).

**Results**

In Experiment 1, all 120 participants were retained for data analysis. In Experiment 2, seven participants were removed from data analysis for answering a control question incorrectly (4), not completing the filler task (2), or because Qualtrics recorded their experiment duration time as exceeding four hours and the participant did not respond to requests to elaborate (1), leaving 284 participants.
**Descriptive statistics for choosing on showups.** Across the three showup identification decisions in both Experiments 1 and 2, choosing rates were low, at 34-42%. Overall, only 4-12% of participants chose on all three showups. Meanwhile, 15-22% of participants rejected all decisions. Less than half of participants (26-47%) chose on at least two showups. See Table 2.1 for choosing and accuracy rates for each experiment.

**Experiment 1: Testing for sequential dependencies.** In order to establish the relation of previous identification decisions and both previous and current target-presence with current identification decisions, we performed separate binary logistic regressions for choosing on the second and third showup. For example, for choosing on the second showup, we entered previous target-presence (absent vs. present on Showup 1), current target-presence (absent vs. present on Showup 2) and previous choosing (yes vs. no on Showup 1) as predictors. For choosing on the third showup, we used previous target-presence, current target-presence (Showup 3), and previous choosing (yes vs. no on Showup 1 and Showup 2) as predictors. Because target-

<table>
<thead>
<tr>
<th>Choosing by showup</th>
<th>Overall Choosing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Showup 1</td>
<td>Showup 2</td>
</tr>
<tr>
<td>Expt. 1</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>.53 (32)</td>
</tr>
<tr>
<td>TA</td>
<td>.25 (15)</td>
</tr>
<tr>
<td>Overall</td>
<td>.39 (47)</td>
</tr>
<tr>
<td>Expt. 2</td>
<td></td>
</tr>
<tr>
<td>TP</td>
<td>.54 (76)</td>
</tr>
<tr>
<td>TA</td>
<td>.25 (36)</td>
</tr>
<tr>
<td>Overall</td>
<td>.41 (114)</td>
</tr>
</tbody>
</table>

**Note:** Displayed under “Choosing by showup” are proportions of participants choosing on target-present and target-absent showups. Displayed under “Overall Choosing” are proportions of participants who chose on zero, one, two, or three showups. TA denotes target-absent showups and TP denotes target-present showups.
presence for the first and second showups did not vary in Experiment 1, target-presence for Showups 1 and 2 were included as a single predictor.

In the initial analyses for Showup 2, we included all main effects in the equation. In the initial analyses for Showup 3, we included all main effects and the current target-presence by previous response (selection vs. rejection) interaction. We then sequentially excluded the interaction if non-significant and any non-significant main effects by order of distance from the current decision. However, given our theoretical predictions, previous choosing was always included in the final model. Although we present the results descriptively here, relevant statistics can be found in Table 2.2.

Choosing behavior on second showup. Only target-presence was a significant predictor in the final model. Participants were more likely to choose when the target was present. However, due to the fact that target-presence for Showups 1 and 2 did not vary within subjects, it is unclear if it is current target-presence, previous target-presence, or both that are associated with choosing behavior for Showup 2.

Choosing behavior on third showup. The current target-presence by previous choosing interaction was significant. Examination of simple effects revealed that only when the current trial was target-present, choosing on Showup 2 predicted choosing on Showup 3 such that participants who chose on the second showup were 5.88 times more likely to also choose on the third compared to participants who had rejected the second showup. In other words, 79% of those who chose on Showup 2 also chose on a target-present Showup 3, while only 39% of those who rejected Showup 2 subsequently chose on a target-present Showup 3.
Table 2.2
Experiments 1 and 2: Final Models of Logistic Regressions Predicting Choosing on Second and Third Showups Based on Previous Choosing and Target-Presence

| Experiment 1 | Showup 2 | Choosing 1 | -0.10 | 0.42 | 0.06 | .813 | 0.40 | 0.91 | 2.06 |
| Target-Presence 1 and 2 | 1.35 | 0.42 | 10.41 | < .001 | 1.70 | 3.87 | 8.79 |
| Constant | -1.17 | 0.32 | 13.14 | < .001 | 0.31 |
| Showup 3 | Choosing 2 | -0.45 | 0.60 | 0.57 | .451 | 0.20 | 0.64 | 2.07 |
| Choosing 1 | 0.23 | 0.42 | 0.32 | .573 | 0.56 | 1.26 | 2.85 |
| Target-Presence 3 | 0.36 | 0.50 | 0.53 | .469 | 0.54 | 1.44 | 3.81 |
| Choosing 2 × TP 3 | 2.23 | 0.89 | 6.31 | .012 | 1.63 | 9.27 | 52.62 |
| Constant | -0.91 | 0.40 | 5.08 | .024 | 0.40 |
| Showup 3, reversed | Choosing 2 | 1.77 | 0.65 | 7.47 | .006 | 1.65 | 5.88 | 20.96 |
| Choosing 1 | 0.23 | 0.42 | 0.32 | .573 | 0.56 | 1.26 | 2.85 |
| Target-Presence, reversed | -0.36 | 0.50 | 0.53 | .469 | 0.26 | 0.70 | 1.85 |
| Choosing 2 × TP 3 | -2.23 | 0.89 | 6.31 | .012 | 0.02 | 0.12 | 0.61 |
| Constant | -0.55 | 0.37 | 3.21 | .137 | 0.58 |
| Experiment 2 | Showup 2 | Choosing 1 | -0.54 | 0.27 | 4.04 | .044 | 0.34 | 0.58 | 0.99 |
| Target-Presence 2 | 0.87 | 0.26 | 11.22 | .001 | 1.44 | 2.39 | 3.40 |
| Constant | -0.92 | 0.21 | 18.70 | < .001 | 0.40 |
| Showup 3 | Choosing 2 | 0.40 | 0.27 | 2.27 | .132 | 0.89 | 1.50 | 2.52 |
| Choosing 1 | -0.17 | 0.26 | 0.43 | .512 | 0.50 | 0.84 | 1.41 |
| Target-Presence 3 | 1.28 | 0.26 | 24.90 | < .001 | 2.17 | 3.59 | 5.93 |
| Constant | -1.04 | 0.23 | 20.22 | < .001 | 0.35 |

Note: Variables were coded as follows. Choosing: non-choosing = 0, choosing = 1; target-presence: TA = 0, TP = 1. Experiment 1. Showup 2: $R^2 = .09$ (Cox & Snell), .13 (Nagelkerke). Model $\chi^2(2) = 11.71$, $p = .003$; Showup 3: $R^2 = .14$ (Cox & Snell), 18 (Nagelkerke). Model $\chi^2(3) = 17.50$, $p = .002$. In order to examine the target-presence by previous choosing interaction, the variable TP 3 was reverse-coded so that TA = 1, TP = 0.

*Showup 3, reversed represents the regression that was conducted using the reverse-coded target-presence variable and reported in results. Experiment 2. Showup 2: $R^2 = .05$ (Cox & Snell), .07 (Nagelkerke). Model $\chi^2(2) = 15.10$, $p = .001$; Showup 3: $R^2 = .10$ (Cox & Snell), 14 (Nagelkerke). Model $\chi^2(3) = 30.07$, $p < .001$. 

Note: Variables were coded as follows. Choosing: non-choosing = 0, choosing = 1; target-presence: TA = 0, TP = 1. Experiment 1. Showup 2: $R^2 = .09$ (Cox & Snell), .13 (Nagelkerke). Model $\chi^2(2) = 11.71$, $p = .003$; Showup 3: $R^2 = .14$ (Cox & Snell), 18 (Nagelkerke). Model $\chi^2(3) = 17.50$, $p = .002$. In order to examine the target-presence by previous choosing interaction, the variable TP 3 was reverse-coded so that TA = 1, TP = 0.
Experiment 2: Testing for sequential dependencies. Analyses for Experiment 2 were analogous to Experiment 1 with the exception that all initial models included the current target-presence by previous response interaction.

Choosing behavior on the second showup. As expected, choosing on Showup 1 was a significant predictor for choosing on Showup 2. However, current choosing contrasted previous choosing, so that participants were 1.72 times less likely to choose on the second showup if they had chosen on the previous one. In other words, 72% who chose on Showup 1 subsequently rejected Showup 2. Meanwhile 62% of participants who rejected Showup 1 went on to reject Showup 2. The lack of significant interaction for current target-presence by previous choosing indicates that this sequential dependency was not affected by the current presence of the target. However, current target-presence was also a significant predictor for choosing.

Choosing behavior on the third showup. For choosing on Showup 3, only current target-presence was a significant predictor.

Discussion

Experiments 1 and 2 were initial tests for sequential dependencies across multiple showup identification decisions in the context of multiple perpetrator crimes. We expected previous responses (choosing) and previous target-presence to be related to current decisions. While we did find some evidence for sequential dependencies in both experiments, effects were not consistent. In Experiment 1, when the current trial was target-present, participants who chose on the second showup were nearly six times more likely to also choose on the third showup compared to those who had rejected the second showup (assimilation). Although we did expect to find an interaction between current target-presence and previous choosing, the interaction operated counter to expectations that target-absent trials would be more likely to show
assimilation between current and previous choosing. By contrast, in Experiment 2, participants who chose on the first showup, were more than twice as likely to not choose on the second trial (contrast), regardless of target-presence. Taken together, results from both Experiments 1 and 2 provide inconsistent evidence for the expected sequential dependencies. This inconsistency is surprising given the theoretical overlap to fields that have robustly produced sequential dependencies, including perception, absolute identification, and, most pertinently, recognition.

In recognition tests, Malmberg and Annis (2012), found sequential dependencies between previous and current responses: A hit on a previous trial increased the probability of a hit on a current trial, but previous hits and false-alarms also increased the probability of false-alarms on a current trial. In essence, participants were more likely to choose on a current trial if they had chosen on a previous one. This effect was replicated with a variety of paired stimuli (e.g., landscape photo pairs, non-word pairs), as well as with a single-item classic recognition test. While the current research retains similarities to these basic recognition paradigms, as well as other contexts in which sequential dependencies have robustly appeared (i.e., perception, categorization tasks; Holland & Lockhead, 1968; Hsu & Yang, 2012), the eyewitness paradigm also presents differences that may explain the inconsistent results reported here.

Consequently, we considered potentially important differences that may explain the inconsistent results reported here. First, our experiments focused on the recognition of faces, rather than images, words, or non-words. While sequential dependencies have been found in categorization for facial expression (Hsu & Yang, 2012), they have not been tested specifically for face recognition. Some evidence in perceptual and recognition research gives us reason to suspect that individuals process
and remember faces differently than other non-face images (e.g., detecting minute changes in facial features, Diamond & Carey, 1986; capacity limits in face processing, Bindemann, Jenkins, & Burton, 2007). Given that sequential dependencies in recognition are thought to arise from mnemonic or perceptual processing, such differences in stimuli may be important.

Second, the number of stimuli in our experiment differs greatly from a basic recognition paradigm. In a typical recognition experiment, participants are presented with long lists of words or images, given little time to study these items, and are then tested on those items along with never-before-seen items. Conversely, our experiment only included three perpetrators to study over the course of a 2.5 min mock-crime video. Although we cannot ignore the possibility that there are simply not enough stimuli being studied, and therefore participants are not uncertain enough to rely on previous responses, the maximum average participant accuracy rates of 65% do suggest that our filler task allowed for sufficient memory decay to induce uncertainty. Meanwhile, sequential dependencies in recognition are thought to be a result of interference from previous trials that affect mnemonic processing during testing. Therefore, it seems more likely that our results reveal a difference during testing rather than a difference during encoding.

A third difference lies in the number of trials during the testing phases. While recognition experiments may have tens or hundreds of test trials, our participants encountered only three. Perhaps this is not a sufficient number of trials for sequential dependencies to arise. Sequential dependencies have been explained through accumulator models, which predict shifts over time based on criterion placement or accumulation starting points (e.g., selective attention, mapping and ballistic accumulation; SAMBA; Brown et al., 2008; Matthews & Stewart, 2009). The
SAMBA model, for example, posits that a participant classifying the loudness of a sound (i.e., *soft* vs. *loud*; Jones, Love, & Maddox, 2006) uses the sound on initial trials to generate a range between which the subsequent sounds are expected to fall. This range establishes how soft the participants can expect a *soft* sound to be and how loud they can expect a *loud* sound to be. When confronted with the task of classifying the sound on the current trial, the observer will compare the sound to the upper and lower range in relation to the loudness of the previous response. Their response will depend upon the strength of the evidence for each of these answers. When a *soft* response is given on the current trial, it is hypothesized that this biases the perception of the sound on the subsequent trial by temporarily reducing the strength of evidence needed to favor another *soft* response. Thus, assimilation arises from the decisional processes: because the *soft* response now has the advantage, the following trial is more likely to reach the threshold to be classified as *soft*. Contrast, however, arises from the perceptual processes: because observers are comparing the current sound to the previous one, any change louder or softer can lead to over- and underestimation of strength of that sound. In this model, assimilation and contrast both appear because the stronger effect (assimilation) eventually decays to give way to the weaker one (contrast; Brown et al., 2008). It is possible that such models require an adjustment period over multiple trials in order to calibrate the upper and lower range of perceptual (and in the case of recognition, mnemonic) processing. As a result, the small number of trials present in our experiment might be insufficient for sequential dependencies to arise.

To address the issues outlined above, Experiment 3 used the recognition paradigm in an attempt to replicate and extend the work of Malmberg and Annis (2012, near-pairs condition) using three different categories of stimuli: photos of
faces, photos of landscapes (places) and words. Accordingly, these concerns were translated into three main goals: (1) to extend previous research by testing for sequential dependencies on overall responding in face recognition, (2) if found, to determine if these sequential dependencies translate to predictable choosing behavior, and (3) to examine whether the strength of these effects vary across the testing phase.

We predicted sequential effects would arise across all three sets of stimuli. If sequential dependencies were observed for responses overall, we predicted that sequential effects would be stronger in the second half compared to the first half of testing blocks and thus also expect to be able to predict choosing behavior in late, but not early, test trials. Should sequential dependencies fail to appear for faces (cf. places and words), it would not only undermine response-interference as a theory for the source of the sequential effects in recognition, but also suggest that sequential dependencies are unlikely to be problematic in the eyewitness context.

**Experiment 3**

**Participants and design**

One-hundred-fifty participants were recruited from online participation platforms. Five participants were excluded for the following reasons: failing two of the four control questions (1), failing to follow instructions (2), and taking a 20+ min break in the middle of the first testing block (2). Participants with other anomalous data (e.g., low activity during the filler task) were flagged; When exploratory analyses to examine hit rates, false-alarm rates, accuracy, and choosing behavior did not reveal any of these flagged participants to be outliers, their data were retained for all further analyses. The average age of the remaining 145 participants was 22 years ($M = 22.14$, $SD = 6.49$).
Participants were randomly assigned to one of three conditions to study paired stimuli of faces, places, or words. For each study-test block, participants viewed 18 paired target stimuli during the learning phase and were tested on the 36 target stimuli and 36 fillers. Each participant took part in two study-test blocks. Participants were compensated with research participation credit if eligible, or otherwise not compensated.

**Method**

**Materials.** For each stimulus type (faces, places, words), a total of 288 individual stimuli were selected. Of those, 144 were used as target stimuli, and the other 144 were used as lures during test. See Figure 2.1 for example stimuli pairs.

**Faces.** Participants viewed paired male and female faces with neutral expressions that were selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). Faces that were particularly distinctive, (i.e., shaved eyebrow, facial piercing, unique hair) were removed. Half of the target face stimuli were presented during the learning phase as same-gender pairs and half as opposite-gender pairs.

**Landscape photos.** Photographs of varied landscapes (e.g., sunsets, mountains, deserts, fields) were selected from the Places Scene Recognition Database (Zhou, Lapedriza, Xiao, Torralba, & Oliva, 2014). Photos with particularly distinctive features (e.g., color filter) were not selected.

**Words.** One thousand nouns were randomly chosen from the 5,000 most frequently used words according to the Corpus of Contemporary American English (Davies, 2008). Words were piloted for recognition by seven non-native English speakers whose nationalities are representative of the student population from which the sample is drawn (two Germans, two Belgians, and three Dutch). These non-native
English speakers were asked to view the list of 890 nouns and remove those words that they did not recognize (i.e., would need to search for or translate). The stimulus pairs were randomly selected and paired from the remaining 813 nouns.

**Procedure.** The procedure followed the procedure of Malmberg and Annis (2012; Experiment 1, near-pairs condition replications), with two exceptions. First, because the study was distributed online, a shape appeared at the end of each encoding block (block 1: blue star, block 2: black arrow), which were later used as control questions for attention. Second, due to availability of faces, participants studied only 36 total pairs (cf. 40 pairs in the original experiment) of the varying stimuli (faces, places, or words).

Participants were provided a link for the computer task. Participants in the face condition, for example, studied 18 paired faces. Each pair was presented on screen for 2 s with a 0.1 s interstimulus interval. Following a 30 s distractor task (Pac-Man), participants were presented with two control questions asking them to indicate the shape and color of the shape presented at the end of the encoding phase. Participants were then tested for their recognition of the previously-studied faces using the self-paced computer task. Participants in all conditions saw 36 target trials plus 36 filler trials of never-before-seen stimuli presented at random, with the constraint that half of the pairs were tested consecutively and the other half were randomized into positions at least seven trials away from their corresponding target trials. Following another 1 min distractor task, this procedure was repeated for a second study-test block. At the end of the experiment, a final question prompted participants to describe the environment in which they completed the experiment (i.e., time of day, location, presence of others).
Results

We focused on two types of analyses to address the three goals of the experiment. First, we conducted within-subjects tests on overall response patterns to replicate and extend those analyses conducted by Malmberg and Annis (2012). Accordingly, we conducted mixed Analyses of Variance (ANOVAs) on conditional hit rates, false-alarm rates, and choosing rates given previous responses and stimulus type (faces, places, words). Because we were interested in how this effect might vary across the testing sessions, we conducted these same analyses on the conditional hit rates and false-alarm rates for the first half and second half of each of the two testing blocks. We refer to the first half of Block 1 as Section 1, the second half as Section 2, and the first and second halves of Block 2 as Sections 3 and 4, respectively.

Second, we conducted between-subjects analyses in order to determine whether overall patterns would be reflected in individual choosing behavior. More specifically, we conducted logistic regressions analogous to those conducted in Experiments 1 and 2 to test whether we could predict choosing behavior on individual trials using target-presence and previous choosing as predictors. Although we present...
the results descriptively here, relevant statistics for within-subjects analyses can be found in Tables 2.3 and 2.4. Statistics for between-subjects analyses can be found in Table 2.5.

**Preparation of data and calculation of contingency rates.** Prior to calculating hit rates, false-alarm rates, and choosing rates, trials with response times faster than 200 ms were removed. This is because 200 ms is the conservative threshold for recorded brain activity in response to human faces, as well as the earliest threshold for our ability to distinguish between familiar and unfamiliar faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Caharel, Ramon & Rossion, 2014). Hit rates were calculated as the proportion of correct answers on target-present trials and false-alarm rates were calculated as the proportion of incorrect answers on target-absent trials. Analyses used hit rates on the current trial \(i\) given that the previous trial \(i-1\) was a hit, miss, false-alarm or correct rejection. Therefore, separate hit rate contingencies were computed for each participant for (a) hits that followed a hit, \(HR_{hit} = (H | i-1 = hit)\), (b) hits that followed a miss, \(HR_{miss} = (HR | i-1 = miss)\), (c) hits that followed a false-alarm and, (d) hits that followed a correct rejection. Analogous false-alarm rates for each participant were computed given that the previous response was a hit, miss, false-alarm, or correct rejection (Malmberg & Annis, 2012). Choosing rates were calculated as overall proportion of choosing (respond yes vs. no) on target-present and target-absent trials.

**Sequential effects as a function of stimulus type.**

**Hit rate contingencies.** We examined whether a hit on the current trial \(i\) was more or less likely given a hit, miss, false-alarm, or correct rejection on the previous trial \(i-1\), and whether this relationship differed for our three types of stimuli: faces, places, and words. Thus we conducted a mixed ANOVA with previous response
being the within-subjects factor, and type of stimulus being the between-subjects factor. There was a significant main effect of previous response. Planned contrasts indicated that a hit on the current trial was more likely if there was either a hit or false alarm compared to a correct rejection or miss on the previous trial. A hit on the current trial was also more likely if there was a correct rejection compared to a miss on the previous trial. The interaction of previous trial by type of stimulus was not significant. Thus, while we found sequential effects for hit rates, these effects did not differ significantly based on whether the stimuli were faces, places, or words. See Figure 2.2 (Panel A).

**False-alarm rate contingencies.** We next examined whether a false-alarm on the current trial \( (i) \) was more or less likely given a hit, miss, false-alarm, or correct rejection on the previous trial \( (i-1) \), and whether this relationship differed for our three sets of stimuli: faces, places, and words. We conducted a mixed ANOVA with previous response being the within-subjects factor, and stimulus type being the between-subjects factor. There was a significant main effect of previous response, indicating that a false-alarm on the current trial was more likely following a hit or false-alarm (cf. miss or correct rejection) on the previous trial (see Figure 2.2, Panel B). There was no significant difference between hits vs. false-alarms or misses vs. correct rejections. The non-significant interaction of previous trial with stimulus category provided no evidence that these effects differed significantly according to stimulus type.

**Sequential effects as a function of section.** There were no significant interactions for test section (Sections 1, 2, 3, and 4) with previous response for hits, misses, false-alarms, or correct rejections. There was a main effect of test section,
Table 2.3
Experiment 3: Results for ANOVAs on Current Hit Rates, False-Alarm Rates, and Choosing Rates Given Previous Responses and Condition

<table>
<thead>
<tr>
<th>Stimulus Type: Faces, Places, Words</th>
<th>df</th>
<th>F</th>
<th>η²</th>
<th>t</th>
<th>d</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit Rate: Hit, Miss, FA, CR</td>
<td>2.25, 321.58</td>
<td>42.57</td>
<td>.229</td>
<td>&lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous Response</td>
<td>2.143</td>
<td>9.00</td>
<td>.112</td>
<td>&lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulus Type</td>
<td>4.50, 321.58</td>
<td>0.70</td>
<td>.010</td>
<td>.611</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>2.47, 355.40</td>
<td>26.59</td>
<td>.156</td>
<td>&lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA Rate: Hit, Miss, FA, CR</td>
<td>2.144</td>
<td>9.67</td>
<td>.118</td>
<td>&lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous Response</td>
<td>4.94, 355.40</td>
<td>0.67</td>
<td>.009</td>
<td>.646</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulus Category</td>
<td>2.73, 321.58</td>
<td>4.85</td>
<td>.042</td>
<td>.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>2.72, 304.12</td>
<td>9.06</td>
<td>.075</td>
<td>&lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing: Choose vs. Not</td>
<td>7.77, 870</td>
<td>0.98</td>
<td>.009</td>
<td>.447</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The top panel displays results for mixed ANOVAs on hit rates, false-alarm rates, and choosing rates with previous response as the within-subjects factor and stimulus type (faces, places, and words) as the between-subjects condition. False-alarm and correct-rejection are abbreviated here as FA and CR, respectively. The interaction between previous response and stimulus category was significant only in analysis of choosing rates. Although sequential dependencies arose within all stimulus categories, the effect was greatest for places, followed by words, and then faces.

Table 2.4
Experiment 3: Results for ANOVAs on Current Hit Rates, False-Alarm Rates, and Choosing Rates Given Previous Responses and Testing Section

<table>
<thead>
<tr>
<th>Test sections: 1, 2, 3, and 4</th>
<th>df</th>
<th>F</th>
<th>η²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit Rate Contingencies</td>
<td>3, 318</td>
<td>41.22</td>
<td>.280</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Previous Response</td>
<td>2.44, 259.02</td>
<td>8.63</td>
<td>.075</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction</td>
<td>7.26, 770.03</td>
<td>0.47</td>
<td>.004</td>
<td>.860</td>
</tr>
<tr>
<td>False-Alarm Rate Contingencies</td>
<td>2.73, 305.63</td>
<td>4.85</td>
<td>.042</td>
<td>.004</td>
</tr>
<tr>
<td>Previous Response</td>
<td>2.72, 304.12</td>
<td>9.06</td>
<td>.075</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Interaction</td>
<td>7.77, 870</td>
<td>0.98</td>
<td>.009</td>
<td>.447</td>
</tr>
</tbody>
</table>

Note. The top panel displays results for repeated-measures ANOVAs on hit rates, false-alarm rates, and choosing rates with previous response (hit, miss, false-alarm, correct rejection) and test section (1, 2, 3, 4) as the between-subjects factors. Sections are the first half of the first study-test block (Section 1), the second half of the first block (Section 2), and the first and second halves of the second block (Sections 3 and 4).
Figure 2.2 Experiment 3: Hit rate and false-alarm rate contingencies. Panel A displays the probability of a hit on the current trial given the previous response (hit, miss, false-alarm, or correct rejection), collapsed across stimulus type (faces, places, words). Previous responses of hit and false-alarm do not significantly differ from each other, but all other comparisons are significant ($p < .001$). Panel B displays the probability of a false-alarm on the current trial given the previous response, collapse across stimulus type. A false-alarm on the current trial is significantly more likely given a previous hit or false-alarm when compared to a previous miss or correct rejections ($p < .001$). Error bars are with standard error.

such that Sections 1 and 3 displayed higher hit rates and lower false-alarm rates than Sections 2 and 4. Section 1 also displayed higher false-alarm rates than Section 3.

**Sequential effects in choosing.** In these analyses, we ask a similar question in a different manner: overall, is choosing (saying yes) on the current trial, more or less likely if you chose or did not choose (said no) on the previous trial? We conducted mixed ANOVAs with previous choosing (choose vs. not choose) as the within-subjects factor and stimulus type as the between-subjects factor. There was a significant main effect of previous choosing, and a significant interaction between previous choosing and stimulus type. Together, these results indicate that choosing on the current trial was more likely if the participant chose (cf. did not choose) on the previous trial and that this effect was weakest for face stimuli.
Predicting choosing on individual trials. Given that we successfully replicated analyses demonstrating sequential dependencies in overall recognition memory, including for faces, we subsequently tested whether those effects would translate to predictable behavior on individual trials over the course of the testing sessions. Therefore, we chose the first three trials of each testing block and the three middle trials of each block (Block 1: trials 1-3 and 71-73; Block 2: trials 1-3 and 71-73). These analyses are of particular interest because they apply analyses from Experiments 1 and 2 to a dataset in which sequential dependencies have already been detected. We consider first three trials of the first block a proxy for the three showup

Table 2.5
Experiment 3: Results of Logistic Regression Predicting Choosing on Second and Third Recognition Test Trials Based on Previous Choosing and Target-Presence

<table>
<thead>
<tr>
<th>Section 1, trial 3</th>
<th>b</th>
<th>SE</th>
<th>Wald</th>
<th>p</th>
<th>95% CI for Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Odds</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing 1</td>
<td>0.57</td>
<td>0.37</td>
<td>2.37</td>
<td>.124</td>
<td>0.86 1.76 3.62</td>
</tr>
<tr>
<td>Choosing 2</td>
<td>0.41</td>
<td>0.36</td>
<td>1.28</td>
<td>.258</td>
<td>0.74 1.50 3.03</td>
</tr>
<tr>
<td>TP 3</td>
<td>1.62</td>
<td>0.38</td>
<td>18.13</td>
<td>&lt; .001</td>
<td>2.39 5.03 10.58</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.97</td>
<td>0.36</td>
<td>7.41</td>
<td>.006</td>
<td>0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 2, trial 73</th>
<th>b</th>
<th>SE</th>
<th>Wald</th>
<th>p</th>
<th>95% CI for Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Odds</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing 71</td>
<td>0.18</td>
<td>0.38</td>
<td>0.22</td>
<td>.640</td>
<td>0.57 1.20 2.52</td>
</tr>
<tr>
<td>Choosing 72</td>
<td>0.79</td>
<td>0.38</td>
<td>4.45</td>
<td>.035</td>
<td>1.06 2.21 4.62</td>
</tr>
<tr>
<td>TP 73</td>
<td>0.98</td>
<td>0.38</td>
<td>6.68</td>
<td>.010</td>
<td>1.27 2.67 5.61</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.58</td>
<td>0.38</td>
<td>17.04</td>
<td>&lt; .001</td>
<td>0.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 3, trial 3</th>
<th>b</th>
<th>SE</th>
<th>Wald</th>
<th>p</th>
<th>95% CI for Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Odds</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing 1</td>
<td>0.42</td>
<td>0.38</td>
<td>1.22</td>
<td>.269</td>
<td>0.72 1.52 3.20</td>
</tr>
<tr>
<td>Choosing 2</td>
<td>-0.32</td>
<td>0.40</td>
<td>0.63</td>
<td>.428</td>
<td>0.33 0.73 1.60</td>
</tr>
<tr>
<td>TP 3</td>
<td>-0.70</td>
<td>0.36</td>
<td>3.69</td>
<td>.055</td>
<td>0.24 0.50 1.01</td>
</tr>
<tr>
<td>Constant</td>
<td>0.98</td>
<td>0.45</td>
<td>4.71</td>
<td>.030</td>
<td>2.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 4, trial 73</th>
<th>b</th>
<th>SE</th>
<th>Wald</th>
<th>p</th>
<th>95% CI for Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Odds</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing 71</td>
<td>-0.55</td>
<td>0.38</td>
<td>2.12</td>
<td>.145</td>
<td>0.28 0.58 1.21</td>
</tr>
<tr>
<td>Choosing 72</td>
<td>0.48</td>
<td>0.37</td>
<td>1.70</td>
<td>.193</td>
<td>0.79 1.61 3.31</td>
</tr>
<tr>
<td>TP 73</td>
<td>-0.66</td>
<td>0.36</td>
<td>3.43</td>
<td>.064</td>
<td>0.26 0.52 1.04</td>
</tr>
<tr>
<td>Constant</td>
<td>0.83</td>
<td>0.39</td>
<td>4.54</td>
<td>.033</td>
<td>2.30</td>
</tr>
</tbody>
</table>

Note: Variables were coded as follows. Choosing: non-choosing = 0, choosing = 1; target-presence: TA = 0, TP = 1. Section 1, Trial 3: $R^2 = .14$ (Cox & Snell) .19 (Nagelkerke). Model $\chi^2(3) = 22.50, p < .001$; Section 2, Trial 73: $R^2 = .07$ (Cox & Snell) .10 (Nagelkerke). Model $\chi^2(3) = 11.33, p = .010$. Section 3, Trial 3: $R^2 = .03$ (Cox & Snell) .05 (Nagelkerke). Model $\chi^2(3) = 4.98, p = .173$; Section 4, Trial 73: $R^2 = .05$ (Cox & Snell) .07 (Nagelkerke). Model $\chi^2(3) = 11.33, p = .050$. 


identification decisions in Experiments 1 and 2. We chose to test the middle trials rather than later trials in order to avoid isolating groups of responses likely to display fatigue effects. Analyses were analogous to Experiment 2 with one exception. Given that there was no current target-presence by previous choosing interaction in Experiment 2, this interaction was not included.

Choosing on the previous trial predicted choosing on the current trial for only one of the four analyses, and current trial target-presence predicted choosing in only trials of Block 1. Thus, despite finding that, in general, hits and false-alarms were more common when participants chose on the previous trial, behavior on previous trials was not a useful predictor of choosing for these sets of individual trials.

**Discussion**

Experiment 3 sought to replicate and extend previous work in sequential dependencies in recognition memory (Malmberg & Annis, 2012). Except for online data collection and the inclusion of additional questions to control for attention, the procedure followed the near-pairs condition in Malmberg and Annis’ (2012) Experiment 1. We expected that sequential dependencies would arise for recognition responses for all three types of stimuli, that these effects would be stronger in later portions of testing, and that this would be reflected in the capacity to predict current choosing from previous choosing in later, but not earlier, test trials.

As expected, the probability of a hit in the current trial \((i)\) was higher if the previous response \((i – 1)\) was also a hit compared to if the previous response was a miss. The probability of a false-alarm on the current trial was increased if it was preceded by either a hit or a false-alarm (compared to a miss or correct rejection). Noticeably, this pattern of results did not differ across category of stimuli, meaning we replicated Malmberg and Annis’ results using pictures of places and words, and
extended those results to include face recognition. Taken together, these results demonstrate that the probability of saying yes on the current trial increases any time it is preceded by a yes on the previous trial, a conclusion reflected in the analyses conducted on choosing behavior.

We also conducted analyses to determine whether the relationship of previous response reported above changed over the course of the testing session. Contrary to predictions, the effect of previous response did not vary as a function of test section. Although accuracy displayed fatigue effects across the sections (higher accuracy in the first half of each testing block compared to the second half), the strength of sequential dependencies remained constant throughout. Essentially, sequential dependencies did not change across the length of testing.

Lastly, we tested whether these overall effects of sequential dependencies would translate to predictable behavioral outcomes on specific trials. We found little support for the idea that choosing on a previous trial predicted choosing on the current trial. Rather, while sequential dependencies did arise in overall choosing behavior across the total 288 trials, and even the 72 trials comprising each half of the testing blocks, these effects did not reliably arise as predictable choosing behavior on individual trials.

**General Discussion**

This line of research aimed to answer the question: in making a series of ostensibly independent showup identification decisions for different perpetrators, is the current decision of an eyewitness related to the previous one(s)? In Experiments 1 and 2, we addressed this question within the eyewitness identification paradigm. Participants watched a mock-crime video with three perpetrators and were subsequently asked to make three showup identification decisions, one suspect for each of the perpetrators.
Although we found some evidence for sequential dependencies in both experiments, the effect overall was not consistent. These unexpected results led us to question whether methodological differences between the recognition and eyewitness paradigm could explain the inconsistencies. In particular, we considered whether sequential dependencies would arise for recognition specifically for faces, and whether the number of trials tested influenced the ability to identify these dependencies. Thus, Experiment 3 replicated and extended Malmberg and Annis’ (2012) research for sequential dependencies in recognition decisions to test whether (1) sequential dependencies would also arise for face recognition, (2) these effects could predict choosing behavior on individual trials, and (3) the strength of the above effects varied across the testing session. This approach allowed us to conduct both within-subjects testing to replicate previously reported effects in recognition memory, and the between-subjects modeling applied in Experiments 1 and 2. Experiment 3 showed that sequential dependencies do arise overall for face recognition decisions, that the strength of these effects remains consistent across the testing session, but that these effects do not reliably predict choosing behavior for individual trials. These results and their implications for theory and practice are discussed in turn.

**Sequential dependencies arise for face recognition decisions**

In Experiment 3, we successfully replicated previous research, demonstrating that when participants make a series of yes/no recognition decisions, their responses are affected by the previous trial. A hit on the current trial was more likely given a hit or false-alarm (vs. miss or correct rejection) on the previous trial, and a false-alarm on the current trial was more likely given either a hit or false-alarm (vs. miss or correct rejection) on the previous trial. To confirm this, analyses on choosing behavior established that choosing begets choosing: if participants said *yes* (vs. *no*) on the
previous trial, the probability of saying *yes* on the current trial is increased. Notably, these effects were found for three types of stimuli, including images of faces. Indeed, in our analyses with hit rate and false-alarm rate contingencies, while the overall contingency rates varied depending upon the stimulus type, the relationship between previous and current response did not. Thus, this experiment adds to a growing list of decisions in which sequential dependencies arise, including detection of sound (Jones, et al., 2006), ratings of sweetness in wine taste-tests (Schifferstein & Frijters, 1992), and judgements of frequency in landscape recognition (Annis & Malmberg, 2013).

Experiment 3 suggests that we can rule out the possibility that the inconsistent effects within Experiment 1 and 2’s eyewitness identification paradigm reflect the use of face stimuli rather than the words or landscapes used in previous research. By extension, this means that sequential dependencies could conceivably arise for someone making a series of yes/no decisions in person recognition settings, such as security personnel looking for banned football fans in a stadium or scanning the crowd for known-threats at political events. Indeed, this may be a useful setting in which to study sequential dependencies in applied recognition memory, and to consider in training security agents.

**Effects are consistent across testing**

Next, we tested whether the strength of sequential dependencies varied across the length of the testing session. Accumulator models used to explain sequential dependencies predict shifts over time based on variation in criterion placement or accumulation starting points (e.g., SAMBA; Brown et al., 2008; Matthews & Stewart, 2009). We hypothesized that effects would be stronger in the second half of each testing session compared to the first half of each session. Indeed, Schifferstein and Kuiper (1997), go so far as to remove the first 20 “outlier” responses of their
experiment tasting aqueous solutions because high response variability is greatest in these initial trials. Contrary to expectations, the strength of sequential dependencies remained constant across the length of the testing session.

In sum, our results established that sequential dependencies arise consistently within participants separately from individual differences in criteria. We could therefore be certain that our results replicated previous experiments on sequential dependencies as we transitioned to apply the regression models used in Experiments 1 and 2.

**Sequential dependencies are not reflected as predictable choosing behavior**

We next tested whether these dependencies would also predict behavioral outcomes on the first three and middle three trials of each testing block. The first three trials of the first block are of greatest interest because they best represent the three showup identification trials in Experiments 1 and 2. Consistent with Experiments 1 and 2, and despite detecting sequential dependencies in overall data, we were not able to detect sequential dependencies in individual trials. Given that the strength of sequential dependencies detected by within-subjects analyses did not vary across the testing session, it was not subsequently surprising to find that detecting sequential dependencies on individual trials did not differ. Critically, these results appear to be good news for the eyewitness context. We were originally concerned that multiple identification decisions may give rise to sequential dependencies, and thus affect the integrity of the identification decisions being made. However, this is not the case. If we cannot predict current recognition decisions from previous ones, then there is less reason to believe that dependencies are likely to be problematic for the multiple high-stakes recognition decisions in the eyewitness identification context.
This is not to say that a series of identification decisions cannot possibly be related to each other. Indeed, sequential dependencies are only one way in which the assumption of independence may be violated between multiple decisions. Research demonstrates that confidence leaks across tasks, so that confidence in one’s performance of a previous task is carried forward into a different task, even if it is not cognitively related to the previous task (Rahnev, Koizumi, McCurdy, D’Esposito, & Lau, 2015). Confidence ratings were outside the scope of the current research, but given the well-documented concerns of post-identification feedback effect (Bradfield & Wells, 1989), it is important to investigate how confidence ratings might carry forward in multiple identification decisions. In other words, though we have ruled out one possibility on the relationship between multiple identification decisions, there is more to be investigated.

Conclusion

Neither the use of faces nor differences in the number of trials could explain the contradictory results in Experiments 1 and 2 that we sought to resolve. However, the inability to use previous choosing behavior as a predictor for current choosing in Experiment 3, a data set that we know contains sequential dependencies, still serves to clarify our previous contradictions. We suspect that the discrepancy between detecting sequential dependencies in overall responses using within-subjects analyses and not on individual responses with regression models is an indication of weak effects. The within-subjects ANOVAs provide the statistical power to detect small effects, while the regression models do not. In this case, probabilities of choosing on current trials are heightened by previous choosing over hundreds of opportunities to choose or not to choose, but these effects do not translate to detectable behavioral outcomes of choosing on individual trials. In each of three experiments, it was
sometimes possible to predict current choosing from previous choosing, but not reliably so, and often not in the expected direction.

In summary, sequential dependencies arise in face recognition, and though the accuracy across stimuli and section of testing session may vary, the basic relationship does not change. However, these effects did not translate to individual trials, and we therefore suggest that the integrity of identification and recognition decisions is not likely to be impacted by making the multiple decisions in a row. This is the first paper to systematically explore sequential dependencies in face recognition and particularly in eyewitness identification, contributing to the small, but vital, group of literature that aims to disentangle factors underlying the decreased performance in recognition for multiple faces. It thus contributes towards the eventual goal to offer procedural recommendations adapted to the difficulties present in the administration of identification procedures in the context of multiple perpetrator crimes.
Chapter 3: Face value: Testing the utility of contextual face cues for face recognition

Abstract

The presence of multiple faces at encoding increases memorial demand, but may also provide a naturally-occurring contextual cue to support recognition for faces at test. Across two experiments, we sought to replicate previously-reported enhancing effects in cued face recognition, to investigate mechanisms underlying those effects, and to determine whether such effects extended to encoding conditions involving more than two faces. In Experiment 4, participants studied sets of individual faces, pairs of faces, or groups of four faces. At test, participants in the single-face condition were tested only on those individual faces without cues. Participants in the two and four-face conditions were tested using no cues, correct cues (a face previously studied with the target test face), or incorrect cues (a never-before-seen face). In Experiment 5, participants additionally completed a rating task to promote associative encoding. Hit rates were not affected by either cue type or face encoding condition, but cuing of any kind (correct or incorrect) appeared to provide a protective buffer to reduce false-alarm rates in the two- and four-face conditions through increased sensitivity, but mostly reduced response bias. Our findings provide some evidence that cued recognition techniques could be useful to reduce false recognition, but the inconsistency also warrants further exploration to fully understand underlying mechanisms.

Experiments 4 and 5 are presented together in Chapter 3 because these experiments are being prepared for publication together.
Introduction

The recognition and identification of an unfamiliar face is a difficult task. This task becomes even more challenging as the number of unfamiliar faces to be remembered increases. Indeed, accuracy rates for face recognition and identification drop when there are multiple faces to study, even when participants are only tested on one of those faces (e.g., Clifford & Hollin, 1981). Although divided attention at encoding may play a role in this effect (e.g., Bindemann, Jenkins & Burton, 2007), this does not entirely account for the multiple-face recognition disadvantage. Accuracy decreases even when participants have unlimited time to encode the presented faces (e.g., they are instructed to move on to the identification only once they feel confident they will remember all of the to-be-encoded faces) and when divided attention is controlled for (Bindemann, Sanford, Gillatt, Avetisyan, & Megreya, 2012). The persistence of the multiple-face disadvantage suggests that the drop in accuracy is caused, at least in part, by an increase in memorial demand when attempting to hold two faces in memory for any period of time (Bindemann et al., 2012).

Interestingly, the factor that makes multiple-face recognition so challenging may actually provide a solution to support memory for multiple faces. The presence of multiple faces increases memorial demand, but also provides a naturally-occurring contextual cue that may promote recognition. Early experiments demonstrated the benefits of cues on face recognition when faces were studied as pairs (e.g., Watkins, Ho, & Tulving, 1976); When participants studied pairs of faces, and were later tested on only one of those faces, their accuracy in recognizing that target faces was improved if it was also presented with the other face of the pair. More recent experiments have attempted to investigate the benefits of cuing in the context of

---

7 This term is adapted from Bindemann et al.’s (2012) terminology in which they label a two-face disadvantage due to the fact that they tested one vs. two faces.
eyewitness identification for multiple perpetrator crimes with underwhelming results (e.g., Wells & Pozzulo, 2006). However, there is a paucity of contemporary research systematically examining cued face recognition effects in applied contexts. We sought to replicate previously-reported enhancing effects in cued face recognition, to investigate the mechanisms underlying those effects, and to determine whether such effects could extend to include more than two faces, as many crimes involve more than two perpetrators. To this end, we compared the traditional cued encoding condition in which participants study paired faces, with two additional encoding conditions: a control condition where participants encoded single faces and another experimental condition where participants encoded groups of four faces.

**Cued recognition and faces**

Memory researchers have long known that context matters for retrieval. Contextual cues are often implemented to help individuals recall seemingly-forgotten details in episodic memory (encoding specificity principle; Thomson & Tulving, 1970), including to facilitate eyewitness recall during investigative interviews (context reinstatement; Geiselman, Fisher, MacKinnon, & Hannon, 1986). It has been hypothesized that these cues work because the retrieval of a memory is dependent upon the way it was stored, and an item in episodic memory is, by nature, nested within our experience of the relevant event (Tulving & Thomson, 1973; Polyn, Norman, & Kahana, 2009). Thus, episodic memory is not only tied to temporal markers (when an event occurred), but can also be integrated with memory traces for the context of the event. According to the Context Maintenance Retrieval Model, we naturally recall items in semantic clusters (i.e., related words like book, hardcover, paperback, bestseller) as a result of our long-standing associations between items in our experience, but we also have a tendency to recall items in temporal and source
clusters as a result of the associations we form during study phase between items (Polyn et al., 2009). A contextual cue at retrieval takes advantage of the associative nature of memory whereby, for example, peripheral details of the environment can cue additional pathways for the retrieval of critical details.

Thomson and Tulving (1970) experimentally demonstrated the utility of context for memory retrieval when they had participants memorize pairs of words: a target word and a word that was semantically unrelated (e.g., ocean and piano). During the testing phase, the previously-studied, semantically-unrelated pair word (weak cue) or a semantically-related, but previously-unstudied word (strong cue) was presented in order to elicit the target word. If the target word is “ocean”, a strong cue might be “land”, while the weak cue might be “piano”. Results showed that participants were better able to recall the target word if it was presented along with the weak cue (i.e., the cue with which it was originally encoded), compared to the strong cue, suggesting that the context of encoding the target word mattered more than the strength of the semantic association.

Contextual cues also benefit recognition memory, including when recognizing pictures (Palmer, 1975) and faces (e.g., Watkins et al., 1976). In the field of face recognition, a variety of external contexts have been shown to enhance recognition for target faces, including backgrounds, descriptions, clothing, and other faces (see Davies, 1988 for a review). In the latter category, Winograd and Rivers-Bulkeley (1977) asked participants to memorize pairs of faces during the study-phase—one of which served as the target-face while the other served as the cue. During the test-phase, participants were presented with a target face alongside either a face they had previously studied (correct cue), a face that had not been previously studied (incorrect cue), or no cue at all. In this forced-choice paradigm, recognition
performance for target faces was enhanced by the presentation of correct cue faces and impaired by incorrect cue faces, while performance with no cues fell in-between. Similarly, Watkins and colleagues (1976) demonstrated reduced hit rates for face recognition when the incorrect cues were previously-studied faces that had been paired with another face during study. In other words, simply swapping context, as opposed to introducing new context, also affected face recognition performance.

However, although the above experiments found an enhancing effect of using correct cues, other research has not (e.g. Bower & Karlin, 1974) and while Kan, Giovanello, Schnyer, Makris, and Verfaellie (2007) found that incorrect cuing undermined hit rates, they found no benefit for correct cues (cf. no cues). A number of issues relating to encoding and recognition may underpin these discrepant results. First, the context must be strongly encoded in association with the target for recognition. If the participant did not pay attention to the contextual information, or did not link it with the target information, then presenting the cue at test will not improve recognition (Peris, 1985, as cited in Davies, 1988). Second, it appears that context is useful as a recognition memory cue only when other, stronger cues are lacking (Davies, 1988). Some theories of recognition hold that recognition is comprised of two mechanisms (e.g., Mandler, 1980; Peris & Tiberghien, 1984). The first is the perceptual system that is automatically activated and produces fast answers that hinge on the feeling of familiarity. The second is the cognitive system, which is activated when the first does not immediately provide an answer. The second system is slower and searches for external information, like context, to aid the response. This process is also captured in the rationale of the outshining hypothesis, which contends that we use the most relevant cues available to recognize faces. When our memory trace is strong, that memory outshines the utility of environmental context (Smith &
Vela, 2001). However, for weak memory traces, such as when there was suboptimal encoding or longer retention intervals, context may support memory to improve recognition performance (Mandler, Pearlstone, & Koopmans, 1969).

Cued recognition presents an interesting means of enhancing face recognition, a concept that may prove useful to the applied fields of eyewitness identification or wanted-persons recognition. The second face in a two-perpetrator crime provides a naturally-occurring contextual cue for the eyewitness, and one that is particularly relevant for humans. Given our natural tendency to orient attention towards other human faces, a second face may increase the chance of incidental associative encoding (Di Giorgio, Turati, Altoè, & Simion, 2012; Langton, Law, Burton, & Schweinberger, 2008). Thus, face cues may support eyewitnesses of a multiple perpetrator crime while viewing a suspect lineup. This is not an entirely novel idea; At least two published experiments (Hobson & Wilcock, 2011; Wells & Pozzulo, 2006) and one unpublished dissertation (Dempsey, 2012) have attempted to apply face cuing to support eyewitness memory and adapt identification procedures in the context of multiple-perpetrator crimes. However, none of these attempts provided convincing evidence that this method of cuing memory could aid lineup identification decisions, and it is of interest to understand why this might be. The current research aims to understand the disconnect between a theoretically intriguing mnemonic device and the poorly-understood difficulties in applying such a device; particularly by replicating previous work, extending it to explore boundary conditions, and orienting the effects. These goals are explored further below.

Previous experiments showing inconsistent effects of face cues also expose concerns for the replicability of a cued face recognition effect. Relevant experiments were mostly conducted in the 1970’s, and it is only decades later that we are now
interested in applying this research to an entirely new context in which the studies were not designed to apply (i.e., eyewitness identification). This means that the original studies may have made decisions that were incongruent with the eyewitness context, or not in line with our contemporary methodological standards. For one, the relevant research report methodologies with confounding variables that would now be controlled for, which raises issues for replicability. For example, Winograd and Rivers-Bulkeley (1977) specifically paired male-female pairs according to overt compatibility, with particular care to maintain similar ages between them. Further, participants rated perceived compatibility of the couples, further enhancing the romantic link between the pairs. It is possible that this likeness between the pairings provided contextual information for correct and incorrect cuing. A participant, for example, might correctly recognize a target face of a genial, 20-year-old white male paired with a genial, 20-year-old white female because the pairing enhances recognition. It might also be because the nature of the pairing gives the participant a hint as to the correct answer given prior knowledge that the two faces presented are an intuitively compatible couple. By extension, a participant might be worse at recognizing that same male when paired with an older, more-serious looking female because the participant knows they are not a compatible couple. Likewise, previous experiments using face pairs did not randomize the left-right orientation of those pairs such that if the target face was on the right side of the screen during the encoding phase, it was also presented on the right side during test (Watkins Ho & Tulving, 1976; Winograd & Rivers-Bulkeley, 1977). It may be that the boost in hit rates is sensitive to this spatial context as well as contextual cuing. The recent movement in psychological sciences to replicate previous effects stems in part from a realization that the field now has updated knowledge on methodological issues like sample sizes,
randomization, and experimenter influence that change the way we conduct experiments (i.e., Simons, Nelson, & Simonsohn, 2011). Given the intriguing theoretical rational for the effects, but inconsistent results obtained, the current research aimed to replicate the cued recognition effect.

Furthermore, we sought to place the cued face recognition effects in context by comparing them to traditional recognition memory, whereby single faces are encoded and single faces are tested; for simplicity, we refer to this as “orienting” the effects of cuing because we aim to understand how memory tested in cued recognition paradigm compares to memory tested in the traditional, non-cued recognition paradigm. In the cued face recognition paradigm (where test faces are presented with the same cue face as at encoding, a different cue face, or no cue face), the correct cues may be considered equivalent to the weak cues of Thomson and Tulving (1970), meaning that they are contextually associated as a result of being presented together at encoding. Thus, when researchers have demonstrated an increase in hits for these correct face cues (cf. incorrect or no cue faces), it is taken as confirmation that reconstructing encoding context enhances face recognition. However, the incorrect cues in face recognition cannot be considered a proxy for the strong cues, since unfamiliar faces do not have semantic associations. Instead, the incorrect cues test whether change of context undermines recognition compared to the consistent context of the correct cues, or the absence of any contextual cue. Yet, while the no-cue condition is intended to act as a control group, it represents a change of context because the absence of context in itself is a deviation from the original. Given that hit rates in incorrect and no-cue conditions sometimes, but not always, differ significantly, it is unclear whether previously-reported effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both.
Existing research also exposes questions regarding the potential boundary conditions for such an effect. For example, do any beneficial context effects vary according to the number of additional faces to be encoded? Research on cued face recognition to date has held encoding conditions constant while manipulating the conditions at retrieval (e.g., Watkins et al., 1976; Winograd & Rivers-Bulkeley, 1977). However, pairing two faces at encoding represents the minimum number of faces individuals might encode when attempting to implement cued face recognition. Limiting our consideration to pairs of faces fails to reflect the variability in group sizes that individuals encounter every day, and it is unclear if the benefits identified in previous work extend to conditions in which participants need to encode more stimuli. In other words, if these effects replicate, can they also be useful in situations with more than just two faces? Although encoding additional faces may provide contextual cues to aid subsequent recognition, cognitive load is likely to increase with each additional face requiring encoding. As this load increases, the resources needed to successfully encode the target face and to develop the associations required to support subsequent cued-recognition may be reduced.

These applied and theoretical considerations converge into three questions for the current research: (1) do previously-reported effects of contextual cuing replicate with contemporary methodology? (2) if so, do these effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both? and (3) how do cuing effects vary as a function of the number of cues to be encoded?

Experiment 4 was a first attempt to establish to what extent correct cuing is beneficial or compensating for increased cognitive demand, and also whether this effect endures when cognitive load is increased by viewing multiple (i.e., more than two) faces at once. We expected to replicate previous effects of cued face recognition
in the two-face condition, and that those effects would extend to the four-face condition, though to a lesser extent.

The current research

Replication and orientation of previous effects. In Experiment 4, we sought to replicate previously-reported effects of cued face recognition with pairs of faces while isolating the effects of cued recognition from other influences such as the context effects of placement, or intuitive responses based on overt compatibility of couples (see Winograd & Rivers-Bulkeley, 1977). Therefore, the current research randomized orientation, meaning that faces encoded on the right side of the screen were not necessarily later tested on the right side of the screen. We also randomly paired or grouped photographs of male faces selected from a large database (cf. Watkins et al., 1976; Winograd & Rivers-Bulkeley, 1977).

Next, we sought to orient the effects of cued recognition, and therefore to understand how memory tested in cued recognition paradigm compares to memory tested in the traditional, non-cued recognition paradigm. In particular, we aim to explore whether previously-reported effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both. To accommodate this, the current research included a new control condition in which participants study faces presented individually at encoding and tested without cues, rather than in groups of two or four (see Figure 3.1 for example stimuli and Figure 3.2 for graphic representation of encoding-test phases). This control condition allowed us to examine the impact of number of faces at encoding. It may be that recognition rates without cues are equivalent across single-face and multiple-face encoding conditions. Thus, any benefit of correct cuing, for example, would be an enhancement over general face recognition. By contrast, if the no-cue condition varies between the single-face and
multiple-face conditions, then any benefit of correct cuing would be compensation for increased difficulties.

**Extension of cued recognition.** We also tested a potential boundary condition for this cuing effect. We would expect more faces at encoding to increase cognitive load and potentially interfere with effective encoding. Thus, the current research investigated whether the effects of cued face recognition could be extended to encoding groups of more than two faces. For example, if four faces are studied together, can one of those faces effectively serve as a cue for another?

There are several reasons why cuing with four-face stimuli may not be as effective as with two-face stimuli. First, there is the potential for generalized deficits in memory with the increased cognitive load associated with having more faces (e.g., four vs. two) to encode. Cuing in this case may be helpful, but with a general trend for four-face condition to have decreased accuracy. Or cuing could be less effective because there is diminished memory to support. Second, when faces are always presented at study as pairs, a correct cue also represents a complete cue. However, if four faces are presented during encoding and only one face is used to cue the recognition of the target face, this may represent a partial cue. Therefore, providing only one of the set of the associated cues at test may show a weaker effect. Third, there is the potential that associations between multiple (i.e., more than two) faces may not be encoded equally, and therefore may not be equally effective at retrieval. While two faces viewed side-by-side provide only one association that can be formed at encoding, four faces provide multiple associations that can be formed with varying strengths. Thus, the face chosen by experimenters to be a contextual cue may not be the one that was most strongly associated with the target at encoding. By contrast, it may be that each face is one of several possible cues, meaning that any of the faces
would be sufficient to enhance recognition. To be clear, we do not test for the differences between these three concerns, though we raise them as three ways in which cued face recognition may not extend effectively when simultaneously encoding more than two faces.

In comparing effects across conditions with only single faces and groups of four faces, there were two major methodological concerns. First, we considered how increasing the number of faces on a given trial would impact the allocation of attention at encoding. Bindemann and colleagues (2012) demonstrated that given multiple stimuli, participants generally allocate their attention evenly between the stimuli. Participants did not know which of the two faces would be tested afterwards, but it is unsurprising that identification performance was positively correlated with the amount of time spent studying the appropriate target face. Thus, the amount of attention devoted to any individual stimuli would be reduced if encoding duration remains constant but the number of to-be-encoded stimuli increases. We attenuated this concern by providing comparative encoding durations per face, rather than per trial. Participants in all conditions were allotted 2 s per face, even if those faces are presented in pairs or as a group. Thus, the exposure duration for single-face, two-face and four-face trials were 2, 4, and 8 s, respectively.

Second, in order to compare the single-, two- and four-face conditions, we needed participants to complete the same number of test trials. However, this presents a problem at encoding. A participant that studies 36 trials of individual faces in the encoding phase will generally be tested on 72 trials of individual faces (half old, half new faces) in the test phase. Participants in the paired-face and four-face conditions should therefore also confront 72 test trials that vary old and new faces. However, to test for cued face recognition, the 72 target faces at test require extra faces at encoding
to be used as cues. The natural solution to this is to hold constant the number of encoding trials so that all conditions study 36 trials at encoding. Yet 36 trials in a paired-face condition and a four-face condition consequently mean that a participant studies 72 faces and 144 faces at encoding, presenting extreme differences in memory load. In short, if we hold constant the number of test trials, we cannot also control for both the number of encoding trials and the total number of faces studied. Our solution was to create testing blocks so that participants only every studied 36 faces in each block: 36 individual faces, 18 paired faces, and 9 groups of four faces. They were then tested on 72, 36, and 18 old and new faces, respectively. The two- and four-face conditions repeated this study-test cycle until they had also completed 72 test trials, meaning the two-face condition had two study-test blocks and the four-face condition had four study-test blocks. Breaks in between the blocks were included in order to compensate for the memory load induced by viewing more faces.

**Experiment 4**

Experiment 4 was a first attempt to establish to what extent correct cuing is beneficial or compensating for increased cognitive demand, and also whether this effect endures when cognitive load is increased by viewing multiple (i.e., more than two) faces at once.

**Method**

**Participants.** Of 81 university student participants, three failed to follow instructions and were excluded from data analysis. The remaining 78 participants (60 women, 18 men) were between the ages of 18 and 29 (M = 21.99, SD = 2.05). Participants were compensated with either participation credit or a 5€ voucher.

**Design.** We used a 3(encoding group size: single face, two faces, four faces) x 3(cue type: no cue, correct cue, incorrect cue) x 2(test face status: old, new) mixed
design, with encoding group size as the between-subjects factor. Participants were randomly assigned to one of the three between-subjects conditions. Cue type (no cue, correct cue, and incorrect cue) were randomly ordered and target faces were determined at random. Two versions of the experiment were created in which the order of trials for all conditions was changed, and in the two- and four-face conditions, different target faces and cue faces were selected.

Materials.

Face stimuli. For the experiment, 204 photographs of male faces were used from a database of faces at Flinders University (Adelaide, Australia), the Psychological Image Collection at Stirling (pics.stir.ac.uk), and the AR Face Database (Martinez & Benavente, 1998). Of those, 36 faces were used as target faces and the remaining 168 faces were used as face cues, filler faces during encoding, or filler faces during test. Faces with features that were highly distinctive (e.g., piercings, unique haircuts, facial hair) were removed. All photographs were in full color with a resolution of 300 x 300 pixels. They were presented on a computer screen with a resolution of 1366 x 768 pixels using Microsoft PowerPoint 2010. In the single-face condition, faces were placed in the center of the screen. In the two-face condition, two faces were placed side by side. In the four-face condition, the faces were presented in a 2 x 2 matrix. To create the multiple faces conditions, the faces were randomly grouped into pairs (two-face condition) or groups of four faces (four-face condition).
Procedure. At encoding, participants were asked to memorize a number of faces. Between the encoding and testing phases, participants performed a short arithmetic task for 30 s. At test, participants were instructed to judge a test face, which was indicated to the participant by a green square around the face, as either being old.

Figure 3.1. Example encoding stimuli using images from AR Face database.

Figure 3.2. Experiments 4 and 5: Procedure for encoding and testing. Participants studied single faces, pairs of faces, or groups of four faces during the encoding phase. Participants in the single-face condition were only tested on individual faces (i.e., no-cue trials). Participants in the two-face and four-face condition were tested on trials with no cue and trials with either a correct cue or incorrect cue. Here, we demonstrate in
conditions when, during the test phase, both the target face is present and the cue face is correct. However, participants were also tested on target faces that were absent and with cues that were incorrect. Color coding is used to demonstrate how cues and targets were chosen, but participants did not know which faces during the encoding phase would be used as targets or cues. During the test phase, a box indicated the target face participants should report whether or not they recognized. Targets and cues were chosen at random from the left or right position of the encoding phase, but were always presented on the right and left, respectively, during the test phase.

(see at encoding) or new (not seen before). If the participant indicated an ‘old’ response, they were also asked to indicate their confidence on a scale from 1 (not at all confident) to 7 (very confident). In each condition, only one of the faces presented during the encoding phase was used as a target face at test. Participants in the single-face condition completed one study-test block, consisting of 36 single face stimuli during the encoding phase, and 72 single face trials at test (including the 36 study faces as targets and 36 previously unseen faces; these are referred to as old and new faces, respectively). Since the two-face and four-face conditions contained more faces per encoding trial, multiple blocks with fewer encoding trials were used to guarantee that all participants were exposed to the same number of faces during each encoding phase for each condition. Thus, participants in the single-face conditions viewed 36 encoding trials of single faces; the two-face condition viewed 18 encoding trials of face pairs; the four-face condition viewed 9 encoding trials of groups of four faces.

**Single-face condition.** During the encoding phase, participants viewed 36 individual photos of faces. Each face (all target faces) was shown for a duration of 2 s, followed by an inter stimulus interval (ISI) of 500 ms before the next face appeared. During the testing phase, participants completed 72 single face trials. Presence of the test face was manipulated so that half of the trials were old, and half were new.

**Two-face condition.** During the encoding phase, participants viewed 18 pairs of faces. Each pair (consisting of one target face and one cue face) was shown for 4 s

---

8 Note that due to error in collection, confidence data were only recorded for choosers.
(allowing 2 s per face) with an ISI of 500 ms. No information was provided to participants on which face of the two faces they would be tested, but participants were told to focus on both faces equally during encoding. During the testing phase, participants completed 36 trials. Of these trials, 12 trials were presented with no cue, 12 trials were presented with a correct cue (a face seen during the encoding phase), and 12 trials were presented with an incorrect cue (a never-before seen face). Presence of the target face was manipulated so that half of the trials were old, and half were new. The target face was randomly selected from the pairs of faces, so it could be either a face that was studied on the left or the right side of the screen. At test, the face was always shown on the right side of the screen during test and always demarcated by a green square. This process was repeated to create two study-test blocks.

**Four-face condition.** This condition involved four blocks of trials. During the encoding phase, the participants viewed nine groups of four faces. Each stimulus group (consisting of one target face and three cue faces) was shown for 8 s (allowing 2 s per face) with an ISI of 500 ms. Again, we did not tell participants which face would be the test face, but told participants to focus on all faces equally during encoding. During the testing phase, participants completed 18 trials. Of these trials, six trials were presented with no cue, six trials were presented with a correct cue, and six trials were presented with an incorrect cue. Presence of the target face was manipulated so that half of the trials were old, and half were new. The target face was randomly selected from the groups of faces, so it could be either a face that was studied on the top or bottom left or the top or bottom right side of the matrix of four faces. However, the face was always shown on the right side of the screen during test and always demarcated by a green square. This process was completed four times.
Results

Data were screened for outliers and normality prior to analysis. Shapiro-Wilk tests were used to assess normality for each of the conditions. When normality was violated, both nonparametric (Kruskal Wallis H and Mann-Whitney U) and standard inferential tests (one-way Analyses of Variance [ANOVA]s and t-test) were conducted on hit rates and false-alarm rates. Results did not differ as a result of test, therefore ANOVA and t-tests are reported throughout. Where relevant, we conducted Bayesian analyses to determine whether the data provided evidence of equivalence using JASP (2017) software. We use Jarosz and Wiley’s (2014) interpretations of Bayes Factors as evidence for the alternative hypothesis, in which they report descriptive thresholds provided by Jeffreys (1961) and Raftery (1995) work in Bayes models. Approximate cut-offs are as follows: 1-3 constitutes anecdotal/weak evidence, 3-10 is positive/substantial evidence; 10-20 is strong/very strong evidence; 20+ is very strong/decisive. Descriptive statistics of hit and false-alarm rates as a function of cue type are reported in Table 3.1 and descriptive statistics of sensitivity and response bias can be found in Table 3.2. Inferential statistics are not reported in the text, but are available in Tables 3.3-3.5, and Figure 3.3.

Hit rates were defined as the probability of a correct response given that the stimulus (S) is present (i.e., the trial shows an old face; \( H = P(\text{“old”}) \mid S_{\text{present}} \)). The hit rate was calculated here by dividing the number of correct “old” responses (hit) on present trials by the total number of present trials. The false-alarm rate was the probability of an incorrect response given that the stimulus (S) is absent (i.e., the trial shows a new face; \( FA = P(\text{“old”}) \mid S_{\text{absent}} \)). The FA rate was calculated here by dividing the number of incorrect “old” responses on absent trials by the total number of absent trials.
In order to examine the no-cue condition, we tested hit rates and false-alarm rates across all face conditions (single-, two-, four-face). Because the single-face condition does not have correct- and incorrect-cues, only the two- and four-face groups were considered when comparing these cue-types.

**Orienting the effect of cuing.** First we tested whether previously-reported effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both. To address this, we tested the effect of group size in the no-cue conditions on hit and false-alarm rates, using one-way ANOVAs. This allows us to understand the initial differences in how number of faces at encoding impacts recognition performance.

For hit rates, there were no differences between groups, meaning that there was no evidence of any difference between the single-, two-, or four-face encoding conditions when no cue was presented during test (see Table 3.3 for ANOVA results). For false-alarm rates, there were no significant differences between groups, meaning there was no evidence of any difference for false-alarm rates between the face
Table 3.1
Experiments 4 and 5: Mean Hit Rates and False-Alarm Rates (95% CI) by Cue Type and Face Encoding Condition

<table>
<thead>
<tr>
<th>No. faces encoded</th>
<th>Hit rates by cue type</th>
<th>False-alarm rates by cue type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Experiment 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-face</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Two-face</td>
<td>.71 (.64-.78)</td>
<td>.72 (.64-.80)</td>
</tr>
<tr>
<td>Four-face</td>
<td>.72 (.67-.78)</td>
<td>.74 (.68-.80)</td>
</tr>
<tr>
<td>Experiment 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-face</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Two-face</td>
<td>.72 (.66-.78)</td>
<td>.70 (.66-.74)</td>
</tr>
<tr>
<td>Four-face</td>
<td>.72 (.66-.79)</td>
<td>.73 (.64-.81)</td>
</tr>
</tbody>
</table>
Table 3.2
Experiments 4 and 5: Mean Sensitivity and Response Bias Rates (95% CI) by Cue Type and Face Encoding Condition

<table>
<thead>
<tr>
<th>No. faces encoded</th>
<th>Sensitivity by cue type (d')</th>
<th>Response bias by cue type (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td><strong>Experiment 4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-face</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Incorrect</td>
<td>1.24 (0.98-1.49)</td>
<td>1.29 (1.00-1.58)</td>
</tr>
<tr>
<td>None</td>
<td>1.68 (1.36-2.00)</td>
<td>1.56 (1.30-1.82)</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 5</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-face</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Incorrect</td>
<td>1.50 (1.26-1.74)</td>
<td>1.57 (1.39-1.76)</td>
</tr>
<tr>
<td>None</td>
<td>1.61 (1.36-1.86)</td>
<td>1.44 (1.16-1.72)</td>
</tr>
</tbody>
</table>
encoding conditions when no cue was presented during test (see Table 3.3 for ANOVA results).

While Bayesian analyses did not provide evidence to support the alternative hypothesis (i.e., there are differences between face encoding conditions on hits or false-alarms), neither did they provide compelling evidence to support the null (i.e., there are no differences between face encoding conditions on hits or false-alarms). Indeed, Bayes Factors for the null hypothesis for both hit rate and false-alarm do not exceed 2.45, which is considered anecdotal or weak evidence at best (Jarosz & Wiley, 2014). Therefore, the results of the no-cue condition remain unclear and cannot support statements about orienting the effects of cuing. See Table 3.4 for inferential statistics of Bayesian analyses.

**Replicating and extending effects of cued face recognition.** Next, we asked whether we could replicate previous effects of cued face recognition, and whether these effects vary as a function of the number of faces to be encoded. Therefore, we conducted 2 (encoding condition: two-face, four-face) x 3 (cue type: no, correct, incorrect) mixed ANOVAs on hit rates and false-alarm rates with cue type as the within-subjects variable. As per previous experiments, we expected that correct cues would enhance hit rates compared to no cues and incorrect cues. Although we also expected to see false-alarm rates rise with correct cues (cf. no cues or incorrect cues), we did not expect these to eliminate the enhancing effects of correct cuing for recognition.

**Hit rates.** First, we compared hit rates between the two- and four-face conditions given both correct and incorrect contextual cues (see Table 3.3). There were no significant main effects of cue type and face encoding condition on hit rates; the interaction was also non-significant. Thus, we found no evidence that the number of faces at encoding nor the type of cue presented at test affected participant hit rates.
Bayes analyses (see Table 3.4) provide strong support for the null hypothesis, with the model including both main effects and the interaction have a Bayes factor of 233.55. This gives us confidence to suggest that hit rates do not change in our experiment regardless of encoding or testing conditions. Thus, encoding conditions that control for divided attention and testing conditions that attempt to cue participants do not impact true recognition of previously-studied old faces.

**False-alarm rates.** Next, we compared false-alarm rates between the two- and four-face conditions given both correct and incorrect contextual cues (see Table 3.3). There was a significant main effect of cue that was modified by a significant interaction between face encoding condition and cue. We conducted simple main effects analyses to examine this interaction. Comparing face encoding conditions, participants in the two-face condition had significantly more false alarms when a correct cue was shown compared to participants in four-face condition. There were no other differences between the two- and four-face conditions with varying cues. However, these effects should be interpreted cautiously. While we report the results of the interaction because of the significant $p$-value, Bayesian analyses provide weaker support for the interaction ($BF_{10}^9 = 2.81$; see Table 3.4).

**Sensitivity and response bias.** We computed signal detection statistics to test for effects on sensitivity and response bias ($d'$ and $c$, respectively). In the current research, sensitivity is the capacity to discriminate between old and new faces, while response bias is the tendency to respond *old* or *new* regardless of whether the face is actually old.

---

9 For ease of interpretation, $BF_{10}$ is used to designate the Bayes factor as evidence in favor of the alternative hypothesis and $BF_{01}$ is used to designate the Bayes factor as evidence in favor of the null hypothesis.
Table 3.3

Experiments 4 and 5: Results for ANOVAs and t-tests Comparing Hit Rates and False-Alarm Rates Across Cue Types and Face Encoding Condition

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Description</th>
<th>df</th>
<th>F</th>
<th>η²</th>
<th>t</th>
<th>d</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 4</strong></td>
<td>Orienting the effect of cuing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>One-way ANOVAs (No. faces; 1, 2, 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR</td>
<td>2, 75</td>
<td>1.69</td>
<td>.043</td>
<td>.191</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAR</td>
<td>2, 75</td>
<td>1.81</td>
<td>.046</td>
<td>.171</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Replicate and extend cuing effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed ANOVAs 2 (No. faces: 2v4) x 3 (cue type)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR</td>
<td>2, 100</td>
<td>0.27</td>
<td>.005</td>
<td>.762</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1, 50</td>
<td>0.61</td>
<td>.012</td>
<td>.439</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>2, 100</td>
<td>0.41</td>
<td>.008</td>
<td>.666</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAR</td>
<td>2, 100</td>
<td>4.74</td>
<td>.087</td>
<td>.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1, 50</td>
<td>2.26</td>
<td>.043</td>
<td>.139</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>2, 100</td>
<td>3.23</td>
<td>.061</td>
<td>.044</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up FAR interaction: simple main effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Independent sample t-tests (No. faces: 2v4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No Cue</td>
<td>50</td>
<td>0.15</td>
<td>.04</td>
<td>.881</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>50</td>
<td>2.80</td>
<td>.78</td>
<td>.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incorrect</td>
<td>50</td>
<td>1.21</td>
<td>.33</td>
<td>.234</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 5</strong></td>
<td>Orienting the effect of cuing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>One-way ANOVAs (No. faces; 1, 2, 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR</td>
<td>2, 74</td>
<td>1.10</td>
<td>.029</td>
<td>.337</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAR</td>
<td>2, 74</td>
<td>3.95</td>
<td>.096</td>
<td>.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow-up analysis for main effect of cue (FAR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tukey post hoc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1v2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1v4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.105</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2v4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.799</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Replicate and extend cuing effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed ANOVAs 2 (No. faces: 2v4) x 3 (cue type)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR</td>
<td>2, 98</td>
<td>0.10</td>
<td>.002</td>
<td>.904</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1, 49</td>
<td>0.35</td>
<td>.007</td>
<td>.556</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>2, 98</td>
<td>0.18</td>
<td>.004</td>
<td>.833</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAR</td>
<td>1.78, 87.12</td>
<td>5.89</td>
<td>.107</td>
<td>.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1, 49</td>
<td>.316</td>
<td>.006</td>
<td>.577</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>1.78, 87.12</td>
<td>2.96</td>
<td>.057</td>
<td>.063</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** T-tests are two-tailed. Experiment 4: Participants in the two-face condition had significantly more false-alarms when a correct cue was presented in comparison to the four-face condition. Experiment 5: For the no-cue condition, Participants in the single-face group had fewer false-alarms compared to participants in the four-face group. Across the two- and four-face conditions, the no-cue trials produced the highest false-alarms compared to the correct-cue trials and the incorrect cue trials.
Table 3.4
Experiments 4 and 5: Results for Bayesian ANOVAs Comparing Hit Rates (HR) and False-Alarm Rates (FAR) Across Cue Types and Face Encoding Condition

| Models | P(M) P(M|data) | BF_{M} | BF_{10}(BF_{01}) | % error |
|--------|---------------|--------|------------------|---------|

**Experiment 4**

Orient the effects of cuing

One-way ANOVAs (No. faces; 1, 2, 4)

<table>
<thead>
<tr>
<th>HR</th>
<th>Null model</th>
<th>0.50</th>
<th>0.71</th>
<th>2.45</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Faces</td>
<td>0.50</td>
<td>0.29</td>
<td>0.41</td>
<td>0.41 (2.45)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

FAR

<table>
<thead>
<tr>
<th>Null model</th>
<th>0.50</th>
<th>0.69</th>
<th>2.24</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Faces (1, 2, 4)</td>
<td>0.50</td>
<td>0.31</td>
<td>0.45</td>
<td>0.45 (2.24)</td>
</tr>
</tbody>
</table>

Replicate and extend effects of cuing

Mixed ANOVAs 2 (No. faces: 2v4) x 3 (cue type)

<table>
<thead>
<tr>
<th>HR</th>
<th>Null model</th>
<th>0.20</th>
<th>0.68</th>
<th>8.51</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue</td>
<td>0.20</td>
<td>0.06</td>
<td>0.23</td>
<td>0.08 (12.36)</td>
<td>0.80</td>
</tr>
<tr>
<td>No. Faces (2v4)</td>
<td>0.20</td>
<td>0.24</td>
<td>1.27</td>
<td>0.35 (2.82)</td>
<td>1.27</td>
</tr>
<tr>
<td>Cue + 2v4</td>
<td>0.20</td>
<td>0.02</td>
<td>0.08</td>
<td>0.03 (32.99)</td>
<td>2.91</td>
</tr>
<tr>
<td>Cue + 2v4 + Cue 2v4</td>
<td>0.20</td>
<td>&lt; 0.01</td>
<td>0.01</td>
<td>.004 (233.55)</td>
<td>1.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FAR</th>
<th>Null model</th>
<th>0.20</th>
<th>0.11</th>
<th>0.48</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue</td>
<td>0.20</td>
<td>0.30</td>
<td>1.72</td>
<td>2.83 (0.35)</td>
<td>0.80</td>
</tr>
<tr>
<td>No. Faces (2v4)</td>
<td>0.20</td>
<td>0.08</td>
<td>0.32</td>
<td>0.70 (1.42)</td>
<td>0.80</td>
</tr>
<tr>
<td>Cue + 2v4</td>
<td>0.20</td>
<td>0.22</td>
<td>1.13</td>
<td>2.07(0.48)</td>
<td>2.35</td>
</tr>
<tr>
<td>Cue + 2v4 + Cue 2v4</td>
<td>0.20</td>
<td>0.30</td>
<td>1.68</td>
<td>2.81 (0.36)</td>
<td>2.12</td>
</tr>
</tbody>
</table>

**Experiment 5**

Orient the effects of cuing

One-way ANOVAs (No. faces; 1, 2, 4)

<table>
<thead>
<tr>
<th>HR</th>
<th>Null model</th>
<th>0.50</th>
<th>0.79</th>
<th>3.80</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Faces (1, 2, 4)</td>
<td>0.50</td>
<td>0.21</td>
<td>0.26</td>
<td>0.26 (3.80)</td>
<td>0.03</td>
</tr>
</tbody>
</table>

FAR

<table>
<thead>
<tr>
<th>Null model</th>
<th>0.50</th>
<th>0.31</th>
<th>0.44</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Faces (1, 2, 4)</td>
<td>0.50</td>
<td>0.70</td>
<td>2.28</td>
<td>2.28 (0.44)</td>
</tr>
</tbody>
</table>

Replicate and extend effects of cuing

Mixed ANOVAs 2 (No. faces: 2v4) x 3 (cue type)

<table>
<thead>
<tr>
<th>HR</th>
<th>Null model</th>
<th>0.20</th>
<th>0.71</th>
<th>9.90</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue</td>
<td>0.20</td>
<td>0.05</td>
<td>0.21</td>
<td>0.07 (14.23)</td>
<td>0.83</td>
</tr>
<tr>
<td>No. Faces (2v4)</td>
<td>0.20</td>
<td>0.22</td>
<td>1.12</td>
<td>0.31 (3.25)</td>
<td>1.02</td>
</tr>
<tr>
<td>Cue + 2v4</td>
<td>0.20</td>
<td>0.02</td>
<td>0.07</td>
<td>0.02 (43.48)</td>
<td>5.84</td>
</tr>
<tr>
<td>Cue + 2v4 + Cue 2v4</td>
<td>0.20</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01 (353.33)</td>
<td>2.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FAR</th>
<th>Null model</th>
<th>0.20</th>
<th>0.07</th>
<th>0.30</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue</td>
<td>0.20</td>
<td>0.58</td>
<td>5.50</td>
<td>8.21 (0.12)</td>
<td>0.96</td>
</tr>
<tr>
<td>No. Faces (2v4)</td>
<td>0.20</td>
<td>0.02</td>
<td>0.08</td>
<td>0.27 (3.75)</td>
<td>1.14</td>
</tr>
<tr>
<td>Cue + 2v4</td>
<td>0.20</td>
<td>0.15</td>
<td>0.72</td>
<td>2.16 (0.46)</td>
<td>1.00</td>
</tr>
<tr>
<td>Cue + 2v4 + Cue 2v4</td>
<td>0.20</td>
<td>0.18</td>
<td>0.87</td>
<td>2.54 (0.39)</td>
<td>1.24</td>
</tr>
</tbody>
</table>
Note. HR and FAR represent hit rates and false-alarm rates, respectively. BF_{10} refers to evidence in favor of the alternative hypothesis. In parentheses, BF_{01} refers to evidence in favor of the null hypothesis (BF_{01} = 1/BF_{10}).
### Table 3.5
*Experiments 4 and 5: Results for T-tests and ANOVAs Comparing Sensitivity and Response Bias as a Factor of Face Encoding Condition or Cue Type*

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>F</th>
<th>$\eta^2$</th>
<th>t</th>
<th>d</th>
<th>p</th>
<th>BF_{10}(BF_{01})</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two vs. four-face conditions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (d’)</td>
<td>50</td>
<td>1.54</td>
<td>.131</td>
<td>.55</td>
<td>1.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Bias (c)</td>
<td>50</td>
<td>0.62</td>
<td>.541</td>
<td>0.30</td>
<td>3.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cue type (no, correct, incorrect cue)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d’</td>
<td>2, 60</td>
<td>0.74</td>
<td>.024</td>
<td>.482</td>
<td>0.18</td>
<td>5.71</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>1.60, 47.98</td>
<td>2.17</td>
<td>.067</td>
<td>.135</td>
<td>0.53</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two vs. four-face conditions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (d’)</td>
<td>48</td>
<td>0.06</td>
<td>.949</td>
<td>0.28</td>
<td>3.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Bias (c)</td>
<td>48</td>
<td>0.71</td>
<td>.484</td>
<td>.35</td>
<td>2.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cue type (no, correct, incorrect cue)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d’</td>
<td>2, 54</td>
<td>3.23</td>
<td>.107</td>
<td>.047</td>
<td>1.38</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>2, 54</td>
<td>4.32</td>
<td>.138</td>
<td>.018</td>
<td>3.11</td>
<td>0.32</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* We computed signal detection statistics to test for effects on sensitivity and response bias (d’ and c, respectively). We used t-tests to compare across the two-face and four-face conditions. We used repeated measures ANOVA’s across the no, correct, and incorrect cue conditions. The single-face condition was excluded due to the difference in cue types. *Experiment 5:* Sensitivity changed as a result of cue type such that participants had significantly lower sensitivity on no-cue trials compared to incorrect cues. Response bias changed as a result of cue type such that participants had significantly more positive response bias on no-cue trials compared to correct or incorrect cues.
or new (Macmillan & Creelman, 2005). We used t-tests to compare sensitivity and response bias across the two-face and four-face conditions. Again, the single-face condition was excluded due to the difference in cue types. There were no significant differences between groups, $BF_{10} \leq 0.55$, but there was sufficient evidence to support the null hypothesis for response bias ($BF_{01} = 3.39$). We used one-way ANOVAs to compare sensitivity and response bias separately across cue type (no, correct, incorrect cues). There were also no significant differences between groups $BF_{10} \leq 0.53$, though there was sufficient evidence to support the null hypothesis for sensitivity, $BF_{01} = 5.71$.

**Discussion**

In Experiment 4, we sought to replicate and extend previous research on cued face recognition (i.e., Watkins, Ho, & Tulving, 1976; Winograd & Rivers-Bulkeley, 1977). While previous research investigating cued face recognition has previously used only pairs of faces at encoding, we aimed to test the limits of this effect by adding two conditions (single-face control, four-face condition) to the encoding-phase. Specifically, the single-face condition was added to establish baseline differences in studying one vs. multiple faces, thus orienting the effects of cuing. The four-face condition was added to determine whether the effect would also arise in situations in which people encode more than two faces at a time. At test, participants in the single-face condition were shown individual faces that had either been previously studied (old) or had never been studied before (new). Participants in the two-face or four-face encoding conditions were shown either a target face alone (no cue), or a target face presented alongside a correct or incorrect cue face.

As expected, there were no differences in hits or false-alarm rates for test faces presented without a cue regardless of whether participants studied, one, two, or four
faces at encoding. This is in line with previous research demonstrating equivalent accuracy rates when controlling for divided attention (Bindemann et al., 2012). Contrary to expectations, however, hit rates did not benefit from the provision of correct cues at test, and were not diminished by incorrect cues at test. In fact, Bayesian analyses provide strong support for the null hypothesis— that cuing would have no impact on hit rates. This is an interesting finding and may suggest that true recognition is not influenced by the manipulations that we presented when divided control and memory load are controlled for.

Unlike hit rates, false-alarm rates did show some differences between face encoding conditions and cue type. Participants in the two-face condition produced more false-alarms than participants in the four-face conditions only when a correct cue was shown. However, we view these results with some caution given the weaker evidence emerging from the Bayesian analyses.

**Experiment 5**

The results of Experiment 4 are puzzling given previous research in cued face recognition, but contribute to a field that has historically displayed mixed results dependent upon methodology (Davies, 1988). One explanation that may account for our results is that the association between the grouped faces during encoding was not strong enough. In our experiment, participants were simply instructed to study faces, and while pairs or groups of faces appeared simultaneously on the screen during the study phase, there was no interactive component to encourage participants to encode those faces together. Cuing relies on associative encoding, and presenting items together does not necessarily encourage associative links. Peris (1985; as cited in Davies 1988) demonstrated this contingency when he asked participants to encode photographs of people holding tools with posters in the background in order to
explore differences between interactive vs. independent encoding. Most relevant for the current research, he also tracked participant eye-movements during encoding. When participants were later asked to make identification decisions of the people, either the posters or the tool was manipulated to be used as contextual cues. It was expected that the interactive cues (i.e., tools) would enhance recognition accuracy of the person holding them while the independent cues (i.e., posters) would not. By contrast, results showed it did not matter which of these context cues was presented: accuracy was affected by the context change as long as the participant had fixated on that particular context (i.e., tool or poster).

We may need to purposefully encourage participants to develop some connection between the presented faces. Winograd and Rivers-Bulkeley (1977) called this process ‘unitization’, which they implemented in their research by asking participants to rate compatibility between couples (male and female target faces) shown on the screen. We did not consider this an appropriate manipulation to include because of our additional conditions. This was in part because of the stimulus materials (i.e., we used only male faces) and in part because of the difficulty of asking participants to rate romantic compatibility between four targets in the four-face condition. Furthermore, there is no direct proxy for romantic compatibility for only one face in the single-face condition. Thus, in repeating the above experiment, we included a judgment task encouraging participants to consider the (fabricated) relationship between first impressions of and the memorability of guilty defendants’ faces and sentencing length. In this task, each pair or group of faces presented on screen during the study phase were described as presenting perpetrators of a multi-perpetrator crime, who had been convicted by a jury and sentenced to equivalent prison terms. We asked participants to use their initial impressions of the faces
together to estimate how many years the pair or group had been sentenced to prison. We considered this a suitable alternative as a manipulation because criminals that commit a crime together are seen as a very closely-linked, or highly entitative, group (Likel et al., 2000). Furthermore, because participants must get an impression of all faces in order to estimate a single sentencing length, we expected it would necessitate participants to consider the individual stimuli as a group, and thus actively engage in unitized or associative encoding.

Experiment 5 is a replication of Experiment 4 with an added manipulation for associative encoding, thus the hypotheses were the same as for Experiment 4.

**Method**

**Participants.** Seventy-eight university students (61 women, 17 men) participated in the current experiment. Participants were between the ages of 18 and 43 ($M = 22.28$, $SD = 4.13$). Participants were compensated with either participation credit or a 7.50€ voucher.

**Materials.** The photographs, pairings, and presentation orientation, for Experiment 5 were the same as for Experiment 4.

**Procedure.** The procedure for Experiment 5 was the same as Experiment 4, with one exception. Participants were told that the experiment was to test the supposed relationship between initial impressions of defendant criminality, the memorability of their faces, and the number of years a single convicted defendant or a group of convicted defendants were sentenced to prison. After each stimulus was presented during the encoding stage, participants were asked to rate on a Likert scale how many years (4-10) the person(s) presented had been sentenced to prison by the judge. For example, in the four-face condition, participants were instructed, “*All members of the pair were sentenced to the same number of years in prison. If you*
mark the number ‘4’, it means that the person on the top left was sentenced to 4 years in prison, and the person on the top right, the bottom left, and the bottom right were each sentenced to 4 years in prison.” While the encoding phase automatically advanced after 2 s per face, the ISI was self-paced, and participants could press the spacebar on the keyboard to advance to the next stimulus for all conditions.

This manipulation was chosen as an alternative to relationship compatibility ratings of couples used by Winograd and Rivers-Bulkeley (1977) to unitize faces during encoding. However, this instruction could be plausibly applied to the single-face and four-face conditions as well as the two-face condition.

**Results**

Data were screened for outliers and normality prior to analysis. One outlier was removed from the four-face condition for a high false-alarm rate, which was nearly twice the next-highest false-alarm rate in the four-face condition (.69 vs .36). Shapiro-Wilk tests were used to assess normality for each of the conditions. When normality was violated, both nonparametric (Kruskal Wallis H and Mann-Whitney U) and standard inferential tests (one-way ANOVA and t-test) were conducted on hit rates and false-alarm rates. Results did not differ as a result of test, therefore ANOVAs and t-tests are reported throughout. Bayes analyses were also conducted and reported. See Table 3.1 for hit and false-alarm rates according to cue type. Again, results are reported descriptively and inferential statistics can be found in Tables 3.1-3.4, and Figure 3.3.

**Orienting the effect of cuing.** Again, we first sought to determine whether previously-reported effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both. We therefore examined hit rates and false-alarm rates in the no-cue trials across single-face, two-face, and four-face encoding conditions using one-way
ANOVA (see Table 3.3). This allows us to understand the initial differences between the three conditions altering the number of faces and acts as an anchor around which we can understand the resulting effects of cuing.

Figure 3.3. Experiment 5: False-alarm rates as a function of cue type and face encoding condition. The single-face condition had significantly fewer false-alarms compared to the four-face condition. In the two- and four-face condition, participants produced more false-alarms when no cue was presented compared to when a correct or incorrect cue was presented.

For hits, there were no differences between groups, meaning that there was no evidence that hit rates differed across face encoding conditions when no cue was presented during test. Bayesian analyses showed moderate support for the null hypothesis ($BF_{10} = 3.80$; see Table 3.4). For false-alarms, there was a significant difference between face encoding conditions, though Bayesian analyses provide only weak or anecdotal support ($BF_{10} = 2.28$; see Table 3.4). The single-face condition had significantly fewer false-alarms compared to the four-face condition. The false-alarm rate for the two-face condition fell in the middle, and was neither significantly greater than the single-face and nor significantly smaller than the four-face condition.

In summary, contrary to expectations but in line with Experiment 4, hit rates in the no-cue condition did not differ across the number of faces. Unlike in
Experiment 4, false-alarm rates in the no-cue condition did vary as a function of the number of faces, producing a positive linear relationship: the more faces presented at encoding, the higher the false-alarm rate. While Bayesian analyses provide only weak support for this difference, there is at least a descriptive trend that the four-face condition, and, to some extent, the two-face condition, began at a disadvantage compared to a participant viewing the same number of individual faces. It is with this baseline that we move on to consider the effects of cuing.

**Replicating and extending effects of cued face recognition.** We again examined whether we could replicate previous effects of cued face recognition, and whether these effects would vary as a function of the number of faces to be encoded. We compared hit rates, and separately compared false-alarm rates, between the two- and four-face conditions given contextual cues using a 2 (condition: two-face, four-face) x 3 (cue type: correct, incorrect) mixed ANOVA (see Table 3.3).

For hit rates, there was no significant interaction between condition and cue type, and the main effects of cue type and face encoding condition were not significant. Thus, neither the number of faces at encoding nor the type of cue presented at test impacted participant hit rates. Bayes analyses provide strong support for the null hypothesis ($BF_{10} = 353.33$; see Table 3.4). For false-alarm rates, the interaction between face encoding condition and cue type was not significant. However, the main effect of cue type was significant and Bayes analyses provide positive or substantial evidence to support this ($BF_{10} = 8.21$). Collapsing across encoding face condition, participants produced more false-alarms when no cue was presented compared to when a correct cue was presented or when an incorrect cue was presented. There was no significant difference between the incorrect cue condition and the correct cue condition. The main effect of face encoding condition
was not significant and Bayes analyses provide positive to substantial evidence in favor of the null ($BF_{01} = 3.75$; see Table 3.4).

Contrary to hypotheses, but again in line with Experiment 4, hit rates were not affected by the number of faces or cue type. However, false-alarm rates were impacted by cue type, such that correct and incorrect cues reduced the false-alarm rate in comparison to when no cue was provided. In considering orienting the effect of cuing, both multi-face encoding groups were at a slight disadvantage compared to the single-face group within the no-cue trials. In some cases, cuing reduced false-alarm rates as low as the single-face group on no-cue trials (see Table 3.1 for comparisons). This suggests that contextual cues, regardless of veracity, might provide a protective buffer to false-alarm rates.

**Sensitivity and response bias.** Again, we computed signal detection statistics to test for effects on sensitivity and response bias ($d'$ and $c$, respectively; see Table 3.5). We used t-tests to compare sensitivity and response bias across the two- and four-face conditions, and repeated measures ANOVAs to compare sensitivity and response bias across the no, correct, and incorrect cue conditions. Neither sensitivity nor response bias changed as a result of number of faces ($BF_{10} \leq 0.35$). Sensitivity did change as a result of cue type such that participants had lower sensitivity (i.e., were less able to discriminate old from new faces) on no-cue trials compared to correct or incorrect cue trials. However, Bayesian evidence constitutes anecdotal or weak evidence for this ($BF_{10} = 1.38$). Response bias also changed as a result of cue type ($BF_{10} = 3.11$). Participants had a negative response bias on no-cue trials (i.e., were more likely to be biased to respond *old*) compared to trials with correct and incorrect cues. Thus, contextual cues, whether correct or incorrect, appeared to increase sensitivity and reduce response bias.
Discussion

Experiment 5 was a replication of Experiment 4, but with the addition of a rating task to strengthen the associative encoding of the multiple faces. Consistent with Experiment 4, hit rates (correct responses to target-present trials) did not change as a function of number of faces or cue type. In fact, hit rates for Experiment 5 are remarkably similar to those in Experiment 4 (see Table 3.1). This suggests that the manipulation for associative encoding did not impact true recognition for previously-studied faces or the possible utility of cuing.

However, false-alarm rates did show a marked departure from the previous experiment. When no cue was presented during testing, false-alarm rates in the four-face condition were significantly higher than in the single-face condition, and false-alarm rates in the two-face condition fell between the two. As opposed to Experiment 4, false-alarm rates in the multi-face conditions were higher when no cue was presented compared to when a correct or incorrect cue was presented with a target face. Furthermore, response bias was reduced any time a cue was presented in comparison to no cue. This is contrary to previous work in cued recognition, where correct cues often raise the false-alarm rate (see Davies, 1988). In our case, providing context, regardless of whether that context was correct or incorrect, appeared to provide a protective buffer against an otherwise higher bias to say that a face was old. This protective quality helped to bring false-alarm rates down nearly to the rate of false-alarms in the single-face condition.

General Discussion

Previous research has shown that when faces are studied in pairs, one of those faces can be used to cue the recognition of the other. Across two experiments, we sought to answer the following three questions in this cued face recognition domain: (1) do
previously-reported effects replicate with updated methodology? (2) if so, do these effects reflect a benefit of correct cuing, a detriment of incorrect cuing, or both? and (3) how do cuing effects vary as a function of the number of cues to be encoded?

Experiment 4 provided little evidence with which to answer these questions in either direction, and it is only in Experiment 5, when we added a manipulation to enhance unitization or associative encoding, that effects arose. While we include Experiment 4 as demonstration of the inconsistency of cued recognition effects and as a caution while interpreting the findings of Experiment 5, we focus primarily on the results of Experiment 5 in this discussion.

**Orienting cuing effects**

First we attempt to orient the effect of using face cues to enhance face recognition. Given that context reduced false-alarm rates (thus increasing recognition accuracy), we wondered if this is a result of the benefit of providing context (cf. single faces have no context), or whether it is beneficial only as a means to compensate for having encoded more faces at once. Therefore we first considered the no-cue trials across the three face groups. Results are mixed. Contrary to hypotheses, false-alarm rates in Experiment 5 statistically differed between groups on the no-cue trials. However, Bayes analyses provide only anecdotal/weak support for this finding. These conflicting results mean it is again difficult to draw strong conclusions from our data. There are a few reasons we may have obtained these weak differences between groups in Experiment 5 but not Experiment 4, and these are discussed in turn.

First, it is possible that this difference between groups could be a result of the varying cue types at testing (i.e., the single-face condition was only ever tested with no cues). However, this explanation seems unlikely given that we did not see this
same pattern in Experiment 4, and because false-alarms were higher for the four-face condition, even though the two-and four-face conditions experienced similar testing conditions. Second, although we attempted to control for divided attention by providing 2 s of encoding time per face, it may be that the associative encoding manipulation affected the allocation of attention when participants were encouraged to engage in unitization. We did not track eye movements, but it is possible, for example, that participants split their time evenly between pairs of faces, but not groups of four faces. However, we would expect such differences in encoding to impact the hit rate, which it did not. Third, it may be that the difference in overall set-sizes of the face stimuli (the combined total number of faces studied across all blocks) affected the false-alarm rates between groups. While participants studied a total of 36 faces within a single study-test block, participants in the two- and four-face conditions completed two and four test blocks, respectively. Therefore, participants in the four-face group were required to remember the most number of faces overall (36 faces multiplied by 4 blocks). Increasing set-sizes does decrease recognition accuracy (see Nortje, Tredoux, & Vredeveldt, in press), but we expected the 5 min break between each test-study block to specifically compensate for this additional memorial demand. Furthermore, if set-size was problematic across conditions, we would have also expected to see this difference in Experiment 4 with accuracy decreasing as the number of faces increased. This was not the case; Not only was there little change in accuracy across face encoding conditions in Experiment 4, the only difference showed an increase in false-alarms for the two-face condition rather than the four-face condition. Fourth, it could be a combination of any of these issues which created this statistically significant, but negligible difference.
A limitation of this study was that it was not possible to control for difference in cuing, set size, and encoding group size all at once. Therefore, it may be more valuable to compare recognition rates when each of these variables is held constant in turn. For example, when participants study single, pairs, and groups of four faces, but are only ever presented with the no-cue trials. This would be a logical next step for future research in order to understand the benefits of cued recognition. What we can say is that something about encoding groups of four faces in our experiment was more difficult for participants than encoding single faces or pairs of faces, even when they encoded the same number of faces in a study-test block. However, this difference was also minimal, with very little support from Bayesian analyses. Therefore, when participants were given contextual cues, the benefits of cuing appear to be a compensation for what was a slightly more difficult task of seeing multiple faces at once.

**Replication of cued face recognition effects**

Next, we consider whether we managed to replicate previous effects in cued face recognition. The two-face condition mimicked the original research in cued face recognition, but our methodology differed in two important ways. First, we randomized the left-right placement of the target faces between study and test, such that placement effects could not exert an influence on participant responses. Second, we randomized pairing of faces so that likeness between the faces (i.e., age, impression of personality, etc.) could not provide clues to correct answers. Following cued recognition and the encoding specificity principles (Thomson & Tulving, 1970), we expected to see hit rates increase when a correct cue was presented compared to either an incorrect cue or no cue at all. However, this was not the case. Hit rates were not affected by cuing, but false-alarm rates decreased in response to correct *and*
incorrect cuing. In line with these results, we also found that sensitivity (in this case, the capacity to distinguish an old from a new face) increased and response bias decreased when any context was shown, whether it was correct or incorrect context.

It is interesting to note that the failure to increase hit rates arises in an experiment when face context is isolated from the possibility of placement context and intuitive impressions of couples belonging together. Admittedly, taking advantage of first impression instincts regarding romantic compatibility, as Winograd and Rivers-Bulkeley (1977) did, may offer an ecologically valid means of grouping faces; This is because we make social judgments about groups that may in turn affect our ability to recognize individual members within them (e.g., Lickel et al., 2000; McGuire & Pezdek, 2016). However, the encoding specificity principle asserts that contextual information is more important than semantic information in cuing (Thomson & Tulving, 1970). Thus, if contextual cuing is useful for face recognition, randomized groupings of faces should not theoretically reduce the effect of correct cuing as long as the faces are encoded as context. Thus we would expect to still see increased hit rates as a result of correct cuing, and reduced hit rates as a result of incorrect cuing or no cuing. By contrast, our hit rates were not impacted. While we still saw an enhancement of accuracy as a result of cuing, it was a result of reduced false-alarm rates instead. Thus, cuing did not enhance the true recognition of an old face, but did enhance the ability to reject a new face that was not previously studied.

Cued recognition studies typically find that correct cues increase false-alarm rates, but not enough to outweigh the benefits of cuing context (see Davies, 1988). However, our results diverge from previous patterns of contextual cuing, such that false-alarm rates actually decreased as a result of cuing. We should note that it is not unusual for a manipulation in memory research to affect hits and false-alarms to
different degrees; For example, context reinstatement in eyewitness identification research often inflates the false-alarm rate rather than the hit rate (Shapiro & Penrod, 1986). However, the current research is closely aligned to the previous experiments in cued face recognition. This raises the question why, within the same field of cued recognition research, cues of faces would not present previously-reported risks of increasing false-alarms and why it would conversely reduce the false recognition of new faces. Because both correct and incorrect context reduced the false-alarm rate, and because sensitivity was minimally affected by cue type, it would be difficult to argue that memory was enhanced as a result of cuing. However, response bias did become more conservative any time context was present, meaning that participants needed more evidence to say that a face was old. This was reflected in the reduced false-alarm rate. However, it is not clear why this response bias became more conservative. One possible reason could be that the presence of a cue with a target, whether correct or incorrect, signaled a different task than the presence of a target alone. Should participants become more suspicious of trials with cues, they may need more evidence to response old than new.

**Extending cuing effects to more than two faces**

Lastly we consider whether such cuing effects extend to contexts in which more than two faces are encoded at the same time. While the four-face condition appeared to be at a slight disadvantage compared to the two-face group when no cue was presented, contextual cuing was equally useful in both groups to reduce the false-alarm rate. Signal detection analyses provided further support for this notion, showing that neither sensitivity nor response bias differed between the two- and four-face groups. Thus, cuing was shown to be effective for both pairs of two and groups of four faces.
Limitations

It is important to consider our results within the limitation that we used identical images at study and test. Some researchers justifiably argue that this ignores the natural variability across representations of a face (see Burton, 2013). Because recognition for unfamiliar faces is fragile to even minute deviations, including, lighting, hair-style, image hue, expression, and focal point of the camera (e.g., Jenkins, White, Van Montfort, & Burton, 2011), it is likely that using the same photographs at study and test results in an easier task and, thus, overestimates eyewitness memory performance for person recognition (Bruce, 1982). Although it is true that this method does not provide the most realistic test of cued face recognition, there is little reason to assume this approach undermines the results presented here.

Research in face recognition using the same images at encoding and test phases has produced similar patterns of results to those tested in the eyewitness paradigm where the faces are always different between the two phases (i.e., confidence-accuracy calibration: Sauer, Brewer, & Weber, 2008; Sauer, Brewer, Zweck, & Weber, 2010, Weber & Brewer, 2004). Interestingly, our results are also in line with Wells and Pozzulo’s (2006) test of the novel two-person serial lineup against the traditional simultaneous and sequential lineup procedures: while there were no differences in lineup procedure for accurate identification decisions for target-present lineups, the two-person serial lineup consistently produced fewer false-identifications in target-absent lineups. Although our task provides a basis for testing associative memory of faces, future research should include a more realistic task (i.e., video-to-photo recognition) to be able to generalize to the more complex real world task of face recognition.

Conclusion
Across two experiments, we sought to replicate previous work that has demonstrated the benefits of cued face recognition for paired faces, to understand those findings in comparison to straightforward single-face recognition, and then extend those findings to situations in which participants study more than two faces. We failed to replicate previous research in cued face recognition with face pairs in the sense that the hit rate for true recognition did not increase when correct cues were available. However, we found that any cue (correct or incorrect) could reduce the false-alarm rate and that these effects extend to those studying groups of four faces, as long as the associative encoding was engaged. Furthermore, we demonstrated that cuing likely compensates for the more-difficult task of studying multiple faces at once (cf. single faces), even when divided attention and memory load are controlled for, and that this occurs by reducing response bias. However, the inconsistent effects reported between Experiments 4 and 5 warrant caution in the utility of such an effect in the applied setting for which this research has recently been adopted (i.e., Dempsey, 2012; Hobson & Wilcock, 2011; Wells & Pozzulo, 2006). In order to apply cued face recognition techniques to eyewitness identification procedures, future work should extend such findings in experimental settings that are incrementally closer to the eyewitness identification context, such as using video-to-picture methodology, and using fewer trials (cf. recognition paradigm).

Our results confirm the utility of using other faces as contextual cues to enhance recognition accuracy. However, our work suggests that accuracy is enhanced (1) by a decrease in false recognition rather than an increase in true recognition (decreased false-alarm rate), and (2) a result of a shift in response bias rather than memorial enhancement. This research was the first replication of original cued face recognition findings using contemporary methodological procedures (i.e.,
randomization of face groups and left-right placement), as well as novel research on extending such cued effects to situations in which there are more target faces presented at the same time.
Chapter 4: Thesis General Discussion

This programme of doctoral research explored underlying issues in memory and decision-making that impact eyewitness identification in the context of multiple perpetrator crimes. In particular, the research focused on two factors: the associations between multiple decisions and the associations between memories for multiple faces. Across five empirical studies and one exploratory survey, this thesis reviewed police practice in three EU countries in the context of multiple suspect identification (Chapter 1: Police Survey), tested the independence of multiple identification decisions made successively (Chapter 2: Experiments 1, 2 and 3), and examined the purported utility of contextual face cues for recognizing the faces of multiple perpetrators (Chapter 3: Experiments 4 and 5). A key aim of the research was to examine concerns in current eyewitness identification procedures for multiple perpetrators. In particular, to explore those factors relevant to previously-unsuccessful attempts to create novel identification procedures adapted to the context of multiple perpetrator crimes. In this discussion, an overview of the key findings are presented, followed by theoretical implications for memory and decision-making, and practical implications for researchers in the subject of multiple perpetrator identification and for police in the field. This is followed by an examination of the limitations of the research presented here, and suggestions for future research.

Summary of findings

The thesis started with a survey-based review of police practices in three EU countries: Sweden, Belgium, and the Netherlands. The survey was conducted to (a) inform our understanding of the prevalence and characteristics of multiple perpetrator crimes from the perspective of law enforcement agencies, (b) to discern how agencies
in various countries conduct identification procedures with multiple perpetrators, and (c) to gain insight into how law enforcement agents and eyewitnesses experience the identification process in the context of such crimes. Results highlighted the practices that are similar between countries as a result of established regulations (i.e., using mostly photographic, sequential lineups), and practices that vary between and within countries in the absence of specific protocols to follow (i.e., whether to put multiple suspects in the same or in different lineups). For the purpose of this thesis, three findings from this survey are particularly relevant. First, there are very few regulations concerning multiple perpetrator crimes, both for a country that required certification for officers to conduct identification tests (NL), and for the countries that did not (SWE, BEL). Second, in the absence of such rules, officers appear to apply a set of similar common-sense extensions of procedures they use for single-perpetrator lineups. For example, most officers avoided telling eyewitnesses which of the multiple perpetrators the suspect lineup was created for, even though there is no rule that expressly forbids this practice. Further, while most officers in the three countries created separate lineups for each suspect of the multiple perpetrator crime, a percentage instead favoured placing all suspects in the same lineup. Third, reported practices make it clear that the association between multiple, separate decisions (i.e., making multiple identification decisions), and the association between the multiple memories of faces (i.e., multiple suspects corresponding to different perpetrators in the same lineup) are critical avenues for future research.

The second chapter of the thesis focused on the association between multiple, separate recognition decisions. Previous work in recognition memory has demonstrated that decisions, although ostensibly separate, are not independent (Malmberg & Annis, 2012). In Experiments 1 and 2, we tested the relationship
between previous identification decisions and current choosing behavior in the context of the multiple showup identification decisions for a multiple perpetrator crime using a mock-eyewitness paradigm. Across experiments, evidence for sequential dependencies for choosing behavior was inconsistent: We could sometimes predict choosing from previous choosing patterns and sometimes not. Experiment 3 examined whether methodological differences between the recognition and eyewitness paradigms used in previous research on sequential dependencies could account for the inconsistent findings presented in Experiments 1 and 2. Experiment 3 therefore sought to replicate previous recognition research in sequential dependencies using word and landscape stimuli, to extend these effects to face stimuli, and to examine whether the strength of these sequential dependencies changed as a result of the number of test trials (i.e., beginning vs. middle of experiment). Sequential dependencies were detected in recognition decisions over many trials, including recognition for faces: Overall, the probability of a yes response on the current recognition trial increased if the response on the previous recognition trial was also yes (vs. no). Although the expected dependencies arose across the many trials, these effects did not mean that the outcome of any individual trial could be reliably predicted by the previous one. Given that sequential dependencies did not notably impact observed choosing behavior for any individual trial, it is unlikely sequential dependencies would have a substantial impact on applied identification procedures or eyewitness choosing patterns across the multiple lineups decisions.

The third chapter of this thesis focused on the association between memories for multiple faces seen at the same time. Experiments 4 and 5 examined the purported utility of associative memory for recognizing the faces of multiple perpetrators, investigated the mechanisms underlying those effects, and sought to determine
whether such effects could include more than two faces. These experiments attempted to replicate previous experiments that paired to-be-encoded faces, and then tested whether recognition performance was enhanced when test faces were paired with the correct face cue (the second face of the encoded pair), an incorrect face cue (a never-before-seen face), or no cue was presented next to the test face. We compared this replication group to two new groups with fewer faces (single faces) and more faces (groups of four faces) presented during study. The correct recognition of previously-studied faces (hits) was not affected by face cuing. However, face cuing of any kind (correct or incorrect cuing) appeared to provide a protective buffer to reduce the rate of false recognition of never-before-seen faces (false-alarms), regardless of whether there were two faces or four faces at study. This appeared to be a result of reduced response bias.

**Theoretical implications**

This research was the first to test face recognition decisions for sequential dependencies, which have been consistently reported for a wide variety of tasks, including emotion categorization and wine taste perception (Hsu & Yang, 2013; Schifferstein & Frijters, 1992). Experiment 3 showed for the first time that sequential dependencies can arise for recognition of images of faces. Although recognition accuracy for faces was lower compared to words and landscape images, sequential dependencies arose in similar patterns as for other stimuli. This is particularly interesting because there are many instances exposing how we encode faces differently compared to other stimuli. In contrast to encoding of objects, for example, it is difficult (if not impossible) for us to encode multiple faces at the same time. Rather, it is likely that we suppress the processing of one face in favor of devoting our attentional resources on the other face (Bindemann, Jenkins, & Burton, 2007). Yet
Despite such differences in encoding, faces are subject to the same recognition dependency patterns over many recognition decisions. This is perhaps less surprising when we consider that most recent models explaining the patterns of sequential dependencies posit interference at the level of perception during testing, not encoding. Thus, it is unlikely to matter how an item was encoded, as long as the item was encoded. And if such models are correct, cognitive processes would be affected as memory is accessed during testing, whether that memory is of items, landscapes, or faces. In summary, while it is more difficult to remember faces, it is one more category of stimuli for which sequential dependencies arise across multiple decisions. Therefore, we could hypothesize that other tasks to measure sequential dependency patterns would likely produce similar results when using faces.

This research was also the first to examine sequential dependencies in the field of eyewitness identification, specifically for multiple perpetrator crimes. However, at the time of writing, there has been one other recent study that considers the effect of making multiple lineup decisions, although not within the theoretical framework of sequential effects (Mansour, Beaudry, & Lindsay, 2017). Mansour and colleagues (2017) asked whether multiple-trial experiments are appropriate for eyewitness research; in particular they focused on whether accuracy, choosing, and confidence rates changed over the course of eyewitness trials. Participants watched 24 videos and made 24 lineup identification decisions on target-present and target-absent trials. They found that the number of trials had no effect or a trivial effect on accuracy, choosing, or confidence. This is in line with our results from Experiments 1 and 2 on making multiple showup identification decisions, for which we found little capacity to predict choosing behavior based on previous choosing. Furthermore, in Experiment 3, we found sequential effects for face recognition overall, but still could not predict
choosing behavior on individual trials. Combined, these results suggest that sequential effects in themselves are not likely to greatly impact making multiple identification decisions in a row in an eyewitness context.

Meanwhile, this research was also the first to test the effects of cued face recognition against non-cued face recognition by having the traditional paired face condition to compare with single-face condition. It was also the first to test cued recognition memory when more than two faces were present at encoding. Our research suggests that cuing may help to overcome the multiple face disadvantage, at least in reducing response bias apparently inflating the false-alarm rate. However, it is also important to note that the cuing effect was small and only effective if the faces are encoded as a meaningful group. Importantly, these effects are present also when more than two faces are present at encoding, meaning that cuing effects may be able to be extended to more complex scenarios. This finding is particularly important to consider given that many crimes are committed by more than two perpetrators, and therefore research should take larger groups of stimuli (i.e., faces) into consideration in constructing and exploring theoretical models for memory and decision-making.

The presented experiments in sequential dependencies and cued face recognition were similar in that they produced unexpected results. The predicted effects did not show up or, if they did, they presented in unexpected ways: The robust effects of sequential dependencies did not arise in the eyewitness paradigm for multiple identification decisions and the cued face recognition effects did not arise in the first attempt. While the cued recognition effects did present in the second attempt when a manipulation was added to increase the strength of associative encoding, even these effects were contrary to expectations; Unlike previous work with cued face recognition, contextual face cues in the presented experiment affected false-alarm rather than hit rates. These
surprising results suggest a number of moderators may be at play, including strength of associative encoding. Current theoretical perspectives need to be expanded through research better elucidating the boundary conditions and moderators of these core effects.

**Practical implications**

The research presented in this thesis is in many ways an attempt to act as counter-weight to common sense. Psychology is often believed to be a common-sense science; indeed, in cases when judges bar eyewitness memory experts from testifying, they often rule that the psychologist cannot offer information outside of general common sense (e.g., *State v. Coley*, 2000). While common sense is a useful heuristic in daily life, it does not always lead to accurate beliefs. And while holding inaccurate common-sense beliefs is not desirable, there are rarely dire consequences in daily life for holding them. What does it matter, for example, if people believe that reading by a dim light ruins eyesight despite no empirical evidence; or if they think that vision involves emitting rays of energy at some point before or during the perception process? And while insomniacs may count sheep to fall asleep, they are not at risk for anything more serious than wasting time (Lilienfeld, Lynn, Ruscio, & Beyerstein, 2010). However, the justice system is a field where common sense beliefs are applied by police officers, lawyers, judges, and jurors throughout investigations and trials. All of these legal decision-makers holding common sense beliefs about eyewitness memory have the capacity to make life-changing decisions for victims, witnesses, suspects, and perpetrators. Unfortunately, many of these beliefs are in conflict with empirical evidence. For example, only 41% of surveyed jury-eligible Americans believe that lineup instructions can impact the accuracy of an identification and only 50% know that eyewitness confidence in that identification is highly susceptible to
outside influences (compared to 98% and 95% of memory experts, respectively; Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006). We find similar disagreements between lay people and experts regarding how the presence of a weapon impacts encoding and how race plays a role in identification (Houston, Hope, Memon, & Read, 2013). This is precisely because there are many aspects of memory that are not common sense.

Although it is important to acknowledge some research seeks to isolate the underlying mechanisms of multiple face disadvantage of identification and recognition— including determining whether eyewitnesses mix up face features between the perpetrators (i.e., Megreya & Bindemann, 2012), or whether the disadvantage stems from memorial capacity, or number of comparisons (Bindemann et al., 2012)—most research in multiple perpetrator identification tests common-sense solutions to identification procedures. Lineups for multiple perpetrators have been presented in different orders and with new instructions (i.e., Hobson & Wilcock, 2011), they have been tested using techniques attempting to control for relative vs. absolute judgments (i.e., Dempsey & Pozzulo, 2008; 2013), and combined lineups to present suspects and fillers for both perpetrators at once (i.e., Wells & Pozzulo, 2006). This research tests common-sense solutions for a problem that is not well understood. Therefore, results for these solutions do not provide answers for why they benefit identification accuracy or, more often, why they fail to do so. The a-theoretical nature of such investigations limits the generality of their findings, and such results are therefore difficult to build upon in future research.

Meanwhile, the most useful research explores the problem of the multiple face disadvantage and tests which variables do or do not contribute to it. In doing so, this research systematically works its way through a checklist of variables to find those
variables we should consider when creating novel lineup techniques or protocols. For example, the current thesis tested for sequential dependencies to determine whether decision-making on multiple lineups is impacted simply because of the task of making multiple decisions. Given that these sequential dependencies for recognition did not translate to predictable observable behavior, it is not fruitful for the field of eyewitness identification to find solutions for a variable that does not appear to be problematic in the first place. An example of a common-sense solution would have been to provide instructions that specifically addressed the idea that the multiple lineups are not connected; if these instructions improved identification accuracy, we would not have known if they were compensating for sequential dependencies or, for example, moving criterion-placement by making respondent more conservative overall; if the instructions had not improved accuracy, we would not have known if they did not appropriately address sequential dependencies or if sequential dependencies were not an issue that needed addressing. By contrast, once we identify and understand the variables that contribute to the multiple face disadvantage, we can test for solutions that might compensate the disadvantages posed by these variables. For example, in this thesis I chose to explore cued face recognition in order to address the previously-tested variable of increased memorial demand (Bindemann et al., 2012). Based on the results provided in Chapter 3, it seems that providing any context of the other suspect may help recognition accuracy. While this work requires replication and further tests particularly to elucidate why this inhibits false-alarms but does not improve hit rates, it is a solution that may be worth the time and resources to pursue.

The fact that we did not replicate cuing effects serves as a caution to research building on older studies that have not been more-recently replicated or more
thoroughly-explored. For example, Dempsey (2014), and Wells and Pozzulo (2006) have applied cuing theory to create two different, novel identification procedures to enhance memory for the multiple perpetrators. One presented the face of only the second suspect next to the entire lineup for the first suspect (Dempsey, 2014); the other presented both lineups for both suspects at the same time, although in a sequential-presentation style (Wells & Pozzulo, 2006). In neither case did cuing improve identification accuracy for the eyewitnesses, and results did not provide clues for why. However, there are two important considerations in mining older theoretical work for newer applied solutions. First, cued face recognition is one example of research that is conducted for one purpose and is later used for another. In early manuscripts of cued face recognition, there is no indication that the authors anticipated that their results would be applied in a practical context such as eyewitness identification. The experiments were designed to test associative memory, not to provide a solution for reduced eyewitness memory in the context of multiple perpetrators. This is certainly not to say that solutions for new problems cannot stem from older research. Rather, that it is important to consider how methodological choices pertinent to the original aims might affect the generality of findings to new domains.

For example, in the Watkins and colleagues (1976) and Winograd and Rivers-Bulkley’s (1977) experiments, incorrect face cues were created by mis-matching pairs of faces, or having participants circle the pairs of faces that correctly matched. This is an entirely different task compared to separating old faces from new ones. Also, in the typical recognition paradigm, participants view hundreds of faces and are tested on hundreds of trials in order to determine overall trends in memory. Even if the cued recognition effect is successful in such a paradigm, it is possible that the effects are
too small to consistently arise when participants are only memorizing a few faces and only tested on a few trials. While the experiments presented in this thesis were designed to explore the underlying mechanisms of the cued face recognition effect, and to extend this to situations beyond pairs of faces, this research line still requires several intermediary steps before we can determine whether it has practical value as a solution to multiple perpetrator lineups or to explain the results when we do apply it. For example, does this effect replicate when using videos of targets instead of photographs during encoding? Does the enhancing effect of context cues for recognition consistently arise when participants encode and are tested on very few faces? These are questions that need to be answered before applying them to lineup solutions.

Second, the original experiments on cued recognition with face cues were conducted in the 1970’s, and have only now been revisited in the last decade as the field of multiple perpetrator identification and multiple person recognition have gained some interest. However, the field of psychology has since experienced a methodological and statistical revolution in the face of a replication crisis (e.g., Simmons, Nelson, & Simonsohn, 2011). In other words, we now understand that some classic effects in psychology may be false positive findings—artifacts of an accumulation of issues like hiding null findings in the file drawer, reporting only significant variables, conducting experiments with small sample sizes, failing to protect against experimenter influence, and failing to adhere to stopping-rules in data collection. Because research cannot take for granted that phenomenon will replicate and generalize, experimenters need to be more programmatic in replicating key phenomenon, and then systematically extending these effects across stimulus types or moderating influences. This is a lesson learned in this own dissertation, first in
Chapter 2, when it became clear it was necessary to step back into the recognition paradigm to understand the boundaries of effects of sequential dependencies (i.e., whether they would arise for face stimuli and to what extent), and again in Chapter 3, when the expected cued recognition effects did not replicate. In both cases, it became particularly important to replicate and understand original effects along with extensions of those effects, for which both lines of research require even further investigation.

Police officials and legal decision-makers are the end-users of accumulated research on multiple perpetrator identifications, and therefore particularly important to consider in applied value of such research. To date, police have been using common-sense adaptations to create and administer multiple perpetrator identification procedures. These decisions are sometimes at odds with each other, such as many officers’ decisions to create separate lineups for the multiple suspects, or some officer’s decisions to create one lineup with multiple suspects. On the one hand, putting multiple suspects in the lineup goes against the common-sense extension of the golden-standard for single-suspect lineups, a standard that is in place because it is considered imperative to decrease the probability of misidentifying an innocent suspect. On the other hand, it is not immediately obvious by common sense that such rules should apply when multiple suspects relate to different people (i.e., a man and a woman of a crime). These are reasonable decisions given that there are unlikely to be guidelines specific to multiple perpetrator crimes and that there is little applied understanding of the difficulties that might arise for either lineup technique.

Unfortunately, the research field is not yet ready to provide practical advice to police officers or policy makers regarding the best methods for testing identification for multiple perpetrator crimes. For one, it appears that there is no obvious harm in
placing suspects for the different perpetrators in separate lineups, since sequential dependencies have no substantial effect when eyewitnesses make multiple decisions. Nevertheless, sequential dependencies are not the only way decisions can be linked. For example, people appear to naturally engage in *probability matching* in order to maximize optimal responding (see Vulkan, 2000). In other words, they use an expectation of base rates to inform current decisions. An example of this is when students, knowing that teachers tend to vary position of correct responses on multiple-choice tests, become suspicious after circling too many A’s. This may be an interesting avenue to explore in making multiple recognition decisions. Furthermore, there is still the question of how Alan Crotzer, the case presented in the beginning of this thesis, was misidentified and wrongly convicted when the two other perpetrators were correctly identified. Just as confidence leaks from one task to another, it could be that confidence in police competency could increase when an eyewitness correctly recognizes the first perpetrator, thus giving them the expectation that police likely also got the next guy. Additionally, our results for cued face recognition suggest it may be advantageous to have the multiple suspects related to different perpetrators in the same lineup, since context may help to reduce the possibility of falsely identifying an innocent suspect. However, our results would need to both be replicated in a recognition paradigm and subsequently tested in the eyewitness identification paradigm while controlling that one of the suspects might get picked by chance before we can recommend such an action. Given the dearth of understanding in eyewitness memory and identification for multiple perpetrators, it would be irresponsible to supply advice to police as to the correct applied protocols. At the moment, we can only provide a greater understanding surrounding the protocol decisions already being made.
Methodological considerations and future directions

This thesis takes a multi-faceted approach to consider eyewitness recognition and identification for multiple perpetrator crimes with studies that investigate police practice, memory, and decision-making. This approach allows us to initiate research in a wide range of research areas. For example, we examined eyewitness decision-making for showup identifications and recognition decisions. Although we conclude that sequential dependencies themselves are not significantly problematic in the field of eyewitness identification, there are other ways in which decisions could be linked. For example, future research should explore whether expectations play a role, such as eyewitness expectations for police competence, and whether suspicion can moderate that effect. For example, the post-identification feedback effect is a robust effect in which feedback from an experimenter or authority figure artificially increases confidence for an identification decision, regardless of the actual accuracy of the eyewitness. However, inducing suspicion about the trustworthiness of the source of feedback helps to eliminate the effects of feedback (Neuschatz, et al., 2007).

Specifically, when an experimenter escorting the participant to a location change revealed that the study was being funded by the Tennessee District Attorney’s office in an attempt to prove the accuracy of eyewitness identifications, and that the experimenter was telling everyone that they had picked the perpetrator from the lineup, participants did not show the typical confidence inflation from the confirming feedback. In such a case, if an eyewitness believes police to be competent and the first lineup they recognize a perpetrator, do they carry those expectations forward to assume the next lineups will also have a perpetrator? Does this change if eyewitnesses are induced to be suspicious to rely on this heuristic? Another context in which decisions might be linked is when more than one suspect is presented in the lineup; In
such a case, eyewitnesses must make decisions for multiple suspects at the same time. On the other side of this topic, it is important to consider how sequential dependencies may impact applied jobs that involve making a large number of recognition decisions. For example, officials at football matches on the lookout for banned hooligans, or border control officers matching people to their passport photos. While all member states of the European Union report violence with football matches, the highest rates of football-related violence have historically been seen in England, Italy, Germany, the Netherlands, and Belgium, with at least 10% of supporters classified as “violent” (Carnibella, et al., 1996). As recent as 2014, the European commission has supported attempts to coordinate football leagues and security services to prevent or contain violence at these matches (European Commission, 2014). Spotters may be given books of fans that have been banned from the game, and a larger number of low- to high-risk fans attending the game to identify. With stadiums of tens of thousands of spectators, these spotters scan the crowd, making a high frequency of rapid recognition decisions. As some regions train specialized teams (i.e., Rotterdam police force) to aid recognition, it may be of interest to understand how sequential dependencies may influence their accuracy across these decisions.

This dissertation also contains experiments designed to examine the link between eyewitness memories by testing cued face recognition. Two studies attempted to replicate and extend previous works on cued face recognition in which a single-face condition was compared to the two- and four-face conditions on those trials that did not use a face cue. The single-face condition was considered a control in order to determine baseline differences in recognition accuracy between the three groups. However, participants in both multiple face conditions also encountered other trials that included correct and incorrect cues, and therefore their testing condition
differs from that of the single-face condition. This difference was minimal because although it was statistically significant, there was very weak support for this difference from Bayesian analyses. Furthermore, this difference did not arise in the first experiment, where testing conditions were the same. However, given that this did arise in the second experiment, it may be a valuable comparison for participants to be presented with only no-cue trials, thus keeping testing conditions constant. This would be a logical next step for future research in order to understand the benefits of cued recognition. Another future direction for this field is the aforementioned incremental steps to conduct cued recognition research with videos of faces, and with very few trials, and, lastly, within the eyewitness paradigm for cued face research.

One difficulty encountered in Experiment 5 within the cued-recognition paradigm was the need to purposefully encourage participants to create a meaningful connection between the presented faces when studying pairs of faces or groups of four faces. Winograd and Rivers-Bulkeley (1977) called this process ‘unitization’, which they implemented in their research by asking participants to rate compatibility between couples (male and female target faces) shown on the screen. This would not be an appropriate manipulation to include because of our additional conditions. This was in part because of the stimulus materials (i.e., we used only male faces) and in part because of the difficulty of asking participants to rate romantic compatibility between four targets in the four-face condition. Furthermore, there is no direct proxy for romantic compatibility for only one face in the single-face condition. Thus, in repeating Experiment 4, Experiment 5 included a judgment task aimed to encourage participants to consider the (fabricated) relationship between first impressions of and the memorability of guilty defendants’ faces and sentencing length. In this task, each pair or group of faces presented on screen during the study phase were described as
presenting perpetrators of a multi-perpetrator crime, who had been convicted by a jury and sentenced to equivalent prison terms. Participants were asked to use their initial impressions of the faces together to estimate how many years the pair or group had been sentenced to prison. This was considered an appropriate manipulation because criminals that commit a crime together are seen as a very closely-linked, also known as highly entitative, group (Lickel et al., 2000). Furthermore, because participants were asked to use their impression of all of the faces in order to estimate a single sentencing length, it was expected this judgment would actively engage in unitized or associative encoding.

However, without eye-tracking equipment, it is not possible to know for certain that participants attended to all available faces when making this judgment. It is possible, for example, that they only used one of the faces to determine sentence length for the whole group. If this is the case, the faces themselves would not be integrated in memory as associated items and cuing would not be useful to enhance access to memory. This might explain why cuing did not improve participant hit rates, for example. Nevertheless, given that Experiment 5 was a direct replication of Experiment 4 except for the addition of this judgment task, this explanation is difficult to reconcile with the change in false-alarm rates between experiments. It is unclear how a judgment task at encoding would lead to a shift of response bias at test when cues are employed.

A limitation of the current experiments in face recognition (Experiments 3, 4 and 5) is that images of faces were used at both study and test. Unlike in the eyewitness paradigm (Experiments 1 and 2), where participants viewed a video of perpetrators and then were asked to make identification decisions from photographs of suspects, participants in the recognition paradigm studied photographs of
individuals and were later tested on the same photographs. Some researchers argue that this methodology is problematic for our understanding of face recognition, because participants may be matching images rather than recognizing faces (see Burton, 2013). The critique is that using images of faces at study and test is ignoring natural variability across representations of a face, meaning that research is measuring the recognition of face images rather than the recognition of people. This is a reasonable assertion: To recognize someone, we should be able to see them from different angles, in different lighting, and know who they are. For example, the musician Bob Dylan is easily recognizable to his fans across thousands of photographs at different ages, in different poses, with different background and perspective and clothing and facial expressions. Recognition for familiar faces, like those of family members, friends, celebrities, is robust across these temporal and contextual changes, but recognition for unfamiliar faces is conversely fragile to even minute deviations, including, but by no means limited to, lighting, hair-style, image hue, expression, and focal point of the camera (see Burton, 2013). This methodological issue is pervasive in recognition and identification research. For the current thesis, the large numbers of face images required to test the hypotheses mitigated against the compilation of a bespoke face database. Further, given differences in available databases that prevent combining faces from different sources (i.e., decisions regarding stimulus, cropping, background, camera angle and distance; equipment such as lighting, cameras, and lenses), it was difficult to obtain the number of faces needed for such experiments (i.e., more than 250 photos in each Experiments 3, 4 and 5, 3) when decisions to control other influencing variables (i.e., gender, age, race) further restrict options for faces. Although it is true that this method does not provide the most realistic test of cued face recognition, there is little reason to assume
this approach undermines the results presented here. Research in face recognition using the same images at encoding and test phases has produced similar patterns of results to those tested in the eyewitness paradigm where the faces are always different between the two phases (i.e., confidence-accuracy calibration: Sauer, Brewer, & Weber, 2008; Sauer, Brewer, Zweck, & Weber, 2010, Weber & Brewer, 2004).

Importantly, the current approach is used as a starting point, which should certainly expand to replicate with realistic variation that more closely resembles real life face recognition.

Even with a multifaceted approach, this dissertation only touched on a few of the many research opportunities present in the field of multiple perpetrator identification. It is important, for example, to explore factors of the encoding event, such as how attention is distributed among members of a group of perpetrators, how salience of the perpetrator (i.e., central vs. peripheral perpetrators) impacts which members are remembered, and the degree of similarity between perpetrators (i.e., gang members that look alike). Another is to explore those factors at test that are unique to multiple perpetrator recognition and identification, such as whether the multiple suspects should be placed in the same lineup or whether eyewitnesses should be instructed to search for specific suspects (i.e., this lineup is for the man that held the gun). We began tapping into only some questions in decision-making, memory, and practice that are relevant to eyewitnesses of multiple perpetrator crimes, but there are more variables at encoding, retention, and recall/recognition that are both unique and inherent to multiple perpetrator crimes.

Conclusion

In five experiments and one survey, this thesis examined underlying issues in memory and decision-making that impact eyewitness identification in the context of
multiple perpetrator crimes. This thesis explored a range of topics related to the recognition and identification of multiple perpetrators in three areas: practice, memory, and decision-making. A survey of police methods provided an initial picture of how they approach identification procedures for multiple perpetrator crimes, how lineups are constructed and presented, what kind of instructions are given, and problems that they experience in conducting such procedures. Responses made clear that there are few regulations for identification procedures specifically for multiple perpetrator crimes and that police vary in their decisions on how to create and administer lineups, decisions that should be tested and advised upon by empirical research. The presented research in sequential dependencies for identification decisions demonstrated that there is little concern for the integrity of identification and recognition decisions to be impacted by making the multiple decisions in a row. Lastly, the presented research showed that cued face recognition may be a useful technique to use, and a promising avenue for future research.
References


doi:10.1080/1068316X.2014.915322


Dempsey, J. L., & Pozzulo, J. D. (2013). Children's identification accuracy of multiple perpetrators: Examining the simultaneous versus elimination line-
doi:10.1080/13218719.2012.679124

doi:10.1037/0096-3445.115.2.107


JASP Team (2017). JASP (Version 0.8.1.2)[Computer software].


State v. Coley, 32 S.W.3d 831 (Tenn. 2000).


The output for this paper was generated using Qualtrics software. Copyright ©2016 Qualtrics. Qualtrics and all other Qualtrics product or service names are
registered trademarks or trademarks of Qualtrics, Provo, UT, USA.

http://www.qualtrics.com


doi:10.3758/BF03199666


doi:10.1016/0022-5371(76)90045-1


doi:10.1023/A:1025750605807


Appendix A - Ethical Approval

The police survey presented in this thesis received ethical approval from the University of Maastricht’s Ethical Committee for Psychology (ECP). The five experiments presented in this thesis were conducted under research lines granted to Dr. Melanie Sauerland by the University of Maastricht’s ECP. These experiments were conducted under Dr. Sauerland’s supervision. Below are the letters of favorable opinion for the police survey and the research lines under which the experiments were conducted.
Het Bestuur van de FPN
Universiteit Maastricht
Postbus 616
6200 MD Maastricht

Ethische Commissie Psychologie

Our reference
direct dial
Maastricht
ECP-160_02_01_2016
043-3884008
1-2-2016

Geachte bestuur,

Na kennisname van het protocol "Eyewitness Identification Procedures for Multiple Perpetrator Crimes: International Police Survey" is de ECP tot het volgende oordeel gekomen.

1. Dit onderzoek valt niet onder de werking van de WMO en behoeft derhalve geen verdere goedkeuring van een Medisch-Ethische Commissie.

2. Gegevens die door de ECP gehanteerde testsituaties aangaan, zijn er geen bezwaren tegen uitvoering van de in het aanmeldingsformulier beschreven onderzoeksproject.

De ECP heeft in haar vergadering van 13 februari 2012 besloten dat op alle advertentieteksten, van door de ECP goedgekeurde onderzoeken, het ECP kenmerk vermeld dient te worden.

Opgezien dient veranderingen in de onderzoeksopzet een hernieuwde kennisname door de ECP noodzakelijk maken.

Wij verzoeken de aanvrager ons kenmerk te vermelden bij toekomstige correspondentie.

Met vriendelijke groet,

Prof. Dr. G. Kok,
Voorzitter

Mr. M. Schrijnemaekers,
Ambtelijk secretaris

Cc. Melanie Sauerland

Maastricht University

ECP
Voorzitter: G. Kok
Ambtelijk secretaris: M. Schrijnemaekers
Ethical Approval for Research Line “Study on Eyewitnesses Decision Processes”

10 This research line was originally granted in 2009 and used for Experiment 1. In 2015, it was amended to include non-Maastricht populations, including residents of the U.K. and the U.S. and was used for Experiments 2.
Hi,

I have added you as a researcher into the system. Your study code (ethics code) which you will enter when entering in your new study is ECP.157-10. Once you create your study in the system click the option to email me so I can activate/approve your study. Your login information has been sent in a separate email from the electronic participant system (Sona). I will return your ethics form via your supervisor’s mailbox.

Please make sure your study adheres to these credit guidelines before you send me a request for approval. Please try and provide accurate information regarding the duration of the experiment to match the credit value.

Length of Study Credit Allocation

- 10 – 30 minutes 0.5
- 31 – 60 minutes 1
- 61 – 90 minutes 1.5
- 91 minutes + 2

Best,
Endre
Ethical Approval for Research Line “Eyewitness identification research line”

This research line was granted in 2016 and was used for Experiments 3, 4, and 5
Appendix B - Police Survey Questions

GENERAL INFORMATION
Gender
☐ Male
☐ Female
☐ Other

Age?

How many years of experience in conducting eyewitness identification procedures do you have?

What is your job role?

Jurisdiction?

THE CRIMINAL OFFENCES

Crimes performed by multiple perpetrators referring to crimes involving two or more people are involved in the commission of the offense (for example, four people who all attack a victim or an accomplice who distracts the victim, while the other bag of the victim grabs.)

An identification of multiple defendants refers to an identification procedure performed for a person suspected of involvement in a crime carried out by multiple perpetrators. One or more suspects for each of the unknown perpetrators who actually committed the crime.

---

12 Questions are marked to differentiate between the original questions from Hobson et al.’s (2012) original survey, and questions that have been altered or added.

a denotes original, unchanged questions
b denotes original, but adapted questions
c denotes new questions added to the survey for this thesis
1) Of the crimes have you dealt in the last 12 months, what proportion involved multiple suspect showings? (Please select the box that applies)

- [ ] 0%
- [ ] 10%
- [ ] 20%
- [ ] 30%
- [ ] 40%
- [ ] 50%
- [ ] 60%
- [ ] 70%
- [ ] 80%
- [ ] 90%
- [ ] 100%

2) How many suspects are typically involved in the multiple perpetrator cases you have dealt with? (Please select the box for the category that applies most often)

- [ ] 2-3
- [ ] 4-5
- [ ] 6-7
- [ ] 8-9
- [ ] 10+

3) In the past 12 months, what types of crimes have you dealt with that typically involve multiple perpetrators? (For each category, please estimate how many cases in the past 12 months you have encountered each involving multiple perpetrators. For listed crimes you have not encountered involving multiple perpetrators, enter 0)

- [ ] Robbery
- [ ] Burglary
- [ ] Assault
- [ ] Sexual Assault
- [ ] Homicide
- [ ] Other (Please provide details)

CURRENT PROCEDURES
All questions are referring specifically to the context of multiple suspect identifications unless otherwise stated.
Scenario 1: Two men (A and B) robbed a bank. An employee witnessed the robbery. Two suspects are arrested, both suspected for being perpetrator A. The suspects are called suspect A1 and suspect A2. You are preparing an identification line up.

4a) Choose the option that resembles what you would do in this case.

□ The eyewitness sees one line up, only for suspect A1 or only for suspect A2, not both
□ The eyewitness sees two line ups, one for suspect A1 and one for suspect A2
□ The eyewitness sees one lineup, with both suspects A1 and A2 in the same lineup

4b) In your work with multiple perpetrator crimes, Scenario 1 occurs:

□ Never
□ Sometimes
□ Often
□ Always

Scenario 2: Two men (A and B) robbed a bank. An employee witnessed the robbery. Two suspects are arrested, one suspect for perpetrator A and one suspect for perpetrator B. The suspects are called suspect A1 and suspect B1. You are preparing an identification line up.

5a) Choose the option that resembles what you would do in this case.

□ The eyewitness sees one line up, only for suspect A or for suspect B, not both
□ The eyewitness gets to see two line ups, one for suspect A and one for suspect B
□ The eyewitness gets to see one lineup, with both suspects A and B in the same lineup
5b) In your work with multiple perpetrator crimes, Scenario 2 occurs:
   - Never
   - Sometimes
   - Often
   - Always

CURRENT PROCEDURES
All questions are referring specifically to the context of multiple suspect identifications unless otherwise stated.

6b) In what manner do you present the parades to witnesses in a multiple suspect identification? Select all the options that apply

Lineups:
   - Live lineup (in person)
   - Photo lineup
   - Video lineup

Choose the format of lineups:
   - simultaneous presentation
   - sequential presentation
   - other

Show-ups
   - Live (in person)
   - Photo show-up
   - Video show-up

7c) Are there any procedural requirements or guidelines in place for multiple suspect identifications? (This may include any national or jurisdictional guidelines, written procedures outlined by the law enforcement agency, or training material used for new officers. Such guidelines may include instructions for live lineups vs photo-arrays or video lineups, number of fillers, selection of fillers, and number of suspects presented at one time. Please provide as much detail as possible.)

   Yes (Please provide as much detail as possible)

   No
CURRENT PROCEDURES

All questions are referring specifically to the context of multiple suspect identifications unless otherwise stated.

8a) How do you organize the identification presentations for eyewitnesses in the case of a multiple-perpetrator crime?

- [ ] A witness will only see the lineups when all suspects / photos / videos are available for the identification procedure
- [ ] The witness sees an identification presentation when a suspect / photo / video becomes available
- [ ] Other ____________________________

9b) What instructions do you give to a witness for multiple perpetrator identifications? (Please provide details of instructions given, and at what point during the identification process these instructions are given and if they differ from single-suspect procedures)

10a) Do you ask the witness to look for a specific suspect? (e.g., identify the one who was driving the car)

- [ ] Yes - If yes, on occasions that you do use specific instructions, in what situations do you do this? (Please provide details of instructions and any relevant examples)
- [ ] No

CURRENT PROCEDURES

All questions are referring specifically to the context of multiple suspect identifications unless otherwise stated.

11a) Do you ask the witness to describe the role of the suspect they are identifying?
CURRENT PROCEDURES
All questions are referring specifically to the context of multiple suspect identifications unless otherwise stated.

12c) Do you record all eyewitness identification decisions in a crime with multiple perpetrators? (Please provide details regarding in what circumstances and how decisions are recorded)

- Yes (e.g., all decisions are recorded, including if the eyewitness identifies a filler or a does not make an identification)
- No (e.g., only positive identifications of suspects are recorded)

13c) Do you record confidence for all suspect identifications for multiple suspect identifications?

- Yes (Please provide as much detail as possible: what instructions are given? How long after the identification is confidence recorded? Is anything said to the witness about their identification before asking for confidence? Do witnesses write down their confidence or respond verbally? Is it recorded in the witness' own words or summarized? Is it recorded using numerical scales or anchors?)
- No
14) Who is responsible for constructing the lineups? Is the same person responsible for all suspect lineups in a given case involving multiple perpetrators?

ISSUES WITH CURRENT PRACTICE
All questions are referring specifically to the context of multiple suspect identifications unless otherwise stated.

15a) Do you, as someone who administers identifications, experience any problems with multiple suspect identifications? (Please describe in as much detail as possible)

16a) Do you think witnesses experience any problems with multiple suspect identifications? (Please describe in as much detail as possible)

PERCEPTIONS OF EYEWITNESSES
17c) How do you think eyewitnesses of a multiple perpetrator crime perform in identifications compared to eyewitnesses of a single perpetrator crime? Generally eyewitnesses to crimes committed by multiple perpetrators are ______ compared with eyewitnesses to crimes committed by a single perpetrator.

☐ Worse
☐ As good as
☐ Better

18c) In your opinion, how useful is a witness for you if they identify one, but not all of the suspects presented?

YOUR SUGGESTIONS
19b) Do you have any ideas of how multiple suspect identifications could be improved from the point of view of the police?
Appendix C - Recruitment E-mail for Police Survey

Dear (name),

I am a PhD researcher Maastricht University (NL), and the University of Portsmouth (UK). I am contacting you because you have previously expressed interest in completing a survey regarding eyewitness identification procedures for multi-perpetrator crimes. The purpose of the research is to understand how police handle multi-perpetrator identifications in practice and to determine what, if any, issues arise in such situations.

I appreciate your interest in completing the questionnaire and I am e-mailing to provide you with the link here:

www.link.com

Please feel free to contact me with any questions or concerns you may have.

Lastly, if you have other colleagues who conduct identification procedures and might be interested in completing this questionnaire, we appreciate you forwarding the link to them as well. The survey is available in Dutch, English, and Swedish.

Best regards,
Nina Tupper

To contacts to distribute survey:
Dear (insert name here),

My name is Nina Tupper, I am a PhD researcher for the House of Legal Psychology—a collaboration between the universities of Maastricht (NL), Portsmouth (UK), and Gothenberg (SWE). My research focuses on eyewitness identification procedures in crimes involving multiple perpetrators and I am currently conducting an international police survey to understand how police handle multi-perpetrator identifications in practice and to determine what, if any, issues arise in such situations.

To give you an idea of our aims and our target-participants, the introduction to the survey is below this email. The study is an online questionnaire, so the link could be easily distributed to anyone willing to take the survey, and is available in English, Dutch, and Swedish.

I am contacting to see if you would be able to help me distribute this survey or point me towards a contact with someone who could. I am happy to provide a list of questions if you need more information. Thank you very much for your time and I look forward to discussing this further with you.

Kind regards,
Nina Tupper

International Police Survey:
Our aim:
The purpose of this survey is to inform our understanding of the prevalence and characteristics of multiple perpetrator crimes from the perspective of law enforcement agencies. We further seek to gain insight regarding how agencies in various countries conduct identification procedures (e.g., lineups, photo-arrays, showups) with multiple perpetrators and how law enforcement agents and eyewitnesses experience the process.

Who should complete this survey:
This survey is intended for law enforcement agents with experience in administering identification procedures. In particular, we are looking for those in the department who are assigned the particular role of identification officer (or the equivalent) or, alternatively, the agents with the most expertise in lineup administration.

The survey:
The survey consists of four main sections to address the aims outlined above. Specifically, the sections ask for a) your estimation of both proportion of crimes you have dealt with and type of crimes you have dealt with that involve multiple perpetrators, b) the current procedures for multiple identifications, c) problems experienced while administering such identification procedures and d) your experiences in your interactions with eyewitnesses viewing such lineups. There is also space at the end for you suggestions regarding if/how to adjust or change current procedures. We appreciate as much detail as you are able to provide, including any examples from your own experience or the experiences of others administering identification procedures.
Appendix D - Information Sheet and Consent Form for Police Survey

International Police Survey – Identification of Multiple Perpetrators

Hello and thank you for taking the time to complete this survey.

Our aim:
The purpose of this survey is to inform our understanding of the prevalence and characteristics of multiple perpetrator crimes from the perspective of law enforcement agencies. We further seek to gain insight regarding how agencies in various countries conduct identification procedures (e.g., lineups, photo-arrays, showups) with multiple perpetrators and how law enforcement agents and eyewitnesses experience the process.

Who should complete this survey:
This survey is intended for law enforcement agents with experience in administering identification procedures. In particular, we are looking for those in the department who are assigned the particular role of identification officer (or the equivalent) or, alternatively, the agents with the most expertise in lineup administration.

The survey:
The survey consists of four main sections to address the aims outlined above. Specifically, the sections ask for a) your estimation of both proportion of crimes you have dealt with and type of crimes you have dealt with that involve multiple perpetrators, b) the current procedures for multiple identifications, c) problems experienced while administering such identification procedures and d) your experiences in your interactions with eyewitnesses viewing such lineups.

Confidentiality:
Your participation is completely voluntary and you may terminate your participation at any time. The findings of the study will be interpreted and reported on a group basis without reference to key individual information. All information about individuals will be held in confidence.

We sincerely appreciate your time and valuable input!

Clicking the button to continue indicates that you agree to these conditions and are ready to begin the survey>>>

If you have questions and/or comments on this research, please contact the researcher, Nina Tupper, or her supervisors:
Nina Tupper (nina.tupper@maastrichtuniversity.nl)
Maastricht University and the University of Portsmouth

Dr. Melanie Sauerland (melanie.sauerland@maastrichtuniversity.nl)
Maastricht University

Prof. Lorraine Hope (lorraine.hope@port.ac.uk)
University of Portsmouth
Appendix E - Debriefing for Police Survey

Thank you for completing this survey on eyewitness identification in the context of multi-perpetrator crimes. The purpose of this survey is to inform our understanding of the prevalence and characteristics of multiple perpetrator crimes from the perspective of law enforcement agencies. We further seek to gain insight regarding how agencies in various countries conduct identification procedures (e.g., lineups, photo-arrays, show-ups) with multiple perpetrators and how law enforcement agents and eyewitnesses experience the process.

Thank you for your participation, your time is very much appreciated. If you have any questions, complaints, or concerns about this research, or if you would like a copy of the results of this research, you may contact the researcher, Nina Tupper, or her supervisors.

Nina Tupper (nina.tupper@maastrichtuniversity.nl)
Maastricht University and the University of Portsmouth

Dr. Melanie Sauerland (melanie.sauerland@maastrichtuniversity.nl)
Maastricht University

Prof. Lorraine Hope (lorraine.hope@port.ac.uk)
University of Portsmouth
Table A.1

Chapter 2, Pilot Study: Mean (standard deviation) Age, Distinctiveness, Memorability, Typicality and Similarity Values for Target Faces and Corresponding Innocent Suspect

<table>
<thead>
<tr>
<th>Perpetrator 1</th>
<th>Innocent Suspect 1</th>
<th>Perpetrator 2</th>
<th>Innocent Suspect 2</th>
<th>Perpetrator 3</th>
<th>Innocent Suspect 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Age</td>
<td>23.45 (1.47)</td>
<td>22.91 (2.76)</td>
<td>22.05 (2.05)</td>
<td>23.27 (2.76)</td>
<td>25.14 (1.64)</td>
</tr>
<tr>
<td>Distinctiveness</td>
<td>2.95 (1.00)</td>
<td>2.82 (1.22)</td>
<td>3.14 (0.94)</td>
<td>2.86 (0.99)</td>
<td>3.77 (0.97)</td>
</tr>
<tr>
<td>Memorability</td>
<td>3.00 (1.07)</td>
<td>2.59 (1.14)</td>
<td>2.00 (1.16)</td>
<td>2.73 (0.83)</td>
<td>3.50 (1.01)</td>
</tr>
<tr>
<td>Typicality</td>
<td>1.72 (0.98)</td>
<td>1.72 (1.12)</td>
<td>1.86 (1.08)</td>
<td>1.64 (1.29)</td>
<td>2.00 (1.02)</td>
</tr>
<tr>
<td>Similarity</td>
<td>2.81 (0.80)</td>
<td>2.45 (0.96)</td>
<td>2.45 (0.96)</td>
<td>2.45 (0.91)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Participants were shown each of the photographs (targets and replacements) individually and were asked to estimate age and to rate distinctiveness and memorability on a five-point scale from 1 (not at all distinctive/memorable) to 5 (extremely distinctive/memorable) and to rate deviation from typicality (How much would this face have to be modified to look completely typical/average?) on a scale from 0 (no modification) to 5. Participants indicated how similar they considered the two faces on a scale from 1 (not at all similar) to 5 (very similar). Innocent suspects were rated as statistically non-different to the perpetrator for the following three factors: memorability, distinctiveness, and deviation from typicality. Innocent suspects 2 and 3 significantly differed in age from their respective perpetrators: Suspect 2: t(21) = 2.73, p = .013; Suspect 3: t(21) = -6.41, p < .001. Perpetrators and their corresponding innocent suspects were also rated for similarity. These tests revealed no significant differences between pairs; ps ≥ .162.
Table A.2

Chapter 2, Experiment 3: Hit Rates (standard error) Given Previous Response as a Function of Stimulus and Testing Section

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Hit Rate</th>
<th>FA Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit</td>
<td>Miss</td>
</tr>
<tr>
<td>Face</td>
<td>.58 (.02)</td>
<td>.45 (.03)</td>
</tr>
<tr>
<td>Place</td>
<td>.53 (.02)</td>
<td>.42 (.02)</td>
</tr>
<tr>
<td>Words</td>
<td>.66 (.02)</td>
<td>.51 (.02)</td>
</tr>
</tbody>
</table>

Note. We examined whether the strength of sequential dependencies changed over the course of the testing session by breaking the two testing blocks into a total of four sections: the first half of the first block (Section 1), the second half of the first block (Section 2), and the first and second halves of the second block (Sections 3 and 4).