

2019

Identifying the masked perpetrator: A new hope

Krista D. Manley
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>



Part of the [Cognitive Psychology Commons](#), and the [Law Commons](#)

Recommended Citation

Manley, Krista D., "Identifying the masked perpetrator: A new hope" (2019). *Graduate Theses and Dissertations*. 17506.
<https://lib.dr.iastate.edu/etd/17506>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Identifying the masked perpetrator: A new hope

by

Krista D. Manley

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Psychology

Program of Study Committee:
(Chun) Jason Chan, Co-major Professor
Gary Wells, Co-major Professor
Shana Carpenter
Max Gyll
Christian Meissner

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Krista Manley, 2019. All rights reserved.

DEDICATION

I dedicate this dissertation to all of the strong-willed women in my life, including the three-legged one—you inspire me every day.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	vi
LIST OF TABLES.....	vii
ACKNOWLEDGEMENTS.....	viii
ABSTRACT	ix
CHAPTER 1. INTRODUCTION.....	1
Transfer-Appropriate Processing.....	4
Transfer-Appropriate Processing and Disguise	6
Partial Concealment and Face Processing	10
Challenges to the Face Recognition Literature.....	14
An Alternative Account of Face Recognition.....	16
Meta-memory and Eyewitness Identification.....	21
Mock Juror Perceptions of Eyewitness Identifications.....	25
CHAPTER 2: EXPERIMENT OVERVIEW	28
CHAPTER 3: EXPERIMENT 1.....	31
Method	31
Participants.....	31
Materials and Design	31
Procedure	33
Results and Discussion	35
Target-Preset Lineups	37
Target-Absent Lineups.....	39
Confidence-Accuracy Characteristic (CAC)	40
Compound Signal Detection Model.....	41
Experiment 1 Discussion	42

CHAPTER 4: EXPERIMENT 2.....	49
Method.....	50
Participants and Design.....	50
Materials and Procedure.....	51
Results and Discussion.....	52
Target-Preset Lineups.....	52
Foil Identifications and Incorrect Rejections.....	52
Target-Absent Lineups.....	54
Lineup Preference and Pre-ID Confidence.....	55
Confidence-Accuracy Characteristic (CAC).....	58
Compound Signal Detection Model.....	59
Experiment 2 Discussion.....	59
CHAPTER 5: COMBINED ANALYSES OF THE DATA IN EXPERIMENTS 1 AND 2.....	66
Target-Present Lineups.....	66
Target-Absent Lineups.....	67
Combined Confidence-Accuracy Characteristic (CAC).....	68
CHAPTER 6: EXPERIMENT 3.....	73
Method.....	74
Participants.....	74
Materials.....	75
Design and Procedure.....	76
Results and Discussion.....	77
CHAPTER 7: GENERAL DISCUSSION.....	82
Identification Credibility.....	92
A Comparison to Previous Work.....	94
Applied Implications.....	95
Limitations and Unanswered Questions.....	99
REFERENCES.....	103
APPENDIX A: EXPERIMENTS 1 AND 2 MATERIALS.....	112
Lineup Instructions.....	112
Bike Theft Video Screenshots.....	112

Example of a Bike Thief Lineup (Full-Face, Target-Absent).....	112
Identity Theft Video Screenshots.....	113
Example of an Identity Thief Lineup (Full-Face, Target-Present)	113
Computer Theft Video Screenshots	114
Example of a Computer Thief Lineup (Masked-Face, Target-Absent).....	114
Wallet Theft Video Screenshots	115
Example of a Wallet Thief Lineup (Masked-Face, Target-Present).....	115
 APPENDIX B: EXPERIMENT 3 MATERIALS.....	 116
Case Vignette 1	116
Case Vignette 2	118
 APPENDIX C: PARTICIPANT DEMOGRAPHICS	 120
 APPENDIX D: PILOT LINEUPS.....	 121
Pilot Lineups for Identify Thief	121
Pilot Lineups for Wallet Thief	121
Pilot Lineups for Bike Thief.....	122
Pilot Lineups for Computer Thief.....	122
 APPENDIX E: PRE-ID VIDEO QUESTIONS	 123
Pre-ID Video Questions.....	123
Preference Question	124
Pre-ID Confidence Questions	124
 APPENDIX G: EXPERIMENT 3 QUESTIONS.....	 125
Lineup Credibility Questions.....	125
 APPENDIX H: IRB APPROVAL.....	 126

LIST OF FIGURES

	Page
Figure 1. Experiment 1 Procedure	30
Figure 2. Experiment 2 Procedure	30
Figure 3. Experiment 1 Target-Present Lineup Identification Proportions	46
Figure 4. Experiment 1 Target-Absent Lineup Identification Proportions.....	47
Figure 5. Experiment 1 CAC Curve and Data	48
Figure 6. Experiment 2 Target-Present Lineup Identification Proportions	62
Figure 7. Experiment 2 Target-Absent Lineup Identification Proportions.....	63
Figure 8. Experiment 2 Pre-ID CAC Curve.....	64
Figure 9. Experiment 2 Post-ID CAC Curve and Data.....	65
Figure 10. Experiments 1 and 2 Post-ID CAC Curve and Data	73

LIST OF TABLES

	Page
Table 1. Lineup Pilot Data	45
Table 2. Experiment 1 Data	46
Table 3. Experiment 2 Data	60
Table 4. Experiment 2 Pre-ID Confidence Ratings and Lineup Preference	61
Table 5. Experiments 1 & 2 Combined Data.....	70
Table 6. Experiments 1 & 2 Confidence Ratings and ID Proportions by Trial	71
Table 7. Experiment 3 Factor Loading Data.....	68
Table 8. Experiment 3 Lineup Credibility Data.....	70

ACKNOWLEDGEMENTS

Many people have greatly impacted my life throughout this journey in ways that have both challenged me and supported my growth as a researcher. Dr. Jason Chan is one of these people. I have deep gratitude and respect for him and want to thank him for his guidance and for continuing to ask me, “What does that mean?” As my advisor he has made me a better researcher, presenter, and mentor. He also made me more resilient and gave me the ability to never back down from Reviewer #2. I also owe Dr. Gary Wells thanks for being a wonderful advisor. His mentoring has had “profound implications” on my work as a researcher and he has made me a better writer. Further, he has changed the way I understand our science, the people it impacts, and of course—golf.

I’d also like to thank my committee members: Dr. Chris Meissner, who was always available for advice, welcomed me into his lab, and taught me the power of tact. Dr. Shana Carpenter, who is an inspiration. She has offered me invaluable advice throughout the years that have helped both my career and understanding of research. And Dr. Max Gyll who was always available for statistics advice, offered helpful questions, and a new perspective. I would also like to thank Dr. Stephanie Madon whose honesty and wisdom I hold in the highest esteem.

The best friends I have ever made in my life were made in graduate school and if it was not for them, this dissertation would have never been finished. Dr. Davis, after over a decade of following your steps, I can say I am truly fortunate to call you my friend. Rachel, believe in you like you believe in me and the rest will be easy. To the rest of my squadron, thank you for being my force. Finally, I want to thank my remarkable partner, Allen Beard. He stood by me through the years of distance, stress, traveling, and uncertainty. I am honored to call him my best friend and myself...Dr. Manley-Beard.

ABSTRACT

During a premeditated crime, perpetrators often wear a disguise such as a ski mask to hinder subsequent identification. Ski masks occlude facial features, which impedes holistic processing—an important component of face recognition (Tanaka & Simonyi, 2016). Despite significant advances in eyewitness identification research, there are no evidence-based recommendations for lineup construction for cases involving a masked perpetrator. The purpose of this dissertation was to examine identification accuracy of a masked perpetrator as a function of lineup type (i.e., full-face or masked-face lineups) and target presence (i.e., target-absent or target-present). Participants completed four trials (one per condition). For each trial, participants watched a mock crime featuring a masked perpetrator then were administered a lineup. I predicted superior identification performance from masked-face lineups compared to full face lineups because masked-face lineups provide a perceptual match to a masked perpetrator whereas a full-face lineup does not. The aim of Experiment 2 was to conceptually replicate Experiment 1 with additional measurements. Participants were asked to indicate their lineup preference then predict their ability to make an accurate lineup decision from either a masked-face lineup or a full-face lineup (pre-ID confidence). In both Experiments 1 and 2, identification of a masked perpetrator was superior when participants were administered a masked-face lineup compared to a full-face lineup. Moreover, participants showed no preference and comparable pre-ID confidence for either lineup on the first trial. However, preference and pre-ID confidence ratings shifted to favor masked-face lineups on subsequent trials. Experiment 3 examined how mock jurors perceive an identification made from a masked-face lineup compared to a full-face lineup. Crucially, mock jurors did not discount masked-face lineup identifications compared to full-face lineup identifications. My data suggest that presenting a lineup containing faces that better match the perceptual appearance of the originally encoded target can enhance identification performance.

CHAPTER 1. INTRODUCTION

Imagine in the middle of a busy workday, a convenience store is robbed by a man wearing a black ski mask that only reveals his eyes. He manages to escape the scene, but the store clerk who witnessed the robbery saw the perpetrator's masked face clearly. Police now have a suspect in custody, but the lead investigator on the case has yet to decide if and how to administer a lineup to the witness. Recent findings suggest that identification accuracy increases when the perceptual appearance of members in a lineup match the perceptual appearance of the original target (e.g., a lineup consisting of masked faces because the perpetrator was wearing a mask) compared to when lineup members do not match (e.g., a lineup consisting of unmasked faces; Manley, Chan, & Wells, 2019). Because the perpetrator was wearing a ski mask, administering a lineup in which members are also wearing ski masks should result in a greater likelihood of a correct identification than a lineup in which their entire face is shown.

Now imagine that the investigator asks the witness if he believes he saw the robber well enough to be able to accurately identify him in a lineup. The witness states that he saw the masked man's face clearly and would be able to recognize him in a lineup. Given the witness's confident claim, the investigator administers a lineup consisting of masked faces to the store clerk. The witness confidently identifies the suspect as the masked robber. However, the case goes to court after the suspect pleads not guilty, and both the investigating officer and the witness testify before a jury. The lead investigator is asked to justify the decision to administer a masked-face lineup to the witness –is there enough empirical evidence to warrant the decision to administer a non-typical masked-face lineup? Further, the witness is questioned and asked if he was given a choice or had a preference for a masked-face lineup over the standard full-face lineup. Of course, the witness was not given a choice, but would he have had a preference for a

specific type of lineup, or even known what effect lineup type would have on his identification accuracy? Further, will the mock jurors trust the witness's identification made from a masked-face lineup to the same extent as an identification made from a full-face lineup? The current dissertation aims to answer these empirical questions and address the profound implications they present for the criminal justice system.

Current policy does not address appropriate lineup procedures in the event of a masked perpetrator, such as the hypothetical case outlined above. Rather, the investigator is responsible for the decision of if and how to administer a lineup. By inspecting past cases involving a masked perpetrator (e.g., *Dupuie v. Egeler*, 1977; *State v. Courteau*, 1983; *State v. Fierro*, 1971; *Williams v. State*, 1979;), it is apparent that there is high variability for this particular type of lineup administration decision. For example, in *State v Courteau* (1983) investigators administered a full-face lineup to both witnesses, who were exposed primarily to the perpetrator's disguised face. In another example, witnesses only saw the perpetrator while he was wearing a ski mask and investigators chose to use a show-up where the suspect was unmasked (*Williams v. State*, 1979). However, in *Dupuie v. Egeler* (1977), after witnesses were not able to make a decisive choice from a full-face lineup, investigators chose to have members of the lineup put on a mask similar to the one used in the crime. Subsequently, witnesses made an identification from the masked-face lineup. Thus, these cases demonstrate that investigators have a highly inconsistent approach when a case involves a masked perpetrator. In consideration of these cases, a lineup in which the lineup members are masked (thus, matching the perceptual appearance of the perpetrator) may benefit identification accuracy. However, it is unclear if the witness would be sensitive to which type of lineup (masked-face or full-face) would be more effective for them to make an accurate identification.

Little research has been done on this type of pre-ID meta-memory judgment (i.e., predictions of future memory performance) in eyewitness identification (for meta-analysis, see Cutler & Penrod, 1989; Nguyen, Abed, & Pezdek, 2018). Further, if the witness does make an identification from a masked-face lineup, it is unclear whether jurors would trust this identification to be credible. Indeed there is evidence that even when mock jurors are given an explicit warning (Report of the Special Master, *State v. Henderson*, 2011; The Innocence Project, 2012a; 2012b) about the detrimental impact that disguise has on identification accuracy (e.g. perpetrator was wearing sunglasses, a hat, or mask), mock jurors do not rate an identification of a disguised perpetrator as less credible than an undisguised perpetrator (Jones, Bergold, Dillon, & Penrod, 2017). However, how jurors perceive the credibility of an identification made specifically from a masked-face lineup has yet to be examined.

Given the evidence presented above, it is clear that more research is needed to examine identification of a masked perpetrator. As such, the current dissertation examined three important questions: 1) Is administering a masked-face lineup more beneficial to identification accuracy than a full-face lineup for a masked perpetrator? 2) If so, are witnesses aware of the benefit to memory that a masked-face lineup provides over a full-face lineup? Specifically, do witnesses have a lineup preference? Further, are witnesses able to accurately assess whether or not they can make an accurate identification decision from a masked-face or a full-face lineup? 3) If a witness was administered a masked-face lineup, would mock jurors find the witness's identification less valid than if it was made from a full-face lineup?

In three experiments, I aimed to answer these important empirical questions using mock crimes featuring a masked perpetrator. The perpetrator was always masked because the focus of this dissertation was on how identification accuracy for a masked perpetrator can be improved,

not made worse. Further, I did not include a condition in which the perpetrator was unmasked because investigators would not be likely to administer a masked-face lineup for an unmasked perpetrator. Moreover, identification conditions that are ideal for an unmasked perpetrator have been extensively examined (Shapiro & Penrod, 1986; Wells, 1993) and are not the focus of the presented experiments. Therefore, in the current dissertation the perpetrator was always masked, and witnesses (participant or hypothetical) were administered either a masked-face lineup or a full-face lineup.

Specifically, the purpose of Experiment 1 was to examine whether identification accuracy would benefit from the administration of a masked-face lineup as opposed to a full-face lineup. Experiment 2 investigated witnesses' preferences and meta-memory judgments regarding the two different types of lineups (masked-face or full-face lineup). Finally, Experiment 3 focused on participant-jurors' trust, perceived accuracy, and credibility of a witness's identification when it was made from a masked-face lineup compared to a full-face lineup. In the following section, the potential theoretical mechanisms that provide the motivation, support for the design, and predictions of this dissertation are discussed.

Transfer-Appropriate Processing

I begin by reviewing the principles of transfer-appropriate processing, the finding that memory transfer and performance are best when the processes active during retrieval match the processes that were active during original learning (Morris, Bransford, & Franks, 1977; Tulving & Thomson, 1973); and consider how these principles apply to the hypothetical scenario outlined in the introduction as well as eyewitness identification in general. One way in which transfer-appropriate processing is applied to investigating memory is through context reinstatement—i.e., reestablishing the mental, environmental, or emotional settings at retrieval that were present

during original learning (Smith & Vela, 2001). A large body of research has established that reinstating environmental or emotional context cues during retrieval improves memory performance (for a meta-analysis see Smith & Vela, 2001; for a qualitative review see Chandler & Fischer, 2006). Of particular importance for present purposes, the benefits of context reinstatement have also been reported in the eyewitness literature (Cutler, Penrod, & Martens, 1987a; Cutler, Penrod, O'Rourke, and Martens 1986; Dalton, 1993; Gibling & Davies, 1988; Krafka & Penrod, 1985; Malpass & Devine, 1981; for meta-analysis see Shapiro & Penrod, 1986; Toseeb, Keeble, & Bryant, 2012).

Shapiro and Penrod (1986) found that reinstating originally-encoded context cues during face identification tasks (combined eyewitness identification and face recognition tasks) significantly improves identification performance by increasing correct identifications ($d = 1.91$) and decreasing false identifications ($d = -0.44$). Moreover, context reinstatement is one of the primary components of the Cognitive Interview (Fisher & Geiselman, 1992; for a meta-analysis, see Memon, Meissner, & Fraser, 2010). When witnesses recall details of a crime after being administered a Cognitive Interview, they tend to provide more details when they are asked to think about the original context, compared to witnesses who are not asked to retrieve the original context (Geiselman, Fisher, MacKinnon, & Holland, 1986). However, "context" is not precisely defined and covers a broad array of processes, environmental factors, and/or stimuli types.

Given Shapiro and Penrod (1986) combined the findings of both eyewitness identification and face recognition tasks into their meta-analysis, it is important to note the distinction between the two paradigms. In a face recognition task, participants are presented with a series of faces, one at a time during encoding. After a retention interval, participants are shown a large array of old and new faces and their task is to respond "yes" or "no" depending on

whether they recognize the face from the encoding phase. Face recognition accuracy is based on the rate of hits (i.e., correct “yes” responses to studied faces), false alarms (i.e., incorrect “yes” responses to unstudied faces), misses (i.e. incorrect “no” responses to studied faces), and correct rejections (i.e., correct “no” responses to unstudied faces).

In an eyewitness identification task, participants typically see one face or target in a mock crime (usually presented via slides or video). Following a retention interval, they are asked to identify the target from a lineup that is either target-present or absent. For target-present lineups, identification accuracy is also based on hits, false alarms and misses. However, here false alarms and misses are generally designated as foil identifications and incorrect rejections. For target-absent lineups, identification accuracy is based on correct rejections and foil identifications. Finally, for face recognition paradigms one participant usually provides multiple data points. Typically, in eyewitness identification paradigms, each participant provides one data point. In Experiments 1 and 2 of the present dissertation, I used an eyewitness identification paradigm to examine how reinstating the *perceptual appearance* between encoding and retrieval may benefit memory performance.

Transfer-Appropriate Processing and Disguise

In regard to eyewitness identification, several factors have been found to contribute to decreased memory performance. One in particular applies to the current dissertation: disguise. In their meta-analysis, Shapiro and Penrod (1986) identified the use of disguise as a significant predictor of decreased identification accuracy ($d = 0.71$ for hits, $d = 0.23$ for false alarms). Even small changes in appearance between encoding and retrieval (i.e., transformations), such as the addition (or subtraction) of eyeglasses or a hat, hair style changes, or facial hair changes have been shown to significantly decrease identification accuracy compared to no changes ($d = 1.05$

for hits, $d = 0.40$ for false alarms; Cutler et al., 1987a; 1987b; Kramer & Ritchie, 2016; Metzger, 2001; Righi, Peissig, & Tarr, 2012; Shapiro & Penrod, 1986; Toseeb et al., 2012).

Mansour et al. (2012) investigated the detrimental impact of different disguises on identification accuracy. Disguises included a hat, sunglasses, or a nylon stocking mask. In Experiment 1, participants viewed targets wearing either sunglasses, a hat, both, or no disguise. They were then asked to try to identify the target from a 6-person lineup. All types of disguise decreased identification accuracy compared to when a perpetrator wore no disguise. However, the largest negative impact on identification accuracy occurred when the perpetrator wore both the sunglasses and a hat. In Experiment 2, participants encoded targets wearing a stocking nylon mask pulled over either 1/3, 2/3, or their full face. Results showed that the nylon stocking mask decreased identification accuracy most when it covered 2/3 of the face or the full face (with no difference between these two groups) compared to when 1/3 of the face was covered or no disguise was worn.

It has been well-established that disguise negatively affects identification accuracy. However, surprisingly few studies have attempted to examine what effect transfer-appropriate processing has on identification when disguise is present—even though it was essentially suggested by Cutler (1988): “One plausible procedure for reducing the effects of variables such as retention interval and disguise is to provide subject-witnesses with contextual cues that might enhance their ability to recognize the target.” (p. 235). For example, if a perpetrator were to wear a disguise such as a ski mask during a crime, would reinstating the perceptual processes for the witness by administering a lineup where its members also wore a ski mask improve identification accuracy? Due to limited research on this topic, it is difficult to make concrete conclusions. However, the studies that have considered transfer-appropriate processing in the case of a

disguise have found positive memory transfer (Davies & Flin ,1984; Foley & Foley, 1998; Hockley, Hemsworth, & Consoli, 1999; Terry, 1994; Manley et al., 2019).

In one study, participants studied the faces of people wearing sunglasses or no sunglasses (Experiment 1), or people who had a beard or no beard (Experiment 2; Terry, 1994). Participants were then given a recognition test that included old and new faces with the original disguise (i.e. sunglasses or beard) either intact or removed. Superior recognition performance was found for transfer-appropriate conditions when the disguise was a beard (i.e., when the studied face had a beard, and the tested face had a beard) compared to transfer-inappropriate conditions (i.e., when the studied face had a beard, and the tested face had no beard). However, this transfer-appropriate pattern was not found for faces wearing sunglasses. For faces wearing sunglasses, performance was best when participants encoded faces wearing sunglasses, regardless of whether the same faces wore sunglasses or no sunglasses at test.

In another study, Hockley, Hemsworth, and Consoli (1999) found evidence for transfer-appropriate processing in a face recognition task using faces that were or were not wearing sunglasses. Participants studied a series of faces, some of which were wearing sunglasses and others were not. If the faces had been shown wearing sunglasses at encoding, participants were better able to accurately recognize studied faces when they wore sunglasses at test, compared to when the sunglasses were removed from the faces at test. The opposite was true for faces not wearing sunglasses at encoding. Recognition performance was best for these faces when they were shown without sunglasses at test. This shows that matching the perceptual appearance at encoding elicits face recognition processes at test that are transfer-appropriate which benefited recognition performance.

In a similar study, participants encoded a series of faces which were either disguised (part of the face was covered with a hat, mustache, sunglasses, or a combination of two disguises) or not (Foley & Foley, 1998). Again, recognition performance was best when the perceptual appearance at test matched what was originally studied, compared to when the faces were altered (i.e., not wearing a disguise or wearing a different disguise). Davies and Flin (1984; Experiment 3) also found evidence of a transfer-appropriate pattern using semi-transparent stocking masks that distorted facial features. Participants were shown four target faces who wore a semi-transparent nylon stocking mask or no mask. Participants were asked to choose the four targets from a lineup of 16 members who were either wearing the nylon masks or were unmasked. Recognition performance again was superior when the target faces in the lineup matched what was originally encoded.

Altogether, these studies suggest that the principles of transfer-appropriate processing can be applied to enhance eyewitness identification of a masked perpetrator. In particular, participants are better able to identify a disguised target in a lineup that provides a perceptual match to the appearance of the target (i.e. a lineup where members wear the disguise seen at encoding) compared to a perceptual mismatch. However, few studies have specifically examined this hypothesis when a perpetrator wears a mask that partially conceals his face (e.g. a ski mask that covers all features except the area around the eyes). Given that this type of mask is common during the commission of crimes (Noble, 2013), it is important to investigate how partial concealment affects eyewitness identification as well as how to improve consequential deficits to identification.

Previous studies have examined the effect of disguises such as hats, nylon stocking masks, eyeglasses, and hair changes (both head and facial). Arguably, these disguises are

fundamentally different forms of concealment than a ski mask. For instance, nylon stocking masks *distort* features, whereas hats do not occlude internal features (i.e., eyes, nose, mouth), and sunglasses cover only the eyes. A ski mask that only reveals the eyes, however, leaves only one important internal feature exposed: the eyes. This makes a ski mask as a disguise distinct because the eyes in particular are an important internal feature for accurate face recognition (Fraser, Craig, & Parker, 1990; Haig, 1986). Further, a ski mask may reduce identification accuracy due to several mechanisms such as reduced exposure to distinctive features (e.g., a mole, large nose, or purple hair) or by decreasing the amount of information available at encoding (Brewer, Weber, & Semmler, 2005; Bruce & Young, 2012). Most importantly, a ski mask may also prevent the face from being processed holistically and hide configural relationships amongst various facial features. The resulting partial occlusion may therefore force featural processing of the face (Farah, Tanaka, & Drain, 1995). Consequently, understanding the mechanism by which partial concealment affects eyewitness identification processing may be informed by the face recognition literature.

Partial Concealment and Face Processing

The dominant theory of face perception posits that faces, unlike non-face objects, are processed holistically rather than featurally (i.e. as a whole unit rather than as individual components; Tanaka & Farah, 1993; for reviews of face processing, see Bruce & Young, 2012; Bartlett, Searcy, & Abdi, 2003; Lampinen, Neuschatz, & Cling, 2012; Maurer, Le Grand, & Mondloch, 2002; Tanaka & Simonyi, 2016). Though typically presented as a dichotomy, many researchers support the notion that faces, objects, and words are processed on a holistic-featural continuum (Farah, 1992; Tanaka & Simonyi, 2016) . Faces are thought to be processed at the holistic end of the continuum. Specifically, they are processed primarily as holistic

representations with little need to decompose a face into individual features. Words are processed at the other end of the continuum—first as individual letter units which are then combined to create words and then sentences. Non-face objects, however, fall somewhere in the middle of the continuum and are processed with varying ratios of holistic and featural representations, depending on the object. Face-like objects (e.g., Greebles) may be processed closer to the holistic end of the continuum (but only with extensive practice, Gauthier & Tarr, 1997), whereas other objects may be processed closer to the featural end of the continuum (e.g., houses; Yin, 1969; Young, Hellawell, & Hay, 1987)

Behavioral research that supports the holistic account of face processing includes findings like the face inversion effect (Yin, 1969), change detection for faces (Wilford & Wells, 2010), the composite face effect (Young et al., 1987; Hole, 1994), the part-whole face effect (Tanaka & Farah, 1993; Tanaka & Sengco, 1997), and whole-face interference (Leder & Carbon, 2005). Each of these effects will be described including how they support the holistic account of face processing in the following section.

The face inversion effect is the finding that face recognition accuracy declines when faces are inverted compared to when they are upright—a decline that is less pronounced for objects (Yin, 1969). Yin concluded that faces are likely processed as detailed, whole representations when upright—a process that is disrupted when faces are inverted, thus decreasing recognition accuracy. The face inversion effect provided initial support for the holistic account of face processing. Further supporting the holistic account Young et al. (1987) found evidence that faces are automatically processed holistically by testing participants with facial “composites.” Composites are created by combining the top half of one face with the bottom half of a different face. Participants are asked to try to recognize *either* the top or bottom

half of the composites when the halves are either aligned or misaligned. When the halves are aligned, holistic processing causes an entirely new face to be perceived. Consequently, participants are unable to ignore the automatic holistic representation. Holistic processing thus slows the recognition speed of one of the halves when they are aligned, compared to when the halves are misaligned. Moreover, the face composite effect has been found with both familiar and unfamiliar faces (Hole, 1994; Young et al., 1987).

A study that investigated an application of the holistic account of face processing implemented a change detection task. Half of the participants were asked to indicate if a part of a face or a house had changed (holistic level; Wilford & Wells, 2010). The other half of participants were asked to identify *which part* of the image had changed (featural level). Participants were better at indicating that something had changed on a face compared to a house. However, participants were better at identifying the specific part that had changed on a house compared to a face. This finding indicates that participants are likely more sensitive to facial changes at the holistic level compared to changes at the localized featural level. Further, research shows that certain features are more important for successful recognition than others. Specifically, several studies have indicated that the eyes in particular are an important internal feature for accurate face recognition (Fraser et al., 1990; Haig, 1986). Specifically, using eye tracking, Henderson, Williams, and Falk (2005) found that when participants encoded a face, a majority of fixation time was spent on the eyes. Therefore, it seems that the eyes can be an important feature for successful face recognition (though not solely necessary, as is evident by the sunglass-wearing face accuracy in Hockley et al., 1999).

In addition to behavioral findings, neuroscience research generally supports the theory that faces are largely processed holistically. Specifically, greater activation is observed in areas

of the brain thought to be responsible for face processing (i.e., middle fusiform gyrus and inferior occipital gyrus) when an individual is presented with aligned face composites, compared to when misaligned face composites are presented (Harris & Aguirre, 2008). An increased response in these areas is also found for upright faces, compared to scrambled faces (Kanwisher, McDermott, & Chun, 1997) and inverted faces (Yovel & Kanwisher, 2004).

Importantly, the holistic account of face recognition provides theoretical foundation for why a masked-face lineup might increase identification accuracy of a masked perpetrator compared to an unmasked lineup. There are several face recognition findings that are particularly relevant to the proposed studies. First, Tanaka and Farah (1993) provided evidence of holistic processing using their part-whole paradigm. In this paradigm, participants encode a whole face and then are asked to recognize individual parts of the learned face (such as a nose) either integrated into a whole face, or as isolated parts. Participants were better at recognizing the old features as part of a whole face than they were at recognizing the features in isolation.

Leder and Carbon (2005) investigated the part-whole paradigm by reversing Tanaka and Farah's paradigm. Specifically, participants encoded *features* in isolation and then attempted to identify each feature integrated into a whole face or as isolated parts. Interestingly, here they found increased recognition performance for a feature in isolation rather than as part of a whole face—an effect they characterized as whole-face interference. The authors concluded that the optimal strategy to successfully recognize an individual feature would be to ignore the holistic representation of the face and rely on featural processing—a task that is likely very difficult because perception of a whole face automatically evokes holistic processing (Leder & Carbon, 2005; Tanaka & Simonyi, 2016; Young et al., 1987). Consequently, participants were unable to ignore the holistic representation when they perceived a whole face, even though the existing

memory trace only included a featural representation. Thus, recognition for the isolated features suffered.

Leder and Carbon's results indicated that when participants view a whole face, they are unlikely to spontaneously engage in featural processing. However, people can invoke featural processing if they are given instructions that make featural processing optimal to complete a task (Wells & Hryciw, 1984). Specifically, participants were given either trait-encoding instructions (i.e. rate the honesty of each face, holistic processing) or feature-encoding instructions (i.e. rate size of the eyes or narrowness of the nose, featural processing) then shown whole faces. Participants who received trait encoding instructions performed better at subsequent face *identification* compared to participants who received feature-encoding instructions. However, participants who were given feature-encoding instructions performed better at face *reconstruction* (building a face from individual features) compared to participants who were given trait-encoding instructions. As applied to the current research question, when participants encode a masked face, potential performance decrements found for identification accuracy from a full-face lineup may be due to a mismatch in the perceptual appearance of the faces. A product of this perceptual appearance mismatch is the resulting mismatch in processing requirements. The holistic processing evoked by the full faces causes whole-face interference (Leder & Carbon, 2005). However, for the masked-face lineups the processing mismatch and thus whole-face interference is not present—leading to increased identification accuracy.

Challenges to the Face Recognition Literature

The holistic account of face processing may be the dominant theory to date; however, it has not been without opposition. Burton (2013) has offered several criticisms of the face recognition literature. First, most studies have been conducted using highly controlled stimuli

(i.e., computer generated or manipulated faces), which might not evoke the same face perception processes as real faces. Second, many of these studies have not taken the role of familiarity into account. Familiarity is an important factor to consider because humans are only face recognition experts for familiar faces (see Burton, Wilson, Cowan, & Bruce, 1999; Hole, George, Eaves, & Rasek, 2002) and perform poorly when they encounter unfamiliar faces as evidenced by reduced recognition accuracy and speed (Bruce, 1986; Burton et al., 1999; Ellis, Shepherd, & Davies 1979; Klatzky & Forrest, 1984). However, unfamiliar faces are also likely processed holistically, though not to the same degree as evidenced by lower recognition accuracy.

Third, within-face variability has been largely ignored by face recognition studies. Within-face variability is defined as the normal, subtle changes (e.g., view, facial expression, exposure or lighting, hairstyle changes, aging, pose, etc.) that can make an unfamiliar face look vastly different from one occasion to another. These within-face changes can render recognition of an unfamiliar face difficult, yet do not affect recognition of a familiar face (Jenkins, White, Van Montfort, & Burton, 2011). Burton argued that within-face variability is higher than between-face variability. Therefore, studies that use the same picture of faces at study and test are conflating image recognition with face recognition. Jenkins et al., (2011) suggested that familiarity exists on a continuum and that it varies within each face depending on how familiar a person is with a particular face. These issues that Burton has raised do not negate the holistic account of face recognition, they simply challenge the way in which face recognition studies have been conducted. Consequently, the current dissertation was designed to avoid each of the three problems that Burton raised.

Although the general consensus among researchers is that holistic processing is not synonymous with configural processing, this third type of processing has been discussed at

length in the face recognition literature. Configural processing is generally understood to be the processing of minute differences in the spatial relations between facial features that make each face unique (also called second-order relations, Diamond & Carey, 1986). Burton, Schweinberger, Jenkins, and Kaufmann (2015) argued against the idea that configural processing is important for familiar face recognition. First, configural processing theories are not well defined nor do they specify which spatial relations are important. Second, as a face moves, the spatial relationships among features can change, but recognition does not decrease. Even when configurations are drastically changed, recognition of familiar faces does not change (Hole et al., 2002). This point is further evidenced by a person's caricature (features in a face are exaggerated) sometimes being easier to identify than the actual person (Rhodes, Byatt, Tremewan, & Kennedy, 1997). Third, recognition of familiar faces can be impaired by non-configural changes such as a photographic negative (Galper, 1970). Finally, when the common shape is removed from faces, leaving only texture information (e.g., color, reflectance), face recognition is not impaired. Whereas Burton and colleagues have presented ample evidence to challenge the configural processing account of face recognition, they note that they do not hold similar evidence against the holistic account (Burton et al., 2015, p. 493).

An Alternative Account of Face Recognition

Alternative to the holistic account of face recognition is Valentine's (2001) Face-Space Model of face recognition. Here faces are represented in a multidimensional space where small changes in the configuration of the face can have a large impact on the appearance. Within the face space, faces that are average looking (do not stand out in a crowd) are near the center. As faces become more distinctive (defined by the speed in which a face is picked out in a crowd), they fall further from the center of the face space. Further, faces that are more similar to each

other will be located closer together in the face space. In contrast, the more dissimilar faces are, the further apart they will be in the face space. As such, typical faces will be recognized as faces faster than distinctive faces. However, distinctive faces will be recognized as specific faces faster, because they are further from other faces, thereby reducing noise and confusion.

Valentine describes how recognition versus identification fit into the model by indicating that face recognition is a judgment that a stimulus has been previously encountered, whereas identification requires a judgment of which specific stimulus has been previously encountered. This model is also not without limitations. Importantly, it cannot account for view changes or within-face variability except by claiming there is a separate face space for every view of every face. This idea is meant to account for the delay in recognition when a person is exposed to a previously unseen angle of a previously seen face, as well as changes in view, lighting, and aging.

Assuming a view change includes changes in appearance due to disguise, the face-space model might apply to the current experiments in two ways. First, when participants encode the masked face of a previously unencountered individual, the decrement seen in identification accuracy when participants are administered a full-face lineup might be because the participant does not have a face space developed for the full-view version of the face. However, since they do have a face-space for the masked-face view, they should be able to accurately identify the masked target from a masked-face lineup. However, occlusion of features is not specifically mentioned in Valentine's explanation of the Face-Space Model, so a masked face may or may not be included in view changes.

It is not the aim of the current dissertation to solve the debate between the face recognition models, nor is it the purpose to directly test them. Instead, it is important to note how

the predictions of the current dissertation were borne out of these theories. Primarily, my predictions were based on the holistic account of face processing. Specifically, that faces are automatically processed as primarily holistic representations. For an unmasked perpetrator, a perceptual match would require a full-face lineup because both the encoded face and lineup faces elicit holistic processing. However, when a participant encoded a masked face (that showed only the perpetrator's eyes), featural processing would likely be induced. Therefore, a full-face lineup would be a perceptual mismatch relative to the participant's memory. The full-face lineup would lead the participant to automatically invoke holistic processing, resulting in whole-face interference and decreasing a participant's ability to successfully use featural processing to remember the masked perpetrator (Leder & Carbon, 2005; similar to the face composite effect, Young et al., 1987). Therefore, to potentially capitalize on the featural processing used during the encoding of a masked perpetrator, the lineup should also require featural processing. Using computer generated faces, Manley et al. (2019) applied this theory to an face identification paradigm. The authors predicted that identification accuracy would be best when the perceptual appearance of individuals in a lineup matched rather than mismatched the perceptual appearance of the individual that was originally encoded.

In Experiment 1, participants encoded a target face which was either a full face or a partial face (i.e., a face that was cropped so that a rectangle showing only the eyes was presented). After a 5-min retention interval, participants attempted to identify the target from a three-person, target-present lineup consisting of partial faces or full faces. Results revealed a transfer-appropriate pattern where identification performance was superior when the perceptual appearance of the lineup members matched the encoded target (e.g., encoded a partial face, retrieved from a partial-face lineup) compared to when it was mismatched (e.g., encoded a

partial face, retrieved from a full-face lineup). Importantly, the pattern found in confidence ratings further supported this pattern.

Given these promising results, the authors altered the stimuli in Experiment 2 to reflect more ecologically valid conditions. Specifically, the partial faces were replaced with faces wearing a ski mask that only revealed the eyes. Otherwise, the design and procedure were identical. The results of Experiment 2 conceptually replicated the results of Experiment 1 using masked faces (including confidence data). For Experiment 3, target-absent lineups and the option to choose “Not Present” were added to the design. Participants were shown either an unmasked face or a face wearing a ski mask and were asked to make an identification decision from either a target-absent or target-present lineup that consisted of unmasked faces or faces wearing masks. Importantly, the transfer appropriate patterns replicated for target-present lineups even with the option to reject the lineup. For target-absent lineups, regardless of whether participants had encoded a masked face or a full face, they correctly rejected the masked-face lineup more often than the full-face lineup.

Experiment 4 was designed to find whether the masked-face lineup superiority found for the masked target was due to whole-face interference (processing mismatch) or an alternative account. Specifically, identifying the masked face in a full-face lineup might be more difficult due to the additional irrelevant features included in the full face. That is, the full-face lineups require the participant to compare not only three different sets of eyes, but also noses, lips, and brows (i.e., noise). The masked-face lineup holds all of these irrelevant features constant because they are covered. As such, the detriment to identification performance from a full-face lineup might be a product of feature overload. Therefore, the features of all of the faces used in Experiment 4 were held constant with the exception of the eyes. This was to equate the number

of comparisons that participants would be required to make between the features of the masked faces and the full faces. In all other respects, Experiment 4 was identical to Experiment 3. Results replicated Experiment 3, showing masked-face lineup superiority for a masked target—evidence for a processing-match account.

In conclusion the data from Manley et al. (2019) supported the predictions outlined above. Specifically, when participants encoded a masked face, they identified the target more often from a masked-face lineup than from a full-face lineup. Conversely, when participants encoded a full face, they identified the target less often from a masked-face lineup than from a full-face lineup. This identification pattern was consistently found across four experiments using computer-generated faces. Similarly, confidence patterns paralleled identification patterns. However, this study was not without the limitations that Burton (2013) outlined. The primary issue is that the picture used at encoding and in the target-present lineup were identical. Given this limitation, the results may only apply to picture memory and not eyewitness identification. To test this possibility, the purpose of Experiment 1 was to apply the theories and logic from Manley et al. to an ecologically valid paradigm using new materials (e.g. mock crimes and real faces instead of computer generated target and filler faces).

Given the promising results of Manley et al., one may wonder what a witness might think if they were administered a masked-face lineup. Would a witness think it was a more difficult task and prefer a full-face lineup? Is what a witness believes about her own memory important? Further, how well a witness believes she remembers a perpetrator may change how likely she is to identify an individual in a lineup, testify, or even report what she saw to police. Therefore, it is important to consider what a witness believes about her ability to identify the target in a masked-

face lineup compared to a full-face lineup. I now turn to the meta-memory literature to explore this issue.

Meta-memory and Eyewitness Identification

What people know and believe about their own memory is known as meta-memory (for review see Dunlosky & Thiede, 2013). The ways in which people monitor and control their memory is important to consider in reference to eyewitness identification. This is because what witnesses believe about the quality of their memory will likely influence whether they choose someone in a lineup or not. There are two broad classifications of meta-memory judgments—pre-identification (pre-ID) judgements (i.e., predictions of future memory performance), which are administered prior to a memory test, and post-identification (post-ID) judgments (i.e., assessment of past memory performance), which are administered after a test has been completed (Dunlosky & Thiede, 2013; Chua, Schacter, & Sperling, 2008). The eyewitness literature is replete with studies examining post-ID judgments in the form of confidence ratings due to the serious implications these rating have for perceived identification accuracy. Specifically, mock jurors tend to weigh the reported witness confidence ratings heavily when deciding if an eyewitness’s identification is accurate (Cutler, Penrod, & Dexter, 1990; Cutler, Penrod, & Stuve, 1989; Wells, Lindsay, & Ferguson, 1979). Typically, participant witnesses make an identification decision, then rate how confident they are that they made a correct decision (Juslin, Olsson, & Winman, 1996; for review see Leippe & Eisenstadt, 2007; Wixted & Wells, 2017).

Although the value of confidence-accuracy relationship has been hotly debated in the past, recently Wixted and Wells (2017) have agreed that under “pristine” circumstances high confidence ratings are positively related to accuracy. According to the authors, for these conditions to be met, a confidence rating must be made immediately after the identification

decision and only their initial rating should count. Further, this rating should be uncontaminated by outside influences—including feedback from the administrator. Additional conditions designate that confidence ratings are only pristine when given for a fair lineup, and when lineups are administered in a blind fashion to prevent intentional or unintentional influences on the decision. Confidence judgments made under these pristine circumstances denote a witness's ability to monitor retrieval and thus, are typically highly related to identification accuracy (Wixted & Wells, 2017).

Similar to post-ID judgments, pre-ID judgments could have potentially serious implications. For example, investigators might ask a witness if she would recognize the perpetrator if she were to see him again. If the witness is overconfident in her own memory for the perpetrator's face, the witness may incorrectly identify an innocent suspect. In addition, overconfidence may lead investigators to place unfounded trust in the witness and thus not search for another witness who may have had a better view of the crime. Alternatively, if the witness is underconfident in her memory, investigators might not administer a lineup at all. How eyewitnesses monitor their learning, especially in the case of perpetrator identification, is uniquely important because witnesses cannot go back and re-study to-be-retrieved information. Despite the importance of the above issues, the eyewitness literature has scarcely addressed this type of meta-memory judgment. In the eyewitness literature, the terms “pre-confidence”, “predictive confidence”, or “prospective confidence” ratings have all been used to describe how witnesses monitor and predict their learning prior to an identification task.

In a meta-analysis of nine studies, Cutler and Penrod (1989) concluded that “pre-confidence” judgments are weakly associated with identification accuracy and should not be used when determining whether to administer a lineup to a witness. Several studies since have

also found that people are not able to accurately predict their performance on face recognition and eyewitness identification tasks (Clark & Tunnicliff, 2001; Nguyen et al., 2018; Olsson & Juslin, 1999; Perfect, 2004; Sommer, Heinz, Leuthold, & Schweinberger, 1995). Using a face recognition paradigm, one study found that participants were only able to accurately predict their face recognition performance when the faces were distinctive (Sommer et al., 1995).

Hourihan, Benjamin, and Liu (2012), however, concluded that participants' abilities to make accurate pre-ID confidence judgments are limited to recognition of own-race faces.

Other researchers have found evidence that people can accurately predict their own face recognition performance. Further, these judgments of accuracy are not necessarily contingent upon perceptual properties like distinctiveness (Busey, Tunnicliff, Loftus, & Loftus, 2000). Using an eyewitness identification paradigm, Clark and Tunnicliff (2001) found that participant witnesses were unable to make accurate pre-ID confidence judgments (i.e., they did not perform as well as they predicted they would during the identification task). Alternatively, Nguyen et al. (2018) found that delayed pre-ID confidence ratings were more indicative of accuracy than immediate pre-ID confidence ratings. Yet, even at the highest levels of pre-ID confidence, accuracy was still fairly low ($M = 0.64$, chance accuracy = 0.50). This low pre-ID confidence-accuracy relationship may be attributable, at least partly, to the fact that pre-ID confidence is assessed before witnesses know about task difficulty.

Given that the evidence is not strongly conclusive for whether witnesses are able to successfully monitor and predict how well they learned a face, Experiment 2 included both pre-ID and post-ID meta-memory questions. These questions were aimed at assessing lineup preference given a masked perpetrator. Specifically, I was interested in understanding if and why witnesses may prefer a masked-face over a full-face lineup or vice versa. Pre-ID confidence for

both a masked-face lineup and full-face lineup was also evaluated. Are witnesses' lineup preferences reflected in their meta-memory judgments and are those preferences reflective of actual performance? Is there a level of pre-ID confidence where it is apparent that witnesses prefer one type of lineup over the other? That is, participants who prefer the masked-face lineups over the full-face lineup should have higher pre-ID confidence ratings for identifications made from a masked-face lineup compared to a full-face lineup. Higher pre-ID confidence in this case may occur because the participant understands a match in perceptual appearance between encoding and retrieval supports memory transfer. Alternatively, participants might find the identification task from a masked-face lineup easier than one from a full-face lineup without knowing why. As such, participants may not have a preference until they gain experience making an identification decision from a full-face or masked-face lineup.

In Experiment 2, participants witnessed a mock crime in which the perpetrator is wearing a mask that revealed only his eyes. Next, participants were asked if they would like to see a lineup of faces or a lineup of faces wearing masks. Participants were then asked to rate how likely it is that they would be able to identify the perpetrator in a lineup consisting of faces wearing masks or a lineup consisting of unmasked faces. Finally, participants were administered a lineup then asked to rate their confidence that they made a correct identification decision.

Because the face recognition and eyewitness identification literatures are mixed in regard to pre-ID confidence, my predictions for this study were borne out of the testing effect literature. Evidence from the testing effect literature shows that taking a test enables participants to make more accurate judgments of learning (JOLs) on subsequent trials (Adesope, Trevisan, & Sundararajan, 2017; Finn & Metcalfe, 2014; Koriat, Sheffer, & Ma'ayan, 2002). Given participants completed four trials in Experiment 2, a more accurate meta-memory judgment

could be observed after the first identification trial. However, each identification was made from one of four unique crime scenarios using different perpetrators, so retrieval during a lineup may not have the same calibrating effect on identification decisions as is found in educational contexts.

Not only is it important to consider what witnesses believe about their own memory ability when administered a masked-face lineup compared to a full-face lineup, but it is also important to consider how mock jurors might perceive the credibility of that witness's identification. When an identification of a masked perpetrator is made from a masked-face lineup compared to a full-face lineup, would a mock juror find one identification less credible than the other based on lineup type? In the next section, I will briefly review evidence showing how mock jurors perceive information that affects eyewitness identification.

Mock Juror Perceptions of Eyewitness Identifications

Decades of eyewitness research has established that eyewitness identifications are highly persuasive to jury members, even though eyewitnesses are often incorrect and subject to overconfidence and misinformation (Wells, 1993). In fact, misidentification from eyewitnesses has been found to be the leading cause (72%) of wrongful convictions (Innocence Project, 2018). Given that the variables affecting the accuracy of eyewitness identification have been extensively researched, it is important to examine whether mock jurors understand the factors that decrease identification accuracy.

Research shows that participant-jurors' perceptions of eyewitness identification accuracy are only minimally affected by variables known to substantially decrease identification accuracy (Cutler et al., 1989)—including disguise of the perpetrator (Shapiro & Penrod, 1986). In two studies, mock jurors watched a video of a trial that involved an eyewitness identification. During

the trial, 20 identification factors were discussed during a cross examination of two witnesses. Ten of the identification factors were manipulated (two levels) and ten were held constant (always presented). For example, disguise of the perpetrator, weapon presence, and violence towards the victim were either present or absent from the witnesses' testimony. The primary measure was whether participants rated the defendant guilty or not guilty and the participants' estimate of the probability that the witness's identification was correct or not. Using eligible jurors from the general population (as opposed to the student sample used in Cutler et al., 1989), Cutler et al., (1990) found that mock jurors were not sensitive to several variables that negatively affect identification accuracy (e.g., disguise, biased lineup instructions, similarity of lineup members, and construction and administration of lineups). In both papers, Cutler and colleagues (1989; 1990) concluded that the only variable that mock jurors were able to use in deciding whether to trust an identification, was eyewitnesses' *post-ID* confidence ratings, which they tend to rely on too heavily.

Fortunately, some courts recognize that mock jurors tend to be insensitive to variables that impact eyewitness identification negatively and consequently offer mock jurors either explicit instructions (e.g., Henderson Instructions¹) or expert testimony to increase understanding. However, mock jurors do not take these instructional warnings into account when judging the credibility of a witness statement (Dillon, Jones, Bergold, Hui, & Penrod, 2017; Jones et al., 2017). These findings are concerning and alternative methods to educate jurors should be explored. Further, no research to date has examined whether jurors would find an

¹ Disguises and changes in facial features can affect a witness' ability to remember and identify a perpetrator. The Special Master found that “[d]isguises (e.g., hats, sunglasses, masks) are confounding to witnesses and reduce the accuracy of identifications.” According to the State, those findings are “so well-known that criminals employ them in their work.”

identification of a masked perpetrator made from masked-face lineup more or less valid than those made from full-face lineup.

Experiment 3 aimed to examine this question by using mock case briefings in which the perpetrator was always masked. After reading a case vignette, participants were shown the type of lineup the witness was administered, the identification decision the witness made, and the confidence rating (or confidence rating was absent) that the witness made at the time of the identification. Given previous research has shown that mock jurors do not appropriately discount identifications made under difficult conditions (e.g., disguised perpetrator; Cutler et al., 1989; Cutler et al., 1990), it is unlikely that an identification made from a masked-face lineup would be differentially judged compared to an identification made from a full-face lineup. However, I predicted that participants who are given the witness's confidence rating will likely take these ratings into account when rating the identification's credibility. Specifically, participants might employ an "anchoring and adjustment" strategy (Slavic & Lichtenstein, 1971; Tversky & Kahneman, 1974), where they use the implied value of witness confidence (e.g., pretty sure) as an anchor point, then slightly adjust their own identification credibility rating based on that confidence value. When participants are not provided with a confidence rating from the witness, they may need to rely on their own perception of the difficulty of the identification task. Since they are not making the identification themselves, they may not have a way to assess difficulty. Therefore, participants will likely rate an identification from a full-face lineup similarly to an identification made from a masked-face lineup.

CHAPTER 2: EXPERIMENT OVERVIEW

Experiment 1 used a 2 (Lineup: masked-face lineup, full-face lineup) x 2 (Target: absent, present) within-subjects design. Participants were asked to make four total identifications in which they completed one identification trial from each of the four conditions (masked-face target-present lineup, masked-face target-absent lineup, full-face target-present lineup, full-face target-absent lineup). The order of the trials was presented in a random order for each participant and they were not informed which type of lineup they would see, or that they would be completing four trials.

In Experiment 1, each of the four trials consisted of three phases (see Figure 1). First, participants watched a mock crime in which the perpetrator wore a ski mask that only revealed his eyes (Mock Crime Phase). Second, in the Pre-ID Phase participants watched an unrelated pre-identification video (retention interval). Third, in the Lineup Phase participants were asked to make an identification from a six-person lineup (masked-face or full-face, target-absent or target-present). Throughout the study, participants had the option to choose “not present” for each lineup. If participants chose the “Not Present” option, the same lineup (containing the same lineup members in different positions) was re-administered in which participants were forced to choose (to equate criterion). After the Lineup Phase was complete, the next trial began with a new mock crime video (see Appendix A for examples of materials).

Experiment 2 employed the same design as Experiment 1 (see Figure 2), except that participants answered three pre-ID meta-memory questions between the Pre-ID Phase and the Lineup Phase. They were first asked a preference question about what type of lineup they would prefer given the choice (a masked-face or a full-face lineup). Next, two pre-ID meta-memory questions were asked in a random order about their predicted identification performance, one for

a full-face lineup and one for a masked-face lineup. Finally, they were administered a lineup and asked to make an identification decision.

Experiment 3 used a 2 (Lineup: masked-face lineup, full-face lineup) x 3 (Witness Confidence Rating: low, high, none) mixed design with lineup manipulated within-subjects and witness confidence rating manipulated between-subjects. Here, participants read two case vignettes, one at a time. Each case vignette described a crime in which the perpetrator wore a ski mask and a witness subsequently identified a suspect in a lineup (see Appendix B). In one, the witness identifies a suspect from a masked-face lineup. In the other, the witness identifies a suspect from a full-face lineup. In addition, the hypothetical identifications were accompanied by a verbal confidence rating (high or low), or it included no confidence rating. Participants then rated the credibility and accuracy of a witness' identification based on whether it was made from a masked-face lineup or a full-face lineup.

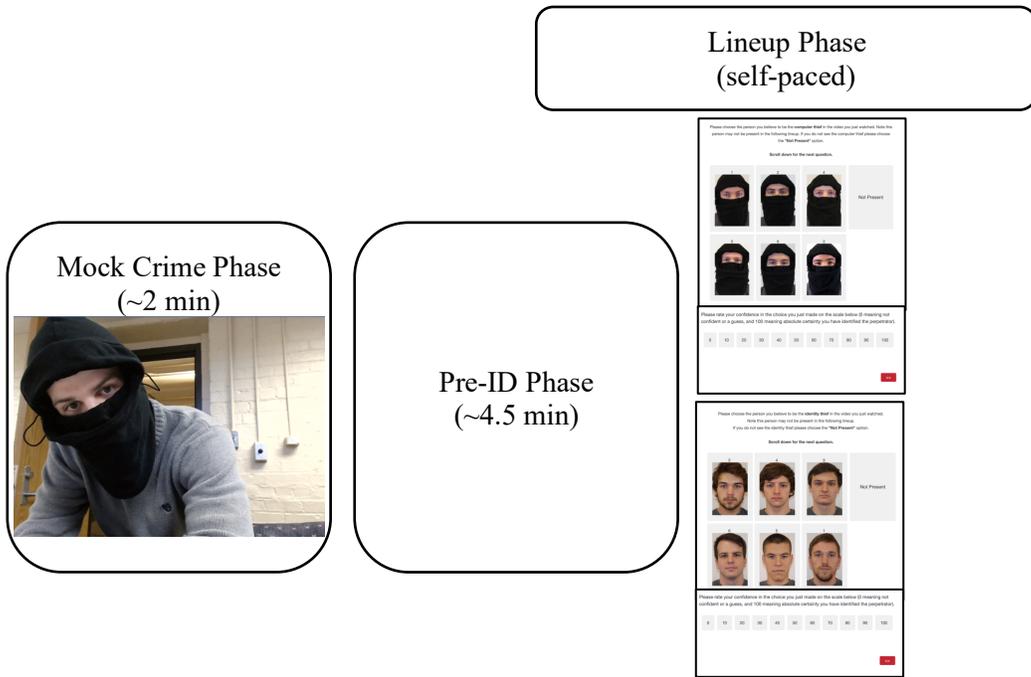


Figure 1. Depiction of a single trial in the Experiment 1 procedure. Each participant completed four trials, one for each of the four conditions.

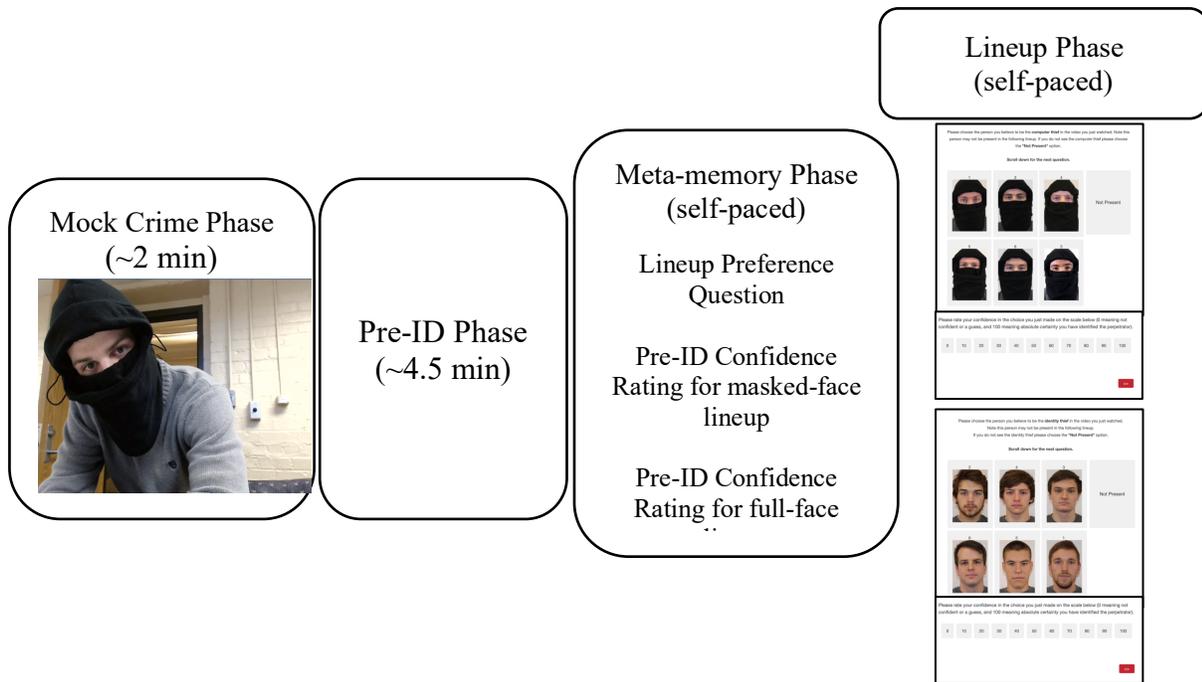


Figure 2. Depiction of a single trial in the Experiment 1 procedure. Each participant completed four trials, one for each of the four conditions.

CHAPTER 3: EXPERIMENT 1

Method

Participants

Participants were 314 undergraduate students from Iowa State University who participated in exchange for course credit. Four participants were excluded from analyses due to video playback problems. Therefore, 310 participants were included in final data analyses. Data from 214 participants were collected in the lab and data from 96 participants were collected online through the university SONA system. Participant demographics including age ($M = 19$, $SD = 1.36$), ethnicity, and gender were collected and can be found in Appendix C.

The sample size was based on an exact proportions McNemar power analysis which used the most conservative odds ratios and discordant pairs from Manley et al. (2019; Experiments 3 and 4). Specifically, the predicted necessary sample size ($n = 282$) was computed using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) with power set at .85, a two-tailed value of $\alpha = .05$, discordant pairs = .53, and an $OR = 1.65$. Data for all three experiments were collected with the approval of the Iowa State University Institutional Review Board (IRB #15-663).

Materials and Design

Four mock crime videos were used, which depicted four different crime scenarios with similar themes (i.e., theft of an identity and money from a study room, theft of a bike from a university bike rack, theft of a wallet and money from a car on campus, and theft of a desktop computer from a conference room). Each crime portrayed a different Caucasian perpetrator ($M_{Age} = 20$) who was wearing a black ski mask that only showed the area around his eyes (see Appendix A). The duration of each mock crime video was approximately equal ($M = 122.25$ s, $Range = 119$ s-125 s). In addition, each masked perpetrator was on screen for approximately

equal amounts of time (total time $M = 77.25$ s; close-up $M = 25.25$ s). The videos were counterbalanced such that each video served equally as often in each of the four conditions (masked-face target-present lineup, masked-face target-absent lineup, full-face target-present lineup, full-face target-absent lineup).

Lineups consisted of six total faces and a “Not Present” option (see Appendix A). They were constructed by selecting 8-12 faces from a large database of photos taken of Iowa State students that matched the physical description of each of the targets. Initial pilot testing was completed to select six filler faces from the 12 faces (see Appendix D). For pilot testing, I used 156 participants from Amazon mechanical-turk who were monetarily compensated for their time (\$1.10 for 10 min total). Participants watched each of the four videos and immediately after each video attempted to choose the target from either a full-face ($n = 81$) or masked-face ($n = 75$), target-absent lineup. Six faces were selected from the middle of the distribution which was based on the combined average from each lineup type (i.e., masked-face identifications and full-face identifications; for descriptive statistics see Table 1). Statistically, the average choosing rates of fillers from the full-face lineups compared to the masked-face lineups were comparable, $ps > .185$. I chose to implement the above method in lieu of a simpler method in which fillers were chosen based on the description of the perpetrator. This was due to the challenge of creating a more detailed description of a masked face (e.g., “he wore a mask and had blue eyes”).

Four retention interval videos were used to serve as *pre-identification* distractor videos. Each of these videos was approximately equal in length ($M = 4:30$ min; Range = 4:26 min – 4:31 min). The pre-identification videos were dialogue-free short films that contained music. Each video was randomly assigned to each trial and was shown only once per participant. Qualtrics was used to program and counterbalance each condition. Participants were randomly assigned to

each counterbalance. A 2 (lineup: masked-face, full-face) x 2 (target: absent, present) within-subjects design was implemented to examine identification rates and confidence ratings.

Procedure

The experiment began with bot checks to ensure humans were completing the online version of the experiment. Additionally, participants were instructed to only complete the experiment on a computer (rather than a mobile device), to eliminate environmental distractors, shut their phones off, and complete the study in one sitting (approximately 45 min). The program collected data to ensure participants were completing the experiment on a computer and not a mobile device (e.g., browser type and version, screen resolution).

All participants began the experiment with the mock crime phase (see Figure 1). They each read the following instructions, “You will watch several short videos, one at a time that depict various thefts. Please pay attention to each of the perpetrators as you may be asked to identify that individual later.” Depending on which video participants viewed, they then read instructions directing them to pay attention to the specific thief (e.g., “You will now watch a short video that depicts the theft of a computer. Please pay attention to the **computer thief** as your memory may be tested later.”).

Next, participants were instructed to immediately put the video on full screen and then watched a video of a mock crime featuring a masked perpetrator. Participants could not move on from the video until 130 s had passed. However, the program continued automatically after 180 s. After each video, an attention check question was asked to measure the time that elapsed between the end of the video and the next phase (e.g., Question: What is six minus two—In word form? Answer: four).

Participants then completed the retention interval and Pre-ID phase where they viewed a ~4.5 min pre-identification video followed by two cued-recall questions about the video (attention check). These questions were the same for every pre-identification video to simplify counterbalancing, but the order in which participants received the questions was randomized (see Appendix E). Afterwards, participants proceeded to the Lineup Phase. First, participants read, “On the next page you will see a lineup. Please choose the person you believe to be the thief in the video you just watched. Note this person may not be present in the following lineup. If you do not see the thief, please choose the "Not Present" option”. These instructions were repeated on the next page where participants were administered a six-person, target-absent or target-present, masked- or full-face lineup. Participants indicated their decision by using the mouse and clicking on their choice.

Once participants made their lineup decision, they were asked to rate their confidence in their decision by selecting a number on a scale of 0 to 100 (0 indicating not sure at all or a guess, 100 indicating absolute certainty (see Appendix E). Finally, they pressed an arrow button (synonymous with a “next” button) to proceed from the lineup and confidence phases. If participants selected the “Not Present” option, they received a forced choice lineup from the same condition as the original lineup with the instructions, “You chose the "Not Present" option. Now as if you were forced to make a choice, please choose the individual who most closely resembles the thief.” Participants were given an unlimited amount of time for their lineup decisions and confidence ratings.

This process repeated until four total trials, each representing a different condition, were completed (approximately 45 min total). The order of the conditions was counterbalanced, and participants were randomly assigned to each counterbalance by the program. Importantly,

participants were not told that they would see four mock crimes or make four identifications. Further, they did not receive any feedback, and they did not know which type of lineup they would receive in advance. Finally, participants were given a short demographics questionnaire, were asked if they knew any of the actors or lineup members, debriefed, and dismissed.

Results and Discussion

Due to the difference in the accuracy measurement (target identifications versus correct rejections), accuracy data from target-absent and target-present lineups were analyzed separately using Cochran's Q (an extension of McNemar χ^2 for related samples; Field, 2009). Data are reported starting with target-present lineups, followed by target-absent lineups, Confidence-Accuracy Characteristic curves, and a compound signal detection theory (SDT-CD) analysis. The stimuli and data for each experiment can be found on the Open Science Framework (OSF; https://osf.io/v6pn2/?view_only=f7f12d30b459446ca7267ff4431eda05).

Confidence-Accuracy Characteristic (CAC) curves were calculated for each condition to examine how suspect (both guilty and innocent) identification accuracy varies at each level of confidence (Wixted and Wells (2017), using the following equation:

$$\text{Suspect ID accuracy} = 100\% \times nSID_{TP-c} / (nSID_{TP-c} + (nFID_{TA-c}/n))$$

Here, $nSID_{TP-c}$ is the number of target identifications made from a target-present lineup (i.e., guilty-suspect IDs) at each level of confidence c . Individual confidence ratings (11 options) were collapsed into four levels (0-40, 50-60, 70-80, 90-100), then adjusted so there were approximately equal cases in each group (see Figures 5, 8, and 9). Additionally, $nFID_{TA-c}$ is the number of foil identifications made from target-absent lineups (i.e., innocent-suspect IDs) at the same level of confidence. Because there is no designated innocent suspect in mock-crime studies such as the one presented here, $nFID_{TA-c}$ is divided by the lineup size in the target-absent lineup

($n = 6$ in the current study) to estimate the innocent suspect frequencies. Suspect identification accuracy (minimum 50%) is then plotted at each level of confidence to show a CAC curve for each condition. This equation is used to examine suspect identification accuracy in an applied setting. As such, foil identifications from a target-present lineup do not contribute to the equation because these would be known innocents and thus, not at risk for wrongful conviction. Instead, I am simply interested in suspect identification accuracy at each level of confidence for masked-face lineups compared to full-face lineups. Simply stated, the CAC is guilty suspect choices \div (guilty suspect choices + (innocent suspect choices \div number of options in the TA lineups)) within each defined level of confidence.

In addition, a compound signal detection model (SDT-CD; see Palmer, Brewer, & Weber, 2010; Palmer & Brewer, 2012) was used to examine whether the lineup conditions had an affect on discriminability. The SDT-CD model analyzes recognition performance for compound decisions. These decisions consist of both detection (i.e. the target is present among an array of stimuli; correctly determining yes for all target-present lineups and no for target-absent lineups) and identification (i.e. the target is correctly identified). SDT-CD accounts for all possible response probabilities (target identifications, foil identifications, and incorrect rejections for target-present lineups; correct rejections for target-absent lineups) which are used to calculate a single d' and response bias (c) value for each condition. Response bias values can range from conservative (i.e., negative) to liberal (i.e., positive).

The integration decision rule was selected when computing discrimination (d') and response bias (c). This rule states that decisions are based on global assessment of the array (i.e., a comparison of the stimulus array against the decision criterion). The alternative to the integration decision rule is the independent decision rule which posits that stimulus is separately

compared to the decision criterion (Palmer & Brewer, 2012). For Experiments 1 and 2, I reported estimates based on the integration decision rule because it best fit the current data ($G^2_{Total} < 5.81$, $ps > .12$). Typically, the integration decision rule fits eyewitness identification data better than the independent decision rule (Palmer & Brewer, 2012). Alpha levels for all analyses were set at .05.

Target-Preset Lineups

Target Identifications

For target identifications, I predicted that participants would be more likely to accurately identify the target from a masked-face lineup compared to a full-face lineup. To assess this prediction, I used a Cochran's Q χ^2 test which revealed a significant difference in target identifications as a function of lineup type, $\chi^2(310) = 6.72$, $p = .010$, $OR = 1.80$, $CI^{95}[1.14, 2.84]$ (see Figure 3). Specifically, participants were significantly more likely to make a target identification from a masked-face lineup ($M = .58$, $SD = .49$) compared to a full-face lineup ($M = .48$, $SD = .50$). This finding is consistent with my predictions. Further, it supports the idea that when a masked perpetrator is involved, identification accuracy increases when a witnesses is provided with a perceptual match at lineup rather than a perceptual mismatch. Finally, participant location (lab, online) did not interact with lineup type or change the pattern of results, $F(1, 616) = 1.21$, $p = .271$.

Foil Identifications and Incorrect Rejections

Separate χ^2 tests investigated differences in foil identifications and incorrect rejections made from masked-face lineups compared to full-face lineups. Here, participants made foil identifications significantly more often from a full-face lineup ($M = .36$, $SD = .48$) compared to a masked-face lineup ($M = .28$, $SD = .45$), $\chi^2(310) = 6.15$, $p = .013$, $OR = 0.62$, $CI^{95}[0.44, 0.78]$.

Lastly, participants made incorrect rejections at a relatively similar rate from a masked-face lineup ($M = .14$, $SD = .35$) compared to a full-face lineup ($M = .15$, $SD = .36$), $\chi^2 < 1$, $p = .637$ (see Figure 3).

Confidence

Linear Mixed Models (LMM) were used to examine differences in confidence ratings associated with target identifications, foil identifications, and incorrect rejections based on lineup conditions (See Table 2). First, an LMM analysis was used to compare target ID confidence ratings made from masked-face lineups compared to full-face lineups. Participants rated the confidence of their target identifications (i.e., correct IDs) from masked-face lineups ($M = 73.74\%$, $SD = 20.85$) significantly higher than those made from full-face lineups ($M = 68.52\%$, $SD = 22.76$), $F(1, 326) = 4.69$, $p = .031$, $d = 0.23$. Gamma correlations were calculated between confidence ratings and target identifications. For both full-face lineups ($r = .29$, $p < .001$) and masked-face lineups ($r = .32$, $p < .001$) confidence was significantly related to target identifications.

Participants also rated their confidence for foil identifications higher from masked-face lineups ($M = 62.64\%$, $SD = 18.52$) compared to full-face lineups ($M = 56.46\%$, $SD = 22.52$), $F(1, 198) = 4.31$, $p = .039$, $d = 0.28$. Confidence ratings for incorrect rejections from masked-face lineups ($M = 64.77\%$, $SD = 26.63$) and full-face lineups ($M = 61.88\%$, $SD = 24.12$) were comparable, $F(1, 90) = .30$, $p = .585$, $d = 0.10$.

An additional 2 (lineup: masked-face, full-face) x 2 (accuracy: correct, incorrect) LMM indicated that participants rated their confidence higher for target identifications ($M = 71.37\%$, $SD = 21.86$) compared to incorrect identifications (i.e. foil IDs and incorrect rejections) ($M = 60.45\%$, $SD = 22.49$) across both lineup types, $F(1, 616) = 34.17$, $p < .001$, $d = 0.47$. Foil IDs and incorrect rejections were combined as “incorrect IDs” because there were very few incorrect

rejections. This analysis also revealed a main effect of lineup type, $F(1, 616) = 8.69, p = .003, d = 0.25$ such that participants were more confident in identifications made from masked-face lineups ($M = 69.35\%, SD = 21.71$) compared to full-face lineups ($M = 63.10\%, SD = 23.47$). The interaction was not significant, $F < 1, p = .986$.

Forced-Choice Lineup Identifications

In addition to the initial lineup decisions, forced-choice identifications made by participants who initially chose the “Not Present” option from the target-present lineups were investigated. Specifically, I looked at the second identification hits and added them to hits from the initial lineup. The purpose was twofold: First, the forced-choice lineup essentially held response criterion constant by not allowing participants to reject the lineup or choose no one; second, this decision was to simulate suboptimal lineup instructions where the witness is discouraged from choosing no one. For masked-face lineups, target identifications increased by 6% ($M = .58$ to $M = .64, SD = .48$). This increase for full-face lineups was slightly higher at 8% ($M = .48$ to $M = .56, SD = .50$). However, even with this change in hit rates, participants still made significantly more target identifications from masked-face lineups compared to full-face lineups, $\chi^2(310) = 5.00, p = .025, OR = 1.50, CI^{95}[1.26, 3.23]$.

Target-Absent Lineups

Correct Rejections and Foil Identifications

Similar to target-present data, I analyzed response accuracy using Cochran’s $Q \chi^2$. For target-absent lineups, an accurate response is to reject the lineup by choosing the “Not Present” option. Numerically, participants correctly rejected full-face lineups ($M = .31, SD = .46$) at a slightly higher rate than masked-face lineups ($M = .25, SD = .43$; see Figure 4). However, this effect did not reach significance, $\chi^2(310) = 3.14, p = .076, OR = 0.72, CI^{95}[0.54, 1.09]$. In a

target-absent lineup, foil identifications are the numerical inverse of correct rejections (e.g., correct rejections + foil identifications = 1), and therefore were not analyzed separately. Finally, participant location (lab, online) did not interact with lineup type or change the pattern of results, $F(1, 616) = .89, p = .345$.

Confidence

For target-absent lineups (see Table 2), confidence ratings for correct rejections made from masked-face lineups ($M = 64.42\%$, $SD = 25.42$) and full-face lineups ($M = 63.96\%$, $SD = 25.15$) were statistically similar, $F < 1, p = .906, d = 0.02$. Gamma correlations between confidence rating and correct rejections revealed a small positive relationship for full-face lineups ($r = .17, p = .037$). However, this relationship was not significant for masked-face lineups ($r = .05, p = .570$).

Similar to target-present lineups, participants were significantly more confident in foil identifications made from masked-face lineups ($M = 63.43\%$, $SD = 20.47$) compared to full-face lineups ($M = 57.34\%$, $SD = 21.84$), $F(1, 445) = 9.28, p = .002, d = 0.25$. In addition, a 2 (lineup: masked-face, full-face) x 2 (accuracy: correct, incorrect) LMM indicated the main effect of accuracy was not significant, $F(1, 616) = 3.57, p = .059, d = 0.15$. Participant rated their confidence in correct rejections ($M = 64.16\%$, $SD = 25.20$) similarly to their confidence in foil identifications ($M = 60.51\%$, $SD = 21.33$). The main effect of lineup and interaction were not significant, $F_s < 2.65, p_s > .104$.

Confidence-Accuracy Characteristic (CAC)

Confidence-Accuracy Characteristic (CAC) curves were calculated (Wixted & Wells, 2017) to investigate suspect identification accuracy at each level of confidence for both masked-face lineups and full-face lineups. Figure 5 shows CAC curves for participants who made a

suspect identification (both guilty and innocent) from masked-face lineups compared to full-face lineups. Because there were too few observations at each of the eleven confidence levels, I reduced the confidence levels from 11 options down to 4 categories which contained roughly an equal number of responses (see Figure 5). The four confidence levels correspond to 0-40, 50-60, 70-80, and 90-100.

As shown in Figure 5, the CAC curves had a positive slope, such that the probability of suspect identification accuracy was higher at high levels of confidence compared to low levels of confidence. Although masked- and full-face lineups produced mostly overlapping CAC curves (i.e., lineup type did not influence CAC curve), participants made more high-confidence guilty-suspect identifications from masked-face lineups (*total* = 57) compared to full-face lineups (*total* = 43). This difference is important and indicates that there are more high quality identifications made from the masked-face lineups compared to the full-face lineups. However, participants also made more innocent suspect identifications from masked-face lineups (*total* = 6) compared to full-face lineups (*total* = 3). That is a 4.5% increase in guilty suspect identifications from a masked-face lineup compared to a full-face lineup, but only a 1% increase in innocent suspect identifications.

Compound Signal Detection Model

A compound signal-detection model (SDT-CD; see Palmer, Brewer, & Weber, 2010; Palmer & Brewer, 2012) was used to compute discrimination (d') and response bias (c). To make inferences, G^2 (similar to χ^2) for the model where d' (or c) was constrained is subtracted from the G^2 for the model where d' was permitted to differ. For example, for d' the G^2 values were equal to 10.4 for the fixed model and 4.35 for the model in which d' was allowed to vary, respectively. The resulting value, 6.05 is then compared to the critical value for $\chi^2(1) = 3.84$. Given 6.05

exceeds the critical value, I can infer that restricting the two lineup conditions to a single d' value significantly impaired model fit. As such, it is unlikely that the two lineup conditions have the same d' value.

Results for d' and c showed that discriminability was higher when participants were administered a masked-face lineup ($d' = 1.55$, $SE = .09$), compared to a full-face lineup ($d' = 1.32$, $SE = .07$; $G^2(1) = 6.05$, $p = .014$; see Table 2). However, response bias was similar and conservative leaning (i.e., negative value) for masked-face lineups ($c = -0.59$, $SE = -.03$) and full-face lineups ($c = -0.49$, $SE = -.03$; $G^2(1) = 1.68$, $p = .195$).

Experiment 1 Discussion

Results from Experiment 1 indicated that identification accuracy for a masked perpetrator increased when the administered lineup provided a perceptual match to the originally encoded face compared to when it did not. Note that these promising results were found only for target-present lineups, as performance was comparable for masked-face and full-face lineups when the target was absent. Intriguingly, participants rated their confidence in identifications made from masked-face lineups higher than identifications made from full-face lineups regardless of accuracy.

Further, participants were more confident in target identifications made from masked-face lineups compared to full face lineups. This finding was reflected in the higher number of high confidence target IDs made from masked-face lineups than full-face lineups. This pattern shows the benefit of matching the perceptual appearance of lineup members to the originally encoded target. The higher confidence ratings for IDs made from masked-face lineups compared to full-face lineups may be an indication that participants have insight into the difficulty of the task. Therefore, it is important to explore what participants believe about their memory in the

case of a masked perpetrator. Further, how lineup preference changes across the trials may reveal valuable insight into differences in the lineup tasks.

Table 1*Average choosing rates of fillers as a function of lineup type in the pilot experiment*

	Masked-Face Lineup	Full-Face Lineup
Identity Thief	.08 (.04)	.11 (.03)
Wallet Thief	.11 (.08)	.10 (.08)
Bike Thief	.10 (.08)	.10 (.09)
Computer Thief	.12 (.05)	.10 (.03)

Note: Values in parentheses represent standard deviations.

Table 2

Identification proportions, confidence ratings, discriminability and response bias for identification decisions in Experiment 1

	Masked-Face Lineup	Full-Face Lineup
Target-Present Lineups		
Identification Proportions		
Target IDs	.58 (.49)*	.48 (.50)
Foil IDs	.28 (.45)*	.36 (.48)
Incorrect Rejections	.14 (.35)	.15 (.36)
Confidence Ratings		
Target IDs	73.74 (20.85)*	68.52 (22.76)
Foil IDs	62.64 (18.52)*	56.46 (22.52)
Incorrect Rejections	64.77 (26.63)	61.88 (24.12)
Target-Absent Lineups		
Identification Proportions		
Correct Rejections	.25 (.43)	.31 (.46)
Foil IDs	.75 (.43)	.69 (.46)
Confidence Ratings		
Correct Rejections	64.42 (25.42)	63.96 (25.25)
Foil IDs	63.43 (20.47)*	57.34 (21.84)
d'	1.55 (.09)*	1.32 (.07)
c	-0.59 (-.03)	-0.49 (-.03)

Note: Values in parentheses represent standard deviations except in the cases of d' and c , for which the parenthetical values are standard errors. An asterisk indicates a significant difference between lineup types.

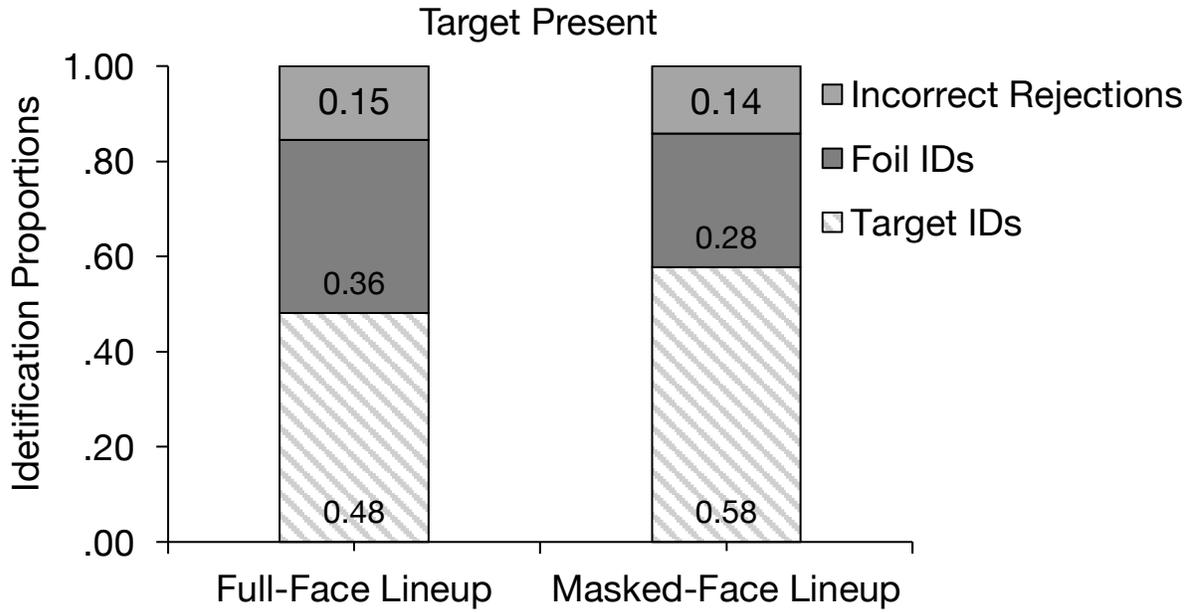


Figure 3. Experiment 1 target-present lineup identification proportions as a function of lineup type.

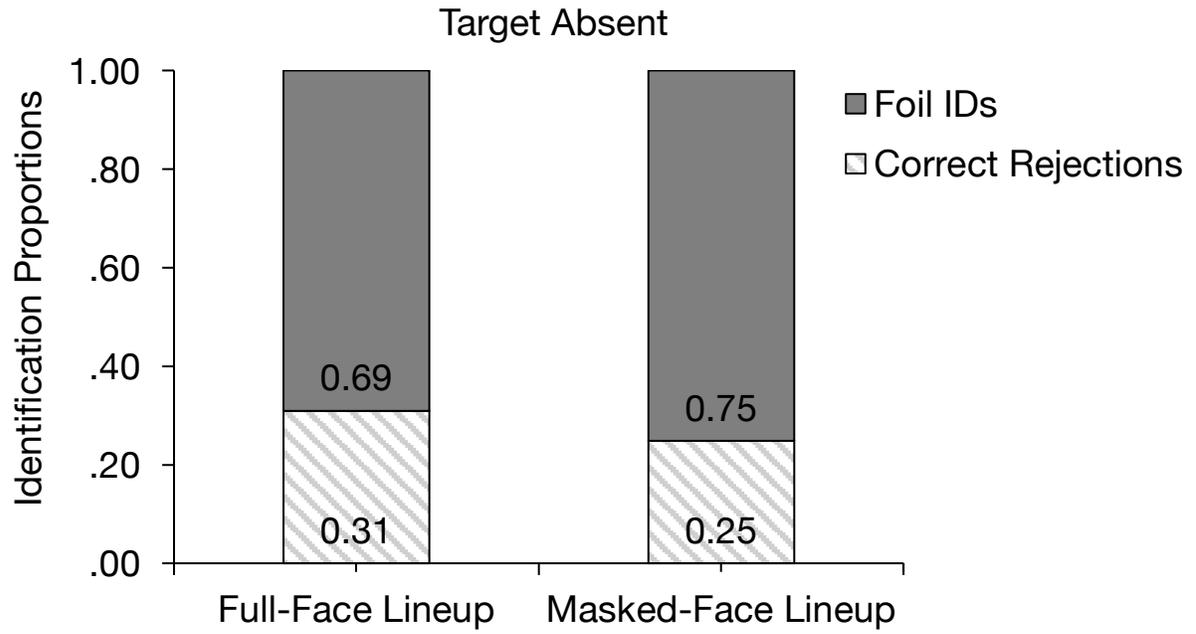


Figure 4. Experiment 1 target-absent lineup identification proportions as a function of lineup type.

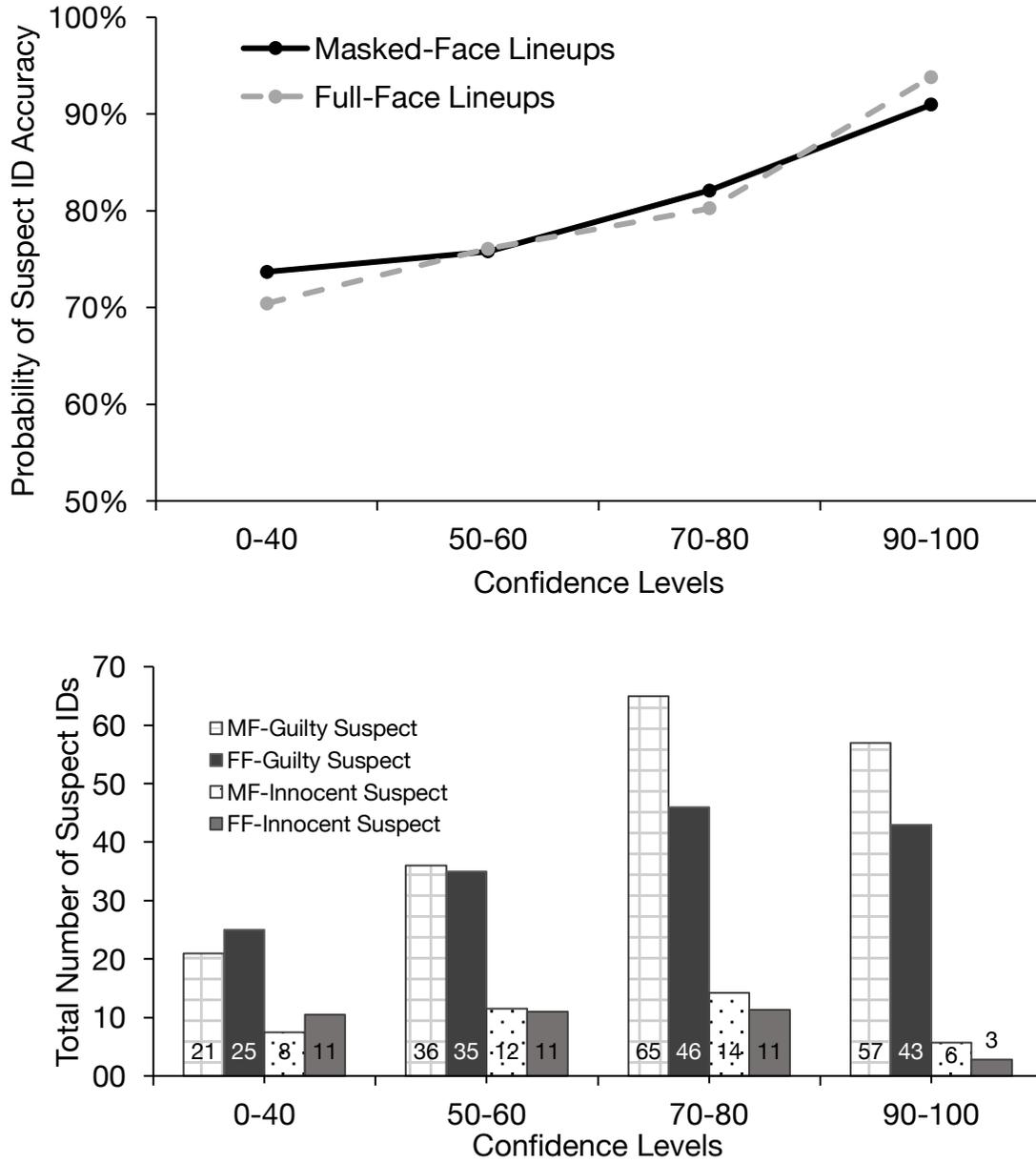


Figure 5. The top panel shows a CAC curve which is the probability that a suspect identification is accurate plotted at each level of confidence in Experiment 1. The bottom panel shows the total number of suspect identifications at each level of confidence that make up the CAC.

CHAPTER 4: EXPERIMENT 2

In Experiment 2, I examined pre-ID meta-memory judgments in two ways: First, participants were asked for their lineup preference (i.e., masked-face lineup or full-face lineup); second, participants were asked to make a pre-ID confidence judgment about their ability to make an accurate identification decision for both lineup types. Given the novelty of the research question, I did not make strong directional predictions; rather, I examined how and why each outcome is theoretically possible.

Producing a pre-ID confidence judgment might encourage participants to retrieve the masked face. This retrieval may benefit participants who are administered a masked-face lineup due to the perceptual match between encoding and retrieval. However, covert retrieval might not benefit participants who are administered a full-face lineup due to a difference in processing and perceptual mismatch between encoding and retrieval. Pre-ID meta-memory judgments might be more accurate for identification decisions made for masked-face lineups compared to full-face lineups. However, past studies have shown that pre-ID confidence ratings are highly inaccurate for lineup decisions (Cutler & Penrod, 1989; Nguyen et al., 2018) thus this rating may not be indicative of accuracy for either lineup type.

Conversely, the relationship between pre-ID confidence and identification accuracy may be more complex. Given the evidence from the word recognition literature (and the multi-trial design), pre-ID confidence ratings may more accurately reflect identification accuracy across trials (Finn & Metcalfe, 2014; Koriat et al., 2002). Specifically, on the first trial participants might not have a lineup preference or insight into their identification ability due to the novelty of the task. However, regardless of whether the first trial is a masked-face lineup or a full-face lineup, pre-ID ratings may become more accurate on the next trial. If participants are

administered a masked-face lineup, they may realize there is a benefit to the perceptual match (or that it just seems easier without knowing why) and thus shift preference ratings and pre-ID confidence to favor masked-face lineups. However, if participants are administered a full-face lineup, they may realize the difficulty of the perceptual mismatch and still shift their preference to favor a masked-face lineup. Further, I did not predict these meta-memory ratings to have a negative impact on ID accuracy. Rather, I expected identification patterns to replicate those in Experiment 1. Given that pre-ID ratings are similar to judgment-of-learning ratings (JOLs) and there is evidence that making JOLs increases learning (Janes, Rivers, & Dunlosky, 2018), identification accuracy could be higher for both lineups in Experiment 2 compared to Experiment 1. However, given differences in the tasks, JOLs (or pre-ID judgments) might have no impact on learning or identification accuracy. Further, I hesitate to make strong predictions based on comparisons between experiments.

Method

Participants and Design

Participants were 334 Iowa State University students who consented to complete the study in exchange for course credit. Data from 21 of the 334 participants were not included in final data analyses because they failed attention checks ($n = 13$), because they knew one of the actors or fillers ($n = 5$), or because total duration spent completing the study was over two times the maximum amount of time it took in-lab participants (2 hours; $n = 2$). Of the final 313 participants, 158 completed the study in the lab and 155 completed the study online through the university SONA system. Demographics can be found in Appendix C.

Like Experiment 1, a 2 (Lineup: masked-face, full-face) x 2 (Target: absent, present) within-subjects design was implemented (see Figure 2) to investigate identification rates and

confidence ratings. Additionally, lineup preference and pre-ID confidence judgments were examined.

Materials and Procedure

All materials and procedures were identical to Experiment 1 with one important exception—immediately before each lineup was administered, participants were asked three questions (see Appendix F). The first question was the lineup preference question, “You are going to see a lineup. The lineup will either consist of unmasked faces or faces wearing ski masks that expose the eyes. The computer randomly determines which version you will see. But if you were given a choice, would you like to see a lineup of unmasked faces or a lineup of faces wearing masks?” Participants chose between “a lineup of unmasked faces” and “a lineup of faces wearing masks.” Next, two pre-ID confidence questions were asked in a random order. One of these question asked: “If you were asked to identify the perpetrator in a 6-person lineup consisting of *faces wearing masks*, that may or may not contain the perpetrator how likely do you think you would be able to make a correct decision? Please rate your answer on a scale from 0-100, 0 indicating no chance of correct decision, 100 indicating absolute certainty of correct decision?” The other pre-ID confidence question asked: “If you were asked to identify the perpetrator in a 6-person lineup consisting of *unmasked faces*, that may or may not contain the perpetrator, how likely do you think you would be able to make a correct decision? Please rate your answer on a scale from 0-100, 0 indicating no chance of correct decision, 100 indicating absolute certainty of correct decision.” After these meta-memory questions, the procedure followed the same steps as Experiment 1.

Results and Discussion

Similar to Experiment 1, lineup decision accuracy data from target-absent and target-present lineups were analyzed separately with one exception. Pre-ID meta-memory data from target-present and target-absent conditions are reported together because these measures were taken prior to lineup administration. Data from Experiment 2 are reported starting with target-present identification accuracy data, followed by post-ID confidence data. Next, target-absent data are similarly presented followed by pre-ID confidence, CACs, and CD-SDT data.

Target-Preset Lineups

Target Identifications

A Cochran's Q χ^2 test was used to examine identification accuracy for masked-face lineups compared to full-face lineups (see Figure 6 and Table 3 for means). Replicating Experiment 1, there was a significant difference in target identifications made from masked-face lineups compared to full-face lineups, $\chi^2(313) = 14.30, p < .001, OR = 1.90, CI^{95}[0.84, 2.11]$. Participants were able to identify the target more often in masked-face lineups ($M = .53, SD = .50$) relative to full-face lineups ($M = .38, SD = .49$). This finding replicates the results from Experiment 1 and confirms my prediction that identification performance would benefit when participants were administered a masked-face lineup compared to a full-face lineup. Finally, participant location (lab, online) did not interact with lineup type or change the pattern of results, $F(1, 622) = .57, p = .453$. This indicates that the data collected from participants in the lab did not significantly differ compared to participants whose data was collected online.

Foil Identifications and Incorrect Rejections

A Cochran's Q χ^2 tests test revealed that participants made significantly more foil identifications from full-face lineups ($M = .41, SD = .49$) compared to masked-face lineups ($M =$

.31, $SD = .46$), $\chi^2(313) = 8.53$, $p = .003$, $OR = 0.58$, $CI^{95}[0.50, 0.83]$. Further comparisons revealed no significant difference in incorrect rejections of masked-face lineups ($M = .16$, $SD = .37$) compared to full-face lineups ($M = .21$, $SD = .41$), $\chi^2(313) = 2.04$, $p = .153$.

Post-ID Confidence

LMMs were used to examine differences in confidence ratings associated with target identifications, foil identifications, and incorrect rejections as a function of lineup type (see Table 3). Analyses revealed a nonsignificant difference in confidence ratings for target identifications based on lineup type, $F(1, 282) = 2.88$, $p = .091$, $d = 0.20$. Specifically, participants rated the confidence in their target identifications similarly from masked-face lineups ($M = 68.24\%$, $SD = 23.53$) and full-face lineups ($M = 63.53\%$, $SD = 22.46$).

Further, gamma correlations revealed a significant relationship between confidence rating and target identifications for masked-face lineups, ($r = .36$, $p < .001$). Confidence rating and target identifications were also significantly correlated for full-face lineups, ($r = .38$, $p < .001$). However, participants rated the confidence in their foil identifications from masked-face lineups ($M = 55.15\%$, $SD = 22.74$) higher than those made from full-face lineups ($M = 48.99\%$, $SD = 22.36$), $F(1, 224) = 4.15$, $p = .043$, $d = 0.27$. There was no significant difference in confidence ratings for incorrect rejections made from masked-face ($M = 55.69\%$, $SD = 21.00$) lineups compared to full-face lineups ($M = 49.08\%$, $SD = 23.17$), $F(1, 114) = 2.52$, $p = .115$, $d = 0.30$.

An additional LMM was conducted to examine whether confidence ratings differed based on lineup type and accuracy. This analysis revealed a significant main effect of lineup, $F(1, 622) = 8.96$, $p = .003$, $d = 0.31$ and a main effect of accuracy, $F(1, 622) = 55.33$, $p < .001$, $d = 0.61$. Specifically, higher confidence ratings were associated with masked-face lineups (62.14% , $SD = 23.72$) compared to full-face lineups (54.54% , $SD = 23.57$) regardless of accuracy. Higher

confidence ratings were also associated with target identifications (66.27%, $SD = 23.16$) compared to incorrect responses (51.75%, $SD = 22.55$) regardless of lineup type. The interaction was not significant, $F < 1, p = .664$.

Forced-Choice Lineup Identifications

When the proportion of hits from the initial lineup were combined with the proportion hits from the forced-choice lineups (for participants who initially chose the “Not Present” option), hits from the masked-face lineups increased by 5% ($M = .58, SD = .49$) and hits from the full-face lineups increased by 7% ($M = .45, SD = .50$). However, the benefit of the masked-face lineup for target identification persisted over the full-face lineup, $\chi^2(313) = 12.08, p = .001, OR = 1.81, CI^{95}[0.90, 2.24]$.

Target-Absent Lineups

Correct Rejections and Foil Identifications

For target-absent lineups, correct rejection rates based on lineup type were analyzed using Cochran’s $Q \chi^2$. There was no difference in correct rejections from masked-face lineups ($M = .38, SD = .49$) compared to full-face lineups ($M = .40, SD = .49$), $\chi^2(313) = .30, p = .587, OR = 0.65, CI^{95}[0.49, 0.84]$ (see Figure 7 and Table 3 for means). Finally, participant location (lab, online) did not interact with lineup type or change the pattern of results, $F(1, 622) = .98, p = .322$.

Post-ID Confidence

I used an LMM to examine differences in correct rejection confidence ratings made from target-absent masked-face lineups compared to full-face lineups (see Table 3). There was no statistical difference in confidence ratings assigned to correct rejections made from masked-face lineups ($M = 58.64\%, SD = 19.43$) compared to full-face lineups ($M = 57.02\%, SD = 23.20$), $F <$

1, $p = .556$, $d = 0.07$. Gamma correlations for each lineup type were calculated between confidence ratings and correct rejections. These revealed that confidence was not significantly related to correct rejections either for full-face lineups, ($r = .10$, $p = .182$), or masked-face lineups, ($r = .04$, $p = .601$). Participants rated their confidence in foil identifications made from masked-face lineups ($M = 57.90\%$, $SD = 21.02$) higher than those made from full-face lineups ($M = 53.33\%$, $SD = 21.73$), $F(1, 382) = 4.38$, $p = .037$, $d = 0.22$.

Finally, a 2 (Lineup: masked-face, full-face) x 2 (Accuracy: correct, incorrect) LMM on post-ID confidence ratings revealed a marginally significant effect of lineup, $F(1, 622) = 3.11$, $p = .078$, $d = 0.17$. Participants rated their confidence in their decisions from masked-face lineups ($M = 58.18\%$, $SD = 20.40$) higher compared to full-face lineups ($M = 54.79\%$, $SD = 22.36$). There were no other significant effects, $F_s < 1.59$, $p_s > .208$.

Lineup Preference and Pre-ID Confidence

Lineup preference and pre-ID confidence data for both target-present and target-absent lineups were analyzed together since these measures were taken prior to lineup assignment (see Table 4). First, a Cochran's Q χ^2 test was used to analyze lineup preference differences as a function of trial (1-4; see Table 6). This analysis revealed a significant difference in lineup preference from Trial 1 to Trial 4, $\chi^2(313) = 47.96$, $p < .001$. Specifically, preference shifted significantly from Trial 1 to Trial 2, $\chi^2(313) = 27.77$, $p < .001$, then did not shift from Trial 2 to Trial 3, $\chi^2(313) = .21$, $p = .651$, or from Trial 3 to Trial 4, $\chi^2(313) = .34$, $p = .558$. Interestingly, on the first trial participants preferred a masked-face lineup ($M = 53.35\%$) at a comparable rate to a full-face lineup ($M = 46.65\%$), $\chi^2(313) = 1.41$, $p = .235$. However, by the second trial, lineup preference shifted to favor masked-face lineups ($M = 71.57\%$) compared to full-face lineup ($M = 28.43\%$), $\chi^2(313) = 58.23$, $p < .001$ and this preference persisted for Trial 3 ($M_{MF} = 70.29\%$, M_{FF}

= 29.71%), $\chi^2(313) = 51.53, p < .001$. and Trial 4 ($M_{MF} = 71.88\%$, $M_{FF} = 28.12\%$), $\chi^2(313) = 59.97, p < .001$.

Pairwise comparisons compared pre-ID confidence for hypothetical masked-face lineups to full-face lineups. Specifically, on Trial 1, pre-ID confidence ratings for a masked-face lineup decision ($M = 53.39\%$, $SD = 21.53$) was similar to pre-ID confidence ratings for a full-face lineup decision ($M = 52.62\%$, $SD = 22.64$), $t(312) = .56, p = .567, d = 0.04$. However, by Trial 2 pre-ID confidence ratings diverge and masked-face lineup ratings ($M = 55.65\%$, $SD = 19.98$) become significantly higher than full-face confidence ratings ($M = 50.80\%$, $SD = 20.55$), $t(312) = 4.23, p < .001, d = 0.24$. This difference was maintained on Trial 3, $t(312) = 5.61, p < .001, d = 0.32$ and Trial 4, $t(312) = 4.55, p < .001, d = 0.26$, with participants rating their pre-ID confidence higher for masked-face lineup decisions ($M_{MF-T3} = 54.44\%$, $SD_{MF-T3} = 20.61$; $M_{MF-T4} = 51.57\%$, $SD_{MF-T4} = 20.44$) than full-face lineup decisions, $M_{FF-T3} = 48.47\%$, $SD_{FF-T3} = 20.45$; $M_{FF-T4} = 46.54\%$, $SD_{FF-T4} = 20.94$). That is, pre-ID confidence ratings were reflected in participant preference ratings.

An additional 2 (Lineup: masked-face, full-face) x 2 (Preference Match: mismatch, match) LMM was conducted to assess if accuracy (target identifications plus correct rejections compared to foil identifications plus incorrect rejections) varied by whether the lineup matched participants' preference. Results indicated a significant effect of lineup, $F(1, 1248) = 4.99, p = .026$, such that participants were able to make an accurate identification (target ID or correct rejection) more often from a masked-face lineup ($M = .45, SD = .50$) than a full-face lineup ($M = .39, SD = .49$). However, preference match did not impact accuracy, as performance was similar whether their preference matched ($M = .43, SD = .50$) or did not match ($M = .41, SD = .49$), $F < 1, p = .756$. Further, the interaction was not significant, $F = 2.64, p = .104$.

Next, a 2 (Lineup: masked-face, full-face) x 2 (Accuracy: correct, incorrect) LMM was conducted on pre-ID confidence ratings. This analysis was intended to find if higher pre-ID confidence ratings were associated with correct IDs (target IDs and correct rejections) compared to incorrect IDs (foil IDs and incorrect rejections) as a function of lineup type. A significant main effect of lineup was revealed, $F(1, 1248) = 7.23, p = .007, d = 0.14$. Specifically, participants were more confident in their ability to make an accurate decision from a masked-face lineup ($M = 53.12\%, SD = 20.51$) compared to a full-face lineup ($M = 50.03\%, SD = 21.12$) regardless of whether they actually made an accurate decision or not. The main effect of accuracy and the interaction was not significant, $F_s < 2.47, p_s > .117$.

Gamma correlations were calculated for each lineup type between pre-ID confidence and identification accuracy (target IDs/correct rejections). For masked-face target-absent lineups, correct rejections negatively correlated with pre-ID confidence ($r = -.15, p = .032$). For full-face target-absent lineups, pre-ID confidence was also negatively correlated with correct rejections ($r = -.18, p = .011$). Both of these correlations indicate that higher pre-ID confidence was associated with a lower likelihood of correct rejections when the target was absent. Further, for masked-face target-present lineups ($r = .04, p = .522$) and full-face target-present lineups ($r = .01, p = .847$), pre-ID confidence was not significantly correlated with target IDs.

Finally, CAC curves were plotted for pre-ID confidence using the same equation as discussed in Experiment 1 to examine suspect identification accuracy. Figure 8 shows CAC curves for suspect identifications (i.e., guilty and innocent suspects) from masked-face lineups compared to full-face lineups at each level of confidence. The CAC curve is essentially flat which indicates that there was no discernable difference between the probability of suspect identification accuracy from a masked-face lineup or a full-face line up controlling for pre-ID

confidence ratings. Further, the flatness of the curve indicates that there is no relationship between pre-ID confidence and subsequent identification accuracy.

Finally, pre-ID confidence ratings significantly correlated with post-ID confidence ($r = .44, p < .001$). This finding indicates that there are individual differences between participants. For example, participants who tended to rate their pre-ID confidence high also tended to rate their post-ID confidence as high. Conversely, pre-ID confidence ratings were not significantly related to accuracy collapsed across lineup types ($r = -.04, p = .164$). This indicates that there was virtually no relationship between pre-ID confidence and subsequent accuracy regardless of lineup type.

Confidence-Accuracy Characteristic (CAC)

Similar to Experiment 1, Confidence-Accuracy Characteristic (CAC) curves were calculated (Wixted & Wells, 2017) for post-ID confidence (see Figure 9) to examine suspect identification accuracy. The CAC curve shows that the probability of suspect identification accuracy was higher at high levels of confidence compared to low levels of confidence. Again, masked- and full-face lineups produced mostly overlapping CAC curves (i.e., lineup type did not influence the CAC curve), but participants did make more high-confidence target identifications from a masked-face lineup ($total = 47$) compared to a full-face lineup ($total = 21$). However, participants also made more innocent suspect identifications from a masked-face lineup ($total = 4$) compared to the full-face lineup ($total = 3$). That is, guilty suspect identifications increased by 8% from a full-face lineup to a masked-face lineup, but innocent suspect identifications only increased by 0.9%.

Compound Signal Detection Model

Again, discrimination (d') and response bias c were calculated using CD-SDT (Palmer, Brewer, & Weber, 2010; Palmer & Brewer, 2012) for each lineup type. Results for d' and c showed that discriminability was higher when participants were administered a masked-face lineup ($d' = 1.54$, $SE = .09$) compared to a full-face lineup ($d' = 1.13$, $SE = .06$; $G^2(1) = 18.05$, $p < .01$; see Table 3). Further, response bias was similarly conservative for masked-face lineups ($c = -0.33$, $SE = -.02$) compared to full-face lineups ($c = -0.30$, $SE = -.02$; $G^2(1) = .26$, $p = .611$).

Experiment 2 Discussion

The purpose of Experiment 2 was to gain insight into participant's beliefs about their own memory for a masked perpetrator. Would participants have a preference for a masked-face lineup or for a full-face lineup? Results indicated that with experience, participants preferred a masked-face lineup over a full-face lineup, which was also reflected in their pre-ID confidence ratings. However, although pre-ID confidence ratings were indicative of preference, they were not indicative of identification accuracy (Cutler & Penrod, 1989; Nguyen et al., 2018). Participants did rate their pre-ID confidence higher for masked-face lineups than full-face lineups. Further, preference match did not affect identification accuracy.

An important finding from Experiment 2 was a replication of the identification accuracy findings from Experiment 1. Specifically, for target-present lineups, participants' identification performance benefited when the lineup matched the perceptual appearance of the originally learned face. For target-absent conditions, correct rejections were comparable for masked-face lineups compared to full-face lineups. Before moving to Experiment 3, I first report analyses for combined data from Experiments 1 and 2.

Table 3

Identification proportions, confidence ratings, discriminability and response bias for identification decisions in Experiment 2

	Masked-Face Lineup	Full-Face Lineup
Target-Present Lineups		
Identification Proportions		
Target IDs	.53 (.50)*	.38 (.49)
Foil IDs	.31 (.46)*	.41 (.49)
Incorrect Rejections	.16 (.37)	.21 (.41)
Confidence Ratings		
Target IDs	68.24 (23.52)	63.53 (22.46)
Foil IDs	55.15 (22.74)*	48.99 (22.36)
Incorrect Rejections	55.69 (21.00)	49.08 (23.17)
Target-Absent Lineups		
Identification Proportions		
Correct Rejections	.38 (.49)	.40 (.49)
Foil IDs	.62 (.49)	.60 (.49)
Confidence Ratings		
Correct Rejections	58.64 (19.43)	57.02 (23.20)
Foil IDs	57.90 (21.02)*	53.33 (21.73)
d'	1.54 (.09)*	1.13 (.06)
c	-0.33 (-.02)	-0.30 (-.02)

Note: Values in parentheses represent standard deviations except in the cases of d' and c , for which the parenthetical values are standard errors. An asterisk indicates a significant difference between lineup types.

Table 4

Pre-ID confidence ratings and lineup preference proportions across trials for masked-face lineups and full-face lineups in Experiment 2

	Trial 1	Trial 2	Trial 3	Trial 4
Lineup Preference (%)				
Masked-Face Lineups	53.35	71.57	70.29	71.88
Full-Face Lineups	46.65	28.43	29.71	28.12
Pre-ID Confidence Ratings (%)				
Masked-Face Lineups	53.39 (21.53)	55.65 (19.98)	54.44 (20.61)	51.57 (20.44)
Full-Face Lineups	52.62 (22.64)	50.80 (20.55)	48.47 (20.45)	46.58 (20.94)

Note: Values in parentheses represent standard deviations.

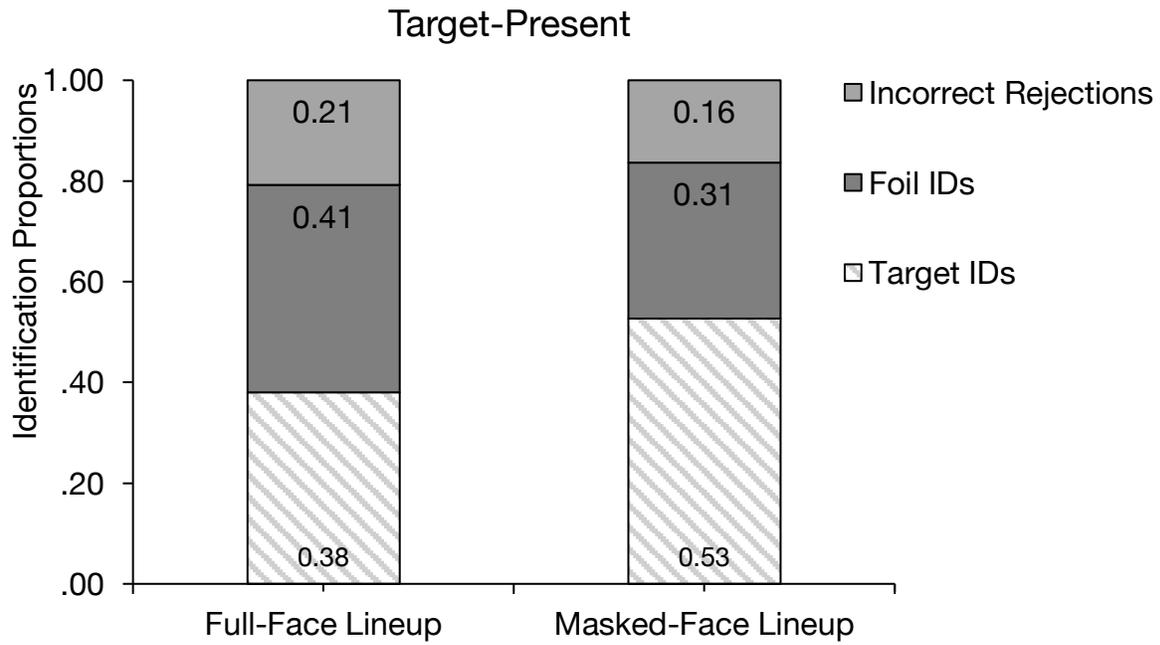


Figure 6. Experiment 2 target-present lineup identification proportions as a function of lineup type.

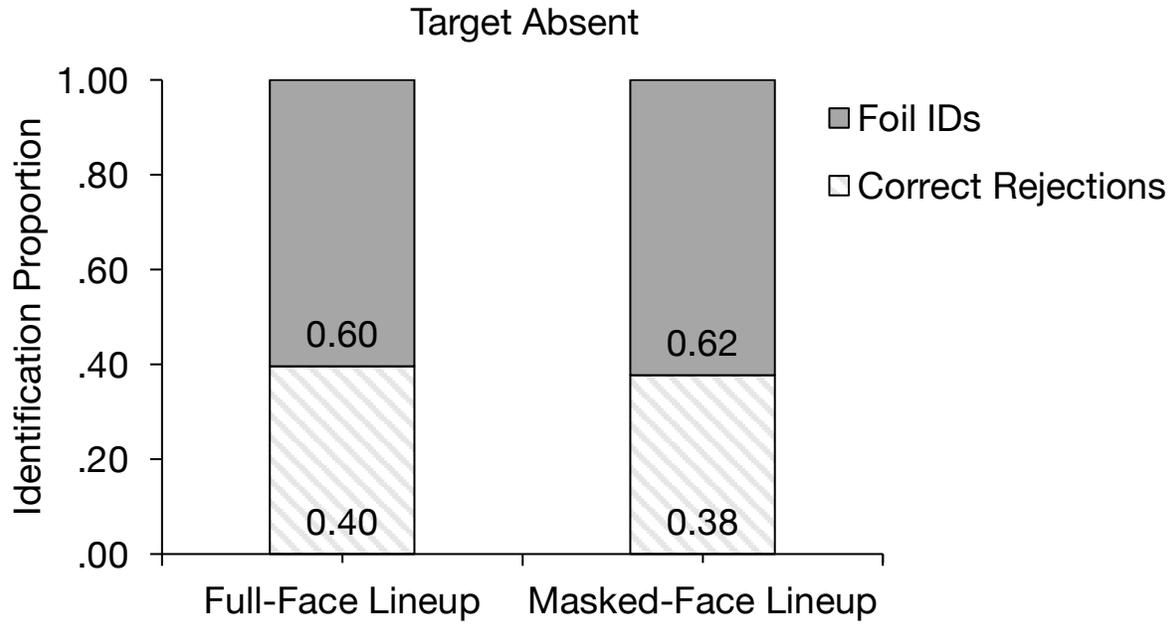


Figure 7. Experiment 2 target-absent lineup identification proportions as a function of lineup type.

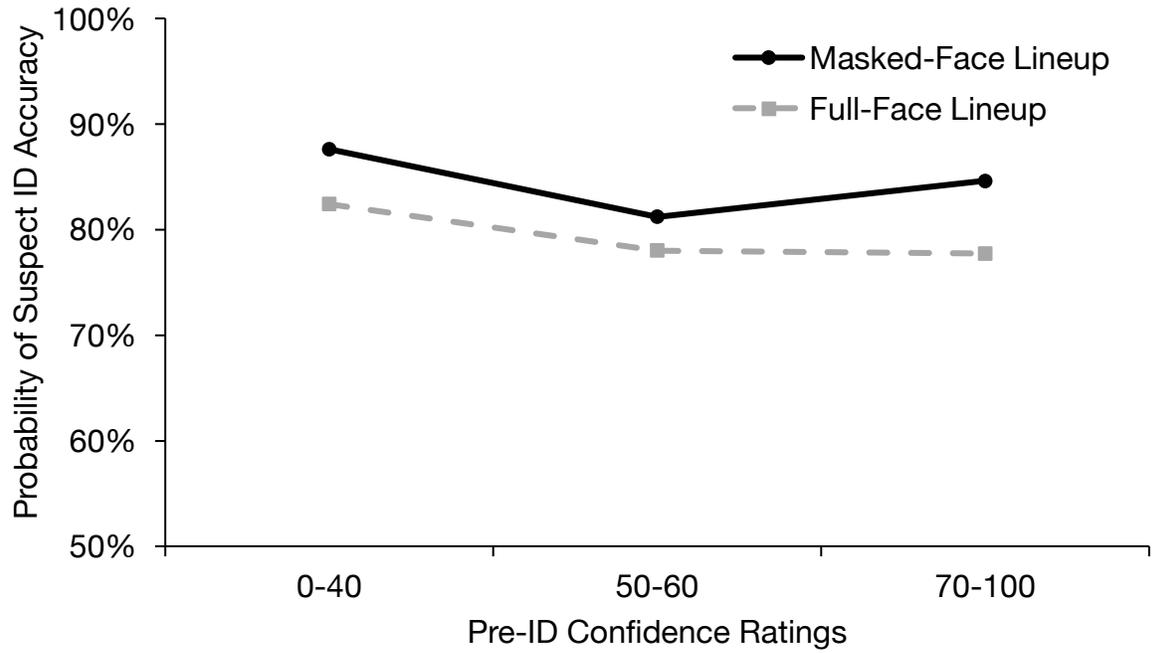


Figure 8. A CAC curve showing the probability that a suspect identification is accurate at each level of pre-ID confidence in Experiment 2.

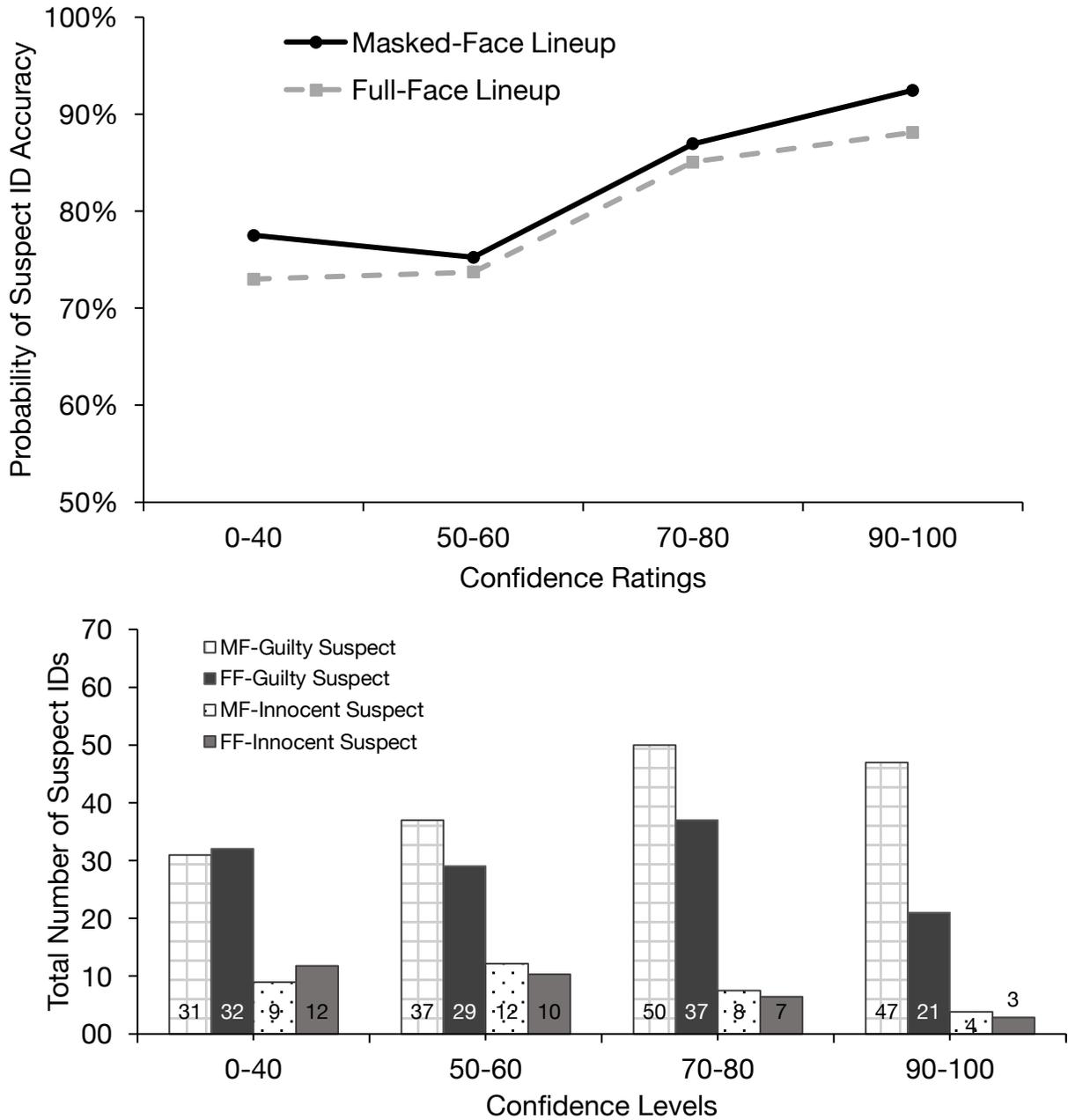


Figure 9. The top panel shows a CAC curve which is the probability that a suspect identification is accurate plotted at each level of confidence in Experiment 2. The bottom panel shows the total number of suspect identifications at each level of confidence that make up the CAC.

CHAPTER 5: COMBINED ANALYSES OF THE DATA IN EXPERIMENTS 1 AND 2

For the sake of increased power, I conducted exploratory analyses on the combined data from Experiment 1 and 2 (see Table 5).

Target-Present Lineups

I used Cochran's Q to examine the difference in target identifications made from masked-face lineups compared to full-face lineups (see Table 5 for means). Analyses from the combined data set revealed that participants were able to identify the target more often from a masked-face lineup ($M = .55$, $SD = .50$) compared to a full-face lineup ($M = .43$, $SD = .50$), $\chi^2(623) = 20.84$, $p < .001$, $OR = 1.74$, $CI^{95}[1.14, 2.17]$. Further, participants made significantly fewer foil identifications from masked-face lineups ($M = .30$, $SD = .46$) compared to full-face lineups ($M = .39$, $SD = .49$), $\chi^2(623) = 14.63$, $p < .001$, $OR = 0.60$, $CI^{95}[0.51, 0.75]$. However, even with the combined data set participants made a comparable amount of incorrect rejections from masked-face lineups ($M = .15$, $SD = .36$) and full-face lineups ($M = .18$, $SD = .39$), $\chi^2(623) = 1.93$, $p = .165$, $OR = 0.81$, $CI^{95}[0.54, 1.29]$.

A 2 lineup type (masked-face, full-face) x 4 trial (trial 1- 4) LMM was conducted on identification accuracy. In addition to the reported effect of lineup type, $F(1, 1238) = 19.19$, $p < .001$, a main effect of trial was revealed, $F(1, 1238) = 3.65$, $p = .012$, such that participants performance increased from Trial 1 ($M = .44$, $SD = .50$), to Trial 2 ($M = .50$, $SD = .50$), and to Trial 3 ($M = .56$, $SD = .50$), then decreased on Trial 4 ($M = .46$, $SD = .50$). The interaction was not significant, $F = 1.78$, $p = .150$. The means for target identifications and post-ID confidence for each lineup type across each trial can be found in Table 6.

I also assessed whether target identifications differed based on lineup type when limited to the first trial. This was to examine if the masked-lineup superiority endured when assessed

between-subjects. The effect persisted when restricted to the first trial. Participants made more target identifications from masked-face lineups ($M = .53, SD = .50$) compared to full-face lineups ($M = .35, SD = .48$), $F(1, 309) = 10.95, p = .001$.

Finally, experiment (1, 2) did not interact with lineup type or change the pattern of results, $F(1, 1242) = .80, p = .371$, such that asking participants to make pre-ID confidence ratings for Experiment 2 did not significantly change the patterns of results. In addition, actor interacted with lineup type, $F(1, 1238) = 4.89, p = .002$. Specifically, for the computer thief ($M_{MF} = .67, SD_{MF} = .47; M_{FF} = .46, SD_{FF} = .50$), the identity thief ($M_{MF} = .50, SD_{MF} = .50; M_{FF} = .37, SD_{FF} = .49$), and the wallet thief ($M_{MF} = .47, SD_{MF} = .50; M_{FF} = .27, SD_{FF} = .44$), participants made more target identifications for masked-face lineups ($M = .55, SD = .50$) compared to full-face lineups ($M = .43, SD = .50$); however, for the bike thief ($M_{MF} = .57, SD_{MF} = .50; M_{FF} = .63, SD_{FF} = .49$), participants made fewer target identifications from masked-face lineups compared to full-face lineups. This last difference was further examined through a post-hoc comparison which found that the difference in target identifications was not significant, $F(1, 308) = .86, p = .355$.

Target-Absent Lineups

Identification accuracy for target-absent lineups was examined using Cochran's Q (see Table 5 for means). Participants made similar quantities of correct rejections from masked-face lineups ($M = .31, SD = .46$) compared to a full-face lineup ($M = .35, SD = .48$), $\chi^2(623) = 2.64, p = .104, OR = 0.67, CI^{95}[0.54, 0.83]$. A 2 lineup type (masked-face, full-face) x 4 trial (trial 1- 4) LMM was conducted on identification accuracy. Here there were no significant effects, $F_s < 1.82, p_s > .142$. In addition, I assessed whether correct rejections differed based on lineup type when limited to the first trial. Similar to the within-subjects analysis, the between-subjects

analysis revealed that participants made similar correct rejections from masked-face lineups ($M = .28$, $SD = .45$) compared to full-face lineups ($M = .31$, $SD = .46$), $F(1, 310) = .36$, $p = .551$. The means for correct rejections and post-ID confidence for each lineup type across each trial can be found in Table 6.

Finally, experiment (1, 2) did not interact with lineup type or change the pattern of results, $F(1, 1242) = .63$, $p = .428$. This indicates that the inclusion of pre-ID confidence ratings for Experiment 2 did not significantly change the patterns of results. In addition, actor interacted with lineup type, $F(1, 1238) = 4.11$, $p = .007$. Specifically, for the bike thief ($M_{MF} = .35$, $SD_{MF} = .48$; $M_{FF} = .50$, $SD_{FF} = .50$), the identity thief ($M_{MF} = .24$, $SD_{MF} = .43$; $M_{FF} = .30$, $SD_{FF} = .46$), and the wallet thief ($M_{MF} = .20$, $SD_{MF} = .40$; $M_{FF} = .26$, $SD_{FF} = .44$), participants made more correct rejections from full-face lineups ($M = .31$, $SD = .46$) compared to masked-face lineups ($M = .35$, $SD = .48$); however, for the computer thief ($M_{MF} = .46$, $SD_{MF} = .50$; $M_{FF} = .36$, $SD_{FF} = .48$), participants made more correct rejections from masked-face lineups compared to full-face lineups.

Given the results of the combined analyses, stronger conclusions can be made that identification accuracy of a masked perpetrator benefits from a masked-face lineup compared to a full-face lineup. These findings in conjunction with the results from Experiment 2 indicate that participants prefer a masked-face lineup over a full-face lineup with experience. As such, the next step is to examine how a person without identification experience might rate the credibility of this type of identification.

Combined Confidence-Accuracy Characteristic (CAC)

Post-ID confidence data were combined from Experiments 1 and 2 to produce higher powered Confidence-Accuracy Characteristic (CAC) curves (see Figure 10; Wixted & Wells,

2017). The combined CAC curve shows that the probability of suspect identification accuracy was highest at high levels of confidence compared to low levels of confidence. Masked- and full-face lineups produced similar, overlapping CAC curves (i.e., lineup type did not influence the CAC curve). However, participants did make more high-confidence target identifications from a masked-face lineup ($total = 104$) compared to a full-face lineup ($total = 64$). However, participants also made more innocent suspect identifications from a masked-face lineup ($total = 10$) compared to the full-face lineup ($total = 6$). That is, guilty suspect identifications increased by 6% from a full-face lineup to a masked-face lineup, but innocent suspect identifications only increased by 0.6%.

Table 5

Identification proportions and confidence ratings as a function of lineup type in the combined data from Experiments 1 & 2

	Masked-Face Lineup	Full-Face Lineup
Identification Proportions (TP)		
Target Identifications	.55 (.50)*	.43 (.50)
Foil Identifications	.30 (.46)*	.39 (.49)
Incorrect Rejections	.15 (.36)	.18 (.39)
Confidence Ratings (%)		
Target Identifications	71.10 (22.31)*	66.31 (22.72)
Foil Identifications	58.70 (22.13)*	52.48 (22.69)
Incorrect Rejections	59.89 (24.08)	54.51 (24.31)
Identification Proportions (TA)		
Correct Rejections	.31 (.46)	.35 (.48)
Foil Identifications	.69 (.46)	.65 (.48)
Confidence Ratings (%)		
Correct Rejections	60.92 (22.11)	60.05 (24.26)
Foil Identifications	60.91 (20.88)*	55.46 (21.85)

Note: Values in parentheses represent standard deviations. An asterisk indicates a significant difference between lineup types.

Table 6

Post-ID confidence ratings and identification accuracy across trials for masked-face lineups and full-face lineups in Experiments 1 & 2

	Trial 1	Trial 2	Trial 3	Trial 4
Post-ID				
Confidence Ratings (%)				
Masked-Face TP (Target IDs)	70.12 (23.09)	76.71 (21.59)	71.31 (20.14)	66.51 (23.86)
Full-Face TP (Target IDs)	60.37 (21.80)	70.00 (21.92)	67.60 (22.35)	65.25 (24.33)
Masked-Face TA (Correct Rej)	62.79 (21.19)	63.00 (21.26)	57.21 (24.72)	60.00 (21.70)
Full-Face TA (Correct Rej)	59.59 (22.26)	61.04 (24.34)	58.09 (25.34)	61.61 (25.17)
Identification Proportions				
Masked-Face TP (Target IDs)	.53 (.50)	.51 (.50)	.64 (.48)	.53 (.50)
Full-Face TP (Target IDs)	.35 (.48)	.49 (.50)	.48 (.50)	.40 (.49)
Masked-Face TA (Correct Rej)	.28 (.45)	.38 (.49)	.27 (.45)	.32 (.47)
Full-Face TA (Correct Rej)	.31 (.46)	.31 (.47)	.30 (.46)	.40 (.49)

Note: Values in parentheses represent standard deviations.

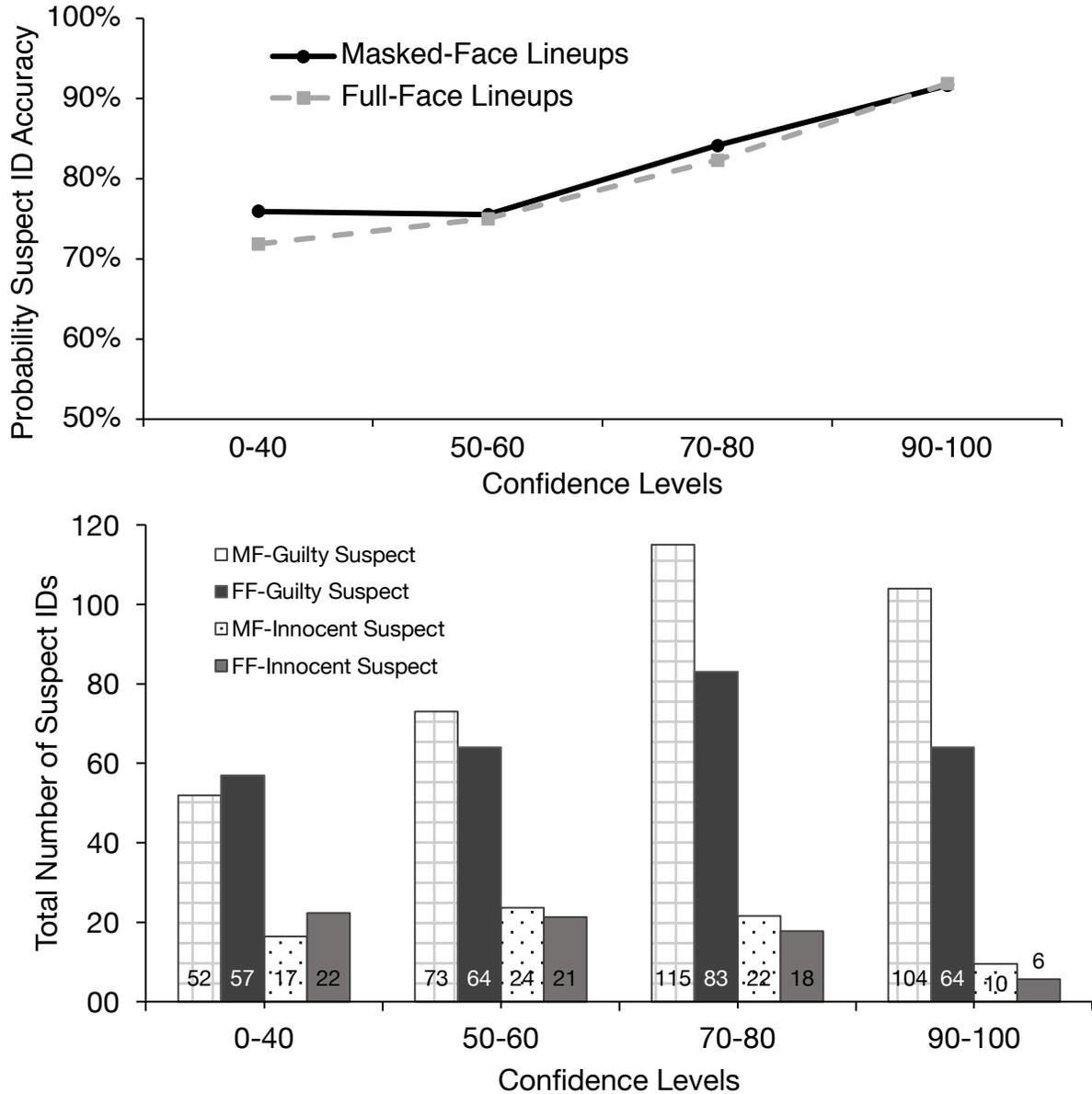


Figure 10. The top panel shows a CAC curve based on data combined from Experiments 1 and 2. The CAC is the probability that a suspect identification is accurate plotted at each level of confidence. The bottom panel shows the total number of suspect identifications at each level of confidence that make up the CAC.

CHAPTER 6: EXPERIMENT 3

One goal of eyewitness research is to improve how policies are implemented. Therefore, it is important to examine how mock jurors might react when faced with an eyewitness identification made from a masked-face lineup, because it might be considered unconventional compared to a full-face lineup. Experiment 3 examined mock jurors' judgment of credibility and trust of an identification of a masked perpetrator made from a masked-face lineup compared to one made from a full-face lineup. In addition, high, low, or none confidence ratings accompanied identification. There are several patterns that could emerge from the data. Given the novelty of the experiment, I outline how each outcome is possible and how specific patterns in the data would could be theoretically explained.

A masked-face lineup offers less facial information than a full-face lineup (i.e., most of the face is occluded by a mask). Because there are more features and thus information when a face is unmasked, it might seem like a witness could make a better decision from a full-face lineup. Similar to findings in object recognition where participants tend to make fewer errors as they are presented with more information (i.e., the lines that make up the object; Biederman, 1987). It is unclear, however, whether this theory (recognition-by-components) applies to face recognition. Findings from the consumer decision-making literature show that people tend to prefer more information, though more information does not necessarily lead to better decisions (Jacoby, Speller, & Kohn, 1974; also see Russo, 1974). As applied to the current experiment, a mock juror might not consider the perceptual appearance of the masked perpetrator matching or mismatching the lineup members. Instead, mock jurors might rely on the quantity of information over the quality and conclude a full-face lineup identification to be more credible than a masked-face lineup identification.

Alternatively, there may be no difference in identification credibility between lineup types. Given the lineup preference results from Experiment 2 (no preference prior to experience with the task), it seems mock jurors might not understand the difference in the difficulty of the task of identifying a masked perpetrator from a masked-face lineup compared to a full-face lineup. In addition, mock jurors might not consider the difference in the amount of information a masked face compared to a full face provides or how this difference may affect identification accuracy. Rather, consistent with past research mock jurors will likely focus primarily on the witness confidence ratings (Cutler et al., 1989; Cutler et al., 1990) and the degraded viewing conditions (i.e., masked perpetrator). In this case, identification credibility will be uniformly low and only vary as a function of witness confidence rating.

Method

Participants

Participants were 172 Iowa State University students and workers from Amazon's mechanical Turk (mTurk). Data from four mTurk participants were not included because these participants chose "0" for every measurement and completed the study in under four minutes ($M = 10.13$ min, $SD = 5.45$). Therefore, the final data set included 86 participants who completed the study in the lab at Iowa State University, 56 participants who completed the study online through mTurk, and 26 Iowa State University students who completed the study online. Sample size ($n = 120$) was calculated using G*Power for a small effect size ($f = .125$; based on the effect of lineup similarity in Cutler et al., 1988) with 90% power, an alpha level of .05 in a repeated measures design.

Materials

Two case vignettes were adapted from real crime reports obtained from law.justia.com. These cases were selected by qualifying length (> 600 words) and criminal act (i.e., robbery). The reports were highly altered to feature a perpetrator who wore a ski mask and a witness who subsequently made an identification from a masked-face or full-face lineup (see Appendix B). The names of the individuals in the original crime reports were changed so that they were Caucasian typical because the faces in the lineups belong to Caucasians. Case vignettes were matched in length ($M = 411$ words, Range = 407 – 415), and each contained reference to the mask five times at approximately equal locations. Mean appearance location of “mask” in the two vignettes were 193.8 and 198.8. Each vignette presented details of a crime including a brief overview of the crime committed, how well the witness saw the perpetrator, persons present during the crime, what items were stolen, and if anyone was injured. The vignettes also contained the “actual lineup” that was administered (either masked-face or full-face) and who the witness chose. Additionally, the vignettes included a statement of the witness’s confidence at the time of the identification. Confidence (low vs. high) was reported using non-numerical language (see Appendix B) or no confidence rating was presented in the vignette (control condition). Order of reports and all conditions were counterbalanced, and participants were randomly assigned to each counterbalance via the computer program. After each report, participants received four questions about the credibility of the identification that the witness made in the vignette (adapted from Wells & Bradfield, 1998; see Appendix G). These questions were presented in a random order to each participant and asked about trust of the identification (trust), guilt of the identified individual (guilt), pertinence of the information presented to the witness (information), and appropriateness of the lineup (lineup) that was administered to the witness.

Design and Procedure

Ratings of identification and lineup credibility were investigated using a 3 (Witness Confidence Rating: none, low, high) x 2 (Lineup Type: masked-face lineup, full-face lineup) mixed design with Witness Confidence Rating manipulated between-subjects and Lineup Type manipulated within-subjects.

To begin the experiment, participants received the same study instructions as those in Experiments 1 and 2 regarding distractions, using a mobile device, and finishing the study in one session (~10 min total). Next, participants read the following instructions: "You will read several short case files, one at a time that depict various crimes. The case files are based on real criminal cases. Please pay attention to the details of the crimes because you will be asked to answer questions about these crimes as if you were a juror hearing the details of these cases." Participants were then given a case vignette that depicted a crime along with a statement at the end of the report that stated the detective administered a lineup to the witness. In one of the reports, the lineup consisted of faces wearing masks, and in the other report, the lineup consisted of unmasked faces. Each was accompanied by a picture of a lineup with a red square indicating who the witness chose (see Appendix B).

For one third of the reports, there was no statement of confidence from the witness. For another third, the reports included a statement of low confidence (e.g., At the time of the identification, the man rated his confidence as low saying, "I think that kind of looks like the man who robbed the casino--but I am not sure" or "I cannot be certain, but I think that might be the guy."). For the final third, the reports included a statement of high confidence (e.g., At the time of the identification, the man rated his confidence as high saying, "I am pretty certain that's

the guy who took our money" or "That is the guy who made us open the safe, I am almost positive.").

Participants were allowed to proceed from the report after 120s, but the program did not automatically proceed without the participant clicking the next button. Next, participants received four questions in a random order (see Appendix G). When participants finished these questions, they received the second case vignette. The procedure for the second case was identical to the first. Finally, participants answered a short demographic questionnaire and were debriefed.

Results and Discussion

Due to the similarity of the credibility lineup questions (see Appendix G), an exploratory factor analysis (EFA) with principal components extraction was conducted to examine if it was appropriate to interpret the data from all four question as a single factor (see Table 7). Based on the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA; Kaiser, 1970), the data was appropriate to factor (*Overall MSA_{FF} = 0.79; Overall MSA_{MF} = 0.82*). The Kaiser-Meyer-Olkin MSA rule states that values over .7 indicate acceptable sampling adequacy. The MSA value for all variables (i.e., trust, information, guilt, and lineup for both full-face and masked-face lineup conditions) was greater than .73 (i.e., good; Hutcheson & Sofroniou, 1999). In accordance with the Kaiser-Guttman rule, one factor was extracted with an eigenvalue greater than one (Kaiser, 1991; *eigenvalue_{FF} = 2.82, eigenvalue_{MF} = 3.00*; see Table 7), which accounted for 71% of the variance for the full-face lineups and 75% of the variance for the masked-face lineups. In addition, all four items for both full-face lineup items and masked-face lineup items achieved high loading (i.e., $\lambda > 0.70$) on the factor.

Communalities represent the amount of variance accounted for in a variable by the factor. Specifically, for each lineup credibility item measured, the communality indicates how much variance in each item is explained by the single factor solution. Given each of the items fell above the minimum threshold of 0.50 for the final communality estimates, it appears that at least half of the variance in these items can be explained by a single factor.

Given the results of the EFA, scores from the four questions regarding each lineup were averaged to create a composite lineup credibility score for each lineup. The effect of lineup type and witness confidence on lineup credibility was analyzed using a 3 (Confidence Rating: high, low, none) x 2 (Lineup: masked-face lineup, full-face lineup) repeated-measures ANOVA. I expected to see a main effect of confidence such that lineup credibility rating was higher as the confidence ratings increased. This prediction is based on findings that mock jurors tend to depend highly on the confidence ratings expressed by the witness when deciding if an identification was accurate (Cutler et al., 1989; Cutler et al., 1990). Indeed, there was a main effect of witness confidence, $F(2, 165) = 3.20, p = .043, \eta_p^2 = .04$, such that participants rated lineup credibility higher when the witness displayed high confidence ($M = 49.57\%, SD = 20.70$) compared to low confidence ($M = 42.79\%, SD = 19.50$) or when a confidence rating was absent ($M = 42.99\%, SD = 19.70$) in the case vignettes.

There was neither a significant main effect of lineup nor an interaction, $F_s < 1, p_s > .623$, $B_{01} = 5.82, B_{01\text{-interaction}} = 5.46$ (see Table 8 for means). Participants rated identifications made from masked-face lineups ($M = 45.89, SD = 20.73$) similarly to identifications made from full-face lineups ($M = 44.35, SD = 19.59$). Further, participants rated high confidence IDs made from masked-face lineups ($M = 49.20, SD = 21.27$) and full-face lineups ($M = 49.96, SD = 20.30$) to be comparable. Further, low confidence identification made from masked-face lineups ($M =$

43.71, $SD = 20.06$) were rated to be just as credible as those made from full-face lineups ($M = 41.88$, $SD = 19.06$). Finally, when no confidence rating was associated with an identification, participants rated the credibility of identifications made from masked-face lineups ($M = 44.78$, $SD = 20.81$) as credible as those made from full-face lineups ($M = 41.21$, $SD = 18.46$).

These results indicate that mock jurors do not find identifications made from masked-face lineups to be less credible than those made from full-face lineups. However, the failure to find a difference in lineup credibility may have been due to lack of ecological intensity as this type of decision might involve more pressure, weight, or contemplation in a real court case.

Alternatively, participants may not have the perspective to understand the challenge a perceptual mismatch creates for an eyewitness. This finding is consistent with preference data from Experiment 2 and show that people might not have insight into the task until they experience it. Rather, participants seemed to anchor to confidence levels and potentially the fact the perpetrator was masked (i.e., the witness had a challenging view of the perpetrator) when determining their lineup credibility rating. Given the overall low ratings, these data might simply suggest that mock jurors believe an identification of a masked perpetrator is difficult regardless of lineup type. Finally, participant location (lab, online) did not interact with lineup type or change the pattern of results, $F(1, 166) = .004$, $p = .951$.

Table 7

Factor loadings for all lineup items following principal components extraction as a function of lineup type for Experiment 3.

Lineup credibility Item	Full-Face Lineup Factor Loadings	Masked-Face Lineup Factor Loadings	Full-Face Lineup Communalities	Masked-Face Lineup Communalities
Trust	0.87	0.89	0.76	0.79
Information	0.74	0.75	0.55	0.56
Guilt	0.83	0.91	0.69	0.83
Lineup	0.91	0.91	0.83	0.82
Eigenvalue	2.82	3.00		

Table 8

Lineup credibility ratings as a function of witness confidence and lineup type for Experiment 3

	Masked-Face Lineup	Full-Face Lineup
Witness confidence ratings	Masked Perpetrator	Masked Perpetrator
High	49.20 (21.27)	49.96 (20.30)
Low	43.71 (20.06)	41.88 (19.06)
Absent	44.78 (20.81)	41.21 (18.46)

Note: Values in parentheses represent standard deviations.

CHAPTER 7: GENERAL DISCUSSION

When a crime is premeditated, perpetrators often don a disguise such as a ski mask to decrease the chances of subsequent identification. However, even though disguise is known to be a challenge for identification (Shapiro & Penrod, 1986), there are no empirically-based recommendations for how to administer a lineup for a masked perpetrator. In this dissertation, I extended one theoretically-based way (Manley et al., 2019) to construct a lineup for a masked perpetrator using a novel paradigm that simulates an actual crime and identification task. In three experiments, I examined identification accuracy of a masked perpetrator, witness meta-memory, and perceived credibility of an identification made from a masked-face lineup compared to a full-face lineup. Predictions for this dissertation were borne out of face recognition theories (Tanaka & Simonyi, 2016) in conjunction with the principles of transfer-appropriate processing and context reinstatement (Morris et al., 1977; Smith & Vela, 2001; Tulving & Thomson, 1973). First, I predicted that matching the perceptual appearance of lineup members to the masked perpetrator would increase identification accuracy. Further, I predicted that this pattern would occur even when ecologically valid materials were employed, rather than static images (e.g., Manley et al., 2019).

In Experiments 1 and 2, participant witnesses watched a mock crime featuring a masked perpetrator and were then administered either a masked-face lineup or a full-face lineup. The administered lineup was either target-absent or target-present. Arguably the most important finding from both experiments indicated that when the target was present, identification accuracy and discriminability were higher when participants were administered a masked-face lineup compared to a full-face lineup. Further, participants made faster, more confident target identifications from a masked-face lineups. These results indicate that providing a lineup where

the perceptual appearance of the members match the perceptual appearance of the perpetrator is beneficial to identification accuracy. To understand the potential mechanism of these results, I turn to the face recognition literature and examine the differences in face processing theories.

Well over thirty years of face recognition research has tested the holistic account of face recognition. In the past, some authors have used the terms “configural” and “holistic” interchangeably (see Richler, Mack, Gauthier, & Palmeri, 2009); however, for the current dissertation the two terms are separated by the distinction offered by Mauer et al., (2002): Holistic processing refers to bonding features together into a perceptual Gestalt, whereas configural processing refers to specific spatial measurements within and between features. Burton (2013) pointed out that the literature still has several issues to address. Though these issues specifically afflict the configural processing account of face recognition (i.e., differences in the spatial relations within and between facial features that make each face unique, also called second-order relations, Diamond & Carey, 1986; Mauer et al., 2002), they might also afflict the face recognition literature in general. For instance, most face recognition studies have used highly controlled, computer-generated or manipulated faces which may not be processed in the same fashion as real faces (e.g., Tanaka & Farah, 1993 used line drawings of faces). In the current study, I used a mock crime video at encoding and new, never-seen photographs in the lineups. Therefore, this common limitation did not affect the current dissertation.

The configural account of face processing asserts that the configural relations are important to face recognition (see Diamond & Carey, 1986; for review, see Maurer et al., 2002). In the current dissertation, the configural relations were not manipulated (e.g., the metric distances between the eyes was the same whether a face was masked or unmasked). Participants did see multiple views of the perpetrator’s masked face, enabling them to encode multiple

configural relations within the perpetrator's face (because these relations change as a person moves; see Burton et al., 2015). Most of the facial features were occluded at encoding (masked face), and similarly occluded in the masked-face lineup at retrieval. However, the additional features presented in the full-face lineup might *change* the configural relationships. Given the distances and interaction between features are part of the configural relationship (Maurer et al., 2002), participants might have perceived the relative distances between the covered and uncovered features to have a different configural relationships. For example, when the eyes are presented alone, they may seem an average distance apart; but when a wide jaw is revealed, the eyes now look narrow *relative* to the jaw. As applied to the current studies, identification accuracy should be higher from a masked-face lineup compared to a full-face lineup when a masked perpetrator was encountered. As such, this account would make the same predictions as the holistic account of face processing when transfer-appropriate processing is intact (i.e., for the masked-face lineup). Specifically, participants encoded a masked perpetrator. As such, the processing used to encode the perpetrator's face would evoke similar processing during retrieval when the lineup contained masked faces. However, the processing would differ during retrieval when the lineup members' faces were unmasked.

The holistic account provides a more thorough explanation for the current results for two reasons. There is no known evidence of configural-processing interference (i.e., the automatic configural representation interferes with featural processing) that would explain the decrease in identification accuracy from a full-face lineup. Further, large changes in configural relations do not decrease face recognition (for review see Burton et al., 2015). Specifically, Hole et al., (2002) asked participants to encode standard faces. Next, participants were tested with images of the same faces that were either stretched along both the x- and y-axes (vastly changing the

configural relations) or not stretched. Participants showed no reduction in recognition speed or accuracy for the stretched faces compared to the not-stretched faces (Hole et al., 2002). Before I delve into how the current results can better be explained by transfer-appropriate processing in conjunction with the holistic account, I discuss the results from the perspective of the Face-Space Model (Valentine, 2001).

The Face-Space model of face recognition (Valentine, 2001) would also not uniquely account for the current results. According to this account, faces are represented in a multidimensional space. Distinctive faces fall far from the center of the space and average faces fall close to the center. Faces that look similar would be close to one another in the face space, meaning people would have a difficult time distinguishing between them. Masked faces should, in theory, all look very similar. Distinctive features might be covered by the mask, leaving less to vary compared to the unmasked versions of the faces. According to this model, as faces are encoded, there is always encoding error. The more similar the faces, the higher the encoding error. Thus, masked faces would not only be located close together in the face space but be associated with higher encoding error (Valentine, Lewis, & Hills, 2016).

Further, encoding a face when it is masked is considered a difficult encoding condition (Shapiro & Penrod, 1986). As such, face recognition should decrease when a face is masked compared to unmasked. Unmasked faces should fall further apart in the space compared to masked faces. Because the unmasked versions contain additional features and thus less encoding error, these faces should be easier to distinguish from one another. However, in the current study, participants *always* studied a masked face. As such, it is more important to consider how a masked versus unmasked face would affect the retrieval of the originally encoded masked face.

Valentine (2001) stated that identification requires a judgment of which specific stimulus has been previously encountered. When participants encode a masked face, a dimension within the face-space must be created to account for each feature. Since only the eyes of the masked perpetrator were encoded, this is the only dimension that has been previously encountered. As such, participants would need to rely on this dimension to identify the perpetrator from the lineup. The masked or unmasked faces that make up the members of the lineup are also encoded when participants encounter the lineup. For the full-face lineups, participants encode new, previously unencountered features which increase the number of dimensions that would be registered into the face-space. However, these new features are irrelevant to the dimension that is important for accurate identification of the masked perpetrator: the eyes. Therefore, the extra features might increase the noise in the face-space and hinder identification accuracy because they are not helpful for accurate recognition of the eyes. If a participant can ignore the noise that the extraneous features from the full-face create, then identification accuracy of the masked-perpetrator should be similar from a masked-face lineup and a full-face lineup. However, it is more likely that since the masked-face lineup introduces less noise than the full-face lineup, identification accuracy should be higher for the former. It is important to note; these explanations were made after the results were already known; and thus, are postdictions. Even so, the Face-Space Model of face processing does not make a clear postdictions of whether a masked-face lineup or a full-face lineup would be more beneficial to identification accuracy. However, the model does give insight into the potential face processing mechanism between a masked and unmasked face

The holistic account of face recognition provides the best explanation of the current results. However, because the purpose of this dissertation was not to test the holistic account per

se, this explanation is indirect. Rather, the purpose of this dissertation was to test a predicted *application* of the account. Based on this account, a masked face cannot be processed holistically because the holistic representation is partially occluded. The mask thus forces featural processing of the exposed features (the eyes). This featural processing is then either matched at retrieval when a masked-face lineup was administered or mismatched when a full-face lineup was administered. When the full-face lineup was administered, results suggest that the holistic representation decreased participants' ability to match the featural representation from memory to the holistic representation presented to them (i.e., whole-face interference; Leder & Carbon, 2005). Whole-face interference led to a decrease in identification accuracy in full-face lineups compared to masked-face lineups.

Alternatively, whole-face interference may have been elicited through slightly different means. Specifically, when participants saw the masked perpetrator, they might have spontaneously generated the remainder of the face. However, when participants were administered a full-face lineup the generated face differed from the real full face. Thus, the mental holistic representation created interference with the real holistic representation and subsequently decreased identification accuracy. Regardless of the type of interference, these findings are consistent with the holistic account of face recognition in conjunction with principles of transfer-appropriate processing (Morris et al., 1977; Tulving & Thomson, 1973).

When the target was absent, identification accuracy was similar no matter if participants were administered a masked-face lineup or a full-face lineup. In other words, when the target was absent, administering a masked-face lineup did not increase foil identifications compared to a full-face lineup. In Experiment 3 of Manley et al., (2019), when participants encoded a masked face, they correctly rejected masked-face lineups more often than full-face lineups. However, in

Experiment 4 of the same study, when participants encoded a masked face, there was no difference in correct rejections as a function of lineup type. Therefore, a priori, it was unclear whether matching the perceptual appearance of lineup members to the perpetrator would impact rejection decisions. Given that rejection decisions (in both target-present and target-absent lineups) were similar for masked-face lineups compared to full-face lineups, it seems the perceptual match had minimal impact on this type of decision. In a legal context, this finding is positive because it suggests that there is no drawback to administering a masked face lineup when the target is absent.

With respect to the confidence-accuracy relationship, the CAC curves and gammas were similar for both lineup types—a finding that indicates that at the highest level of confidence participants were similarly likely to make an accurate suspect identification. However, in both Experiments 1 and 2, participants made numerically more high confidence guilty-suspect identifications from a masked-face lineup compared to a full-face lineup. The increase in guilty-suspect identifications from masked-face lineups were also associated with an increase in innocent suspect identifications. However, the increase in guilty-suspect identifications at the high level of confidence for masked-face lineups compared to full-face lineups was larger than the increase in innocent-suspect identifications. The primary message from this data is that the masked-face lineup is turning witnesses into *better* witnesses compared to the full-face lineup.

This is an important finding due to its application to legal settings. Specifically, confidence is highly persuasive to persons involved (i.e., jurors and investigators) with the investigative process (Cutler et al., 1989; Cutler et al., 1990; Wells, 1993; Wixted & Wells, 2017). As such, high confidence guilty-suspect identifications from masked-face lineups would be considered highly probative evidence to the investigative team in criminal proceedings

because investigative officers like to operate under only high certainty conditions (0 or 100, not in-between; Leippe & Eisenstadt, 2007). Assessing confidence ratings *after* the participant has made an identification is only one way to measure what witnesses believe about their own memory. As such, in Experiment 2 participants were also asked to rate their confidence for a subsequent identification decision.

In Experiment 2, in addition to examining identification accuracy, I assessed participant witnesses' meta-memory or their ability to monitor and control what they remembered and will remember in reference to the identification task. As in Experiment 1, identification accuracy was higher for masked-face lineups compared to full-face lineups in Experiment 2—an important replication of the effect. However, unlike post-ID confidence, pre-ID confidence was not *predictive* of accuracy (i.e., the slope of the CAC plot was essentially flat for both lineup types and the correlation was not significant between pre-ID ratings and accuracy). This finding is consistent with research that has found no relationship between pre-ID confidence and accuracy (Cutler & Penrod, 1989; Clark & Tunnicliff, 2001; Nguyen et al., 2018; Olsson & Juslin, 1999; Perfect, 2004; Sommer, Leuthold, & Schweinberger, 1995). Given that most people do not have experience with such an identification task (and four trials is likely not enough to master the task), this finding is not surprising. Research from the testing effect literature has shown that participants are typically overconfident when they predict their future performance on an unexperienced test. With practice, participants eventually become better calibrated and then underconfident in their test performance (Koriat et al., 1985). However, the tests in these studies usually contain multiple trials within each test (Adesope et al., 2017; Finn & Metcalfe, 2014; Koriat et al., 2002). Here, participants had one trial per test, and few trials total. As such, participants might not have had enough practice for the pre-ID confidence to calibrate.

Given novices tend to have trouble predicting their performance on a task they have never experienced (Kruger & Dunning, 1999), lineup preference data allowed for insight into the benefit of experience with the lineup task. On the initial trial, participants indicated no preference for either a masked-face lineup or a full-face lineup—until they had experienced the difficulty of identifying a masked perpetrator from a lineup. After this point, the majority of participants shifted to favor a masked-face lineup. This pattern was also reflected in pre-ID confidence ratings. Specifically, participants initially predicted their ability to make an accurate lineup decision to be similar from both a masked-face lineup and a full-face lineup. But once they had experienced the lineup setting on the first trial (regardless of whether it was a full-face or masked-face lineup), pre-ID confidence ratings increased for masked-face lineups and decreased for full-face lineups. Taken together, it seems that with identification experience, participant witnesses correctly believed that the masked-face lineup would lead to more accurate identification performance. This ultimately led participants to indicate a preference for a masked-face lineup. These results also challenge the idea that participants might perceive an identification decision from the masked-face lineup to be a more difficult task compared to a full-face lineup identification (Manley et al., 2019). Whereas people might believe more information leads to better decisions (Jacoby et al., 1974), people also tend to prefer easier tasks over more difficult ones when their performance is under scrutiny (Elliot & Dweck, 1988). Participants were likely aware their lineup decision performance was being measured. Further, a witness would know the high stakes of their identification performance in a criminal scenario. Thus, both witnesses in the lab and witnesses in a criminal scenario would likely prefer the easier task. Based on the shift in lineup preference to favor masked-face lineups, participants likely

perceived a masked-face lineup identification to be an easier task than a full-face lineup identification.

Given the profound implications of an under-confident or over-confident witness, these meta-memory findings are concerning. Because witnesses are unable to accurately assess whether they would be able to make a correct identification decision involving a masked perpetrator based on lineup type, investigators may choose to not administer a lineup to the witness—especially if the witness was underconfident. This decision could lead to a case going unsolved, leaving the perpetrator free to re-offend. Alternatively, if the witness was overconfident, investigators might administer a lineup which could lead to the witness identifying an innocent suspect who would then be at risk for wrongful conviction. Thus, one important recommendation to emerge from this dissertation is that scientists who study eyewitness memory should educate investigators that pre-ID confidence is not a reliable indicator of future identification accuracy.

The accumulation of results from Experiments 1 and 2 indicate that there is a replicable benefit to administering a masked-face lineup to a witness who saw a masked perpetrator commit a crime. This finding holds promise for application to legal scenarios. Specifically, the increase in target and foil identifications found for masked-face lineups compared to full-face lineups is not accompanied by a decrease in incorrect or correct rejections. An increase in target and foil identifications accompanied by a decrease in rejections would indicate that a masked-face lineup simply results in a liberal criterion shift (i.e., participants are not more accurate, just more likely to choose). Rather, for target-present lineups there was a pronounced increase in target identifications, a decrease in foil identifications, and no change in incorrect rejections. Further, identification decisions from masked-face lineups across both studies have been associated with

a similarly conservative response bias (negative value) when compared to full-face lineups. Yet, masked-face lineup decisions were associated with higher discriminability (d') than decisions from full-face lineups. Even when participants were forced to make an identification, target identification proportions still favored masked-face lineups over full-face lineups. The purpose of presenting witnesses with a second, forced-choice lineup (when the “not present” option was selected) was to equate participants’ decision criterion.

Importantly, identification accuracy patterns remained stable when data from Experiments 1 and 2 were combined. Identification accuracy of a masked perpetrator was higher from a masked-face lineup compared to a full-face lineup (i.e., a masked-face lineup superiority effect). This combined analysis indicated that the magnitude of the masked-face lineup superiority effect is moderate ($OR = 1.74 \sim d = 0.3$; Chen, Cohen, & Chen, 2010). However, this effect denotes that for target-present lineups, a witness is almost twice as likely to identify the target from a masked-face lineup as from a full-face lineup. Further, a witness is far less likely to identify a foil from a masked-face lineup as from a full-face lineup. In a judicial context, this difference could be critical. However, before policy recommendations can be made, the limitations of the masked-face lineup superiority effect should be further examined.

Identification Credibility

The benefits of administering a masked-face lineup are clear, but the question remained as to how these modified lineups would be judged by triers of fact. This is important to consider from an applied perspective because if a juror finds masked-face lineups to be less credible than full-face lineups, she may be more likely to devalue an identification of a masked perpetrator which could sway the verdict of guilt. Given that the ultimate goal of this type of research is to

improve judicial practices and change policy for the better, investigating how jurors perceive evidence is key.

In Experiment 3, participants read two case vignettes about a crime featuring a masked perpetrator and at least one witness. A witness identifies a suspect from a either masked-face lineup or a full-face lineup which was accompanied by a high or low confidence statement, or the confidence statement was absent. Participants then rated the credibility of the identification. Importantly, I found that mock jurors did not judge an identification made from a masked-face lineup to be less credible than an identification made from a full-face lineup. Unsurprisingly, participants did rate identifications associated with high confidence to be more credible than those associated with low confidence or when confidence was absent. This finding is consistent with previous work showing that people tend to weigh confidence ratings heavily when determining the credibility of an identification (Cutler et al., 1989; Cutler et al., 1990; Wells et al., 1979). The mock juror's ratings from Experiment 3 is further supported by the CAC curves from Experiments 1 and 2. Specifically, within the same level of confidence, it seems to not matter what procedure a witness was given. Therefore, it may be more beneficial for jurors to focus on the confidence ratings of the witness rather than the lineup procedure. Though this may be oversimplifying the problem as reflected in the breakdown of the data from the CACs.

It is unclear, however, if identification credibility ratings were superficially low ($M = 45.12\%$, $SD = 20.15$) due to the perpetrator being masked. For instance, participants may discredit the witness's view of the perpetrator's face due to the mask; thus, the participant rated credibility to be low regardless of which lineup was administered. Additional research should be conducted to establish how mock jurors would rate the identification credibility of an unmasked perpetrator using these materials. There are few studies that specifically compare the credibility

of an identification made of a masked versus unmasked perpetrator (although disguise was one of the twenty identification factors manipulated in Cutler et al., 1990). Therefore, this additional condition would enable an appropriate comparison to the identification credibility of a masked perpetrator.

Future studies also might consider having mock jurors complete a demonstration where they attempt to identify a masked perpetrator from a masked-face lineup and a full-face lineup. This experience with the task might lead them to better understand the difficulty of the task (similar to how the pre-ID meta-memory measures changed with experience in Experiment 2). As such, a demonstration might have the same impact on jurors' understanding of the identification task from a masked-face compared to a full-face lineup.

A Comparison to Previous Work

One aim of Experiments 1 and 2 was to construct a lineup for a masked perpetrator using a novel paradigm that mimics an actual crime and lineup identification. Many face recognition studies have not accounted for the within-face variability that real faces possess (Burton, 2013). Rather, the same static images are used during both the learning and test phases. This might be more representative of picture recognition rather than face recognition. This dissertation, however, used a mock crime video. The video showed multiple angles of the perpetrator's masked face during the learning phase and a new image that participants had never seen during the test phase. Importantly, the transfer-appropriate patterns from Manley et al. (2019), which were found using highly-controlled, computer-generated faces, were also found here using real faces. Given the differences in the materials and procedure between the two studies, direct comparisons should be interpreted cautiously. However, it is important to compare the patterns of results between Manley et al. and those obtained here (Experiments 1 and 2).

One notable similarity is that the pattern of results remained consistent when participants were asked to identify a masked face. That is, the benefit of administering a masked-face lineup for a masked perpetrator persisted even after adjusting for ecological validity. Overall, identification accuracy was lower with these materials compared to the materials used in Manley et al., but the important patterns remained the same. Specifically, that participants were better able to identify a masked perpetrator from a masked-face lineup compared to a full-face lineup when the target was present; and rejection accuracy did not decline in target-absent lineups as a result of masked-face lineups compared to full-face lineups. Another similarity between the two studies was that participants were more confident in target identifications than incorrect identifications for target-present lineups. I now turn to the applied implications of the current dissertation.

Applied Implications

Another important contribution of this dissertation is how it applies and impacts critical issues within the criminal justice system. A mask can make the task of identification difficult and has been shown to negatively impact accuracy (Shapiro & Penrod, 1986). Accurate identification is vital to the criminal justice system, given that an eyewitness identification is one of the strongest forms of evidence (Wells, 1993); yet *misidentification* is a leading contributor to wrongful convictions (Innocence Project, 2012). Investigators and researchers alike know there is a problem with eyewitness identification accuracy. It is also known that disguise is a contributing factor to eyewitness misidentification (Shapiro & Penrod, 1986). However, Manley et al. was the first study that attempted to address the issue from a system variable perspective (i.e., a controllable change can be made within the criminal justice system; Wells, 1978) by manipulating and testing a novel lineup procedure.

The current study provides preliminary evidence that administering a masked-face lineup might be a good alternative to a full-face lineup for a masked perpetrator. Further, I now have evidence that mock jurors would not discount a masked-face lineup identification compared to a full-face lineup identification. These findings are promising in light of their potential applications in legal contexts. Not only could masked-face lineups lead to accurate identification of a masked perpetrator, these data suggest that this type of lineup would not lead to a consequent increase in foil identifications regardless of whether the culprit is present in the lineup. In the lab, for a target-present lineup there is an established target and a 1/6 chance of a chooser selecting the correct person. However, there is a 5/6 chance of a chooser selecting a foil. Given the difference in chance rates for a target identification compared to a foil identification, the high rates of target identifications and low rates of foil identifications from masked-face lineups found here are favorable. However, in a judicial context the base rate is not as clear and there is no ground truth. That is, it is not known whether the suspect in a lineup is innocent or guilty. This is why a witness's identification can be so influential in the absence of corroborating evidence. It is also why it is so important that witnesses are given the tools to be as accurate as possible prior to making an identification. The goal of research that examines investigative procedures is to provide legal actors (be they investigators, jurors, etc.) with empirically-based recommendations to improve the likelihood of obtaining an accurate eyewitness account. The more accurate an eyewitness's account is the less likely it is that investigators make a mistake. As such, research that helps improve upon legal practices and thus accuracy are vital, and the current dissertation provides one manner in which to do so.

An additional application of the current dissertation is to the identification of individuals not involved in a crime who wear head coverings or face coverings. For example, both women

and men from Muslim cultures wear attire (i.e., a Hijab, Niqab, or a Burqa) that occludes part or most of the face. During the identification process, many of these individuals are asked to remove their coverings (Hauser, 2018). However, there may be an alternate option: an identification where the person is wearing the same head or face covering. Similar to the masked-face lineup, if a witness sees a person who is wearing a hijab, identification accuracy should increase if the person is still wearing a hijab at retrieval compared to if the person has removed the hijab. This would enable officials to accurately identify people who wear head coverings for religious reasons without violating their human rights.

This idea could be applied to other-race identifications especially in consideration of head coverings and masks. The other-race (or cross-race) effect is the robust finding that people are better able to identify individuals from their own race compared to individuals from an other race (for review see Meissner & Brigham, 2001). The proposed mechanism of this effect is rooted in differential face processing of same and other race faces. Specifically, people that show the other-race effect tend to process faces of their own races more holistically than faces of other races (Tanaka, Kiefer, & Bukach, 2004). Further, people who are susceptible to the other race effect tend to encode race as a facial feature, ignoring individuating information that might increase recognition (Meissner, Brigham, & Butz, 2005).

However, people who are less susceptible to the other-race effect tend to process both own-race faces, and other-race faces holistically (Tanaka et al., 2004). Researchers have suggested that lower susceptibility to the other-race effect might be a result of higher exposure to other-race faces compared to individuals who were highly susceptible to the other-race effect (Meissner & Brigham, 2001; Tanaka et al., 2004); however, exposure does not always lead to a reduction in the other-race effect (Chiroro & Valentine, 1995; Ng & Lindsay, 1994). The

accumulated research suggests then that other-races faces might be processed similarly to a masked face. For example, Caucasian and Asian participants were asked to identify the top or bottom half of a composite face that was either same race or other race. Results showed that recognition was more disrupted when the composite stimulus was a same-race face compared to an other-race face (Michel, Rossion, Han, Chung, & Caldara, 2006). These findings suggest that same race faces are processed more holistically than other-race faces.

Given these results, it is important to consider their implications for the current dissertation. Specifically, would matching the perceptual appearance of lineup members to an masked-face of an other-race have the same benefits as shown here for predominately same-race identifications? A mask would force a witness to process only the exposed features of the other-race face which would not enable them to only focus on race-specific features. Of course, holistic processing of faces would lead to better identification performance of both other- and same-race faces. However, if witnesses see a masked perpetrator, holistic processing is disrupted. Therefore, if witnesses encounter a masked perpetrator of a same or other race, the mask forces witnesses to use only featurally processing for both faces. It would be interesting to examine if administering a masked-face lineup for an other-race masked perpetrator would equate other-race face identifications to same-race face identifications. Witnesses would no longer be able to only use the race specific feature when encoding an other-race masked face. Of course, this assumption might be contingent upon the race specific feature.

Moreover, if witnesses are administered a full-face lineup, whole-face interference might not affect identification performance for an other-race lineup to the extent it affects it for a same-race face lineup. Given that people tend to process other-race faces using a race specific feature, identification performance for full-face lineup might be worse than performance for a masked-

face lineup due to the race-specific feature interference rather than whole-face interference. How the current findings might apply to the other-race effect is an important implication to consider given the evidence that other-race identifications have resulted in wrongful convictions (e.g., Ronald Cotton). Thus, consideration of masked perpetrator identification when the perpetrator is an other race warrants attention from researchers.

Limitations and Unanswered Questions

Although administering a masked-face lineup is better than administering a full-face lineup for a masked perpetrator in a laboratory setting, more research needs to be done prior to making a policy recommendation. For instance, participants in this study had a good view of the masked perpetrator. The perpetrator's face was well-lit, the video was in high definition, and the view of the perpetrator's face was close and from multiple angles. Further, the time between viewing the mock crime and when the lineup was administered was very short (i.e., 5 min). In an investigative setting, the interval between a witnessed event and an identification can range from minutes to years (Semmler, Dunn, Mickes, & Wixted, 2018). The limitations of this retention variable should be examined in future work.

Another important issue to consider is that of familiarity (Burton, 2013). Specifically, that people are face recognition experts of familiar faces and surprisingly poor at recognizing unfamiliar faces (Burton, 2013; Burton et al., 2015). Again, this issue is important to account for when testing face recognition especially in relation to eyewitness identification. The statistics on how often a witness is familiar with the perpetrator is unknown; however, it is important to account for familiarity when examining eyewitness memory. In this dissertation, none of the participant witnesses were familiar with or knew the perpetrator actors in any of the videos.

Given that some crimes, such as sexual assault are more commonly committed by an acquaintance, future research should examine identification of a familiar masked perpetrator.

In the case from the introduction (*Dupuie v. Egeler, 1977*), witnesses only saw the masked face of the perpetrator, yet investigators administered a full-face lineup to the witnesses. It was only after the witnesses could not make an identification that investigators administered witnesses a masked-face lineup from which they immediately made an identification, Thus, knowing what impact initial exposure to full-faces (i.e., potential whole-face interference) has on subsequent identification accuracy from a masked-face lineup is important to investigative application. It is unclear how prior exposure to a full-face lineup may affect identification accuracy from a masked-face lineup. Specifically, exposure to the full faces may lead to whole-face interference (Leder & Carbon, 2005). The interference caused by the holistic representation of the face could render the representation of the eyes less accessible and decrease identification accuracy in the masked-face lineup. Conversely, whole-face interference might be limited to active exposure to the unmasked faces and not effective when the witness is not actively viewing these unmasked faces.

Another potential limitation of the current dissertation is that I used only one type of mask. Although a ski mask is a popular method of disguise, there are other types of concealment used by perpetrators (e.g., stocking mask, sunglasses and hats, full-coverage masks, etc.). Depending on the degree of occlusion and/or distortion of the facial features, the lineup proposed by Manley et al. and in this dissertation may have limited utility. For instance, masks that completely occlude the face differ from disguises that distort the face. If no part of the face can be seen, it cannot be processed neither holistically nor featurally. As such, there is no way to construct a lineup for this type of mask. A similar problem applies to a nylon stocking mask. It is

unclear whether a nylon mask would disrupt holistic processing because it does not occlude any part of the face. A nylon stocking mask distorts rather than occludes facial features, which changes the shape of the features. Further, it cannot be guaranteed that the mask will distort the features in the same manner each time the mask is put on the perpetrator's face. However, Davies and Flin (1984) did find evidence of a transfer-appropriate pattern using nylon stocking masks. Face identification was superior when the target faces in the lineup matched what was originally encoded compared to when the faces in the lineup did not match. Future research should investigate the theoretical mechanisms of these findings and focus on conceptual replication. Replication is vital (Zwann, Etz, & Lucas, 2018), especially when findings are potentially applicable to the criminal justice system. For at least one common type of disguise, however, the masked lineup is effective at increasing identification accuracy. Nevertheless, this line of research should move closer to applied settings before policy recommendations can be made, results from laboratory studies are promising, and importantly—replicable.

Given that eyewitness misidentification contributes heavily to wrongful convictions, it is important to examine methodologies that may increase identification accuracy. The ultimate goal of eyewitness identification research is to encourage investigative approaches that lead to a decrease in misidentifications and an increase in accurate identification decisions. One way to do this is to leverage cognitive theories and principles to inform better lineup construction strategies. In the present dissertation, I showed that when a witness encounters a masked perpetrator, administering a masked-face lineup increases identification accuracy compared to a full-face lineup. Moreover, with experience, the majority of participants actually prefer a masked-face lineup over a full-face lineup. Further, I showed that mock jurors do not discredit an identification of a masked perpetrator from a masked-face lineup compared to a full-face lineup.

The idea to match the perceptual appearance of lineup members to the perceptual appearance of the originally encoded face was vaguely considered decades ago (Cutler, 1988). With the advances in eyewitness identification research, this idea warrants more serious consideration—as it could prove a powerful way to bring guilty perpetrators to justice.

REFERENCES

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*, 659-701.
- Bartlett, J. C., Searcy, J. H., & Abdi, H. (2003). What are the routes to face recognition? In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects and scenes: Analytic and holistic processes* (pp. 21-52.). New York, NY: Oxford University Press.
- Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*, 115-147.
- Brewer, N., Weber, N., & Semmler, C. (2013). A role for theory in eyewitness identification research. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Volume II: Memory for people* (pp. 201-218). Mahwah, NJ: Erlbaum.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.
- Bruce, V., & Young A. (2012). Recognizing faces. *Face Perception* (pp. 253-314). New York, New York: Psychology Press.
- Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology, 66*, 1467-1485.
- Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science, 10*, 482-496.
- Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*, 243-248.
- Busey, T. A., Tunnickliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*, 26-48.
- Chandler, C. C., & Fisher, R. P. (1996). Retrieval processes and witness memory. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition (2nd ed.)*. Memory (pp. 493-524). San Diego, CA: Academic Press.
- Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation, 39*, 860-864.

- Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology*, *48A*, 879-894.
- Chua, E. F., Schacter, D. L., & Sperling, R. A. (2008). Neural correlates of metamemory: A comparison of feeling-of-knowing and post-ID confidence judgments. *Journal of Cognitive Neuroscience*, *21*, 1751-1765.
- Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behavior*, *25*, 199.
- Cutler, B. L. (1988). Context reinstatement and eyewitness identification. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 231 - 244). Hoboken, NJ: John Wiley & Sons Ltd.
- Cutler, B. L., & Penrod, S. D. (1989). Forensically relevant moderators of the relation between eyewitness identification accuracy and confidence. *Journal of Applied Psychology*, *74*, 650-652.
- Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior*, *14*, 185-191.
- Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987a). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, *72*, 629-637.
- Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987b). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior*, *11*, 233-258.
- Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior*, *12*, 41-55.
- Cutler, B. L., Penrod, S. D., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behaviour*, *1*, 113-134.
- Dalton, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory & Cognition*, *21*, 223-234.
- Davies, G., & Flin, R. (1984). The man behind the mask—disguise and face recognition. *Human Learning: Journal of Practical Research & Applications*, *3*, 83-95.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107-117.

- Dillon, M. K., Jones, A. M., Bergold, A. N., Hui, C. Y., & Penrod, S. D. (2017). Henderson instructions: Do they enhance evidence evaluation? *Journal of Forensic Psychology Research and Practice, 17*, 1-24.
- Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24*, 58-61.
- Dupuie, William L.,Petitioner-appellant, v. Egeler, Charles E. ,Warden, Respondent-appellee, 552 F.2d 704 (6th Cir. 1977).
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*, 5-12.
- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*, 431-439.
- Farah, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain specificity in visual object recognition. *Current Directions in Psychological Science, 1*, 164-169.
- Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*, 628-634.
- Field, A. (2009). *Discovering statistics using SPSS*. Third Edition. Thousand Oaks, CA, Sage.
- Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction, 32*, 1-9.
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The Cognitive Interview*. Springfield, IL: Charles C Thomas Publisher.
- Foley, M. A., & Foley, H. J. (1998). A Study of Face Identification: Are People Looking Beyond Disguises? In D. L. Best, & M. J. Intons-Peterson (Eds.), *Memory distortions and their prevention* (pp. 29-47). Mahwah, NJ: Erlbaum.
- Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception, 19*, 661-673.
- Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science, 19*, 207-208.
- Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision research, 37*, 1673-1682.

- Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *The American Journal of Psychology*, 385-401.
- Gibling, F., & Davies, G. (1988). Reinstatement of context following exposure to post-event information. *British Journal of Psychology*, 79, 129-141.
- Haig, N. D. (1986). Exploring recognition with interchanged facial features. *Perception*, 15, 235-247.
- Harris, A. M., & Aguirre, G. K. (2008). The effects of parts, wholes, and familiarity on face-selective responses in MEG. *Journal of Vision*, 8, 4-4.
- Hauser, C. (2018). Women forced to remove hijabs for mug shots settle with New York City. Retrieved April 9, 2019. <https://www.nytimes.com/2018/02/28/nyregion/muslim-hijab-nypd.html>
- Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, 33, 98-106.
- Hockley, W. E., Hemsworth, D. H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition*, 27, 128-138.
- Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, 23, 65-74.
- Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, 31, 1221-1240.
- Hourihan, K. L., Benjamin, A. S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition*, 1, 158-162.
- Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Thousand Oaks, CA, Sage.
- Innocence Project. (2012a). New Jersey Supreme Court Issues New Jury Instructions that will Greatly Improve the Way Courts Handle Identification Evidence. Retrieved November 28, 2017. <https://www.innocenceproject.org/new-jersey-supreme-court-issues-new-jury-instructions-that-will-greatly-improve-the-way-courts-handle-identification-evidence/>
- Innocence Project. (2012b). Henderson Instructions. Retrieved November 28, 2017. <https://www.innocenceproject.org/wp-content/uploads/2017/06/NJ-Jury-Instruction.pdf>
- Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, 11, 63-69.

- Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review*, *25*, 2356-2364.
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313-323.
- Jones, A. M., Bergold, A. N., Dillon, M. K., & Penrod, S. D. (2017). Comparing the effectiveness of Henderson instructions and expert testimony: Which safeguard improves jurors' evaluations of eyewitness evidence? *Journal of Experimental Criminology*, *13*, 29-52.
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1304-1316.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302-4311.
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401-415.
- Kaiser, H. F. (1991). Coefficient alpha for a principal component and the Kaiser-Guttman rule. *Psychological Reports*, *68*, 855-858.
- Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory & Cognition*, *12*, 60-70.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: judgment of learning exhibit increased underconfidence-with-practice. *Journal of Experimental Psychology: General*, *131*, 147-162.
- Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, *49*, 58-69.
- Kramer, R. S., & Ritchie, K. L. (2016). Disguising Superman: How glasses affect unfamiliar face matching. *Applied Cognitive Psychology*, *30*, 841-845.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121-1134.
- Lampinen, J. M., Neuschantz, J. S., & Cling, A. D. (2012). *The psychology of eyewitness identification*. New York, New York, Psychology Press.

- Leder, H., & Carbon, C. C. (2005). When context hinders! Learn–test compatibility in face recognition. *The Quarterly Journal of Experimental Psychology*, *58*, 235-250.
- Leippe, M., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Vol. 2. Memory for people* (pp. 377-426). Mahwah, NJ: Erlbaum.
- Malpass, R. S., & Devine, P. G. (1981a). Guided memory in eyewitness identification. *Journal of Applied Psychology*, *66*, 343-350.
- Malpass, R. S., & Devine, P. G. (1981b). Eyewitness identification: lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482-489.
- Manley, K. D., Chan, J. C. K., & Wells, G. L. (2018). Do masked-face lineups facilitate eyewitness identification of a masked individual? *Journal of Experimental Psychology: Applied*, *in press*.
- Mansour, J. K., Beaudry, J. L., Bertrand, M. I., Kalmet, N., Melsom, E. I., & Lindsay, R. C. L. (2012). Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. *Law and Human Behavior*, *36*, 513-526.
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*, 255-260.
- McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*, 38-41.
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3-35.
- Meissner, C. A., Brigham, J. C., & Butz, D. A. (2005). Memory for own-and other-race faces: A dual-process approach. *Applied Cognitive Psychology*, *19*, 545-567.
- Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, *16*, 340-372.
- Metzger, M. M. (2001). Which transformations of stimuli are the most disruptive to facial recognition? *Perceptual and Motor Skills*, *92*, 517-526.
- Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychological Science*, *17*, 608-615.

- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519-533.
- Noble, A. (2013). Ski masks, a fashion accessory to crime in D.C.: Clearly are problem, but ban not feasible. Retrieved November 28, 2017.
www.washingtontimes.com/news/2013/jan/30/ski-mask-is-fashion-accessory-to-crime-in-dc/
- Ng, W. J., & Lindsay, R. C. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology*, *25*, 217-232.
- Nguyen, T. B., Abed, E., & Pezdek, K. (2018). Postdictive confidence (but not predictive confidence) predicts eyewitness memory accuracy. *Cognitive Research: Principles and Implications*, *3*, 32.
- Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identifications? *Journal of Applied Psychology*, *84*, 42-49.
- Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247-255.
- Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology*, *16*, 387-398.
- Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, *18*, 157-168.
- Report of the Special Master, State v. Henderson, No. A-8-08. (2011, June 18). Retrieved from <http://www.judiciary.state.nj.us/pressrel/HENDERSON%20FINAL%20BRIEF%20.PDF%20%2800621142%29.PDF>
- Rhodes, G., Byatt, G., Tremewan, T., & Kennedy, A. (1997). Facial distinctiveness and the power of caricatures. *Perception*, *26*, 207-223.
- Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research*, *49*, 2856-2861.
- Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition*, *20*, 143-169.
- Russo, J. E. (1974). More information is better: A reevaluation of Jacoby, Speller and Kohn. *Journal of Consumer Research*, *1*, 68-72.

- Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied*, 24, 400-415.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, 100, 139-156.
- Slavic, E., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgement. *Organizational Behavior and Human Performance*, 6, 649-744.
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta analysis. *Psychonomic Bulletin & Review*, 8, 203-220.
- Sommer, W., Heinz, A., Leuthold, H., Matt, J., & Schweinberger, S. R. (1995). Metamemory, distinctiveness, and event-related potentials in recognition memory for faces. *Memory & Cognition*, 23, 1-11.
- State v. Courteau*, 461 A.2d 1358 (R.I. 1983).
- State v. Fierro*, 489 P.2d 713, 107 Ariz. 479 (A. Z.1971).
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46, 225-245.
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, 25, 583-592.
- Tanaka, J. W., & Simonyi, D. (2016). The “parts and wholes” of face recognition: A review of the literature. *The Quarterly Journal of Experimental Psychology*, 69, 1876-1889.
- Terry, R. L. (1994). Effects of facial transformations on accuracy of recognition. *The Journal of Social Psychology*, 134, 483-492.
- Toseeb, U., Keeble, D. R., & Bryant, E. J. (2012). The significance of hair for face recognition. *PloS One*, 7, 1-8.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, 80(5), 352.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- US Court of Appeals for the Sixth Circuit - 552 F.2d 704 (6th Cir. 1977)

- Valentine, T. (2001). Face-space models of face recognition. In M. J. Wenger & J. T. Townsend, (Eds.), *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges* (pp. 1-25). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.
- Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, *69*, 1996-2019.
- Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, *36*, 1546-1557.
- Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571.
- Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, *12*, 338-344.
- Wells, G. L., Lindsay, R. C., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology*, *64*, 440-448.
- Wilford, M. M., & Wells, G. L. (2010). Does facial processing prioritize change detection? Change blindness illustrates costs & benefits of holistic processing. *Psychological Science*, *21*, 1611-1615.
- Williams v. State* 395 N.E.2d 239 Supreme court of Indiana No. 1278S281 (I.N. 1979).
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*, 10-65.
- Yin, R. K. (1969) Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141-145.
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747-759.
- Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron*, *44*, 889-898.
- Zwaan, R., Etz, A., Lucas, R., & Donnellan, M. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, E120.

APPENDIX A: EXPERIMENTS 1 AND 2 MATERIALS

Lineup Instructions

“On the next page you will see a lineup.

Please choose the person you believe to be the [wallet, bike, identity, computer] thief in the video you just watched.

Note this person may not be present in the following lineup.

If you do not see the wallet thief please choose the "Not Present" option.”

Bike Theft Video Screenshots



Example of a Bike Thief Lineup (Full-Face, Target-Absent)

Please choose the person you believe to be the **bike thief** in the video you just watched. Note this person may not be present in the following lineup. If you do not see the bike thief please choose the "Not Present" option.

Scroll down for the next question.



Please rate your confidence in the choice you just made on the scale below (0 meaning not confident or a guess, and 100 meaning absolute certainty you have identified the perpetrator).

0 10 20 30 40 50 60 70 80 90 100



Identity Theft Video Screenshots



Example of an Identity Thief Lineup (Full-Face, Target-Present)

Please choose the person you believe to be the **identity thief** in the video you just watched.
Note this person may not be present in the following lineup.
If you do not see the identity thief please choose the **"Not Present"** option.

Scroll down for the next question.

2 	4 	3 	Not Present
6 	5 	1 	

Please rate your confidence in the choice you just made on the scale below (0 meaning not confident or a guess, and 100 meaning absolute certainty you have identified the perpetrator).

0 10 20 30 40 50 60 70 80 90 100



Computer Theft Video Screenshots



Example of a Computer Thief Lineup (Masked-Face, Target-Absent)

Please choose the person you believe to be the **computer thief** in the video you just watched. Note this person may not be present in the following lineup. If you do not see the computer thief please choose the "Not Present" option.

Scroll down for the next question.



Please rate your confidence in the choice you just made on the scale below (0 meaning not confident or a guess, and 100 meaning absolute certainty you have identified the perpetrator).



Wallet Theft Video Screenshots



Example of a Wallet Thief Lineup (Masked-Face, Target-Present)

Please choose the person you believe to be the **wallet thief** in the video you just watched. Note this person may not be present in the following lineup. If you do not see the wallet thief please choose the "Not Present" option.

Scroll down for the next question.



Please rate your confidence in the choice you just made on the scale below (0 meaning not confident or a guess, and 100 meaning absolute certainty you have identified the perpetrator).



APPENDIX B: EXPERIMENT 3 MATERIALS**Case Vignette 1**

On the night of January 19, 2015, Mr. Daniel Kreuger entered the Royal Palace Casino, walked across the lobby and out the back exit. He repeated this behavior once again, before entering the casino a third time and walking to the bar, where he ordered a “Crown and Coke” from the assistant manager and bartender, Ms. Sarah Wisk. When Wisk told Kreuger that he would be charged \$6.00 for the drink, Kreuger declined and asked for “just a Coke” instead. Kreuger sat at the bar sipping a Coke for approximately 20 minutes before an individual entered the casino wearing a ski mask with a hole for his eyes. He also wore black leather gloves, dark denim pants, a red flannel shirt, and a black leather jacket. Upon entering, the security guard, Ms. Clara Mentzer, stopped him and told him he must remove the mask to be inside the casino or he needed to leave. The masked man struck Mentzer in the head with the butt of a hand gun, grabbed her by the shirt, and hurled her to the floor. He then began to shout at Mentzer, demanding to know where her gun was located and for her to give it to him. The masked individual then started pointing his gun at Mentzer as well as at random patrons and staff inside the casino. At this point Kreuger moved to the main casino doors to act as a lookout. The robber then led Mentzer at gunpoint to the back of the casino behind the bar to the room where the safe was located. When they went by the bar, the gunman demanded Wisk join them in the safe. Wisk and Mentzer opened the casino safe for the masked robber. While the gunman and two hostages were inside the safe, Kreuger proceeded to rob items from the other casino customers. Meanwhile in the safe Mentzer and Wisk attempted to back out of the safe and shut the door which would have effectively locked the robber inside. However, Wisk tripped on the way out which alarmed the robber. The robber shot three bullets into Wisk’s leg and chest. The two

bandits fled in a dark SUV that had been backed into a handicapped parking space just outside the main entrance. Mentzer then called the police. Krueger was arrested later that day but would not reveal the name of his accomplice. On January 21st, Detective Tracy Dark prepared a lineup based on the victims' description of the robber. She administered a 6-person lineup consisting of **faces wearing masks [unmasked faces]** to Mentzer. The following individual marked in red reflects Mentzer's decision.

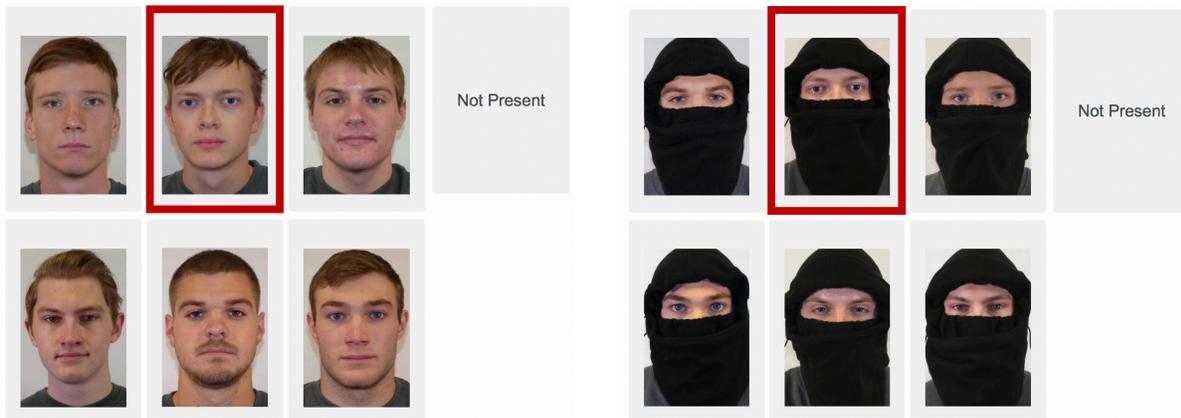
In the low and high confidence conditions, one of the following sentences would be included in the report:

LOW:

At the time of the identification, she expressed low confidence in her decision saying, "I think that kind of looks like the man who robbed the casino--but I am not sure."

HIGH:

At the time of the identification, she expressed high confidence in her decision saying, "That is the guy who made us open the safe, I am almost positive."



Depending on the lineup condition, one of the two above lineups accompanied the report.

Case Vignette 2

On June 13, 2015, two roommates hosted a high stakes poker game at their apartment in North Andover. The apartment was on the second floor of a two-family home. The poker room was in the rear of the apartment and was accessible by a rear door. The poker game was a regular event that attracted eight to ten friends on average. Each card player entered the game with one hundred dollars or more, with the option to reenter the game with more cash if he lost his initial stake. On the night in question, the poker game started sometime after 9 P.M. with a small group that, around 10:30 P.M., had grown to eleven card players. Among this group was Christopher Evans, known as "Shorty." After losing his money, Evans stayed in the apartment, where the victims observed him sending text messages on his cellular telephone. Sometime after Evans was out of the game, a masked man entered the apartment. The mask covered the assailant's entire face except for the eyes. The masked man was armed with a hand gun and demanded that the card players empty their pockets and place their cellphones on the table. After collecting the items, the assailant bound the victims' hands. Initially, Evans pretended to be a victim and, as with the others, the masked robber bound his hands and demanded his cash. Later as events progressed, Evans announced that he "set [the robbery] up" and that he was "hungry [for money]." Evans then started to argue with the armed robber over whether they should shoot the other card players because they were witnesses. The confrontation escalated. Evans shoved the masked robber who fell into a table. The robber got to his feet and shot Evans four times in the chest. The robber then ran out of the apartment. After the robber escaped, two of the victims freed themselves and, from a window in the apartment, observed the robber getting into a dark blue Mitsubishi Galant automobile bearing Massachusetts license plates. The victims got into a vehicle and pursued the robber until they reached an entrance to Route 495. At that point, they

abandoned the chase and returned to the apartment, where they were met by Detective Daniel G. Bullock of the North Andover police department. Detective Bullock commenced his investigation based on the victims' descriptions of the suspect and the getaway vehicle. Two days after the robbery, Detective Bullock prepared and showed a 6-person lineup consisting of **unmasked faces [faces wearing masks]** to the owner of the apartment. The following individual marked in red reflects the apartment owner's decision.

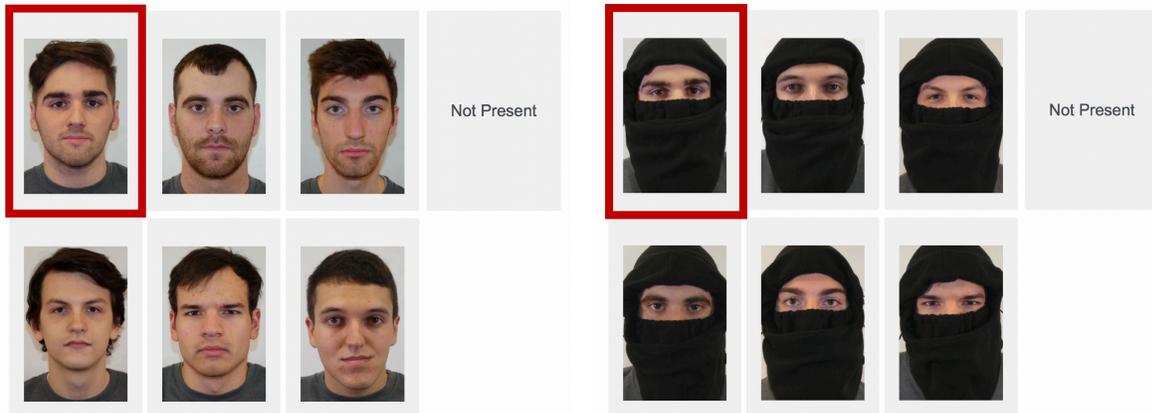
In the low and high confidence conditions, one of the following sentences would be included in the report:

LOW:

At the time of the identification, the man made his decision with low confidence stating, "I cannot be certain, but I think that might be the guy."

HIGH:

At the time of the identification, the man rated his confidence as high saying, "I am pretty certain that's the guy who took our money."



Depending on the lineup condition, one of the two above lineups accompanied the report.

APPENDIX C: PARTICIPANT DEMOGRAPHICS

Table A-C

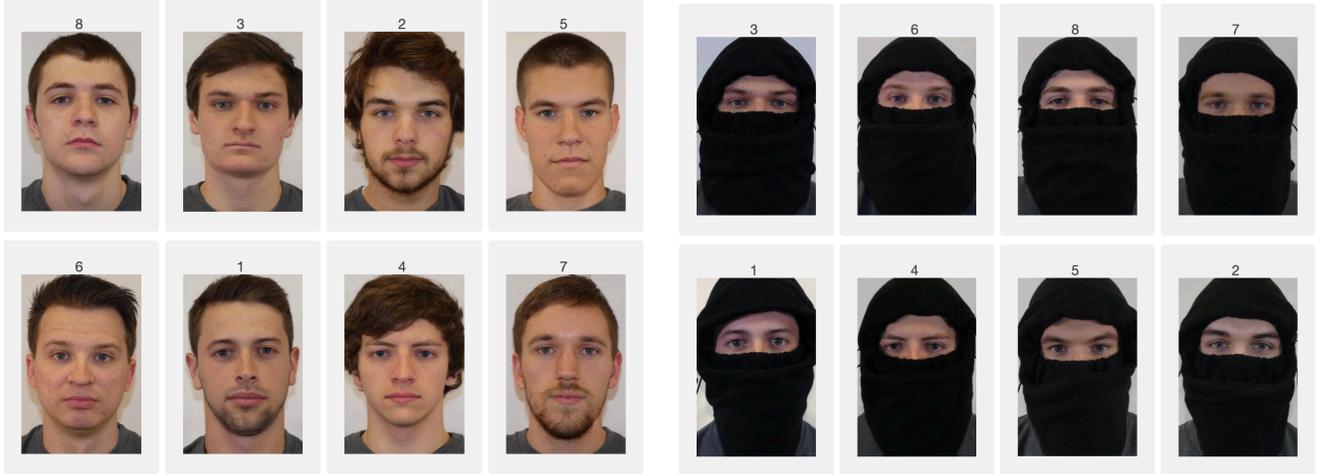
Participant demographics for each experiment

	Experiment 1	Experiment 2	Experiment 3
Age			
Average	18.99 (1.36)	19.06 (1.89)	21.03 (5.90)
Range	18-29	18-43	18-60
Ethnicity			
American Indian or Alaska Native	<1%	<1%	0%
Black or African American	5%	3%	7%
Chose Not to Respond	<1%	<1%	2%
East Asian	2%	2%	1%
Hispanic or Latino/a	6%	4%	9%
Other	2%	2%	<1%
South/Southeast Asian	3%	2%	3%
West Asian/Middle Eastern	0%	<1%	0%
White/Caucasian	80%	84%	78%
Gender			
Man	38%	37%	46%
Woman	60%	62%	54%
Other	<1%	<1%	0%
Chose Not to Respond	<1%	<1%	0%

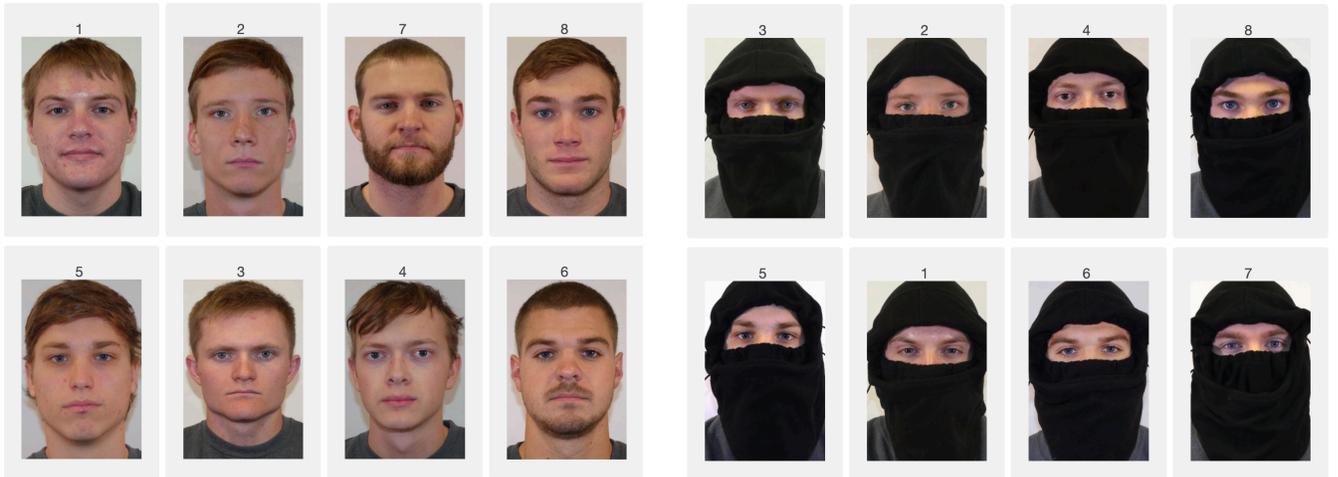
Note: Values in parentheses represent standard deviations.

APPENDIX D: PILOT LINEUPS

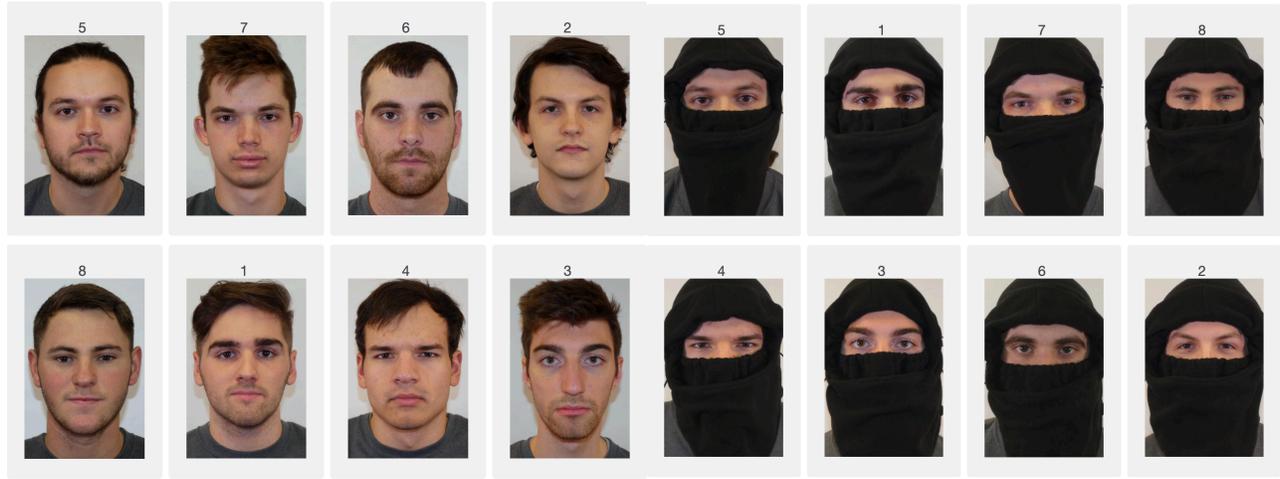
Pilot Lineups for Identify Thief



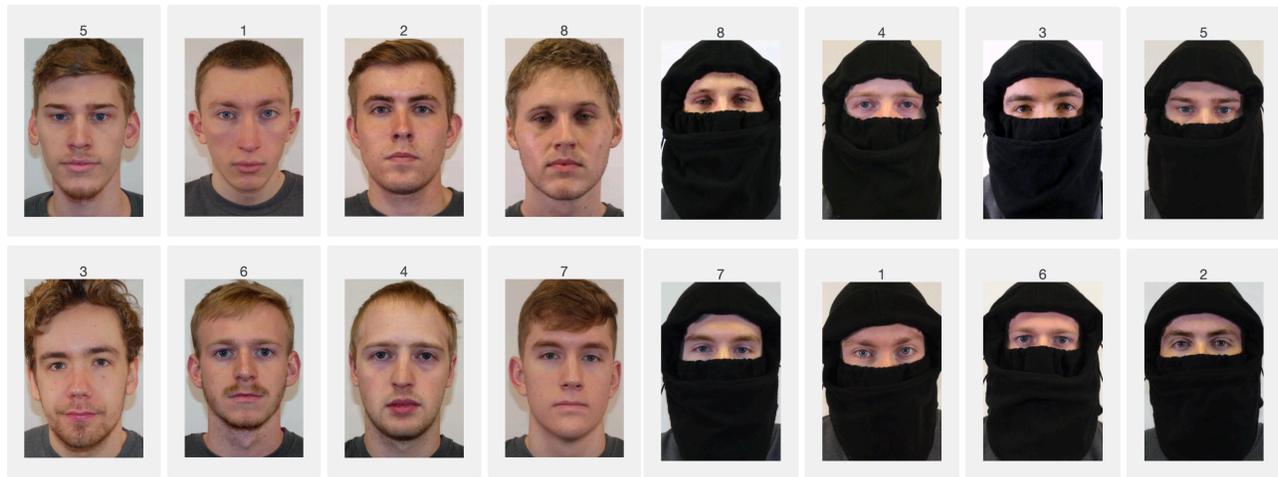
Pilot Lineups for Wallet Thief



Pilot Lineups for Bike Thief



Pilot Lineups for Computer Thief



APPENDIX E: PRE-ID VIDEO QUESTIONS

Pre-ID Video Questions

1. “What was the overall message of the video (in a few words)?”
2. “What was one lesson that the main character learned by the end (in a few words)?”

APPENDIX F: EXPERIMENT 2 META-MEMORY QUESTIONS

Preference Question

You are going to see a lineup.

The lineup will either consist of **unmasked faces** or **faces wearing ski masks that expose the eyes**.

The computer randomly determines which version you will see. But if you were given a choice, would you like to see a lineup of unmasked faces or a lineup of faces wearing masks?

A lineup of faces wearing masks

A lineup of unmasked faces

Pre-ID Confidence Questions

If you were asked to identify the perpetrator in a 6-person lineup consisting of **faces wearing masks**, that may or may not contain the perpetrator how likely do you think you would be able to make a correct decision?

Please rate your answer on a scale from 0-100,
0 indicating no chance of correct decision,
100 indicating absolute certainty of correct decision?

0 10 20 30 40 50 60 70 80 90 100

If you were asked to identify the perpetrator in a 6-person lineup consisting of **unmasked faces**, that may or may not contain the perpetrator how likely do you think you would be able to make a correct decision?

Please rate your answer on a scale from 0-100,
0 indicating no chance of correct decision,
100 indicating absolute certainty of correct decision?

0 10 20 30 40 50 60 70 80 90 100

APPENDIX G: EXPERIMENT 3 QUESTIONS**Lineup Credibility Questions**

1. **Trust:** “Do you think that an identification by this eyewitness ought to be trusted?”
(Trust) with the scale 0-100: 0 (definitely should not be trusted), 100 (definitely should be trusted)

2. **Information:** “To what extent do you feel that the eyewitness had a good basis (enough information) to make an identification?”
(Information) with the scale 0-100: (no basis at all) to 100 (a very good basis)

3. **Guilt:** “If you were a juror, how likely is it that you would find the identified suspect guilty?” (Guilt) with the scale 0-100: 0 (not likely to find guilty) to 100 (very likely to find guilty)

4. **Lineup:** “To what extent do you feel that the eyewitness received an appropriate lineup procedure that enabled them to make an accurate identification?” (Lineup) with the scale 0-100: 0 (not at all appropriate), 100 (completely appropriate).

APPENDIX H: IRB APPROVAL

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
2420 Lincoln Way, Suite 202
Ames, Iowa 50014
515 294-4566

Date: 03/30/2018

To: Jason C Chan

From: Office for Responsible Research

Title: Movie Insight

IRB ID: 15-663

Submission Type: Continuing Review

Review Type: Expedited

Approval Date: 03/28/2018

Date for Continuing Review: 04/20/2020

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, including the **recruitment materials and informed consent documents that have the IRB approval stamp.**
- **Retain signed informed consent documents** for 3 years after the close of the study, when documented consent is required.
- **Obtain IRB approval prior to implementing any changes** to the study.
- **Inform the IRB if the Principal Investigator and/or Supervising Investigator end their role or involvement with the project** with sufficient time to allow an alternate PI/Supervising Investigator to assume oversight responsibility. Projects must have an [eligible PI](#) to remain open.
- **Immediately inform the IRB of (1) all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- **Stop all human subjects research activity if IRB approval lapses**, unless continuation is necessary to prevent harm to research participants. Human subjects research activity can resume once IRB approval is re-established.
- **Submit an application for Continuing Review** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

IRB 03/2018

- Please be aware that IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **IRB approval in no way implies or guarantees that permission from these other entities will be granted.**
- Please be advised that your research study may be subject to [post-approval monitoring](#) by Iowa State University's Office for Responsible Research. In some cases, it may also be subject to formal audit or inspection by federal agencies and study sponsors.
- Upon completion of the project, transfer of IRB oversight to another IRB, or departure of the PI and/or Supervising Investigator, please initiate a Project Closure to officially close the project. For information on instances when a study may be closed, please refer to the [IRB Study Closure Policy](#).

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.