How Systematically Increasing Estimator Variables
Affects the Confidence-Accuracy Relationship

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Psychology

by

Amber M. Giacona
Sam Houston State University
Bachelor of Science in Psychology, 2016
Sam Houston State University
Master of Arts in Psychology, 2018

August 2023
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

____
James M. Lampinen, Ph.D.
Dissertation Director

____
Douglas Behrend, Ph.D.
Committee Member

____
William Levine, Ph.D.
Committee Member

Abstract

Researchers have used signal-detection theory-based approach to show that when police use proper practices with eyewitnesses, highly confident witnesses will be highly accurate even when viewing conditions may be suboptimal (Wixted & Wells, 2017). This is referred to as the pristine conditions hypothesis. There have been multiple, and often contradictory, studies that have investigated the relationship between viewing conditions and memory degradation (Giacona et al., 2021; Grabman et al., 2019; Lockamyeir et al, 2020; Semmler et al., 2018). In the current study, I systematically manipulated five estimator variables (lighting, distance, retention interval, exposure duration, and race) as either suboptimal or optimal to further investigate this relationship. I found that, as expected, overall memory strength decreased as the number of suboptimal estimator variables increased. Next, I assessed CAC curves for the number of suboptimal estimator variables and found that the pristine conditions hypothesis holds, except when all five variables are suboptimal, at which point high confidence does not equal high accuracy. Additionally, these results did not hold for when base rates were low. Similarly, when collapsing across viewing type, it was found that under low base rates, high confidence did not equal high accuracy when the conditions were suboptimal. While this research found a lot of support for the pristine conditions hypothesis, it also established important boundary conditions for when this hypothesis is not valid. Further research is still needed to continue to address the confidence-accuracy relationship.

Keywords: eyewitness, CAC, pristine conditions, estimator variables

Acknowledgements

I would like to thank and acknowledge my committee chair, mentor, and the initial reason I came to the University of Arkansas, Dr. James Lampinen. It has been a long five years, but I would not have this knowledge without you. I would also like to acknowledge my committee members, Drs. Bill Levine and Doug Behrend. Thanks for your helpful comments and willingness to meet over the summer!

I also want to acknowledge that I received partial funding from a Howell's Grant to support my research. I want to give a special thanks to the countless research assistants that have helped with my many projects throughout the years.

Lastly, I want to acknowledge all my friends and support system that have helped me get where I am. Without their dark humor and sarcasm, I would not have gotten through grad school.

Dedication

I dedicate this dissertation to Stormy and Pete Giacona. None of this would have been possible for me without your hard work and sacrifice. Thanks for giving me the tools for success.

I love you, Mom and Dad.

Contents

How Systematically Increasing Estimator Variables Affects the Confidence-Accuracy Relationship

It has been argued that when non-suggestive police practices are used, highly confident witnesses will be highly accurate, even when the witnessing conditions are poor (Wixted et al., 2015; Wixted & Wells, 2017). I have referred to this as the "pristine conditions hypothesis" (Giacona et al., 2021). Tests of the pristine conditions hypothesis almost always involve manipulating only one variable or a small set of variables at a time. However, real-life crimes are very complex and often involve witnessing conditions that are problematic in multiple ways.

Basic Concepts in Eyewitness Identification Research

An eyewitness is a person who sees a crime occur or who sees an event that is relevant to determining who committed the crime, even if they did not see the crime itself. For instance, an eyewitness might personally see a thief stealing a victim's wallet, or they might see a person running away from a location where a person's wallet was stolen. An eyewitness can be a victim of the crime or a bystander who sees the crime (Kassin, 1984). Members of the general public can be eyewitnesses, as can members of law enforcement (Vredeveldt & van Koppen, 2016). They can also have prior familiarity with the suspect in question or be unfamiliar with them (Vallano et al., 2019). Eyewitnesses provide a variety of different types of information including information about what happened (Ebbesen & Rienick, 1998), person descriptions (Meissner et al., 2007) and importantly for this paper, eyewitness identifications (Lampinen et al., 2012). An early study estimated that there are close to 80,000 criminal cases per year in the U.S., in which eyewitness identification is crucial evidence (Goldstein et al., 1989).

The reliability of eyewitness identifications can be influenced by a range of different variables. In 1978, Wells proposed that these variables could fruitfully be characterized as *system* *variables* and *estimator variables*. A system variable is a factor that can affect the accuracy of an eyewitness identification and that is under the control of law enforcement. For instance, the instructions police give to witnesses can affect how the witness responds, and those instructions are under the control of law enforcement (Malpass & Devine, 1981). An estimator variable is a variable that can affect accuracy, but that law enforcement has no control over. For instance, whether a weapon was present during a crime can affect witness accuracy, but police have no control over whether a weapon was present during the crime (Fawcett et al., 2013). Wells (1978) argued that system variables are more important for researchers to study, because they allow for systematic reforms in the criminal justice system. However, others have argued that estimator variables are equally important to carefully study (Deffenbacher, 2008).

There are two main identification procedures used by police: lineups and showups. In both types of procedures, there is an individual who the police have come to think might have committed the crime. This person is called a *suspect.* There are no set standards for how much evidence needs to exist for someone to become a suspect (Wise et al., 2011). In some cases, there may be almost overwhelming evidence against a person before they become a suspect. In other cases, a person may be a suspect based merely on a hunch on the part of law enforcement. The suspect in a criminal investigation may be guilty or may be innocent. Identification procedures in which the suspect is guilty are called culprit present (or target present) identification procedures. Identification procedures in which the suspect is innocent are called culprit-absent (or target absent) identification procedures. Of course, in actual cases, whether an identification procedure is culprit-present or culprit absent is not known. That question is the whole focus of the police investigation. But in experimental studies, researchers can set up situations in which it is known whether the suspect is guilty or innocent, allowing them to test the accuracy of witnesses in both culprit present and culprit absent procedures.

A showup is a procedure in which the witness is shown a single individual (the suspect) (Neuschatz et al., 2016). In practice, most showups involve the live presentation of a suspect to a witness shortly after the crime, although occasionally a witness may be shown a single photograph or video of a suspect and occasionally a showup may occur after a considerable delay. Showups are recognized as a suggestive and error prone procedure by both legal (Bertelsman, 2022) and scientific experts (Wells et al., 2020). For that reason, it is often recommended that showups only be used in exigent circumstances (Lee, 2004).

A lineup is a procedure in which a suspect is shown with a number of fillers (or foils), which are known innocent individuals that match the suspect's description. Like showups, lineups can also be live or through photographs (or by videos in the United Kingdom; Fitzgerald et al., 2021), though the number of fillers slightly differs between these procedures. The number of fillers used also differs by country (Fitzgerald et al., 2021). In the United States, the number of fillers used can vary by jurisdiction, but the modal number of fillers in both live lineups and photographic lineups is five (Police Executive Research Forum, 2013). However, there is more variability in live lineups, presumably owing to the fact that finding actual individuals who match the witness's description to stand in as fillers is more difficult than finding a set of photographs that match the witness's description. The number of people in the lineup is referred to the lineup's *nominal size.* However, there is also the lineup's *functional size*, which refers to the number of plausible choices in a lineup. For instance, a witness may describe a perpetrator as a tall, thin, blonde woman in her 20s. If the lineup contains six individuals (nominal size), but two of those individuals have brown hair not blonde, and another individual is clearly in her 40's, then the lineup would really only have three plausible choices (functional size) for the witness. Ideally, the nominal size and functional size should be equal, but that is often not the case (Valentine & Heaton, 1999). Therefore, the fairness of a lineup, referenced in more depth later, is dependent upon this functional size.

Lineups also vary in terms of the procedures police use to present the lineup members. Lineups can be shown simultaneously, where all the lineup members are presented at the same time, or they can be shown sequentially, where each photograph is shown one at a time consecutively (Lindsay & Wells, 1985). In current practice, a majority of police agencies use simultaneous lineup procedures, although a substantial minority of police agencies administer lineups sequentially (Police Executive Research Forum, 2013). Early research found that sequential lineups outperform simultaneously lineups (Wells & Turtle, 1987). However, more recently, there has been debate about the relative merits of the two procedures, with some researchers arguing that sequential presentation results in more accurate performance and some researchers arguing that simultaneous presentation leads to more accurate performance (Mickes et al., 2012; Steblay et al, 2011).

Why Are Lineups Better Than Showups?

It is widely agreed that lineups are superior to showups when it comes to identification accuracy (Gronlund et al., 2012; Steblay et al., 2003; Yarmey et al., 1996). A recent whitepaper released by the American Psychology Law Society concluded, “There is no debate among eyewitness scientists about the fact that lineups produce better outcomes than do showups...” (Wells et al., 2020, p. 7). Although there is agreement that lineups are better than showups, there are competing theories for why this is the case. The two competing theories are diagnostic feature detection theory and differential filler siphoning theory.

Diagnostic feature detection theory proposes that there are diagnostic and nondiagnostic features. According to the theory, fillers help focus witness attention on the features that are most diagnostic (Wixted & Mickes, 2014). For instance, if a perpetrator had a shaved head and everyone in the lineup had a shaved head, that feature is no longer diagnostic, since it is not a dimension along which the lineup members vary. According to this account, the presence of fillers causes witnesses to focus on the non-shared features among the lineup members and to place less weight on the shared features, increasing accuracy.

Differential filler siphoning, on the other hand, proposes that fillers function by drawing errors away from innocent suspects (Smith et al., 2018; Wells et al., 2015). This tendency of fillers to draw choices away from the suspect is called filler siphoning. In a fair culprit absent lineup, there is no reason to think that the innocent suspect will be a better match to the witness's memory than any of the individual fillers. Because of that, some witnesses who would have identified the innocent suspect in a showup, will instead pick one of the known innocent fillers when presented with a lineup. This will result in a lower innocent suspect identification rate for lineups. Identification of a filler is a mistake, but it is a relatively harmless mistake, because fillers are known innocents. Thus, identification of a filler prevents the more dangerous error of mistakenly identifying an innocent suspect. Filler siphoning will also happen to some degree for culprit present lineups – a filler may just happen to match the witness's memory than the guilty suspect does – but it will be less frequent, especially in situations where memory is strong. This is why the theory is called *differential* filler siphoning. Filler siphoning is expected to happen in both culprit absent and culprit present lineups but will be more frequent in culprit absent lineups.

In support of diagnostic feature detection theory, Wooten et al. (2019) tested showups versus lineups with 3, 6, 9, or 12 photos. Their hypotheses were that quality of the fillers should matter over the quantity of the fillers. Furthermore, they predicted the same findings would be found in ROC and CAC analyses, which are explained more in-depth later in this paper but are the current standard of analysis in eyewitness literature. They found full support in the ROC analyses that all lineup identifications had higher discriminability compared to the showup condition, and there were no differences among the lineup sizes (Wooten et al., 2019). They concluded that this showed that it is the quality and proportion of non-diagnostic features and diagnostic features rather than the number of fillers in a lineup that had the biggest impact on the reliability of the procedure.

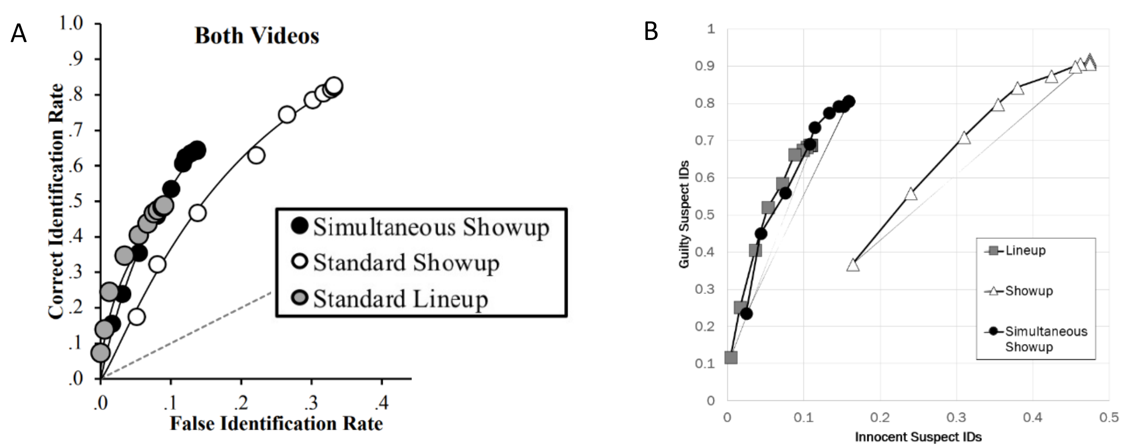
A study by Colloff and Wixted (2020) attempted to pit diagnostic feature detection theory and differential filler siphoning theory against each other by creating a situation in which diagnostic feature detection was possible, but in which filler siphoning was not possible. They did so by having participants view crimes in which the perpetrator had a distinctive feature. For instance, in one of the videotaped crimes, a perpetrator with a black eye committed an act of vandalism. If the lineup only contained one person that had a black eye, the feature would be diagnostic because it would be a salient feature that would set apart the suspect from the fillers. However, in a lineup where everyone has a black eye, the black eye would be nondiagnostic, as that feature would no longer help the eyewitness to discriminate between the suspect and the fillers. Witnesses were later shown a standard six-person lineup, a standard photographic showup, or a new kind of identification procedure that the authors referred to as a *simultaneous showup*. In a simultaneous showup, there is one suspect outlined in a red box, and five fillers that participants are told are examples of what innocent people might look like. Witnesses are told that they are not allowed to select any of the fillers. Their only task is to indicate whether the suspect shown in the red box is the guilty perpetrator. This is like a simultaneous lineup in that numerous fillers are presented in tandem with the police suspect; however, this shares procedures with a showup in that 1) the eyewitness knows who the police suspect is, and 2) the only question being asked is whether the one person outlined in red is the culprit. Colloff and Wixted (2020) argued that the simultaneous showup procedure ensured that witnesses were able to compare and contrast lineup members, allowing them to discover which features were diagnostic and which were non-diagnostic, while precluding fillers from siphoning any choices away from the suspect. They found that discriminability was the same for lineups and simultaneous showups, although the simultaneous showups responses were more liberal (see Figure 1a). They also found that both lineups and simultaneous showups outperformed showups. They concluded that the difference in discriminability in lineups compared to showups must be due to diagnostic feature detection, because filler siphoning is not possible given the experimental instructions (i.e., fillers could not be overtly chosen).

In recent research (Giacona & Lampinen, 2021), I have argued that this procedure fails to account for the possibility of "mental filler siphoning." Mental filler siphoning is the idea that, even though participants are not allowed to actually select one of the fillers, people might still be trying to determine mentally which photograph provides the best match to their memory and may reject the suspect if one of the fillers provides a better match to memory than the suspect does. In Colloff and Wixted's (2020) original publication, they considered this possibility but quickly dismissed it. To test the idea of mental filler siphoning, I conducted two experiments. In Experiment 1, we replicated Colloff and Wixted's simultaneous showup condition using their own materials. After making their judgment, we asked participants if the presence of the other photographs affected their decision. If they said "yes", we asked them to describe in their own words how the other photographs influenced their decision. The descriptions were then coded for evidence of diagnostic feature detection (e.g., "they made me realize that I should focus on the eyes") or differential filler siphoning (e.g., "one of the other photos looked more like the guy than the suspect did").

In the second experiment, we fully replicated all the conditions from Colloff and Wixted (2020). After participants in the simultaneous showup condition made their judgment, we again asked them if they had been influenced by the other photographs. If they said "yes", we again asked them how the other photographs influenced them, this time giving them a choice from a list of alternative options. Two of the options were designed to capture diagnostic feature detection, two were designed to capture filler siphoning, and then other options had nothing to do with either of these.

Figure 1

Red Box ROC Curve



Note. Figure A (left) shows results from Colloff and Wixted (2020, Experiment 3). B (right) shows preliminary results shown at AP-LS (Giacona & Lampinen, 2021).

For our open-ended project, we had 266 overall judgments, and of those, 135 (~51%) indicated that the photographs influenced the participants' decision. Of those responses, 34% were coded as reflecting evidence of filler siphoning (e.g., *There was an individual not boxed in* *red that was the real person from the video so I knew they had the wrong suspect*), and 31% were coded as reflecting evidence consistent with diagnostic feature detection (e.g., *If I had seen just one picture of a person with that face tattoo/scar then I would have said that was the guy. However, seeing multiple faces compared with that face tattoo/scar made me further analyze the face and after that it seemed to me that was not the man I saw*). An important note is that there were other responses we categorized, but those have not been included in this analysis.

In our direct replication, we replicated the basic finding that the lineup condition and the simultaneous showup condition did not differ from each other in terms of discriminability and that both outperformed the standard showup condition (see Figure 1b). We had 346 judgments in the simultaneous showup condition[1], and in 193 (56%) of those, participants indicated that the photographs influenced their decision. Of those 56%, we measured how often participants selected response options consistent with diagnostic feature detection (e.g., *the other pictures (not in the red box) helped me to focus on what features were most relevant for the judgement* or *if the other pictures (not in the red box) were not there, I would have just focused on the fact that suspect (in the red box) had a black eye/scar like the perpetrator did*) or filler siphoning (e.g., *one or more of the other pictures (not in the red box) looked more like the perpetrator than the suspect (in the red box) did* or *if I had a choice, I would have picked one of the other pictures (not in the red box)*). Overall, we found filler siphoning in 31% (n = 132) of tasks, and it was slightly more common than diagnostic feature detection with 24% (n = 101).

More research and data collection are needed before any conclusions can be drawn from this research. However, it seems that most participants' conscious thinking points to either the filler siphoning theory or the diagnostic-feature-detection theory. People appear to use both

---

[1] This is an ongoing study and the results I report are those I described in my conference presentation of the preliminary data.

strategies. We cannot definitively say that one of these strategies or the other accounts for the lineup /showup differences observed in past research. But our primary goal was being able to show that this new identification task (i.e., the simultaneous showup) cannot exclude mental filler siphoning as potentially explaining lineup/showup differences.

The jury is still out on which of these accounts best describes the reason for lineup/showup differences. Most importantly, what all of these studies and theories have in common is that lineups are superior to showups (Colloff & Wixted, 2016; Smith et al., 2016; Wooten et al., 2019). This is important because regardless of the underlying theories for why these differences are occurring, showups are inferior to lineups and should not be used regardless of their convenience unless there is an exigent circumstance that absolutely necessitates it. Indeed, guidelines published by the International Association of Chiefs of Police (2010), the world's largest organization of police executives, recommended that showups should be avoided whenever possible.

Errors in Eyewitness Identification

Types of Errors

Eyewitness identification is not always accurate; witnesses sometimes fail to identify guilty suspects in culprit-present lineups, and they sometimes mistakenly identify innocent suspects from culprit-absent lineups. There are two types of errors: misses, which are failing to identify a guilty suspect, and false alarms, which are mistakenly identifying an innocent suspect. Both have costs for society; however, it has been argued that false alarms carry more societal costs (Blackstone, 1766; de Keijser et al., 2014; Wells et al., 2012). When a witness fails to identify a guilty suspect, that increases the probability that the guilty perpetrator will remain free to commit additional crimes (Wells et al., 2012).

Case Studies

Unfortunately, when an eyewitness identifies an otherwise innocent individual from a lineup, it has two costs: the first is that an innocent person is wrongly put behind bars, and the second is that a guilty perpetrator is free to commit more crimes. This happened in the cases of Marvin Anderson, Ronald Cotton, and Jonathan Irons, which I will discuss here, but also in countless others who will remain unnamed.

Marvin Anderson

The story of Marvin Anderson is heartbreaking (Innocence Project, 2021). When this case took place, Marvin was just an African American man, dating a White woman, while training to be a firefighter. However, in 1982 in Ashland, Virginia, a 24-year-old White woman was sexually assaulted by an African-American man who told the victim that she looked like his girlfriend. The investigators knew that at the time Marvin Anderson was in an interracial relationship with a White woman, and that led them to suspect Mr. Anderson. They obtained a color photograph of him from his firefighter application and presented it along with several black and white mugshots of young, African American males. Unsurprisingly, the victim chose Mr. Anderson's photograph and selected him out of a live lineup later. Based on this, Mr. Anderson was convicted. In 1986, a man name John Otis Lincoln came forward and admitted, under oath and in open court, that he had committed the rape and that Mr. Anderson was innocent. However, the court refused to release Anderson. Fifteen years later, in 2001, there was finally an investigation into Mr. Lincoln's claims; the DNA evidence was tested, and it excluded Mr. Anderson and matched Mr. Lincoln. However, it was not until the following year, 20 years after the initial incident, that Marvin Anderson was issued a full pardon.

Ronald Cotton

In 1984 in North Carolina, a woman named Jennifer Thompson, 22-years-old at the time, was raped by a man who broke into her room and held a knife to her throat (Jones, 2012; National Registry of Exonerations, 2019). Ms. Thompson was determined to remember her attacker's face and tried to memorize it during the attack. Ms. Thompson gave a very detailed description of her attacker and completed a face composite with police. She was shown a photo lineup that had Ronald Cotton, who she chose. She was later shown a live lineup and chose Mr. Cotton out of it. Thompson stated that she was sure it was him. Unfortunately, as Mr. Cotton was the only individual that was repeated between lineups, this increases the chances of incorrectly choosing him. Mr. Cotton was in prison for 11 years before he released due to DNA evidence and the real perpetrator, Bobby Poole, was correctly implicated (Jones, 2012).

Jonathan Irons

In 1997, a man, Stanley Stotler, was attacked by a gunman who shot him in the head. Mr. Stotler survived through emergency brain surgery, but could not describe the attacker other than that he had been a Black man. Police officers learned that Jonathan Irons, a 16-year-old at the time, was in the predominantly White neighborhood, and though he had multiple witnesses present to confirm his alibi and no physical evidence tying him to the crime, he was arrested. Mr. Stotler, while still in the hospital, was shown a lineup and chose Mr. Irons, but only after expressing being unable to pick someone and being told by offers to take his "best guess" (National Registry of Exonerations, 2021). Along with other evidence, most notably misleading or false forensic evidence and official misconduct, there were issues with the eyewitness account. Dr. Lampinen, who studies eyewitness memory, noted that Mr. Iron's photo in the lineup was suggestive. In his photo, Mr. Irons' head "was twenty-five percent larger than the average of the other photos" (National Registry of Exonerations, 2021). Additionally, the police telling the witness to guess is another suggestive practice. On July 1, 2020, after serving 22 years in prison, Mr. Irons was finally released from prison with all charges dropped.

Quantitative Findings

The Innocence Project is an organization committed to helping exonerate individuals on the basis of running DNA evidence. So far, there have been 375 total exonerations since the first in 1989 (Innocence Project, 2022). Of these exonerees, 60% were African American, and on average, they spent 14 years in jail before being released. Of these cases, 165 true perpetrators were identified, but unfortunately not before committing "154 additional violent crimes, including 83 sexual assaults, 36 murders, and 35 other violent crimes while the innocent sat behind bars for their earlier offenses" (Innocence Project, 2022). What the majority of these cases have in common is that 69% of them involved mistaken eyewitness identifications, the most prevalent contributing factor to these cases.

Another resource tracking exonerations is the National Registry of Exonerations supported by the University of California Irvine Newkirk Center for Science & Society, the University of Michigan Law School, and the Michigan State University College of Law. Whereas the Innocence Project solely focuses on cases where DNA evidence exonerations individuals, this database lists any cases of exoneration. As of the writing of this paper, there are 3,268 total exonerees, and 857 cases list mistaken eyewitness identification as a contributing factor. One potential reason for why mistaken eyewitness identifications are more prevalent in the Innocent Project database compared to the National Registry of Exonerations database could be the type of case involved. The majority of DNA cases involve sexual or violent crimes, including rape and murder, whereas the other cases include all crimes, even victimless ones.

Field Studies vs. Experimental Studies

There are two main study types used in eyewitness literature: field research and experimental studies. Field studies can be conducted using archival data or experimental methods. Field studies are hard to assess because determining the ground truth is difficult. In a field study one knows whether the witness identified the suspect, identified a filler or rejected the lineup. But researchers do not know, for each case, whether the suspect is factually guilty or factually innocent. Initial recommendations for policy and procedures set forth by experts in the field (Technical Working Group, 1998) called for more research to be conducted in field settings. When the recommendations were first put forth in 1998, there was only one field study (Wright & McDaid, 1996). However, more recently, Wells et al. (2020) were able to review 11 published field studies. They compared their findings to Clark et al. (2008), who used a meta-analysis of 94 laboratory studies to assess the errors observed by eyewitnesses in an experimental setting.

Figure 2

Wells et al. (2020)

Authors	Number of possible IDs	ID of suspect	IDs of filler	No ID	Suspect %	Filler %	No ID%	% Making an ID	Suspect ID rate among all IDs	Filler ID rate among all IDs
Behrman and Davey (2001)	58	29	14	15	50.0%	24.1%	25.9%	74.1%	67.4%	32.6%
Behrman and Richards (2005)	461	238	68	155	51.6%	14.8%	33.6%	66.4%	77.8%	22.2%
Horry, Halford, Brewer, Milne, and Bull (2014)	833	382	149	302	45.9%	17.9%	36.3%	63.7%	71.9%	28.1%
Horry, Memon, Wright, and Milne (2012)	1,039	406	273	360	39.1%	26.3%	34.6%	65.4%	59.8%	40.2%
Klobuchar, Steblay, and Caligiuri (2006)	178	63	20	95	35.4%	11.2%	53.4%	46.6%	75.9%	24.1%
Memon, Havard, Clifford, Gabbert, and Watt (2011)	1,044	456	437	151	43.7%	41.9%	14.5%	85.5%	51.1%	48.9%
Valentine, Pickering, and Darling (2003)	584	237	121	226	40.6%	20.7%	38.7%	61.3%	66.2%	33.8%
Wells, Steblay, and Dysart (2015)	494	132	75	287	26.7%	15.2%	58.1%	41.9%	63.8%	36.2%
Wixted, Mickes, Dunn, Clark, and Wells (2016)	348	114	104	130	32.8%	29.9%	37.4%	62.6%	52.3%	47.7%
Wright and McDaid (1996)	1,561	611	310	640	39.1%	19.9%	41.0%	59.0%	66.3%	33.7%
Wright and Skagerberg (2007)	134	78	28	28	58.2%	20.9%	20.9%	79.1%	73.6%	26.4%
Overall sum	**6,734**	**2,746**	**1,599**	**2,389**						
Weighted means					**40.8%**	**23.7%**	**35.5%**	**64.5%**	**63.2%**	**36.8%**

Note. Some studies reported data that included identifications by witnesses who knew the culprit (prior familiarity) and those data are excluded from Table 1.

Note. The above shows statistics from 11 field studies and was reprinted from Wells et al. (2020). Copyright American Psychological Association, 2020. Reprinted with permission.

Figure 2 shows the findings of Wells et al. (2020) for filler IDs, suspect IDs, and no IDs. There were 6,374 total lineups in the field studies taken from different jurisdictions and countries. One main question was: are real world witnesses to crimes more cautious about making a mistaken identification than experimental participants? Previously, Mecklenburg et al. (2008) argued that “participants in laboratory experiments make a lot of mistaken identifications only because the witnessed events are not real and the consequences of mistaken identification are trivial” (Well et al., 2020, p. 5). To assess Mecklenburg et al.’s claim, results from Clark et al. (2008) should be addressed first. They found that in laboratory studies, 21.2% of eyewitnesses chose an innocent filler in a lineup when the culprit was present, and 34.6% chose an innocent filler when the culprit was absent. Similarly, Wells et al. (2020) found that in field studies of actual police lineups, 23.7% of witnesses selected an innocent filler, and in the subset of actual cases in which the witness made an identification, 36.8% identified an innocent filler.

When taken together, Wells et al. (2020) was able to show that the field study innocent filler rate (23.7%) was similar to the laboratory filler rate (27.9%). This suggests that contrary to previous beliefs (Mecklenburg et al., 2008; Technical Work Group, 1998), the filler identification rate is high both in laboratory and field studies, and this is non-trivial. This is a very real problem that needs to be addressed and cannot simply be dismissed due to the difference in laboratory versus real world conditions.

Variables that Affect Identification Evidence

Being an eyewitness and making a decision on whether someone is guilty or not is no easy task. As noted above, Wells (1978) broke down eyewitness issues into two main groups: system and estimator variables. His paper grouped large themes of eyewitness issues into those that the criminal justice system has control over (i.e., system variables) and those that the criminal justice system does not (i.e., estimator variables). Originally, Wells (1978) argued that system variables are more important because they allow researchers to impact police practice in a systematic way. The idea was that researching these systematic variables would be able to help the most individuals with the longest-lasting changes. System variable research has led to the development of best practice guidelines published by major scientific (National Academy of Sciences, 2014; Wells et al., 1998; Wells et al., 2020) and criminal justice professional organizations (International Association of Chiefs of Police, 2010; Major Cities Chiefs of Police, 2015; Technical Working Group for Eyewitness Evidence, 1999). However, Deffenbacher (2008) argued that estimator variables are equally important, and more research should focus on looking at how they impact witness accuracy.

System Variables

System variables are variables that can be controlled by the criminal justice system and can thus be held constant from case to case (Wells, 1978). These variables are especially important as they are variables that can be controlled between each crime. Unfortunately, not every recommendation is used by each police department (Police Executive Research Forum, 2013).

Pre-Lineup Admonishment

Pre-lineup admonishment refers to the instructions given to the eyewitness before engaging in the identification procedure (Greene & Evelo, 2015). Witnesses generally want to be helpful when being interviewed by the police. They are also likely to assume that officers would not waste their own time by setting up a lineup unless they thought they actually had a guilty suspect. Thus, the witness is likely to think that they are supposed to choose someone from the lineup. The problem with this line of reasoning is that the actual purpose of a lineup is to determine, not which member of the lineup is the suspect, but *if* the suspect is guilty. This misunderstanding of their proper role in the lineup requires that witnesses be properly instructed to overcome this misunderstanding.

In the 1970's, a series of prominent wrongful conviction cases came to public attention in the United Kingdom, all involving mistaken identification. This led to the empanelment of an expert committee led by High Court Judge Lord Patrick Devlin (Devlin, 1976). The Devlin Committee made a number of recommendations concerning police and judicial practice in eyewitness identification cases. Among their findings, they noted that "... witnesses may tend ... to make an identification on parade because they feel that that is what is expected of them. We have considered various ways of relieving the pressure on witnesses of this type and conclude that the best way is for the officer in charge of the parade to tell the witness expressly that the person he saw may or may not be on the parade" (Devlin, 1976, p. 164).

These findings led to the development of lineup procedures in which witnesses are provided with pre-lineup instructions that minimally tell them that the person who committed the crime may or may not be present in the lineup. A seminal paper by Malpass and Devine (1981) investigated biased lineup instructions and unbiased lineup instructions. Instructions were considered biased if they stated that they "believe the person who…[did the crime]…is present in the lineup" and asked them to choose someone from choices 1-5, all of which were possible suspects (Malpass & Devine, 1981, p. 484). The unbiased instructions said that the perpetrator may be in the lineup, but that "it is also possible that he is not in the lineup" and offered them a selection to choose if they do not believe the perpetrator is in the lineup, making the choices 0-5, with 1-5 being possible suspects and "0" being not present. (Malpass & Devine, 1981, p. 484). The two most important findings from this study showed there were high rates of choosing when the perpetrator was present, whereas when the perpetrator was absent, it depended on the condition. Biased instructions led to higher choosing rates than unbiased instructions when the perpetrator was absent.

Overall, this study was very important as it added a novel instruction that is now a recommended by practice by the American Psychological Association (Division 41), U.S. Department of Justice, International Association of Chiefs of Police, Major Cities Chiefs Association, American Bar Association (Criminal Justice Section), Innocence Project, and the National Academy of Sciences. Malpass and Brigham (1981) were able to show that by using a proper lineup admonishment that indicated that the suspect may or may not be present, false identifications were reduced while correct identifications remained the same. This is important because sometimes there may be procedures where doing what is deemed fair by researchers may inhibit correct identifications; however, according to Malpass and Devine's (1981) finding, there is no downside to using proper admonishments. Additionally, a metanalysis of 18 studies conducted by Steblay (1997) found that incorrect identifications were reduced by using the proper admonishment, but correct identifications remained unaffected. Specifically, they found that "biased instructions produced a moderate effect on accuracy in target-absent lineups. . ., but minimal effect in target-present lineups" (Steblay, 1997, p. 283). This was considered a prevalent view and was echoed by Wells et al. (1998, p. 615) when they summarized Steblay's (1997) findings by saying "A recent meta-analysis of instruction effects shows that the "might or might not be present" instruction has the effect of reducing identifications when the perpetrator is absent from the lineup while having no effect on identifying the perpetrator when the perpetrator is in the lineup."

However, the idea that correct identifications were not affected or reduced by implementing proper pre-lineup admonishment was reexamined by Clark (2005). He conducted a meta-analysis and found that while incorrect identifications were reduced with the addition of the pre-lineup instructions, so were correct identifications. Clark (2012) summarizes the response of the field to minimally accept this finding:

> … that correct identification rates "might be slightly harmed" by unbiased instructions (Wells et al., 2006, p. 62) and that biased instructions, "sometimes result in a higher proportion of culprit selections," (Brewer & Palmer, 2010). In a case recently decided by the New Jersey Supreme Court (*State of New Jersey v. Larry R. Henderson*, 2010), the claim was made that, "the loss in accurate identifications [due to unbiased instructions] pales in comparison to the drop in mistaken identifications," (Scheck et al., 2010). In its opinion, the Court noted the effects of biased instructions on false identifications, but made no mention of the effects for correct identifications. (p. 241)

There are many reasons for this difference in the reanalysis of Steblay's (1997) data. One was that there seemed to be an unpublished study (Hall & Ostrom, 1975, as cited by Clark, 2005) that had an effect sized thrice as much as any other study and was driving the conclusions. Clark (2005) also claims that for several studies, the correct identification rate was already at or near the ceiling rates. The third reason the author gave was the "reconceptualization of bias" for two studies. The most important takeaway is that both Clark (2005) and Steblay (1997) agree that biased instructions inflate false identifications, and although the results are mixed, the correct identifications seem either unaffected or minimally so and seem that that the cost would be worth utilizing unbiased instructions.

Fair Lineups

A fair lineup is one in which the fillers provide plausible alternatives to the suspect (Malpass et al., 2007). Typically, fair lineups are achieved by ensuring that all lineup members match the witness's description equally well (Wells et al., 1993). Multiple studies have found that fair lineups are superior to biased lineups (Smith et al., 2016; Smith et al., 2022). Similar to the pre-lineup admonishments, research has found that there is relatively no difference in accuracy of biased and fair lineups when the suspect is present, but that fair lineups reduce false alarm rates when the suspect is innocent. This is another win-win, in that it is simultaneously reducing innocent identifications while maintaining high rates of correct identifications.

General rules for constructing a lineup indicate that the lineup should consist of plausible fillers and no filler or suspect should stand out from the other lineup members (Malpass et al., 2007). There are also two, primary procedures that have been proposed for selecting fillers for a lineup: matching a suspect's similarity or matching a witness' description (Lampinen et al., 2012; Malpass et al., 2007). Wogalter et al. (1993) found that matching fillers to a suspect's perceptual similarity is the primary method used by police officers. This can be achieved by officers/researchers independently judging if fillers are the similar or by using independent raters to decide which fillers are most similar to the suspect. In the match to description approach, fillers are chosen to match the witness's description, but otherwise are left free to differ from the suspect. This allows all fillers/suspects to be equally good matches to the witness because the important pieces of information to the witness are preserved and kept equal, while the other differences between faces will help the witness distinguish between the fillers and the suspect (Luus & Wells, 1991). Additionally, Clarke and Tunnicliff (1992) found that when fillers are chosen by matching the suspect, it actually increases false identifications.

Lineup bias and lineup size are two different measures that have been used to assess lineup fairness (Malpass, 1981; Malpass & Devine, 1983; Malpass et al., 2007). Lineup bias is relatively easy to spot. If there is a scar mentioned and only one individual has a scar, then that is a biased lineup (Malpass et al., 2007). While not all biases are equally easy to spot, the idea is to limit overt lineup bias as much as possible. The most common evaluation of lineup size is Tredoux's *E* (Tredoux, 1998, 1999). This number essentially illustrates how many lineup members are actually good alternatives. For instance, in a six-person lineup, a Tredoux's *E* of six is considered perfect. That would mean that each lineup member is equally as good of a fit as any other lineup member. Most lineups that are considered very fair are in the 4-5 range.

To compute Tredoux's *E,* researchers use a *mock witness paradigm* (Wells et al., 1979). The idea behind the mock witness paradigm is that, in a fair lineup, a person who did not view the crime should not be able to pick out the suspect at greater than chance levels (e.g., 1/6) based on the description alone. In eyewitness identification experiments, the mock witness paradigm typically involves three steps. First, the perpetrator being used in the experiment is shown to a small set of individuals under conditions like those to be used in the main experiment. Those participants are then asked to describe the perpetrator in their own words. Based on their descriptions, a *composite description* is then created by combining all features that are mentioned by some pre-established proportion of the participants (e.g., any feature mentioned by half of the participants). For instance, in a recent study, we found that for a particular perpetrator the composite description was "young White male with blonde hair." In the second phase, fillers are selected for the lineup. This process involves going to a photo data base and finding other individuals who match the composite description. The selection of fillers requires some judgments to be made. For instance, the suspect in the above example might have a particularly light shade of blonde hair. In selecting fillers, one would want to select fillers who not only match the description (e.g., blonde hair) but also have a similar shade of blonde hair so that the suspect does not stand out in the lineup. In the final stage, the composite description along with the lineup is shown to a new group of participants who did not witness the crime and who have never seen the suspect. These participants are asked to select who they think they suspect is based on the description. In a perfectly fair lineup, everyone will be chosen an equal number of times (about 16-17%). However, sometimes there is something that may make one or two individuals stand out, thus lowering the Tredoux's *E.* In previous lineups I have conducted, one person, unbeknownst to me, had uncharacteristically bushy eyebrows that stood out. Although this was not in the lineup description, it was enough to eliminate him as a filler. Researchers may have to repeat the process of using and finding fillers until they have achieved a sufficiently fair lineup.

Lineup Administration

There are three types of lineup administration: blind, blinded, and double-blind (Kovera and Evelo, 2017). Blind lineups are conducted with a lineup administrator that is aware of who the suspect is, but the eyewitness is blind to who the police believe the suspect to be. In blinded administration, the eyewitness does not know who the suspect is, and the lineup administrator does know who the suspect is, but they are blinded to which photo the eyewitness is looking at. One way of achieving this is called the *folder shuffle method*. In the folder shuffle method, the photos are placed in individual manilla folders and shuffled so that the administrator is unaware of which folder contains the suspect. The folders are given to the witness, and the witness is asked to look at the photos one at a time, without allowing the lineup administrator to see who they are looking at. Lastly, in a double-blind lineup, neither the eyewitness nor the lineup administrator are aware of who the police suspect is. This is achieved by having a personnel member not associated with the case administer the lineup.

Blind lineups are much more suggestive than double-blind lineups. Unfortunately, administrators that know who the suspect is can intentionally or unintentionally suggest, coerce, and telegraph who the suspect is for the eyewitness to choose (Kovera & Evelo, 2020). Even if unconsciously doing so, nonblind administrators tend to smile when they are looking at the suspect or may even remind them that they need to look closely when they are on the suspect. These are behaviors that eyewitnesses pick up and influence their behavior.

In an early study by Phillips et al. (1999), 50 participants acted as eyewitnesses, and 50 participants acted as lineup administrators. There were two lineups per pair, and in one of the lineups, the administrator was told who the suspect was and in the other the administrator was not told who the suspect was. When the administrator knew who the suspect was, for sequential lineups only, they found an increased rate in mistaken identifications of innocent suspects, compared to when the administrator did not know who the suspect was. Interestingly, they did not find that the witnesses felt any more pressure to choose someone. This still illustrates that nonblind administrators should not be present when a lineup is being conducted. Furthermore, in an article by Kovera and Evelo (2020), they discuss how both correct suspect identifications and mistaken suspect identifications are more common in nonblind lineups, however the effect is much larger for mistaken suspect identifications. Nonblind administrators influence choosing by eyewitnesses in general based on who the administrators believe the suspect is. They also found that nonblind administrators are more likely to record a doubtful or uncertain suspect identification and less likely to do so when it is a filler.

In these cases, the task was relatively no risk as it was staged, and they were still influenced by administrators. If the impact of the stress and severity of real-world cases is considered, it is most likely that the pressure to choose and the body language and behaviors admitted from the administrator would only be stronger. This is particularly concerning as the eyewitness' job is not to choose the person the administrator wants, but their job is to decide if the correct suspect is present in the lineup. Although double-blind administration continues to get pushback from departments claiming they do not have the resources to implement it or old-fashioned officers want to use their gut instinct to judge an eyewitness' decision, a double-blind lineup is still the best lineup procedure to implement (State Bar of Michigan, 2012). This sentiment is further found by Findley (2008), where he states, "and the double-blind procedure does not cost anything in terms of lost valid identifications of the guilty." Essentially, many researchers agree that there is no-cost to correct suspect identifications and a decrease in mistaken identifications when proper procedures concerning double-blind lineups are used (Clark, 2012). While researchers do currently acknowledge that double-blind procedures may slightly decrease correct identifications, it is still the prevalent recommendation for conducting lineups (Kovera & Evelo, 2020). Furthermore, Kovera and Evelo (2017, 2020) demonstrated two issues with these conclusions: 1) the increase in correct identifications is from using prohibited tactics and suggestions, and 2) the there is a greater increase in false identifications than there is a reduction in correct identifications. Plainly, the increase in identifications from not using double-blind procedures are highly suggestive in a way that is not legally permitted, and the supposed benefits from not using double-blind lineups is less impressive than the very real dangers they cause (e.g., false identifications of innocent suspects).

Additionally, previous recommendations by Wells et al. (2020) have included encouraging police to implement double-blind lineups. A new preprint by Seale-Carlisle et al. (2022) also investigated expert opinions from researchers in the field and found that 99% believed that double-blind lineups are beneficial and should be used.

Estimator Variables

Estimator variables are variables that cannot be held constant by the criminal justice system and vary from case to case (Wells, 1978). These variables have been widely studied; however, there has been a recent proposal by Wixted and Wells (2017) that has downplayed the effects estimator variables and reasserted the claim system variables should carry more weight. Regardless, there have been decades of research supporting how much estimator variables can impact eyewitness identification. Below, I will discuss different kinds of estimator variables including distance, lighting, retention interval, weapon focus effect, and own-race bias. The list of potential estimator variables is very large, and so my review is limited to some of the estimator variables that have been more widely studied.

Distance and Lighting

In this section, I discuss the effects of distance and lighting on eyewitness identification. The variables are related because both affect the number of photoreceptors that represent the face. Distance affects the number of photoreceptors representing the face because as a face gets further away, the size of the image on the retina gets smaller (Lampinen et al., 2012). Lighting affects the number of photoreceptors representing the face because under bright lighting, vision occurs through the activation of cones, but under dim lighting vision occurs through the activation of rods (Barbur & Stockman, 2010). Cones have smaller receptive fields than rods, making rod vision less detailed. Additionally, multiple studies manipulated both lighting and distance together, so it makes sense to discuss them together.

Wagenaar and Van der Schrier (1996) conducted one of the first experimental studies of the effects of lighting and distance on eyewitness identification. They simulated seven different distances by adjusting the size of the facial images shown to participants to match the apparent size of faces at different distances. They also varied the brightness of the lighting used to view the facial images. Brightness is measured using the lux scale, in which 1 lux is the approximately equal to the light cast by a single candle at one meter away.[2] They used nine different lux levels (unit of measure = lx) ranging from 0.3 lx, which corresponds to a night with full moon, to 3000 lx, which is daylight with clouded weather. Importantly for their findings, 10 lx equals an urban area with bright street lights, and 30 lx equals a badly illuminated room. Faces were shown under one of 63 different combinations of illumination and simulated distance and after each face they were presented with a target present or target absent lineup. Both decreased distance and increased illumination led to higher rates of accuracy. More specifically, they suggested what they called a "Rule of Fifteen" (Wagenaar & Van der Schrier, 1996, p. 329). Their rule of 15 was based roughly on the Blackstone ratio (Blackstone, 1766) which holds that it is better for 10 guilty people to go free than for one innocent person to go to jail. Specifically, they argued that a good summary of their findings was that the ratio of correct identifications to false identifications fell below 15:1 if either the distance exceeded 15 meters or the illumination fell below 15 lux.

In a later field study, Lampinen et al. (2014) used actual target individuals seen from real-life distances rather than simulated distances. They manipulated outdoor distances from six, different distances ranging from 5-40 yards. They found a linear relationship in that accuracy declines fairly consistently as distance increases, with no sharp drop offs at these distances. Specifically, as distance increased, hits decreased and false alarms increased. Although accuracy never fell to zero, it was fairly poor by 20 yards and began to approach zero by 40 yards. Extrapolating out from the regression lineup, accuracy would reach zero at a little over 50 yards.

[2] When the lux scale was first developed, it was measured using standardized candles that were produced under precise conditions.

Nyman et al. (2019), in a study humorously titled "A Stab in the Dark," investigated what the maximum distance was where an eyewitness would be considered unreliable in low light. In particular, they wanted to also use actual distance and lighting rather than simulated distance and lighting, which had not been done together prior to this study. They manipulated three levels of illumination: starlight (0.7 lx), twilight (10 lx) and office space (300 lx). They also manipulated eight distances ranging from 6-20 meters, with a 2-yard increase between the points. They found, like previous studies, further distances and lower lighting negatively impacted eyewitness accuracy. They concluded that the low light (0.7 lx) and 20m condition fell to chance levels, which is the same as guessing, and therefore hold zero diagnostic value. Additionally, they discussed that because the encoding conditions were relatively optimal compared to real eyewitness cases (i.e., short retention interval, long encoding time, etc.) the results may actually be worse for real-world crimes.

It is important in future work to investigate the boundary conditions for these effects, the interactive effect of multiple estimator variables, and to explore which estimator variables carry more weight than others. For instance, if distance is far and the lighting is dim, then it might not matter if a suspect has a weapon, as it would be too hard to see or factor in. Therefore, I would be interested in seeing if lighting and distance carry the most weight, and which of these would be considered most important. Establishing these boundaries or this hierarchy would help jurors make decisions.

Retention Interval

Retention interval is the time between the encoding event (e.g., crime) and the identification task (e.g., lineup). Previously a conference presentation by Giacona al. (2020) attempted to classify different estimator variables from the National Registry of Exonerees

(2022). From 202 cases that had information available about retention interval, we found that 17.% were the same day, 24.8% were within a week, 23.3% were within a month, 17.3% were 1-6 months, 5% were 6-12 months, and 11.9% were over a year. This finding is consistent with earlier research. For instance, Steblay et al. (2014) conducted a field study of actual police lineups and found that the median time delay between the crime and the lineup was about two weeks. It is noteworthy that researchers typically use retention intervals of less than an hour, but the real-life retention intervals are typically much longer.

Sauer et al. (2010) examined the effect of retention interval by showing individuals a target (in-person) and then either showing a lineup immediately after or after a delay (M = 23 days later, R = 20-50 days later). Consistent with prior research, accuracy was superior in the immediate condition rather than the delayed. The delayed condition also had greater levels of overconfidence and a decrease in accuracy as compared to the immediate condition. They did find that when the data was broken into confidence bins, there was no significant difference in performance between the conditions in the high confidence bin. While it makes sense that longer retention intervals lead to less accurate eyewitnesses, Wixted et al. (2016) argued that highly confident eyewitnesses remain accurate even after nine months. In general, research suggests that longer retention intervals lead to less accurate witnesses, but perhaps there is something special about highly confident witnesses; something we will discuss later.

Weapon Focus Effect

The weapon-focus effect refers to the finding that eyewitness memory is impaired when a weapon is present during a crime (Pickel, 2007). The weapon-focus effect has been studied for its effect on eyewitness identification for years (Loftus et al., 1987; Steblay, 1992; Valentine et al., 2003). Participants in studies with a weapon, or other surprising object, often focus on the weapon and therefore do not encode the information from the perpetrator's face. In a classic study by Loftus et al. (1987), participants were shown a slide show with either a cashier being shown a gun or a check. After a 15-minute retention interval, participants were given multiple choice questions about the crime and were then shown a culprit present lineup. Participants in the gun condition did significantly worse on the lineup and multiple-choice test than the participants in the check condition. Eye-tracking data showed they focused more the gun than the check, suggesting that the presence of the gun drew the participant's attention away from other aspects of the event.

In a study conducted by Erickson et al. (2014), they wanted to test if weapon-focus effect applied to weapons or anything that would be unexpected. Participants viewed slides from the perspective of a bartender. On the final slide a woman is shown holding either an expected object (i.e., a glass), an unexpected but non-threatening object (i.e., a rubber chicken), or a weapon (i.e., a gun). Their findings showed that both the weapon and unusual item group were less accurate than the control group. This finding matches other researchers (Pickel, 1998) that also found that weapon-focus effect may be due to the presentation of an unusual item and not specific to a weapon. Erickson et al. (2014) also found that the weapon focus effect occurred even if the weapon and the perpetrator's face were not shown in the same slide, suggesting that the weapon focus effect is not merely due to drawing eye movements away from the face, but may also affect pre or post encoding processes.

Carlson et al., (2017) manipulated whether a weapon was shown or concealed and only suggested to be present. There were three conditions in the experiment. In the visible weapon condition, the video depicted a man robbing a woman with a gun that was clearly visible. In the concealed weapon condition, the robber's hand was in his jacket pocket, and he gestured as if there was a gun inside of his pocket. In the no weapon condition, the robber demanded the victim's money but did not have a weapon, nor was there any suggestion of a concealed weapon. The visible presence of the gun reduced identification accuracy relative to the no weapon condition, but the concealed weapon condition did not significantly differ from the no weapon condition.

Own-Race Bias

The own-race bias (ORB) or cross-race effect (CRE) is the finding that witnesses are superior at recognizing members of their own racial or ethnic group than members of other racial or ethnic groups. To date, there have been 375 DNA exonerations according to the Innocence Project (2022), and 69% of those were due to eyewitness misidentification. Of those cases involving an eyewitness, 42% involved a cross-race effect where the witness and suspect were of different races. One of the most cited phenomena in eyewitness psychology is the ORB. Meissner and Brigham (2001) conducted a metanalysis of 39 studies. They found strong and consistent support for the claim that participants had a higher proportion of correct identifications for own-race faces than for other-race faces and lower false alarm rates. This phenomenon is the highest consensus sub-areas of eyewitness research and has been continually replicated (Brigham et al., 2007; Wright et al., 2001; Wright et al., 2003).

Importance of Understanding Eyewitness Confidence

Courts

Suppression hearings are hearings where attorneys can argue to suppress evidence from being presented to the jury (Vishney, 2016). Suppression hearings arise out of the gatekeeping role of the judge (Findley, 2012). Evidence that is too unreliable, lacks foundation, is hearsay, is lacking in probative value, or where the prejudicial value tends to outweigh the probative value can be excluded by the judge. Defense attorneys will often make motions to exclude eyewitness identification evidence on the grounds that it was obtained in a suggestive manner.

The U.S. Supreme Court addressed the question of when eyewitness identification evidence should be suppressed in the case of *Neil v. Biggers* (1972). *Neil v. Biggers* dealt with a sexual assault case in Tennessee. Biggers had been accused of sexually assaulting a woman and threatening to kill her and her child with a butchers knife. Several weeks after the attack, the victim was brought to the police station, where she was shown the suspect in a showup identification procedure. As I described above in detail, showups are more suggestive and error prone than lineups and this problem was potentially made worse by the long-time delay and the presence of a weapon during the original crime. The victim in the case identified Biggers and said she was highly confident. At trial she testified, “That I have no doubt, I mean that I am sure that when I -- see, when I first laid eyes on him, I knew that it was the individual, because his face -- well, there was just something that I don't think I could ever forget.” The jury voted to convicted Biggers and he was sentenced to 20 years in prison.

Biggers claimed that his due process rights were violated because the procedure used to obtain the identification was suggestive and that the trial court should have excluded the showup identification evidence from being heard by the jury. He appealed all the way up to the Supreme Court, which eventually ruled against him in a 5-3 decision. The Court agreed that the showup procedure was suggestive, but not so suggestive that it was a violation of Biggers due process rights. They developed a two-pronged test for whether an eyewitness identification should be excluded. The first question asks: “Is the identification procedure a suggestive one?” If the answer to this is “Yes,” then the court asks whether the identification is still likely to be reliable "under the totality of circumstances." The Supreme Court listed a set of factors that it held that courts should consider in this analysis.

> The factors to be considered in evaluating the likelihood of misidentification include the opportunity of the witness to view the criminal at the time of the crime, the witness' degree of attention, the accuracy of the witness' prior description of the criminal, the level of certainty demonstrated by the witness at the confrontation, and the length of time between the crime and the confrontation. (*Neil v. Biggers,* 1972)

Thus, *Neil v. Biggers* established that even when police investigators question a witness in a suggestive manner, the identification can be allowed if, under all the circumstances, the chance of a mistaken identification was limited. Key among the factors the Court listed was the certainty of the witness.

In *Manson v. Brathwaite* (1977), the court further confirmed *Neil v. Biggers* (1972). The case involved Jimmy Glover, an undercover Connecticut State Police officer, who was taking part in a drug sting operation. Officer Glover bought heroin from a drug dealer through the open door of an apartment, a $20 transaction that Officer Glover stated lasted about three minutes. Officer Glover later described the drug dealer to another police officer. Based on the description, the other officer had a suspicion that the drug dealer might by Brathwaite. He printed out a picture Brathwaite and left the picture at Officer Glover's office. When Officer Glover returned to the office, he looked at the picture, and identified Brathwaite as the drug dealer. Based mostly on this identification, Brathwaite was convicted. He appealed the conviction, arguing that the identification was needlessly suggestive and that it violated his due process rights. As in *Neil v. Biggers*, the Supreme Court decided that an identification from a suggestive identification procedure should not necessarily be excluded, as long as there are other factors that would tend to make the identification reliable. They said that "reliability is the linchpin in determining the admissibility of identification testimony" and that the question to be addressed in an exclusion hearing is whether, under the totality of circumstances, the procedures used create "a very substantial likelihood of irreparable misidentification." They once again said that an important part of this determination is how confident the witness is. These Supreme Court cases mean that even when the police use suggestive procedures to obtain identification evidence, the identification is unlikely to be excluded if the witness expresses a high degree of certainty.

Bringing Charges

Under the American Bar Association Criminal Justice Standards (2022), prosecutors are only supposed to bring a case against a defendant if (1) they believe the person is guilty and (2) they believe they will be able to prove beyond a reasonable doubt to a jury that they defendant is guilty. Prosecutors tend to believe that high eyewitness confidence is an important sign that the identification is likely to be reliable, and they believe this more than defense attorneys do (Pezdek & Obrien, 2014). Additionally, an analysis of actual cases showed that prosecutors are more likely to bring a case against a defendant if the eyewitness is high in confidence than if the witness is low in confidence (Flowe et al., 2011). Thus, eyewitness confidence is important to understand, because the confidence of the eyewitness plays an important role in determining whether charges will even be brought against a suspect.

Plea Bargaining

Legal scholars have found that the majority of criminal cases are settled by using plea bargaining (Clarke, 2001 as cited by Kramer et al., 2007), and defense attorneys have a tendency to use plea bargaining for quick conclusions to cases (Blumberg, 1979 as cited by Kramer et al., 2007). To further examine the relationship between strength of evidence and plea bargaining, Kramer et al. (2007) surveyed criminal defense attorneys, both private attorneys and attorneys from the public defender's office. They manipulated potential sentence length, likelihood of conviction/strength of the evidence, and the defendant's wishes about the plea bargain to find the likelihood of the attorney to recommend a plea bargain. One of their major findings was that defense attorneys are less likely to recommend plea bargains when the evidence is weak, and more likely to recommend a plea bargain when evidence is strong. Therefore, Kramer et al. (2007) were able to demonstrate that plea bargaining is influenced by the strength of evidence against a defendant.

To the degree that an eyewitness is confident, it puts the defendant in a disadvantageous position with regards to any plea agreement. A study by Kutateladze et al. (2015) investigated the role of evidence in plea bargaining in felony drug cases processed by the New York County District Attorney's office. Specifically, they found that "punitive charge offers when they had audio/video evidence, eyewitness identification(s), …" (Kutateladze et al., 2015, p. 431). This illustrates how important eyewitness identification evidence is viewed by the prosecution, to the extent that it could negatively affect a defendant's ability to negotiate a plea bargain.

Jury Decision Making

Witness confidence is one of the strongest predictors of a jury verdict (Topp-Manriquez et al., 2014; Kabzińska, 2015). Juries tend to side with witnesses regardless of seeing ability, contradictory statements, or contradictory levels of confidence. In one study, witness confidence was found to influence jury decision making even when the witness was legally blind (Loftus, 1974). In a study by Brewer and Burke (2002), when the witness has made contradictory statements, jurors still heavily weighted their statement and were more likely to be believed if the witness was confident. Additionally, even if a witness changes confidence over time, such as when a now confident witness has previously said they were low in confidence, jurors are still more likely to believe the suspect is guilty if the witness' final confidence judgment is high (Key et al., 2022).

Theoretical Accounts of Relationship Between Confidence and Accuracy

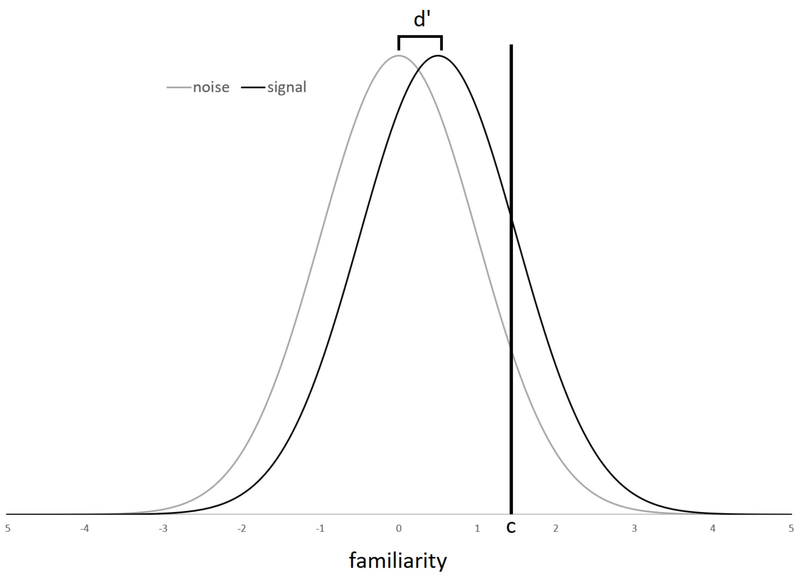
Metacognition is the ability to know what we do or do not know with previous definitions even describing it as "thinking about thinking" (Livingston, 2003). Fleming and Lau (2014) use a more comprehensive definition; "The ability to recognize one's own successful cognitive processing, in e.g., perceptual or memory tasks, is often referred to as metacognition" (abstract). Yeung and Summerfield (2012) conducted a review in which they discussed metacognition with respect to confidence and error monitoring. It has been found that individuals are quite good at producing confidence levels that relate to task performance, even when no explicit feedback is given. There are different models used throughout metacognitive studies that also are used in eyewitness identification literature. The two specific ones I will be discussing are: signal-detection theory and dual process theory.

Signal Detection Theory

Signal-detection theory (SDT) is a common explanation of the metacognitive functions of eyewitness identifications. SDT is a general theory of how people distinguish between signal and noise. The simplest situation in which SDT is applied are old/new recognition tests. In an old/new recognition test, participants are presented with a list of items, some of which were previously studied (i.e., targets) and some of which were not previously studied (i.e., lures). If the correct answer is no, one can respond correctly with "no"— a correct rejection, or incorrectly with "yes"— a false alarm (FA). If the correct answer is yes, one can respond correctly with "yes"— a hit (H), or with "no"— a miss. SDT provides a model for understanding how participants respond to old/new recognition tests and other tests of recognition memory, which is illustrated in Figure 3.

Figure 3

Illustration of Equal Variance Signal Detection Theory.



	Item is Old	Item is New
Say Old	Hit	False Alarm
Say New	Miss	Correct Rejection

When applied to recognition memory, SDT assumes that encountering an item on a recognition memory test results in a sense of familiarity. The sense of familiarity is on a continuum from weak to strong. The sense of familiarity may be produced because one has previously encountered the item (i.e., it is a signal). On the other hand, the sense of familiarity may be produced in response to an item that was not previously encountered (i.e., it is noise). Across individuals and across trials, the feeling of familiarity is normally distributed for both the signals and the noise, however, on average signal trials have higher familiarity than noise trials. There will always be overlap between noise and signal, but the ability to tell them apart is discriminability (Gronlund & Benjamin, 2018). The difference in average familiarity between the signal and noise trials, measured in standard deviation units, is a measure of discriminability called d’. Large values of *d’* indicate that the person is good at distinguishing signal from noise (memory is good). Low values of *d'* indicate that the person is poor at distinguishing signal from noise (memory is bad).

In SDT, it is assumed that there is some overlap between the signal and noise distributions. Because of that, SDT assumes that participants set a response criterion in order to make these judgment calls. If an item on a memory test produces a sense of familiarity that is above the criterion, it will result in a "yes" response (i.e., a hit or false alarm). If the sense of familiarity is below the criterion, it will result in a "no" response (i.e., a miss or correct rejection). The criterion is set by numerous factors that can influence the degree to which a person is willing to answer yes or no. A more stringent (i.e., conservative) set of criteria will result in fewer incorrect acceptances on noise trials (fewer false alarms) but also fewer correct acceptances on signal trials (fewer hits), when compared to a looser (i.e., liberal) set of criteria (Lee & Penrod, 2019). A common assumption in SDT is that people set response criteria partly based on their perception of the relative costs of different kinds of errors (Lynn & Barrett, 2014).

The eyewitness identification task is more complex than an old/new recognition task, and so the signal detection models that have been developed to explain eyewitness identifications are more complex as well. A number of different signal detection models of eyewitness identification have been proposed that make slightly different assumptions (Duncan, 2006; Lee & Penrod, 2019; Smith et al., 2018; Wixted & Mickes, 2014; Wixted et al., 2018). The simplest signal detection model of eyewitness identification claims that when a witness is shown a culprit present lineup, each of the fillers has some level of familiarity associated with it and the guilty suspect as also some level of familiarity associated with it. The familiarity of the fillers can be thought of as being randomly drawn from a noise distribution. The familiarity of the guilty suspect can be thought of as being randomly drawn from a signal distribution, with the average familiarity of the signal distribution being *d'* higher than the average familiarity of the noise distribution. When a witness is shown a culprit absent lineup, the familiarity values of all the lineup choices (fillers and innocent suspect) are drawn from the same noise distribution. Witnesses are assumed to have a response criterion. If all the familiarity of all the lineup choices are below the response criterion, the witness rejects the lineup. If one or more lineup member produces a familiarity value above the criterion, the lineup member with the highest familiarity will be chosen. There is currently debate about whether this model provides the best explanation of eyewitness identification, or if more complex theories are needed (c.f., Colloff et al, 2018; Smith et al., 2018).

Dual Process Theories

Signal detection theory assumes that memory operates on a single dimension. Items on memory tests produce some level of familiarity, and the participant determines their answer based on that single dimension. Dual process theories, on the other hand, claim that there are two different types of processes that can be used in making memory judgments. There are a number of different dual process theories that have been proposed (e.g., Mandler, 1980; Jacoby, 1991; Reyna & Brainerd, 1995).

Pennycook (2017) describes the dual process theory as autonomous or non-autonomous. Thompson (2009) describes two memory systems, that can be described as "fast and automatic or slow and conscious" depending on which system is referenced. The first system, called Automatic System 1 (S1), is the automatic system, whereas System 2, or S2, is the more analytic system (Stanovich & West, 2002). The basic theory is that much decision-making is handled by the automatic process and does not need further processing. When further processing is needed, that is when the S2 system comes into play. For example, Pennycook (2017) asked people to consider 2x0=? versus 22172x71=? in arithmetic. The first problem triggers an immediate and autonomous response, whereas the second problem takes more analytic processing. The answer can be reached, but for the vast majority of individuals, it will not trigger an automatic answer and will instead require greater effort. There are many proposed theories and instances of when and how the S2 is activated or engaged that still warrant research (Thompson, 2009); however, applying this process to memory (and eyewitness) judgments would be beneficial.

The most commonly made distinction in dual process models is between recognition that is based on *recollection* and recognition based on *familiarity*. Yonelinas (2001) proposed that recollection and familiarity differ in terms of 'the three C's': Consciousness, Control and Confidence. Consciousness refers to "remembering" or "self-knowing," where the encoding event is consciously reexperienced. The consciousness component of recollection refers to the fact that recollection involves a kind of re-experiencing of the event that one becomes consciously aware of. Another distinguishing factor is intentional control, where being able to recollect an event should lead to accurate discrimination of items from different sources, whereas familiarity would not be able to discriminate from the different sources. In terms of confidence, recollection is thought to drive the performance in high-confidence responses, and conversely, familiarity supports a wider range of memory confidence responses. Yonelinas (2002) found that familiarity is most affected by forgetting over short spans, fluency manipulations, and changes in response criterion, whereas recognition is most manipulated by semantic encoding, attentional division, responding speed, and novel learning.

There have been some attempts, namely by Meissner et al. (2005), to apply this dual-process theory to memory judgments. In their study, they used Yonelinas' (2002) distinction between recollection and familiarity as the two parameters for memory performance. They hypothesized that reliance on familiarity would be reduced by using sequential lineups. Eyewitness identification researchers often analyze their data using an SDT approach, but in their final experiment, Meissner et al. (2005) collected data using a remember-know-guess procedure to assess the dual-process account. They found that simultaneous lineups were associated with more familiarity judgments in than sequential lineups, and responses in sequential lineups were less influenced by familiarity, which is in line with previous dual-process theory research (Meissner et al., 2005).

Eyewitness Metacognition

Eyewitness studies have shown that eyewitnesses tend to be overconfident (Wixted & Wells, 2017). For instance, in one study, Brewer et al. (2002) found that witnesses who said they were 100% confident were correct less than 80% of the time. As discussed prior, there are two main theories SDT and dual process theory. Both of these are important for understanding the relationship between eyewitness confidence and eyewitness accuracy.

The idea behind these signal-detection studies in eyewitness context is that individuals are able to metacognitively decide how much of an impact different variables would have and adjust their confidence accordingly. For example, if there was a study about what color hair someone had, it is common sense to know that the individual would be better when it is bright versus when it is dark. Therefore, the individual should know that they see better when it is bright than when it is dark, and they will be more confident in what they saw if it was observed during the day than at night. They would be more liberal with their criterion in bright settings because there is a better view, whereas they will be more conservative with their choosing in dark settings because they know they have to be even more sure of their decision due to the bad lighting.

While this theory is good and continues to be a dominant theory, there are some issues. One such issue is that it relies on the participant/witness/individual to make the appropriate decision. While some variables, such as distance and lighting, might be relatively easy to understand, other variables such as stress and weapon-focus effect may have more nuanced answers and contradict people's perceptions. This is something to consider.

Eyewitness research has found that quicker judgments tend to be more correct than those made after a longer time scrutinizing (Sporer, 1993). This may mean that eyewitness judgments tend to be more automatic and may not warrant deeper processing. Typically, accurate eyewitnesses are faster at making identifications than inaccurate eyewitnesses (Brewer et al., 2006; Dunning & Perretta, 2002; Weber et al., 2004). While there is debate on whether there is an exact time boundary or not, it seems that quicker responses have higher accuracy than slower responses. This seems to support the dual process theory that automatic judgments tend to be quicker. Specifically, when a witness is correctly identifying someone, it is a very quick, and seemingly automatic, process. Whereas, when they have to scrutinize and take longer, this indicates that they may not be as automatic or correct.

Measures of the Confidence Accuracy Relationship

There is some controversy surrounding which analyses are best for measuring eyewitness identification in lineup procedures. In this section, I review how the confidence-accuracy relationship has been measured by eyewitness researchers. It turns out that these methodological differences had a large impact on the conclusions that researchers drew about confidence and accuracy.

Point Biserial Correlation

Early on, eyewitness researchers studied the confidence / accuracy relationship by calculating point biserial correlations between how confident witnesses were and their accuracy (Sporer et al., 1995, for a review). The point biserial correlation is just a specific case of the Person product-moment correlation, where eyewitness accuracy was coded as 0 for incorrect (false alarm or miss) and coded as 1 for correct (hit or correct rejection) (Wixted & Wells, 2017). The correlation between confidence and accuracy, when using a point biserial correlation could be anywhere in the range of -1 to +1, with lower absolute values indicating a weaker relationship. It was not usual for early researchers to describe effects in terms of the proportion of variance explained, with that value being the square of the point biserial correlation. For instance, an article by Lindsay et al. (1981) described it this way, "The strength or weakness of the individual accuracy-witness confidence relationship is contained in the proportions of variance on the individual accuracy dimension accounted for by witness confidence" (p. 82).

Many researchers have shown that overall, the relationship between confidence and accuracy using point biserial statistic ranges from low (Wells & Murray, 1984) to medium (Lindsay, Red & Sharma, 1998; Penrod & Cutler, 1995; Sporer et al., 1995). These findings led many in the field to conclude that there was only a weak relationship between confidence and accuracy at best (Lindsay, Red & Sharma, 1998; Penrod & Cutler, 1995; Sporer et al., 1995; Wells & Murray, 1984). However, an important limitation of the point-biserial correlation was outlined in Juslin et al. (1996). They demonstrated that even when the confidence accuracy relationship is perfectly calibrated – i.e., when witnesses are 100% confident, they are accurate 100% of the time, when witnesses are 90% confident they are accurate 90% of the time, etc. – the point biserial correlation can still be very low (Juslin et al., 1996). This is because the point biserial correlation is measuring the degree to which a continuous variable (confidence) and a dichotomous variable (accuracy as measured 0 or 1) can be fit to a straight line – generally they cannot be. In order to explore this deeper, Juslin et al. (1996) generated simulated data, and analyzed data where confidence and accuracy were perfectly calibrated, but different distributions: uniform, unimodal, and bimodal. They found that the correlations varied from low to high correlation (r =.58, r =.30, r =.78 respectively).

Calibration Analyses

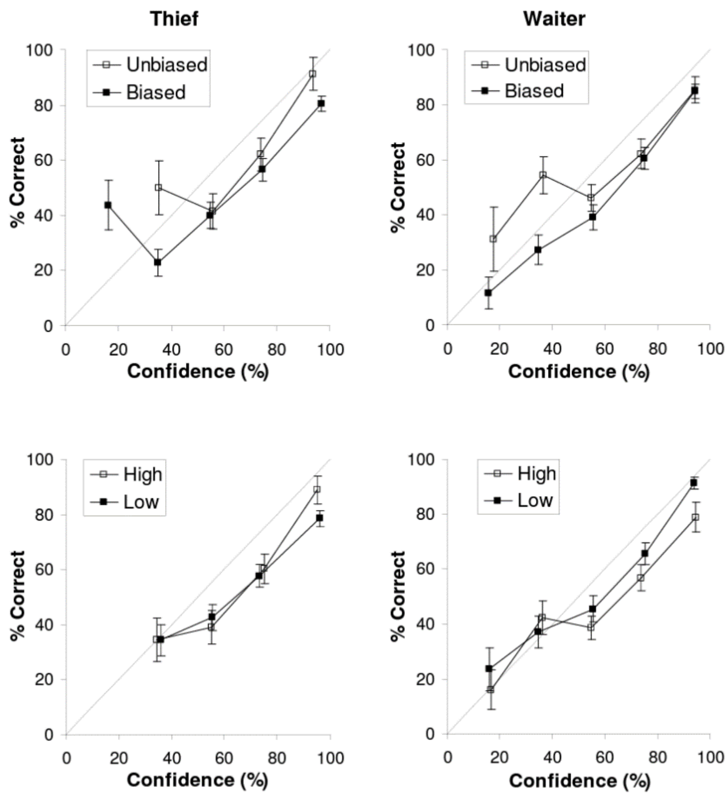
Juslin et al.'s findings led researchers to conclude that the use of the point biserial correlation was flawed. Instead, Juslin et al. recommended the use of calibration analyses to view the relationship between confidence and accuracy. A calibration analysis separates participant responses into different confidence bins (e.g., 90-100, 70-80, 50-60, etc.). For each confidence bin, the accuracy is calculated by taking the proportion of correct suspect IDs from culprit present lineups and dividing by the proportion of correct IDs from culprit present lineups plus the proportion of mistaken IDs from culprit absent lineups.

Accuracy = *correct IDs/(correct IDs + incorrect IDs)*

If witnesses were perfectly calibrated, then the observed accuracy of witnesses in each confidence bin would equal the expected accuracy for that confidence bin, which is usually operationalized as the midpoint of the confidence bin. If witnesses, in general are less accurate than one would expect based on their confidence, then witnesses are over-confident. If witnesses, in general, are more accurate than one would expect based on their confidence, then witnesses are underconfident.

Figure 4

Example of Calibration Curves



Note. The original figure was published in Brewer & Wells (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. Copyright American Psychological Association. Reprinted with permission.

Calibration analyses typically involve plotting calibration curves showing accuracy as a function of the confidence bins. Figure 3 shows a calibration curve from Brewer and Wells (2006) in which they compared biased lineup instructions with unbiased lineup instructions. The X-Axis shows confidence bins and the Y-Axis shows how accurate witnesses are in each of those confidence bins. The diagonal line shows what would be expected if witnesses were perfectly calibrated. As can be seen, for witnesses in the unbiased condition, accuracy for low confidence judgments are more accurate than would be expected based on witness confidence. Graphically, this is seen because those data points are above the diagonal line. For witnesses who are higher in confidence, witnesses are less accurate than one would expect based on their confidence. This can be seen in Figure 4, where those data points are below the diagonal line, which indicates predicted performance.

Calibration analyses also involve calculating a number of summary statistics. For instance, C represents how close accuracy is to confidence. It is analogous to calculating the sum of squared errors. For each confidence bin, the difference between confidence in that bin and accuracy in that bin is calculated and then that difference is squared and this squared difference is multiplied by the number of observations in that confidence bin. The sum of these values is then divided by the overall number of observations. In the equation below, N is the total number of observations, T is the number of confidence bins, n_t is the number of observations in confidence bin t, a_t is the accuracy in confidence bin t and c_t is the confidence midpoint for confidence bin t.

$$C = \frac{1}{N}\sum_{1}^{T} n_t(a_t - c_t)^2$$

Another statistic is the O/U statistic, which measures whether witnesses are, on average, over or under confident. It is calculated in the same way as C, except that the difference between accuracy and confidence in every bin is not squared, allowing the statistic to reflect whether accuracy is, on average, higher than one would expect based on confidence, lower than one would expect based on confidence, or about equal to what one would expect based on confidence.

$$O/U = \frac{1}{N}\sum_{1}^{T} n_t(a_t - c_t)$$

Confidence Accuracy Characteristics (CAC) Curves

The main issue with this calibration analyses is that they treat every filler identification and "error" equally. But recall that when police officers conduct a lineup, they typically have a particular person who is their suspect and who may or may not be guilty. The other people in the lineup are fillers and are known by the police to be innocent. If a witness identifies the suspect, and the suspect is innocent, the suspect is in a great deal of trouble. But if the suspect is innocent, and the witness picks a filler, that filler is not in any trouble at all. The police know that the filler is innocent. Therefore, when a witness incorrectly chooses a filler that is not the suspect, this is less of an issue than if they incorrectly choose an innocent suspect. In the first scenario, police officers and experimenters know the witness incorrectly chose; however, in the second scenario, an innocent person would have been incorrectly chosen, which could lead to an innocent suspect behind bars. So, while picking an innocent suspect and picking an innocent filler are both errors, the errors have dramatically different costs associated with them.

Based on this reasoning, Mickes (2015) developed a graphical technique she called confidence accuracy characteristic (CAC) curves. CAC curves very similar to calibration curves, however in a CAC analysis only mistaken suspect identifications are counted as errors. Therefore, instead of including all errors, the only errors that are included are selections of innocent suspects. The formula for each accuracy point on the CAC curve is as follows (Mickes, 2015):

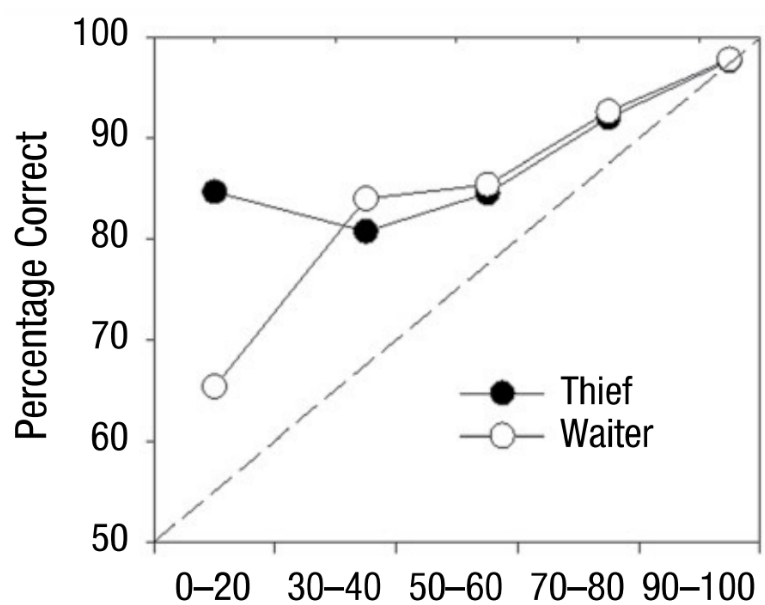
$$suspect\ ID\ accuracy\ (A) = \frac{\#\ correct\ suspect\ IDs}{(\#\ correct\ suspect\ IDs + \#\ incorrect\ suspect\ IDs)}$$

This formula illustrates that for each level of confidence (indicated by "A" in the formula), the proportion of correct suspect identifications is divided by the proportion of correct suspect identifications plus the proportion of incorrect suspect identifications (filler IDs are not counted). Furthermore, Mickes (2015) suggests that if a lineup is fair but there is no "designated suspect" (i.e., one where there is a specific lookalike culprit being used) dividing all selections made in the culprit absent lineup by the lineup size would be an appropriate way to estimate the proportion of innocent suspect identifications to be expected from the procedure.

Below, in Figure 5, is an example of a CAC curve from Brewer and Wells (2006). Just like a calibration curve, the X-Axis shows the different confidence bins. The Y-Axis shows the percent correct specifically of suspect identifications. As with a calibration curve, the dashed line shows a perfect confidence accuracy relationship. As can be seen, confidence and accuracy were related, and also highly confident witnesses were very accurate in their suspect identifications.

Figure 5

CAC Curve Brewer and Wells (2006)



Note. CAC curve based Brewer and Wells (2006) as reproduced in 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*(1), 10-65. Copyright © 2022 by Association for Psychological Science. Used by permission via fair use guidelines.

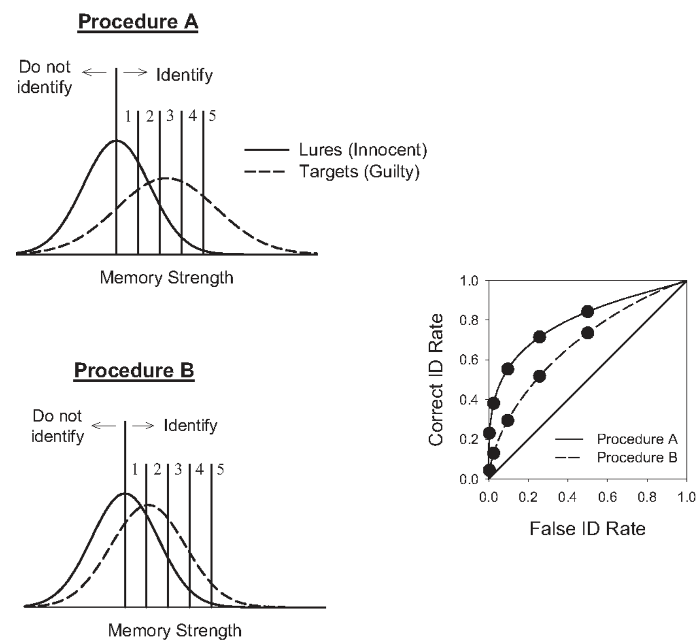
Receiver Operator Curves (ROC) Analyses

In an influential paper, Wixted and Mickes (2012) suggested that eyewitness memory should stop using current methods and “embrace receiver operating characteristic analysis.” As I described above, in signal detection theory a participant’s response is a function both of their ability to distinguish between signal and noise (i.e., discriminability) and their response criterion. ROC analyses were developed as a way of examining discriminability while controlling for response criterion. The idea is to plot the hit rate at a variety of different response criteria allowing the researcher to determine which procedure produces the better outcome after equating the response criteria. An ROC curve plots the hit rate as a function of the false alarm rate as response criterion is varied. There are two main ways response criterion can be varied. Some researchers experimentally manipulate the response criterion by varying the payoffs (e.g., Curran et al., 2007) or the ratio of targets and lures on the study list (Ratcliff et al., 1992). More commonly, participants provide confidence judgments with the assumption that different levels of confidence indicate different response criteria (e.g., Yonelinas, 1994).

For instance, in a confidence-based ROC curve, the hit rate is indicated on the Y-Axis and the false alarm rate on the X-Axis at different confidence cutoffs (Mickes et al., 2015; Rotello & Chen, 2016). It is easy to visualize which condition is superior as the best ROC curve outcomes should be in the top left corner, where false alarm rates are lowest, and hit rates are highest. Ideally, the best procedures minimize incorrect identifications of innocent suspects and maximize the correct identification of guilty suspects. If you are comparing two conditions then (see Figure 6), you can compare hits rates while holding false alarm rates constant. If one condition has a higher hit rate at each of the different false alarm rates, it is said to *dominate* the other condition and is considered to be the superior condition (Fawcett, 2004).

Figure 6

ROC Analyses for Two Different Hypothetical Eyewitness Identification Procedure



Note. For any particular False ID rate, Procedure A is associated with a higher correct ID rate than procedure B, leading to the conclusion that it dominates procedure B and is the better procedure. The figure is from Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262. Copyright American Psychological Association. Used with permission.

Although ROC analyses are common in memory and perception research, there has been criticism of applying ROC methods in eyewitness identification. For instance, Wells et al. (2015) argued that instead of breaking down witness responses into identifying a suspect, identifying a filler, or not choosing anyone (AKA rejecting the lineup”), ROCs combine filler picks and rejecting a lineup as one category. They argued that this is conceptually misleading because rejecting a culprit absent lineup is a correct choice but selecting an innocent filler from a culprit absent lineup is an incorrect choice. But lineup ROC analyses lump them into the same category. Lampinen (2016) argued that the procedure with the better theoretical discriminability does not necessarily produce the higher ROC curve in lineup ROCs and also that the higher ROC sometimes has a lower expected utility, making it a worse procedure from an applied perspective (Lampinen et al., 2019; Smith et al., 2019).

Which Measure Do We Use?

A number of different measures have been proposed to measure eyewitness performance, and there is debate about which procedure is best to use. This has led some to claim that there is a ‘crisis’ in eyewitness identification research (Seale-Carlisle et al., 2022) or that researchers have been using the ‘wrong’ measures leading to mistaken conclusions (Gronlund et al, 2015; Rotello et al., 2015). Some parts of the debate has concerned which measures best capture the underlying memory abilities of witnesses, while other parts of the debate has concerned which measure best captures the applied benefits and costs of different procedures (Wixted & Mickes, 2012). Although these debates are important, Lampinen et al. (2019) showed that a large subset of cases are trivial from a measurement perspective – meaning any reasonable measurement technique will provide the same results. In later work, Lampinen et al. (2021) compared published research in which ROC analyses were used and showed that in more than 90% of cases ROC analyses and diagnosticity (i.e., Hits/FAs) produced the same pattern of results and in more than 80% of cases ROC analyses and difference scores (i.e., Hits – FAs) all produced the same pattern of results. In another analysis, they randomly generated pairs of ROC curves with randomly selected *d’* values for each curve and random chosen degrees of truncation for each curve, and again showed that, in the vast majority of cases, ROC analyses, diagnosticity and difference scores all produced the same pattern of results.

Mickes (2015) proposed that it is most appropriate to use ROC curves when measuring system variables and CAC curves when measuring estimator variables. She suggested that ROC analyses are most important for system variables because they can “simultaneously maximize correct IDs while minimizing incorrect IDs." This is important because ROCs contribute information that would affect systematic changes, whereas CAC curves are used more useful for evaluating a witness's testimony. For example, CAC curves, when paired with estimator variables, typically look at how confidence affects accuracy in different conditions. While it is true to say that individuals have worse memory when the exposure duration is shorter, it would also be true to say that, based on CAC curves, witnesses are that are highly confident are still highly accurate even with the difference in exposure duration (Mickes, 2015).

In short, Lampinen et al. (2021) has shown that most cases are trivial; the exact analysis used will not matter for the majority of cases. Additionally, Mickes (2015) has suggested that ROC curves are best for system variables, and CAC curves are best for estimator variables. However, this remains a contentious debate in the eyewitness community, and often, it is common for researchers to use both ROC and CAC but keep on in the supplemental materials section.

Pristine Conditions Hypothesis

One of the first studies to find that high eyewitness confidence can be associated with very high accuracy was a study by Brewer and Wells (2006). Although previous research had found relatively weak correlations between accuracy and confidence through point-biserial correlation, Brewer and Wells (2006) opted to use the calibration approach. In their study, they manipulated fair versus biased instructions, high or low filler similarity, and whether the target was present or absent from the lineup. Using 1,200 participants, they were able to analyze their results using a point-biserial correlation and a CA approach. When analyzing their results using point-biserial correlations, they found modest effects as most. However, when they used calibration analyses, they found that higher levels of confidence indicated greater accuracy than lower confidence across all conditions. In particular, Brewer and Wells used a measure of diagnosticity at each level of confidence. Diagnosticity is the ratio of correct suspect identifications to mistaken suspect identifications. Thus, a diagnosticity of 2.00 would mean that when a suspect is identified it is twice as likely that the suspect is guilty as it is that the suspect is innocent (assuming a 50% *a priori* probability of guilt). When participants were 90-100% confident, diagnosticity reached 38.31 for one of the lineups and 20.39 for the other lineup. This means that for one lineup, when a suspect was identified it was 38.31 times more likely that the suspect was guilty than they were innocent. For the other lineup, it was 20.39 time more likely that the suspect was guilty than they were innocent.

More recently, researchers have proposed that under certain conditions (known as pristine conditions) eyewitness that are highly confident will tend to be highly accurate (Wixted & Wells, 2017). When using the term "pristine conditions," Wixted and Wells were referring to testing conditions for police lineups that reflect optimal practices. More specifically, Wixted and Wells defined pristine conditions as meeting these criteria: (1) the lineup should include only one suspect, (2) the suspect should not stand out in a lineup, (3) the witness should be instructed that the offender might not be in the lineup, (4) the lineup should be administered using a double-blind procedure, and (5) the statement of confidence should be obtained at the time of the identification (Wixted & Wells, 2017). It is also important to note that the pristine conditions claim was only applied to adult witnesses and not children.

A great deal of research has been conducted that is consistent with the pristine conditions hypothesis. For instance, in Wixted and Wells (2017) re-analyzed data from 20 previous studies and, in each one, suspect identification accuracy for one highly confident witnesses was quite high (typically >95% accuracy). However, there have been some exceptions where highly confident witnesses are not as accurate as would be expected based on the pristine conditions hypothesis, especially under conditions where witnessing conditions are very poor (Giacona, Lampinen et al., 2021; Grabman et al. ,2019; Lockamyeir et al., 2020). I described some of this research in more detail below.

Proponents of the pristine conditions hypotheses claim that individuals adjust their criterion for identification based on their knowledge of the situation, including estimator variables (i.e., variables that cannot be controlled such as lighting and weapon presence; Wells, 1978). For example, if the crime took place in a dark area, then the eyewitness would know that that is a factor that would affect identification and would adjust their confidence accordingly. This would mean that witnesses who are highly confident in this situation, are highly confident even after adjusting their high confidence response criterion to make it more stringent. Although this is a reasonable idea in many situations, it assumes that witnesses have meta-cognitive awareness of how different witnessing variables affect memory. However, survey research has shown that members of the general public often have incomplete or inaccurate knowledge of how memory works in general (Simons & Chabris, 2011) and how eyewitness memory works, in particular (e.g., Benton et al., 2006). For instance, only about 1/3 of the general public is aware that the presence of a weapon can impair eyewitness memory (Benton et al., 2006) despite research demonstrating the effect (Carlson et al., 2017). Another instance is that a substantial proportion (i.e., around 40%) of people think that stressful situations are invariably associated with better memory (Schmechel et al., 2005), despite findings that high levels of stress can impair identification accuracy (Deffenbacher et al., 2004). Knowing that individuals do not always have the right metacognition about estimator variables in itself would mean that not all individuals would be able to correct their confidence in the correct direction. Essentially, if an eyewitness thinks weapons would make identification easier, they will loosen their judgments for what a high confidence decision is when in actuality, they should be more stringent.

Only Contemporaneous Confidence Predicts Accuracy

The pristine conditions hypothesis specifically has to do with the confidence expressed by the witness immediately after the identification. This is important because confidence obtained later in the process, such as when the witness is in court, is typically not as reliable (National Academy of Sciences, 2014). Confidence can be malleable, and how confident a witness is can vary over time. This fragile metric is very susceptible to post-identification feedback and confidence inflation over time. Consider the following true story from *Missouri v. Hutching* (1996, p. 202):

> Eyewitness to a crime on viewing a lineup: "Oh, my God . . . I don't know . . . It's one of those two . . . but I don't know . . . Oh, man . . . the guy a little bit taller than number two . . . It's one of those two, but I don't know."
>
> Eyewitness 30 min later, still viewing the lineup and having difficulty making a decision: "I don't know . . . number two?"
>
> Officer administering lineup: "Okay."
>
> Months later . . . at trial: "You were positive it was number two? It wasn't a maybe?"
>
> Answer from eyewitness: "There was no maybe about it . . . I was absolutely positive."

This eyewitness was clearly hesitant and uncertain when she initially made her identification; how is it then that she was absolutely positive at trial? The answer lies in confidence inflation and confirmation feedback.

In a study conducted by Wells and Bradfield (1998) that was aptly titled "Good, You Identified the Suspect," researchers investigated the relationship between feedback and confidence. Participants were either given confirming feedback (e.g., “good you identified the suspect”), disconfirming feedback (e.g., “you identified X, but the correct answer was Y”), or no feedback. After the feedback, participants were asked how confident they were at the time they made their identification. When participants received feedback it distorted their memory for how confident they had initially been. There was an asymmetry in their findings, however, in that the confirming feedback elicited a greater effect than the disconfirming feedback (a finding replicated by Douglass & Steblay, 2006). They also found that in addition to influencing how confident they were in their decision, post-identification feedback also influenced other judgments such as how good of a view they got and how clear the details of the suspect’s face were. Wright and Skagerberg (2007) conducted a similar study using 134 real eyewitnesses in the United Kingdom and found comparable results. Wixted et al. (2018) have recently argued that the most common pattern in DNA exoneration cases is witnesses who are highly confident in court, despite the fact that their confidence was low during the initial identification.

Another study by Smalarz and Wells (2014) involved a two-phase process that had independent evaluators rate how believable witnesses were. In the first phase, participants watched a video and were then given a target present lineup, where 84% of witnesses correctly identified the suspect and a force-choice target absent lineup, where everyone incorrectly identified a filler. Half of these witnesses were provided with confirming feedback: “good, you identified the suspect.” The other half were not. Then, participants were brought into another room and where they were recorded and interviewed by an administrator that was blind to their condition. In the second phase, independent evaluators watched the videos and indicated how believable they thought the witness was. Researchers found that participants watching the no feedback condition could distinguish between accurate and inaccurate eyewitnesses. However, when viewing the participants in the confirming feedback condition, independent evaluators rated the mistaken and accurate witnesses as equally believable. This is problematic in that providing witnesses with feedback removed features that allowed participants to distinguish accurate from mistaken witnesses. This is particularly problematic given the emphasis on the credibility of highly confident eyewitnesses in current literature. Steps should be taken by police departments to ensure no confirming feedback is relayed to the witness before their confidence is taken. Furthermore, any confidence taken after confirming or disconfirming feedback is contaminated and would violate the pre-conditions set by the pristine conditions hypothesis.

Theories for High Accuracy Despite Suboptimal Estimator Variables

To reiterate, the pristine conditions hypothesis essentially holds that as long as lineups are fair, double-blind, and have one suspect, witnesses are instructed the suspect might not be present, and confidence is taken contemporaneously with the lineup, then high witness confidence is indicative of high witness accuracy. As previously mentioned, Wixted and Wells (2017) conducted a review of 20 studies that found that the pristine conditions hypothesis holds in a myriad of situations. Using the two theories we discussed earlier, dual process theory and SDT, I will discuss the implications for the pristine conditions hypothesis.

Signal Detection Theory

The claim that high confident identifications will be highly accurate, regardless of the strength of the witness's memory, requires some explanation from a signal detection point of view. According to signal detection theory, different levels of confidence reflect familiarity exceeding different response criterion (see Figure 7). If the response criterion does not shift when *d'* changes, then the relative proportion of hits to false alarms would have to decrease when *d'* decreases. For instance, in the example shown in Figure 7, there are virtually no false alarms made with the highest level of confidence and a small number of hits. If *d'* decreases, and the criteria stay in the same spots, the number of high confidence false alarms would not change but the number of high confidence hits would decrease. The only resolution to this would be for the high confidence response criterion to change when *d'* changes. But how that happens is not clear.

Figure 7

Illustration of How Confidence Judgments Are Made in Signal Detection Theory

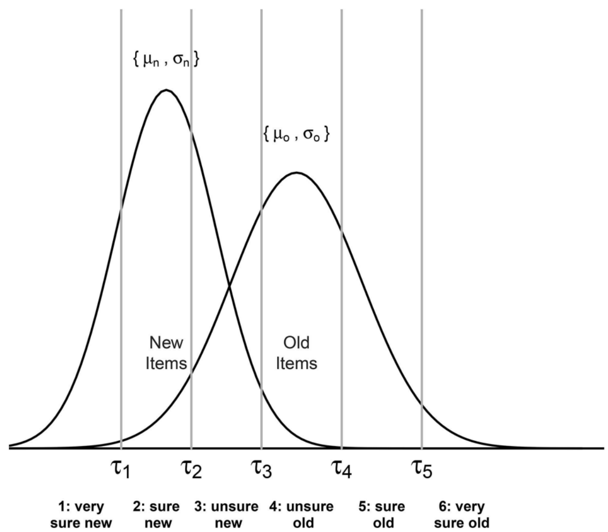


Figure 1. The Gaussian signal detection theory model.

Note. Reproduced from Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review*, *122*(3), 542. Copyright American Psychological Association. Used with permission.

Semmler et al. (2018) investigated the SDT account for the relationship between estimator variables and the confidence-accuracy relationship. In particular, the researchers made the distinction between discriminability of memory strength (*d'*) and positive predictive value (PPV), which is the probability that a suspect identification made by an eyewitness is correct. In their findings, Semmler et al. (2018) show that when looking at exposure duration, retention interval, and distance, poor estimator variables do decrease discriminability; however, they do not decrease PPV of high confidence IDs. They argued that this occurs because as witnessing conditions become poor, response criteria become more stringent, following a constant likelihood ratio model (Stretch & Wixted, 1998)

It is possible that the pristine conditions hypothesis will no longer hold when memory strength becomes sufficiently weak (Semmler et al., 2018). This may be true, in particular, if *d'* is so small that the response criterion would have to be set to an implausibly high level in order to achieve the accuracy rates required by the pristine conditions hypothesis.

Dual-Process Theory

Prior evidence by Mandler (1980) and Yonelinas (2002) have shown that recognition memory involves two processes, and there is a distinction between familiarity and recollection. According to this account, familiarity can be continuous in that there are varying levels of familiarity, whereas recollection either does or does not happen. Wixted and Mickes (2010) proposed an alternative account that attempted to incorporate SDT concepts into a new "continuous" dual-process theory that merges the dual-process theory and SDT concepts. According to this view, both familiarity and recollection are continuous processes.

Mandler (1980) used a "butcher-on-the-bus" anecdote, where seeing a man on a bus is so familiar that it spurs the viewer to scour their memory for why they have this familiar feeling. This familiarity feeling induces high rates of confidence that they have seen this person before, but this happens without the additional recollection of who he is or where the person is. This demonstrates that there are separate recollection and familiarity processes, and people can have strong confidence for familiarity even without explicit recollection.

Further, Nguyen et al. (2017) investigated the cross-race effect using a dual-process approach. Previous research holds that discrimination accuracy is higher for same-race faces than for cross-race faces. However, according to the continuous dual-process theory by Wixted and

Mickes (2010), if confidence is equated, then the race of the participant should not affect the recognition accuracy. Nguyen et al. (2017) found that when confidence was high, both same- and cross-race faces were equally reliable. These results post that the confidence-accuracy relationship may be less dependent on the processing type (e.g., recollection or familiarity) as confidence ratings are expressed by memory strength rather than memory quality.

Exceptions to the Pristine Conditions Hypothesis

More recently there have been a couple of studies that do not support the pristine conditions hypothesis. For instance, Giacona et al. (2021) created a study investigating the effects of multiple variables on the confidence-accuracy relationship. Participants (N = 2,191) viewed a video in which a victim is robbed by two perpetrators. In one condition, the estimator variables were poor in multiple ways: long distance, brief duration, head covering, high stress, weapon present, long retention interval. In the other condition, the estimator variables were good in multiple ways: i.e., short distance, long duration, no head covering, low stress, no weapon, brief retention interval. Participants were then presented with a culprit present lineup for one of the robbers and a culprit absent lineup for the other robber using pristine lineup procedures (i.e., fair lineup, proper pre-lineup instructions, blinded administration, contemporaneous confidence). Assignment of robber to culprit present or culprit absent lineup was counterbalanced across participants as was the order of the lineups. After making their judgment, participants indicated their confidence on a 0%-100% scale in increments of 10%.

Using signal detection theory and the pristine conditions hypothesis, participants should have been able to correctly adjust their criterion based on their viewing condition. Thus, they would still be appropriately calibrated in that those with high confidence should still have high accuracy. In our results, we found that high confidence was reasonably high for both conditions;

however, there was a significant difference in high confidence accuracy between the good viewing conditions and the poor viewing conditions. When the base rates dropped to 35%, the poor viewing condition was significantly less accurate than 90%, showing that this effect may not be replicated in real-world scenarios. This difference was significant and demonstrated that high confidence does not equal high accuracy when there are multiple, suboptimal estimator variables. In terms of SDT, this might show that it is difficult for participants to appropriately change their criterion when there are multiple, competing variables that exist. This research was particularly important as there are many researchers that have minimized the effect that estimator variables have on accuracy (Wixted et al., 2018; Wixted & Wells, 2017).

Grabman et al. (2019) investigated the effect of decision-time, justifications, and face recognition ability on the confidence-accuracy relationship. They found important instances where the high confidence identifications did not equal high accuracy. The first caveat was decision-time. Though the highest confidence (100%) IDs were in the first few seconds were almost always accurate, the accuracy falls to 75% and 50% for six seconds and 20 seconds, respectively. This means that decision-time is still important to consider as an increase in time to make a decision also increases the high confidence errors. Similarly, familiarity justifications and being a “poor face recognizer” in general also lead to higher rates of high confidence misidentifications. Grabman et al. (2019) demonstrates important times when the pristine conditions hypothesis does not hold the high confidence- high accuracy relationship.

Multiple Estimator Variables

Almost all eyewitness identification studies manipulate a single estimator or only a small number of estimator variables (Barbur & Stockman, 2010; Carlson et al., 2017; Lampinen et al., 2012; Loftus et al., 1987; Steblay, 1992; Valentine et al., 2003; Wagenaar & Van der Schrier, 1996); most of which we have discussed previously. This is reasonable as experimenters want to maximize control in their research studies. More variables lead to less control. Additionally, as variables included increases, so will the number of conditions. This means that cell sizes will become small and low powered. Thus, it would take a large number of participants and overall resources to conduct research with a large number of estimator variables.

Early research by Cutler and Penrod explored multiple estimator variables; however, there has been less current research on the topic. In particular, Cutler et al. (1987) reviewed 14 variables, broken into sub-categories: pre-stimulus manipulations, stimulus variables, storage variables, and retrieval variables. There were three pre-stimulus manipulations where the participants were either prepared to see a crime or not (expectation), told to focus on the face or focus on objects in the video (facial vs. non-facial elaboration), or given instructions to count the number of “the” words spoken in the video or not (distraction). The stimulus variables affected the actual videos where the number of bystanders, exposure time, violence threatened, weapon visibility, use of disguises, and crime location were varied. Storage variables were the retention interval and the participants’ exposure to mugshots before the identification task. Finally, retrieval variables were the lineup type, lineup instructions, and reasons given for confidence. They used a fractional factorial design, where not all variables were fully crossed, but they still ended with 64 versions of video tapes. They had no less than two participants per cell, but that corresponded to at least 32 participants per cell for each two-way interaction. Overall, they had 165 participants that participated. They found that disguises, weapon presence, non-facial elaboration, and biased lineup instructions impaired performance. Interestingly, they found that longer retention intervals led to improved lineup performance. The main issues with this study, and their others like it, are the incomplete design, low sample size, and focus on erroneous variables. Another issue was how they calculated confidence using correlation rather than the updated CAC and ROC methods used today. Overall, this is still a very impactful study that has added to the literature, especially the literature emphasizing the importance of estimator variables.

A major issue of current research that while research focuses on manipulating one or two estimator variables, real-life eyewitness cases often have multiple estimator variables that are deficient or suboptimal in multiple ways. This poses a problem in that recommendations to the criminal justice system are based on research conducted that do not accurately reflect real-life scenarios. Like Cutler et al. (1987), more studies need to effectively review multiple variables, especially using the newer analysis to incorporate confidence.

Archival Research

Using the National Registry of Exonerations, we have been collecting data on the frequency of different estimator variables of each case that had a mistaken witness identification (filtered “MWID” in the database) and how often multiple estimator variables occurred. We have coded data ~850 exoneration cases for weapon presence, weapon type, length of exposure, retention interval, distance, lighting, cross-race, cross-gender, and age. Additionally, we have broken down information per each witness in each case and obtained information about system variables, including type of identification task, how many lineups were shown per witness, etc. Data coding was conducted by two independent coders with a third coder to break ties.

Preliminary data were analyzed after only 341 cases were concluded (Giacona et al., 2020). While most cases had missing data, we analyzed as much information as was provided. Cases typically involved strangers (n = 328; 79.3%), a weapon present (n = 340; 65.3%), a moderate exposure time of (n = 273; 57.1%), poor lighting (n = 197; 39.6%), very close distance (*n* = 236; 78.4%), a cross-race effect (*n* = 83; 62.7%), and a cross-gender effect (*n* = 331; 48.0%). The most common retention interval was within a week (*n* = 202; 24.8%); however, the data was pretty evenly spread across conditions. Additionally, we coded for what we deemed "poor" estimator variables (short durations, long retention intervals, poor lighting, far distance, cross-race, cross-gender, stranger, and weapon present). We found that only five cases had no poor estimator variable conditions (*n* = 341, 1.5%), and on average, cases had 3.44 poor estimator variables. Although we have yet to finish coding, preliminary data has shown that the typical exoneration case had multiple, poor estimator variables.

While previous research has dismissed viewing conditions, our research shows that multiple estimator variables impact mistaken eyewitness identification cases. While these results are preliminary and we cannot imply causation, it may be that multiple, poor estimator variables may increase the likelihood of mistaken eyewitness identifications.

Base Rates and Field Research

Another important consideration for eyewitness identification is the prevalence of guilt in lineups. This is called a base rate. Often times, individuals forget to use base rates in their daily decision-making, and this can lead to base rate neglect. In a famous study by Casscells et al. (1978 as cited by Sloman et al., 2003), researchers asked the question "If a test to detect a disease whose prevalence is 1/1000 and has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?" to 69 Harvard Medical School students and staff. Most people responded with 95%, whereas the actual answer would be approximately 2%. This just illustrates how difficult it can be for people to consider base rates in their daily life.

Here I provide an example of a base rate to demonstrate the process. One example is using real world information about mammograms from the American Cancer Society (2017). Suppose a woman in her 40s receives a positive mammogram result. The base rate is about 12.5% prevalence, and mammograms are about 87% accurate. What is the probability that she will have a breast tumor? A lot of people will respond with higher probability than the actual result, which is 51.1% likelihood that her positive result is actually positive. Furthermore, if we lowered the base rate down to about 1% prevalence for the exact same scenario, the likelihood that she actually would have a tumor would decrease down to only 6.3%. Most people get this wrong because they are focused on the overall percentage of it being 87% accurate and less on the prevalence rate. A key takeaway, however, is that when base rates are low, false positives outnumber true positives.

As shown, this can be very difficult for people to incorporate into their daily decision-making. Even more so, this has further implications for eyewitness identification. An easier way to think about it for eyewitness studies is thinking of it as a percentage. For lineups in particular, base rates would be the likelihood that the lineup shown actually has a guilty suspect in it. In most laboratory studies, researchers use a base rate of 50% percent for guilt. What this means is that half of all lineups shown are known to be guilty, and the other half of all lineups are known to be innocent. But what about the real world? This has more variability in it from case to case, detective to detective, and police station to police station. The base rate of guilt can be determined by many factors. One way is to see what threshold of corroborating evidence is needed before putting a suspect in a lineup. Here are two hypothetical differences for high base rates and low base rates. In Case A, suppose there was a suspect that matched an eyewitness description; police had also found fingerprints that matched the suspect's at the crime scene, and there was CCTV footage of the suspect committing the crime. In Case B, there was a suspect that matched an eyewitness description, but they had no other evidence. When eyewitnesses in Case A and Case B are presented the lineups, Case A will have a higher base rate of guilt as there is more corroborating evidence, whereas Case B will have a much lower base rate of guilt. Currently, in the United States, putting a suspect in a lineup is at the discretion of the lead investigator. Although many researchers made a unified recommendation that there needs to be a higher threshold of guilt before placing suspects in lineups (Wells et al., 2020), there have been no changes.

Wixted et al. (2016) conducted a field study with the Houston Police Department (HPD) robbery division in order to attempt to find a potential base rate of guilt. The study recruited 45 HPD police investigators that administered 717 photo lineups, where the witness had not seen a lineup with the suspect and the suspect and witness were strangers. From the lineups that fit their criteria, they were able to produce a model that estimated that the HPD had a 35% base rate of guilt. While this is only one field study from one department, it still shows a low base rate. This means that lineup identifications ultimately do not typically contain a guilty person, and the majority of lineups instead contain an innocent suspect.

Further, a recent study by Katzman and Kovera (2022) attempted to teach base rates to police officers in an effort increase the base rates of guilt in the officers' lineups. They had 279 officers read a case scenario, pretending to be the lead investigator. Half of the participants received education on base rates and how they affect lineup accuracy before reading the cases. However, they did not find that the education affected the officers' judgments on placing suspects in a lineup, and most officers were willing to place suspects in a lineup regardless of the strength of the evidence. They concluded that these officers had base rate neglect regardless of the education received. This is particularly concerning given that Wixted et al.'s (2016) data showed in the real world it is suspected that the base rate may be as low as 35% for lineups.

Pristine Conditions and Multiple Estimator Variables

There are numerous reasons to think that pristine conditions hypothesis would not hold up in cases where estimator variables are deficient in multiple ways. The first reason is that if multiple estimator variables are deficient, the overall memory strength would be very weak. Once memory strength has decayed so much, it would not matter how pristine the conditions are, memory would be too weak to make an identification. At some point, memory would be so weak that highly confident recognition judgments would no longer be indicative of high accuracy (Stretch & Wixted, 1998).

As previously mentioned, a paper by Grabman et al. (2019) equated *d′* for a group of good face recognizers and poor face recognizers, by making the encoding conditions better for the poor face recognizers and worse for the good face recognizers. Even after equating for *d′,* they found that high confidence accuracy differed between the two conditions. This showed an important limitation to the bounds of the high confidence- high accuracy relationship. It also showed that low memory strength cannot be the only factor affecting high confidence accuracy.

Another reason references the SDT model. Multiple estimator variables would require a witness to integrate a large amount of information in making a decision some of which might make people adopt a more stringent criterion, some a less stringent criterion. How would people appropriately adjust their confidence if there are multiple, competing variables? Research has not been conducted that demonstrates how multiple estimator variables would affect criterion shifts. The results could be additive in that the more negative variables, the more they adjust their criterion. They could be sub-additive, in that if there is one negative variable, they will adjust their criterion; however, if there are any additional variables, this would not cause any further criterion shift. Or, they could be super-additive in that multiple, deficient estimator variables would cause an over-adjustment from the eyewitness. Additionally, it is unknown what effect, if any, the availability heuristic would have. The availability heuristic just refers to the idea that people evaluate the probability of events by how easy related instances are evoked (Tversky & Kahneman, 1973). It is plausible that that how heavily the witness weights the positive or negative variables may be influenced by what they remember or think of first, or could be influenced by any prompting questions from authority members such as police, prosecutors, and defense attorneys.

All of this means that it is important to begin doing research with largeish numbers of estimator variables manipulated. Indeed, the National Academy of Sciences (2014) recommended that more needs to be understood about the additive and interactive effects of large numbers of estimator variables:

> In view of the complexity of the effects of both system and estimator variables and their interactions on eyewitness identification accuracy, better experimental designs that incorporate selected combinations of these variables (e.g., presence or absence of a weapon, lighting conditions, etc.) will elucidate those variables with meaningful influence on eyewitness performance, which can, in turn, inform law enforcement practice of eyewitness identification procedures. (p. 105)

Given that many cases involve multiple deficient estimator variables, researchers should be cautious in claiming that high confidence is always associated with high accuracy, since these claims are based on situations where only one or a small number of variables are manipulated.

The Current Study

In this study, I investigated how eyewitnesses would adapt meta-cognitively between their expectations of what would happen when they need to keep track of multiple estimator variables. In practice, SDT states that people are able to shift their criterion if they know the correct way to adjust it. For instance, if the lighting is better, people should know that they can be more liberal with their criterion, and if the lighting is worse, then they should know that they would need to be more conservative with it. However, what happens when there are multiple, suboptimal estimator variables, and some of these estimator variables are competing intuitively? Inherently, there are always multiple estimator variables; lighting, distance, stress, etc. will always be factors. Typically though, research has focused on showing multiple, optimal estimator variables and only one or two suboptimal estimator variables. What if it is a really close distance, but it is really dark? Would those different estimator variables cancel each other out? If there are multiple, negative estimator variables, is there an additive effect? Do people adjust their confidence more appropriately or less appropriately? I believe that the dual processing theory would constitute multiple estimator variables as a System 2 issue and would therefore result in more incorrect answers or needing more effort to appropriately answer. However, there are proponents of signal detection theory and pristine conditions hypothesis that believe that as long as system variables are pristine, high confidence would still equal high accuracy (Wixted & Wells, 2017).

In the current study, participants were shown 12 images of different individuals with a combination of multiple estimator variables manipulated to be purposefully suboptimal or optimal. Participants were asked to complete a lineup, including a contemporaneous statement of confidence, before being asked to complete demographic information and filling out questionnaires. This research sought to answer how multiple estimator variables affect the confidence-accuracy relationship when there are multiple, conflicting variables.

Hypotheses

There have been contradictory outcomes when manipulating multiple estimator variables and/or looking at memory degradation (Giacona et al., 2021; Grabman et al., 2019; Lockamyeir et al, 2020; Semmler et al., 2018). I proposed different, possible outcomes:

1. Pristine Conditions- If the pristine conditions hypothesis holds, then regardless of how many estimator variables are manipulated, then high confidence should be equal to high accuracy, even under low base rates (Semmler et al., 2018; Wixted & Wells, 2017).
2. Weak Memory Strength- It has been proposed that at a certain point, when viewing conditions are too bad, memory strength will be too decayed to make an ID regardless of confidence (Grabman et al., 2019; Lockamyeir et al., 2020; Stretch & Wixted, 1998).
3. Additive- The results could be additive in that the more negative variables, the more they appropriately adjust their criterion.
4. Super-Additive- The presence of multiple, deficient estimator variables would cause an over-adjustment from the eyewitness.
5. Sub-Additive- If there is one negative variable, they will adjust their criterion; however, if there are any additional variables, this would not cause any further criterion shift.

Method

Design

The design of this project conformed to a 2(Lineup: culprit absent, culprit present) x 2(Distance: short, long) x 2(Lighting: bright, dark) x 2(Exposure Duration: long, short) x 2(Retention Interval: short, long) x 2(Race: same race, other race) mixed factorial design.

Lineup and Race were manipulated within participants, and all other variables were manipulated between participants. Dependent variables included proportion of suspect identifications, filler identification, and lineup rejections as well as confidence judgments.

Because this design is very large (64 cells), and because the focus of my research question is on whether the overall quantity of suboptimal estimator variables matters, not so much what those estimator variables are, my analytic plan for analyzing the confidence data involved combining cells in the design that have the same number of suboptimal estimator variables. As in past research, confidence data was analyzed with CAC curves. Each CAC curve combined data from target present and target absent lineups. The central analyses involved comparing the high confidence points on six, different CAC curves created by combining cells in which the same number of estimator variables were suboptimal. For instance, the CAC curve for the *one variable condition* combined data from all the cells where one estimator variable was suboptimal but all the other estimator variables were optimal (e.g., the lighting is dark, but the distance is close, the duration is long, the retention interval is short, the identification is same race). Follow-up exploratory analyses were also performed on the impact of specific variables.

Participants

Overall, 1,066 participants were recruited through Connect by CloudResearch (connect.cloudresearch.com). Data from 1,047 participants were used for analysis in this study. Participants were excluded if they missed both attention checks (n = 1) or were unable to finish the study (n = 6). The completed lineups from participants who completed some, but not all, of the lineups were still included in the study so long as they completed the attention checks (n = 4). Additionally, participants were excluded if they reported having technical difficulties (n = 12).

Full demographic information can be found in Table 2, but my sample size was predominantly White with an average age of 39 years.

Table 1
Demographics

Demographic	***N***	**%**
Race/Ethnicity		
White	661	63
Black or African American	136	13
Hispanic or Latino/a/x	65	6
Asian or Asian American	102	10
Native American or Alaskan Native	5	<1
Pacific Islander or Hawaiian Native	1	<1
Middle Eastern, Arab, or North African	3	<1
Selected more than one	73	7
Prefer not to answer	1	<1
Gender		
Female	504	48
Male	521	50
Non-binary	17	<2
Gender fluid	3	<1
Prefer not to answer	2	<1
	M (SD)	**Range**
Age	39.2 (12.7)	18-85

Sample Size Justification

I established my sample size using the sample sizes from Giacona et al. (2021). Giacona et al. compared a condition in which estimator variables were optimal in multiple ways with a condition in which estimator variables were suboptimal in multiple ways. Because the number of high confidence identifications is likely to be lower when viewing conditions are suboptimal, they over-sampled the condition in which viewing conditions were suboptimal compared to the conditions where viewing conditions were optimal at a ratio of 10 to 1. Total sample size needed to detect a difference in high confidence accuracy of 5% was then established through bootstrapping.

Table 2

How Data Are Aggregated

# of Variables	Lighting	Distance	Retention	Exposure	Race
Zero	+	+	+	+	+
One	-	+	+	+	+
	+	-	+	+	+
	+	+	-	+	+
	+	+	+	-	+
	+	+	+	+	-
Two	-	-	+	+	+
	-	+	-	+	+
	-	+	+	-	+
	-	+	+	+	-
	+	-	-	+	+
	+	-	+	-	+
	+	-	+	+	-
	+	+	-	-	+
	+	+	-	+	-
	+	+	+	-	-
Three	-	-	-	+	+
	-	-	+	-	+
	-	-	+	+	-
	-	+	-	-	+
	-	+	+	-	-
	-	+	-	+	-
	+	-	-	-	+
	+	-	-	+	-
	+	+	-	-	-
	+	-	+	-	-
Four	+	-	-	-	-
	-	+	-	-	-
	-	-	+	-	-
	-	-	-	+	-
	-	-	-	-	+
Five	-	-	-	-	-

Note. Positive signs (+) indicate an optimal variable, whereas negative signs (-) indicate a suboptimal variable.

In the present research, my primary analysis will compare conditions as a function of how many estimator variables are suboptimal. I will be aggregating data into different conditions based on how many suboptimal estimator variables are present at a time, and Table 2 shows every combination of each variable and how it will be categorized. The control condition, known as the “Zero Estimator” condition, will have zero suboptimal estimator variables present, whereas the “Five Estimator” condition will have all five suboptimal estimators present. As race is manipulated within participants, only the four between-subjects variables will be used to calculate sample size. I am going to use the sample size from Giacona et al.’s multiple poor viewing condition for the “Four Variable” condition, and the sample size estimate from Giacona et al.’s multiple good viewing condition as the “Zero Variable” condition. Doing so would suggest that I need 2,000 observations for the Four Variable condition and 200 for the Zero Variable condition. This means if I estimated for the one, two, and three, I would need approximately 650, 1,100, and 1,550 participants respectively. I also estimated a 15% error rate to ensure I met my sample size goal. My entire justification can be found in Table 3.

Table 3

How Data Are Categorized for Sample Size and Analysis

Variables	Estimated *N*	*N* + 15% (Error)	/6 IDs per Participant
Zero	200	230	39
One	650	747.5	125
Two	1,100	1,265	211
Three	1,550	1,782.5	298
Four	2,000	2,300	384
Total			**1,057**

Note. The “variables” column refers to the number of negative estimator variables in each condition. In the last column, any number greater, regardless of decimal point, was rounded up to the next whole number. I do not include the race variable in this breakdown, because race of photograph is a repeated measures variable that does not affect sample size.

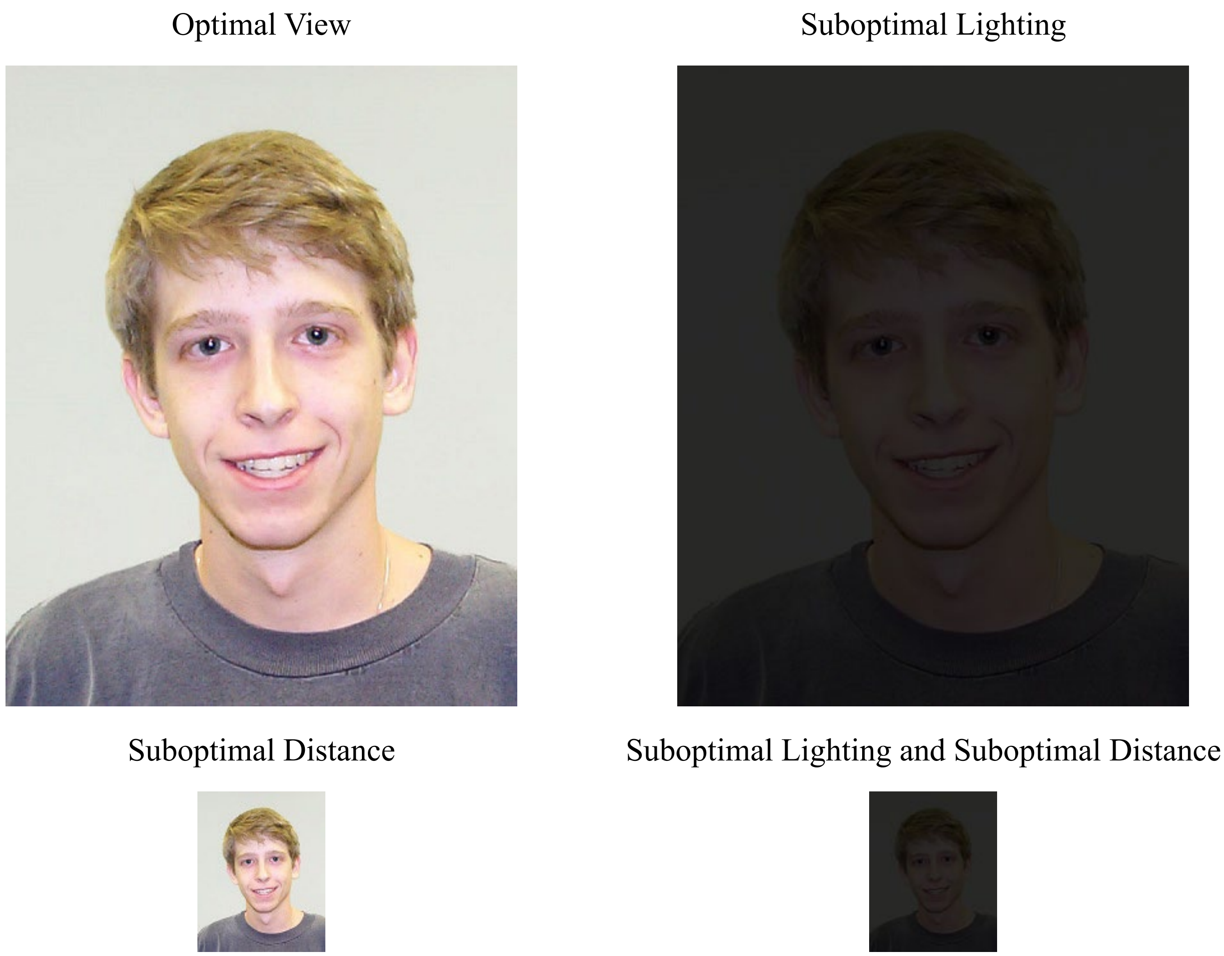
Materials

Materials were taken from Dobolyi and Dodson (2013). These materials were obtained from Dobolyi and Dodson's (2013) open science project (https://osf.io/j25yc/?view) and are used with their permission. At study, participants were shown headshots of 12 men (six Black and six White) who are shown smiling. To manipulate apparent distance, photographs were reduced to 25% of the original size (Lampinen et al., 2015) of 500 pixels by 400 pixels to the reduced size of 125 pixels by 100 pixels. To manipulate apparent lighting, image exposure was adjusted to -5.5 using Photoshop. An example of these photos can be found in Figure 8. Duration was manipulated by showing each picture for 20 seconds or 2 seconds. Retention interval was manipulated by showing the lineups either 90 seconds after the study phase or 12 minutes after the study phase. Additionally, two photos were presented before and after the target photos to act as buffers to the primacy and recency affects. They mirrored whichever condition they were in.

For each man participants study, there were two lineups (24 lineups altogether). The target present lineups will include alternate pictures of the men shown in the study phase, in which the man has a neutral expression, along with five description matched fillers with neutral expressions. The target absent lineup were created by taking the target present lineup, and replacing the target's face with one additional filler. In each lineup type, placement of the photographs was randomized for each participant. Dobolyi and Dodson (2013) normed these materials and showed that they were fair lineups as measured by an average Tredoux's *E* of 4.66 (95% CI = 4.26, 5.07), which is considered fair (p. 347). Each participant saw twelve lineups in a random order. For each race, half of the lineups were in the target present format and half in the target absent format. Assignment of lineups to target present or target absent conditions was counterbalanced across participants. Each lineup also had a "not present" option in the lower left corner. Following each lineup, participants indicated their confidence on an 11-pt (0-100%) scale, where 0% indicates "no confidence at all" and 100% indicates "complete confidence."

Figure 8

Example Buffer Photo



The filler task consisted of viewing either one or eight DRM lists (Deese, 1959; Roediger & McDermott, 1995). The specific lists used were *fruit, high, window, slow, chair, music, bread,* and *sweet*. The lists were comprised of 15 words each, appearing for two seconds each, and then there was a minute allocated for recall of the words. Each list lasted 90s, so the filler task was last either 90s or 720s (12 minutes); the long filler task was only utilized in those conditions that needed a long retention interval.

There were two attention checks asked during the survey. One asked "*What color shirt were the people in the lineups wearing?*" and the other asked "*To ensure that you are paying attention, please select 'red' to the following question*" followed by the question: "*What color is the sky?*" Demographic data was collected concerning participant age, gender, and race/ethnicity.

Procedure

The study was approved as exempt by University of Arkansas' IRB (Protocol # 2208419622). Participants completed an informed consent. Participants were randomly assigned to one of the 16 different between subject conditions created by crossing Duration, Lighting, Distance, and Retention Interval. They were shown 12 photographs (half Black targets, half White targets) in a manner matching the condition they were assigned to. The photographs showed the target individuals smiling. Photographs were presented in a random order. Participants then engaged in the DRM task for the length of time matching the condition they were assigned to. After the filler task, participants were shown twelve lineups in a random order, one corresponding to each target individual. Prior to each lineup, participants received fair pre-lineup instructions (i.e., "Please read the following instructions: 1. You are about to see some photographs. 2. If you see one of the people from the photos you saw before, please select their picture -- if you don't recognize anybody, please select the 'not present' option. 3. The person whose photo you saw may or may not be present in the photographs. 4. It is as important to clear innocent people as it is to identify guilty people. 5. Regardless of your decision, the police will continue to investigate the crime. 6. After your decision you will be asked to indicate your confidence."). Half of the lineups for each race were target present lineups, and half were target absent lineups. Following each lineup, participants were asked to indicate their confidence. Participants were then asked to complete the attention check questions and demographics. Participants were then debriefed.

Results

This analysis plan was preregistered on OSF (https://doi.org/10.17605/OSF.IO/9H3AV). This project will result in two main types of data: lineup and confidence.

Lineup Data

Values of *d'* and Criterion *C* were calculated for each of the 32 cells of the design and can be found in Table 4 (Mickes et al., 2014). Bootstrapping was used to establish the standard error of these signal detection measures (Provenzano et al., 2023). The correlation between the number of estimator variables that are suboptimal and the values of *d'* and *C* were calculated using the Pearson product moment correlation. Higher scores of *d'* indicate better discriminability of participants in that they are able to recognize previously seen faces from novel foils. The strong, negative relationship, $r = -0.65$, $p < .001$, between number of suboptimal estimator variables and *d'* suggests that participants have worse discriminability as the number of suboptimal estimator variables increases. Higher values of C indicate a bias of participants to respond, “Not Present,” whereas lower values of *C* indicate a bias towards choosing a face (filler). There was a moderate, positive relationship between *C* and the number of variables present, $r = 0.49$, $p = .003$, such that increases in the amount of poor estimator variables present indicated a tendency of participants to respond, "Not Present."

Table 4

d' and C for Each Cell

Condition	# Variables	*d'*	*C*
0	0	1.04	0.85
1D	1	1.04	1.18
1Ed	1	1.19	1.00
1L	1	0.75	1.16
1R	1	1.03	1.02
1Ri	1	0.89	1.12
2DEd	2	0.86	1.11
2DR	2	1.07	0.98
2DRi	2	1.16	0.91
2EdR	2	1.08	1.05
2LD	2	0.35	1.31
2LEd	2	0.28	1.40
2LR	2	0.62	1.08
2LRi	2	0.69	1.14
2RiEd	2	0.63	1.41
2RiR	2	1.01	1.06
3DEdR	3	0.64	1.13
3DRiEd	3	0.80	1.03
3DRiR	3	0.63	1.07
3LDEd	3	0.27	1.21
3LDR	3	0.35	1.24
3LDRi	3	0.45	1.20
3LEdR	3	0.37	1.22
3LRiEd	3	0.74	1.13
3LRiR	3	0.48	1.17
3RiEdR	3	0.87	1.19
4DRiEdR	4	0.80	1.13
4LDEdR	4	0.19	1.25
4LDRiEd	4	0.31	1.28
4LDRiR	4	0.50	1.21
4LRiEdR	4	0.44	1.24
5	5	0.34	1.34

Note. The naming conventions are based on which variables are poor (or suboptimal). Ri, Ed, D, L, and R stand for "Retention Interval," "Exposure Duration," "Distance," "Lighting," and "Race."

Analyzing Confidence

I analyzed confidence using CAC analyses (Mickes, 2015), focusing particular attention on the accuracy of confidence in the 90-100% confidence bin. CAC analyses involve plotting the accuracy of suspect identifications as a function of confidence across multiple confidence bins (e.g., 90-100% confident, 70-80% confident, 50-60% confident, etc.). In a CAC analysis, accuracy is defined as follows:

$$Accuracy = \frac{BR * CSID}{BR * CSID + (1 - BR) * MSID}$$

In the equation, CSID denotes the proportion of correct suspect identifications from target present lineups. MSID denotes the proportion of mistaken suspect identifications from target absent lineups. Because target absent lineups in experiments do not necessarily specify any particular person as the "suspect," MSID rate is typically estimated by taking the overall filler identification rate from target absent lineups and dividing that value by six (Clark et al., 2008). BR denotes the base rate; it indicates what proportion of lineups are target present lineups. In an experiment, the base rate is typically 50%, but in actual criminal investigations, the base rate may not be known. However, one researcher estimated that approximately 35% of real world lineups include a guilty suspect (Wixted et al., 2016). For that reason, I created CAC curves under three assumptions: 50% base rate, 35% base rate, and 65% base rate (Giacona et al., 2021).

Because of the large number of cells in my design, analyzing the data will require aggregating the data. I describe the two approaches I used for doing so below.

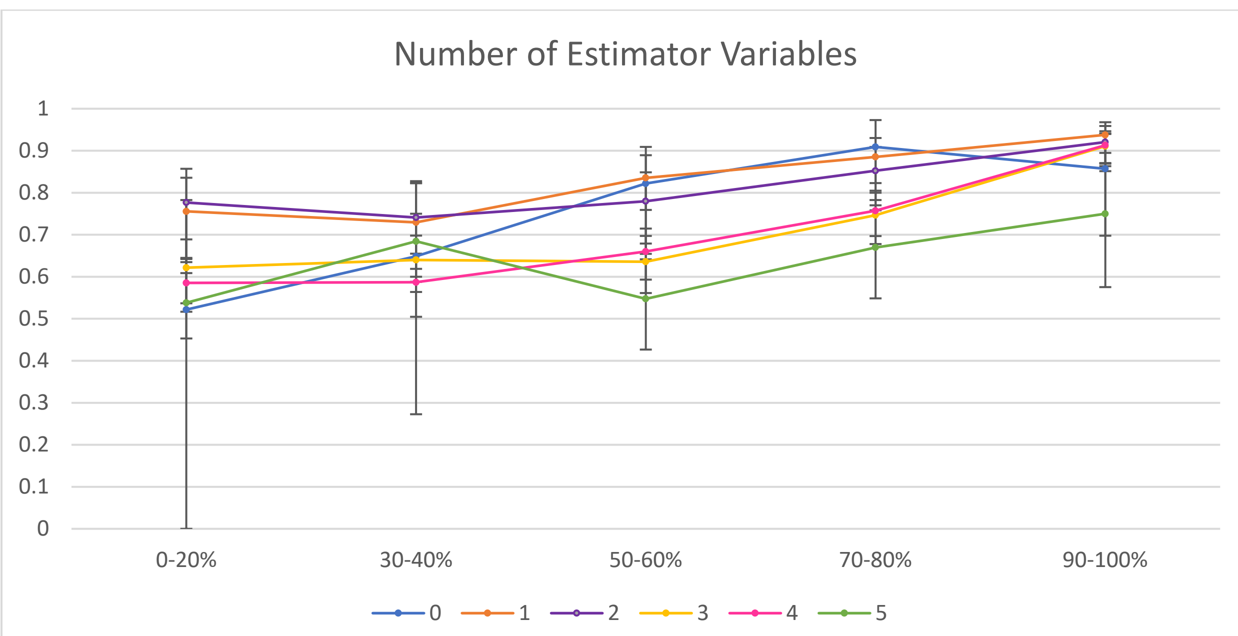
Aggregating Across Number of Suboptimal Estimator Variables

In the first approach, I created six CAC curves by combining data from conditions that have the same number of suboptimal estimator variables (see Table 4 for a complete list). For instance, the CAC curve for the “One Estimator” condition included the following five cells from my experimental design:

1. Illumination is dim, distance is close, duration is long, retention interval is short, target and participant are the same race.
2. Illumination is bright, distance is far, duration is long, retention interval is short, target and participant are the same race.
3. Illumination is bright, distance is close, duration is brief, retention interval is short, target and participant are the same race.
4. Illumination is bright, distance is close, duration is long, retention interval is long, target and participant are the same race.
5. Illumination is bright, distance is close, duration is long, retention interval is short, target and participant are different races.

Figure 9

CAC Curves of Data Aggregated by Number of Estimator Variables

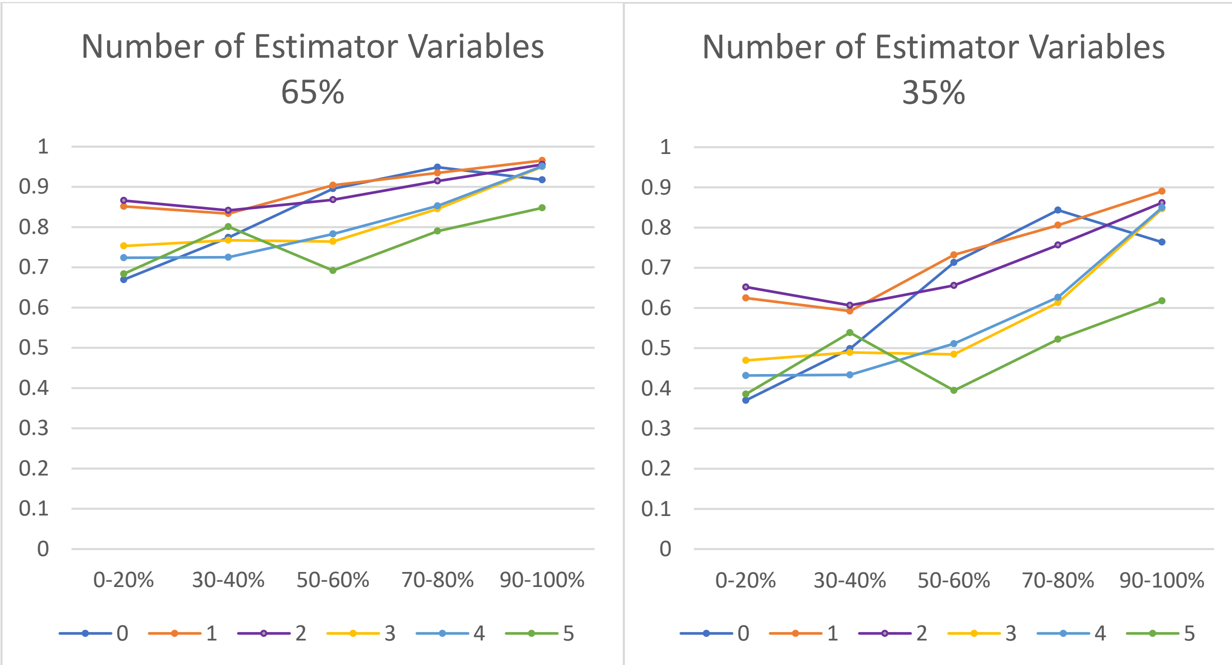


The bootstrapping approach described by Seale-Carlisle and Mickes (2014) was used to generate 95% confidence intervals around each point on the CAC curves. Non-overlapping confidence intervals were taken to indicate significant differences between conditions. Of particular interest for my hypothesis was whether there are non-overlapping confidence intervals for the CAC points in the 90-100% confidence bin.

In general, my findings show that for most conditions, high confidence suspect identifications were highly accurate (> .90). Importantly though, for the condition where all the variables were suboptimal, the accuracy of high confidence suspect identifications was significantly lower (0.75) than the minimum threshold of what is "remarkably accurate" (.90). Additionally, Figure 10 shows the same results but under a low (35%) and high (65%) base rate. Interestingly, even with the high base rates, highly confident suspect identifications were not highly accurate in the condition where all the estimator variables were suboptimal. Furthermore, when base rates are low, no condition showed the pattern of highly confident identifications being 'remarkably accurate,' and only the one negative variable condition came close (0.89).

Figure 10

CAC Curves of Data Aggregated by Number of Estimator Variables for Different Base Rates



Note. Base rates of 65% are a theoretical base rate, and the 35% base rate is based on Wixted et al.'s (2016) study for the proposed real world rate.

The second approach I used to analyze confidence was creating CAC curves comparing the effects of the individual estimator variables. This involved aggregating across all conditions in which a particular estimator variable was suboptimal and comparing it with all cells in which that variable was optimal. For instance, I created a CAC curve that combined all 16 cells in which the lighting was dim and compared that to a CAC curve that combined all 16 cells in which the lighting was bright. This resulted in comparisons between five pairs of CAC curves:

1. Lighting bright vs. Lighting dim
2. Distance close vs. Distance far
3. Duration long vs. Duration brief
4. Retention interval short vs. Retention interval long

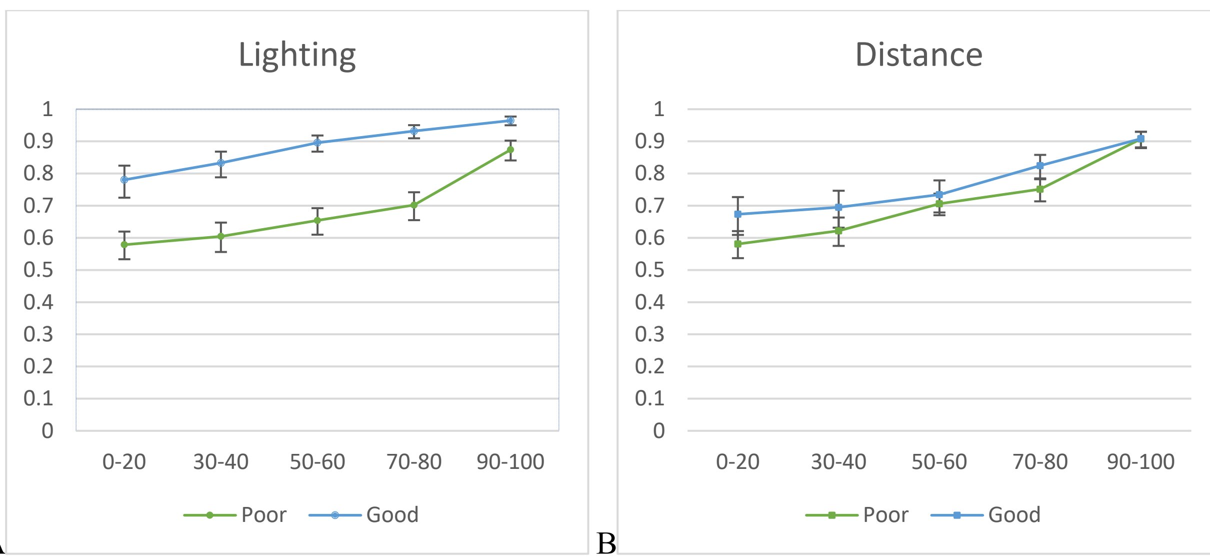
5. Race same vs. Race different

As with the first approach to aggregation that I described, I again used the bootstrapping approach to generate 95% confidence intervals around each point on each the CAC curve to see which variables were non-overlapping, which was used as evidence of a significant difference. Of particular interest for my hypothesis was whether there were non-overlapping confidence intervals for the CAC points in the 90-100% confidence bin.

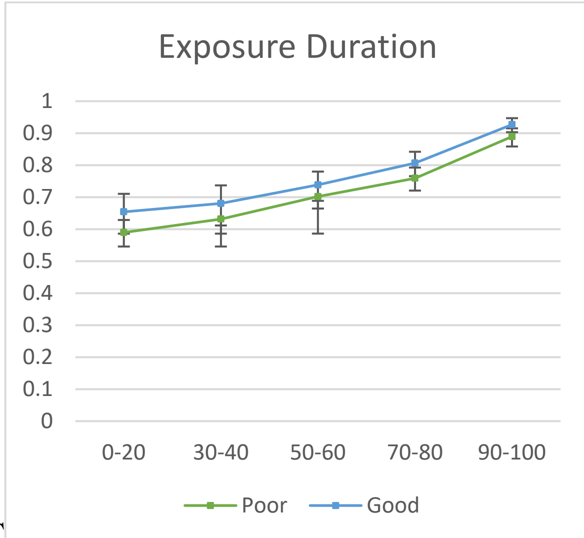
As shown in Figure 11, only one variable, lighting (A), had a significant difference between conditions in that good lighting outperformed bad lighting at every level of confidence. For the high confidence condition, the suboptimal lighting accuracy was still relatively high with 0.87 accuracy with 95%CI [0.90, 0.84]; however, this falls below the commonly accepted threshold of "remarkably accurate."

Figure 11

CAC by Variable Type for Suboptimal and Optimal Conditions

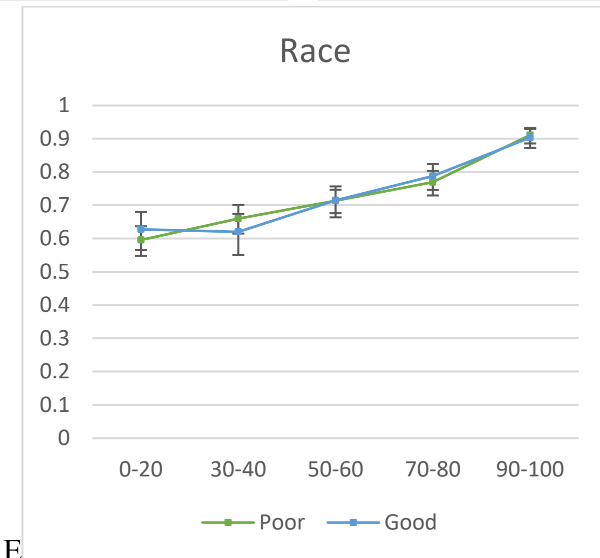


A B



C

D

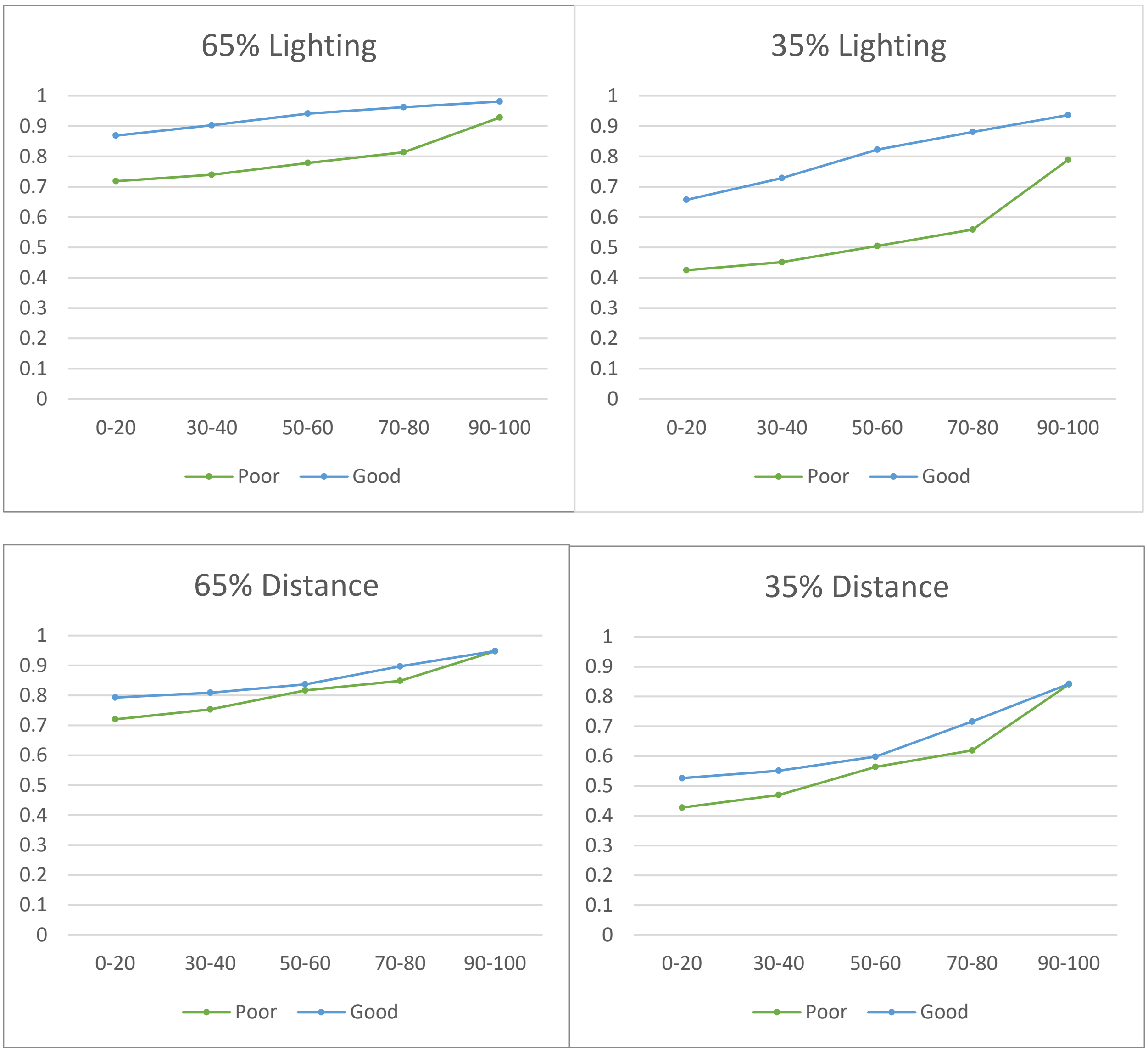


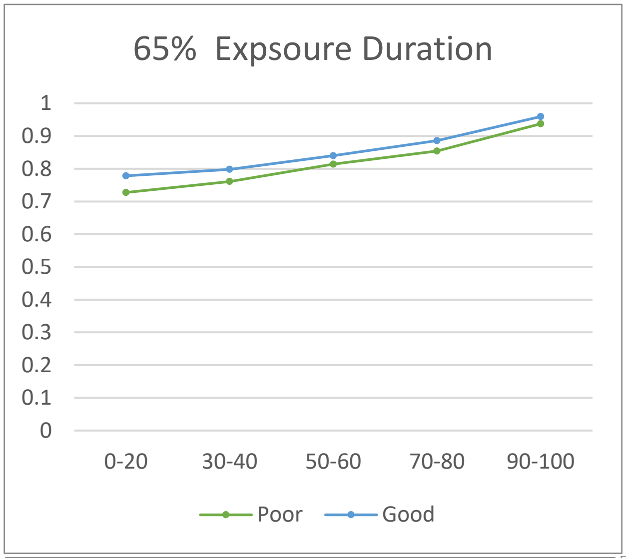
E

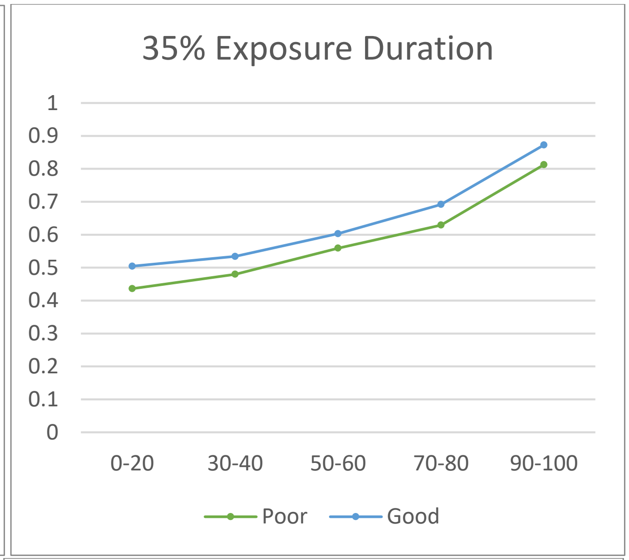
Additionally, I conducted CAC curves under a 65% and 35% base rate. As previously stated, a study conducted showed that when base rates were low (35%), high confidence still equaled high accuracy (Wixted et al., 2016). Figure 12 shows the difference in conditions. For high base rates, all conditions maintained the idea that high confidence equals high accuracy. But for low base rates, overall, lighting was the only condition where high confidence, in the optimal lighting condition specifically, still was an indicator of high accuracy. All other variables resulted in accuracy less than 90%; however, all but two conditions (the suboptimal lighting condition and suboptimal exposure duration condition) maintained accuracy above 80%.

Figure 12

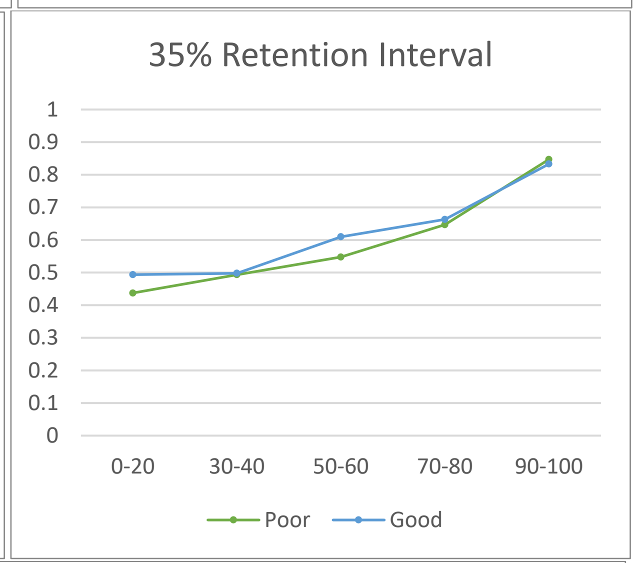
CAC by Variable Type per Base Rate

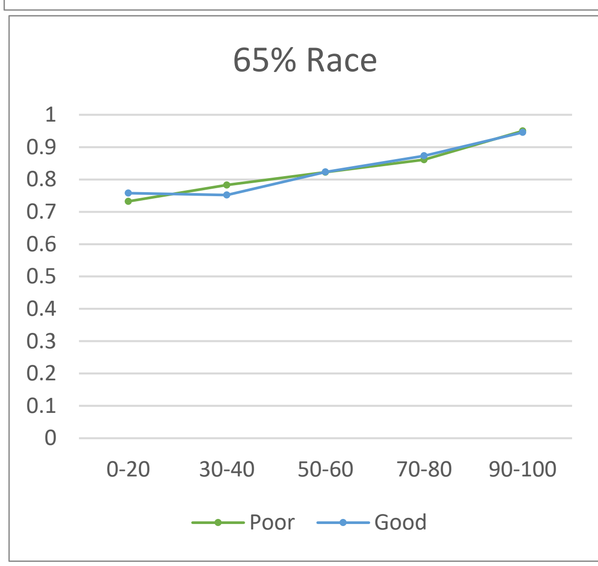


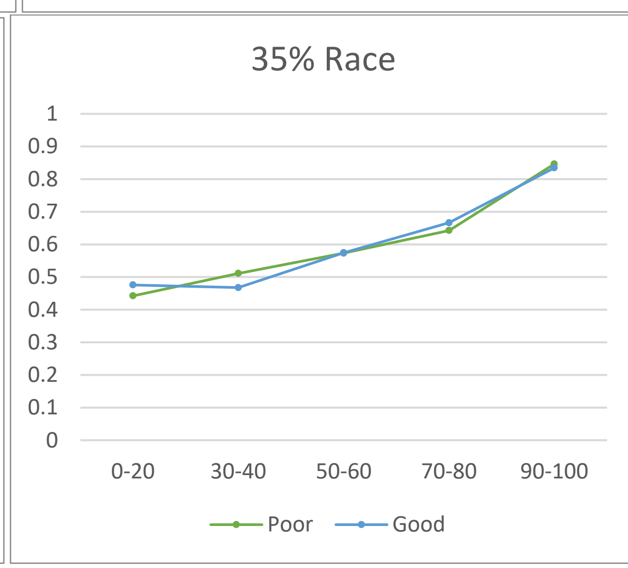

65% Expsoure Duration
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0-20
30-40
50-60
70-80
90-100
Poor
Good


35% Exposure Duration
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0-20
30-40
50-60
70-80
90-100
Poor
Good


65% Retention Interval
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0-20
30-40
50-60
70-80
90-100
Poor
Good


35% Retention Interval
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0-20
30-40
50-60
70-80
90-100
Poor
Good


65% Race
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0-20
30-40
50-60
70-80
90-100
Poor
Good


35% Race
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0-20
30-40
50-60
70-80
90-100
Poor
Good

Regression Analyses

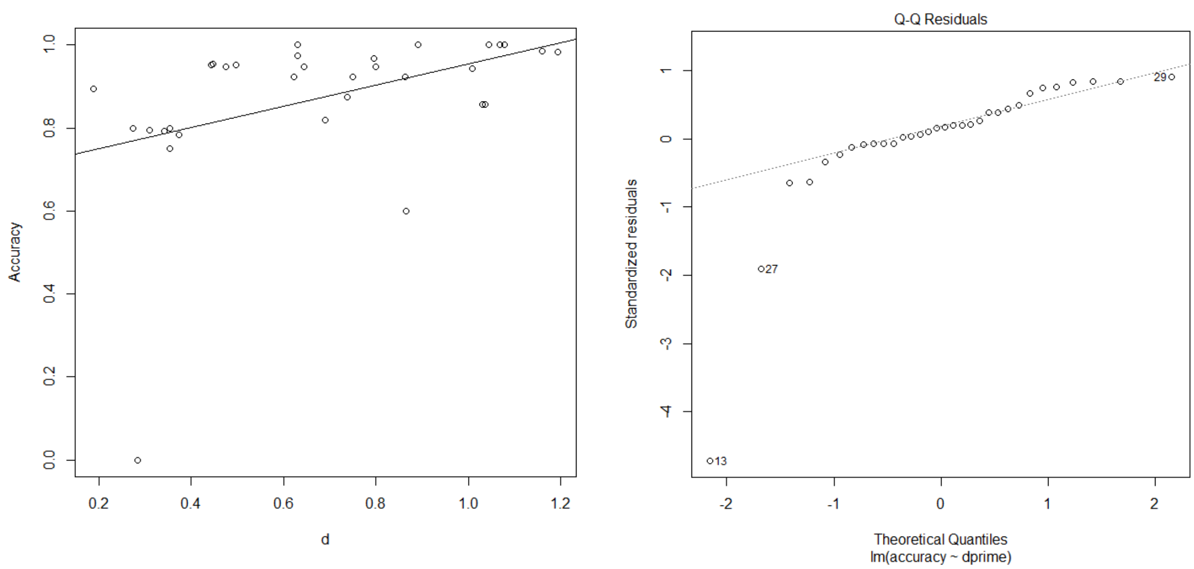
For each of the 32 cells of my design, I have an estimate of memory strength (*d'*) and response criterion (*C*) shown in Table 4. I was also able to generate high confidence accuracy (i.e., accuracy of the 90-100% confidence bin). One issue that will be addressed more thoroughly in the discussion was the issue that when the data were divided by each condition, there were sometimes too few high confidence identifications for stable estimates.

One hypothesis about the confidence-accuracy relationship is that the pristine conditions hypothesis will fall apart if memory becomes sufficiently poor (Giacona et al. 2021; Mickes & Wixted, 2020; Semmler et al., 2018). To examine this possibility, I conducted a regression analysis in which high confidence accuracy in the 32 cells is the dependent variable and *d'* is the predictor variable. Because when memory is good, high confidence accuracy may be very close to ceiling, I also had previously planned the square of *d'* as a predictor. However, after we tested for linearity and normality, it was determined that a simple linear regression could be performed. Figure 13 shows the linearity and normality plots, showing that a simple linear regression is an appropriate analysis to run.

$$\text{High confidence accuracy} = b_0 + b_1 * d'$$

Figure 13

Scatterplot and Q-Q Plot for Accuracy and d'



Note. Scatterplot (left) and Q-Q Plot (right) were used to address linearity. Though there were two significant outliers, this illustrated that overall our data were linear and not curvilinear.

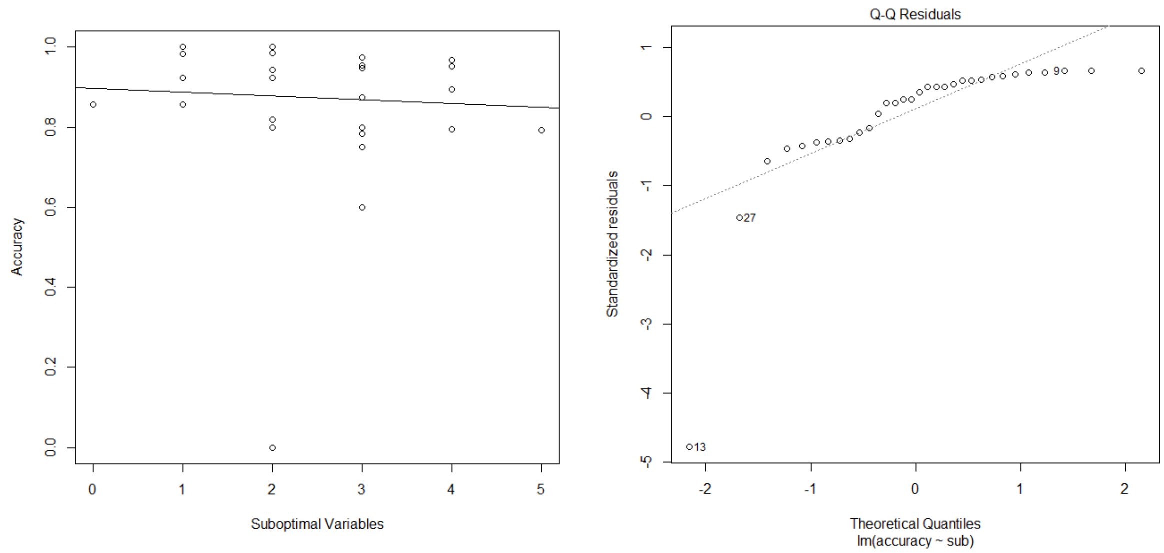
If it is true that high confidence accuracy starts to break down as discriminability gets very poor, then high confidence accuracy should be related to overall memory strength as measured by *d*'. This finding was supported with the model being statistically significant, $F(1, 30) = 6.175$, $p = 0.019$. It was found that *d'* significantly predicted high confidence accuracy, $\beta = 0.41$, $t(30) = 2.458$, $p = 0.019$.

A second hypothesis is that, regardless of discriminability, when the number of suboptimal estimator variables is high, it will make it difficult for witnesses to exercise judgment about how to adjust their high confidence criterion. This would predict that the number of suboptimal estimator variables will be negatively related to high confidence accuracy. A similar analysis was conducted using a number of suboptimal estimator variables as the predictor variable. Similarly, I found that this regression was also linear (see Figure 14).

$$\text{High confidence accuracy} = b_0 + b_1{*}\text{subop}'$$

Figure 14

Scatterplot and Q-Q Plot for Accuracy and Suboptimal Variables



Note. Scatterplot (left) and Q-Q Plot (right) were used to address linearity. Though there were two significant outliers, this illustrated that overall our data were linear and not curvilinear.

This analysis was to determine if the pristine conditions hypothesis holds up even when multiple estimator variables are suboptimal. This is an important applied question, because as I noted in the introduction, it is not uncommon for multiple estimator variables to be suboptimal in actual cases. This regression was also nonsignificant, $F(1, 30) = 0.102$, $p = 0.751$, meaning that the number of suboptimal variables do not predict high confidence accuracy, $\beta = -0.06$, $t(30) = -0.32$, $p = 0.751$, which supports the pristine conditions hypothesis.

Finally, I planned to combine these analyses. The purpose of this analysis was to determine if multiple suboptimal estimator variables reduce high confidence accuracy, even after controlling memory strength.

$$\text{High confidence accuracy} = b_0 + b_1{*}\text{subop}' + b_2{*}d'$$

These results were significant, $F(2, 29) = 4.761$, $p = 0.016$. A closer look confirms our earlier findings that *d'* significantly predicted high confidence accuracy, $\beta = 0.65$, $t(29) = 3.06$, $p = 0.005$, and that the number of suboptimal variables did not significantly predict high accuracy, $\beta = 0.36$, $t(29) = 1.716$, $p = 0.097$.

Discussion

Every year in the United States there are approximately 80,000 criminal cases in which eyewitness identification is crucial to determining the outcome of the trial (Goldstein et al., 1989). The testimony of an eyewitness can provide important information to the jury. However, in some cases, witnesses can be mistaken, leading to the wrongful conviction of an innocent defendant (Wells et al., 2012). Eyewitness identification researchers have tried to address this problem in two ways (Wells, 1978). The first is to develop a set of guidelines for how police should conduct lineups. This is called the system variable approach and has led to the development of science based best practice recommendations published by professional scientific (National Academy of Sciences, 2014; Wells et al., 1998; Wells et al., 2020) and law enforcement organizations (International Association of Chiefs of Police, 2010; Major Cities Chiefs of Police, 2015; Technical Working Group for Eyewitness Evidence, 1999). The second is to identify factors that may influence witness accuracy but that are outside of the control of law enforcement in order to gauge how reliable an identification is likely to be. These are called estimator variables and include factors like lighting (Wagenaar & Vander Schrier, 1996), distance (Lampinen et al, 2014), weapon focus effect (Pickel, 2007), the cross-race effect (Meissner & Brigham, 2001) and so on. There has also been a long term interest in post-dictors of accuracy that, while not causing differences in accuracy, may provide an indication of accuracy (Lampinen et al., 2012). Recently, Smith and Wells (2023) coined the term reflector variable to describe these indicators of reliability. Reflector variables include things like how quickly the witness makes their decision (Sporer, 1993; Wells, 2020) and, of most interest to this project, the expressed certainty at the time of the identification (Brewer & Wells, 2006; Wells, 2020).

A major question of interest in recent years has been the interplay of these three types of variables. One proposal that has gained a great deal of traction in recent years is what I have called the pristine conditions hypothesis (Giacona et al., 2021). The proposal was developed in an influential article by Wixted and Wells (2017). It held that as long as police follow a set of recommended best practices (i.e., system variables), designed to limit the suggestiveness of the identification attempt, highly confident witnesses (i.e., reflector variables) will tend to be highly accurate, even if the witnessing conditions are poor (i.e., estimator variables). From the stand point of signal detection theory, the proposal was that as witnessing conditions get worse, witnesses will take the conditions into account, and adjust their high confidence response criterion by making it stricter – shifting it further towards the right tail of the distributions. This will reduce the number of high confidence suspect identification, but those that do occur will still be highly accurate.

A number of studies have confirmed this general pattern. For instance, Wixted et al. (2016) found that when retention intervals become quite long that the number of high confidence suspect identifications drop, but those that remain are still highly accurate. But there have been some recent studies that have failed to confirm the pristine conditions hypothesis (Giacona et al., 2021; Lockamyeir et al., 2020). Studies that have confirmed the pristine conditions hypothesis have almost always manipulated either a single estimator variable or at most a small number of estimator variables. But in the introduction, I review evidence that in many criminal cases, the estimator variables may be suboptimal in multiple ways (e.g., dark, long distance, wearing a hoodie, high stress, etc.). The question addressed in this dissertation was the degree to which the pristine conditions hypothesis holds up under those circumstances.

I considered two reasons for why multiple estimator variables might make it difficult for witnesses to adjust their high confidence response criterion. The first was that the mental effort required to track multiple variables, and cognitively adjust for them, may be beyond the capacity of most witnesses. How would a witness take into account situations in which some variables are good (i.e., I was close) and other's poor (i.e., but it was dark out and he had a hoodie on)? This would imply that, even holding discriminability constant, the number of estimator variables that are suboptimal should predict the accuracy of high confidence suspect identifications. The second reason why multiple suboptimal estimator variables might interfere with the pristine conditions pattern is simply that the more estimator variables are suboptimal, the worse memory will be. For example, one would expect that if lighting is poor but the distance is close discriminability will be better than if lighting is poor and distance is far. According to this view, multiple suboptimal estimator variables may be expected to decrease high confidence suspect identification accuracy only to the degree that it decreases discriminability overall. Additionally, some researchers (Grabman et al., 2019; Lockamyeir et al., 2020; Stretch & Wixted, 1998) have proposed that at some point, as memory discriminability gets poor enough, high confidence accuracy will decline.

The current study sought to answer how multiple estimator variables affect the confidence-accuracy relationship when there are multiple, conflicting variables. To do so, I systematically manipulated and crossed five estimator variables. Each estimator variable had two levels – optimal or suboptimal. This resulted in a design with a total of 32 unique cells. For each cell of the design, I was able to determine the overall memory discriminability for that cell as measured by *d'*, the number of estimator variables that were suboptimal, and the accuracy of the high confidence suspect identifications. Given the above theoretical discussion, there were three possible outcomes of this experiment. First, it was possible that the pristine conditions hypothesis would hold in its entirety (Semmler et al., 2018; Wixted & Wells, 2017). That is, regardless of the number of estimator variables that were suboptimal, and regardless of the discriminability (*d'*) observed in a particular cell of the design, highly confident suspect identifications will remain highly accurate. Second, it was possible that the pristine conditions hypothesis would hold when only a small number of estimator variables were suboptimal, but that when larger numbers of estimator variables were suboptimal, the pristine conditions hypothesis would no longer hold. That is, when the number of poor estimator variables increase, high confidence suspect identification accuracy would significantly decrease. Importantly, this second hypothesis would hold even after statistically controlling for *d'*. Third, it was possible that the pristine conditions hypothesis would hold when *d'* was high or moderate, but that when *d'* got sufficiently low, high confidence suspect identifications would become significantly less accurate. Importantly, this hypothesis is that it is not so much the number of suboptimal estimator variables that matters, but simply how strong memory is.

First, I addressed the relationship between memory strength and criterion with overall accuracy. I found that when the number of suboptimal estimator variables were increased, memory strength decreased. This was expected that overall memory strength would get worse as viewing conditions are worse. Additionally, I found that as the number of suboptimal estimator variables present increased, participants were more likely to say that the target was "Not Present" rather than choosing someone. This criterion shift shows that individuals are more likely to have a more conservative responding as viewing conditions worsen.

Next, I collapsed all conditions based on how many estimator variables were poor resulting in six possible situations (zero estimator variables are suboptimal to all five estimator variables are suboptimal). I then conducted CAC analyses on the six possible cases. These analyses were conducted to gather information on whether an increase in the number of suboptimal estimator variables would lead to a breakdown of the confidence-accuracy relationship. Given the focus of my research question, the data point of most interest was the high confidence range (90-100% confident). The key findings are shown in Figure 9. Most of the CAC curves show the pristine conditions pattern, with highly confident witnesses being highly accurate. The exception to this is the condition in which all five estimator variables were poor. In this condition, high confidence accuracy drops to around 75%, meaning that, assuming a 50% base rate of guilt, the probability of a highly confident suspect identification being incorrect is about 1 in 4.

The likely estimate of an identification is also influenced by the base rate of guilt (Wixted et al., 2016). I analyzed the effect of base rate by creating CAC curves under different base rate assumptions. These results are shown in Figure 10. Recall that some research has suggested that police lineups containing the guilty suspect only about 35% of the time (Wixted et al., 2016). My results showed that pristine conditions hypothesis was not upheld in a single condition when they were weighted to have a 35% base rate. This again suggests that the pristine conditions hypothesis is not the best fit for explaining these data. However, the all negative condition did not reach high accuracy in the high confidence group even when base rates were 65%. This, again, lends support to the weak memory strength hypothesis.

I also collapsed the data into variable type, where all the good and all the poor conditions of one variable (e.g., lighting) are combined for a comparison. Overall, although some conditions were on the lower end of the what constitutes high accuracy, only one condition had a significant difference between the poor and good conditions: lighting. Good lighting was significantly more accurate than poor lighting at every confidence level. However, whenever base rates were high (65%) both the poor and good lighting conditions were highly accurate for the highly confident bin. On the other hand, when base rates were low (35%) only the good lighting condition remained highly accurate for the high confidence bin. Every other condition fell below the highly accurate threshold. This may reflect something unique about lighting or may simply reflect that my manipulation of poor lighting was particularly strong compared to my manipulation of the other estimator variables. This again lacks support for the pristine conditions hypothesis because the high confidence and high accuracy relationship did not remain under assumptions of low base rates, contrary to the hypothesis (Wixted et al., 2016).

I next addressed the question of whether the number of poor estimator variables impacted the confidence accuracy relationship or if it was simply memory discriminability that impacted the confidence accuracy relationship. To determine this, I calculated high confidence accuracy for each of the 32 cells of the design. I also calculated memory discriminability as indexed by *d'* for each of the 32 cells. The analysis also noted the number of estimator variables that were poor. I then conducted three analyses to see what was the best predictor of high confidence accuracy: memory strength or the number of poor estimator variables. The results showed that number of suboptimal estimator variables did not significantly predict high confidence accuracy, but discriminability, as measured by *d',* did. Thus, a clear conclusion of this research is that as witnessing conditions deteriorate, and consequently memory gets weaker, eventually the high confidence-high accuracy relationship begins to break down. These results are illustrated in the scatterplots in Figure 11.

Limitations

As in all research, there are some limitations to the present work. One limitation concerns the balance of number of high confidence suspect identifications in each cell of the design. Research addressing the pristine conditions hypothesis is complicated by the fact that the main question of interest is a conditional probability of accuracy, given that a witness was highly confident. This means that the researcher needs to make an estimate, in each cell of the design, as to what proportion of identifications are going to be made with high confidence. The process I used involved making these estimates based on the aggregation strategy described above, but that necessarily meant that individual cells sometimes had a relatively small number of high confidence judgments. This limitation does not provide any challenges to the validity of the main findings, as its main statistical impact is to add to noise in estimates of accuracy in some cells of the design.

Other limitations to make note of are the participants sampled and the materials used. I used a procedure and materials developed by Dobolyi and Dodson (2013) in which participants view 12 static photographs and then complete 12 lineups. Using photographs over videos for the encoding procedure was convenient, but limits the ecological validity of the findings. Given the very large number of cells in the design, it was decided that using some sort of repeated measure design with existing materials would be most expedient. Additionally, any time data is collected online through a survey site, there is the bias in the data that these individuals are primed for research and are self-selected to complete these surveys. However, the demographics reported in this paper are very similar to what would have been collected at my university.

Conclusion

Overall, these results are extremely important. This study was one of only a few that has conducted research on the confidence-accuracy relationship using multiple estimator variables. Furthermore, it is the first one to systematically manipulate five estimator variables and attempt to see the differences in memory degradation by estimator amount. There were several instances where, like previous research (Giacona et al. 2021; Mickes & Wixted, 2020; Semmler et al., 2018), the pristine conditions hypothesis did not hold up. Though there were still instances where it did, it is important to note possible boundary conditions for this theory. For this study, the hypotheses most supported was the memory strength hypotheses (Grabman et al., 2019; Lockamyeir et al., 2020; Stretch & Wixted, 1998). These analyses add to the growing and often contradictory literature on memory degradation (Giacona et al., 2021; Grabman et al., 2019; Lockamyeir et al, 2020; Semmler et al., 2018). Further research should still be conducted to continue to shape the narrative on this conversation.

References

American Cancer Society. (2017). https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf

Barbur, J. L., & Stockman, A. (2010). Photopic, mesopic and scotopic vision and changes in visual performance. *Encyclopedia of the Eye*, *3*, 323-331.

Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology*, *20*(1), 115-129.

Blackstone, W. (1766). *Commentaries on the laws of England* (Vol. 4). Oxford: Clarendon Pr.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*(1), 11-30. https://doi.org/10.1037/1076-898X.12.1.11

Brigham, J. C., Bennett, L. B., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 257–281). Lawrence Erlbaum Associates Publishers.

Carlson, C. A., Dias, J. L., Weatherford, D. R., & Carlson, M. A. (2017). An investigation of the weapon focus effect and the confidence–accuracy relationship for eyewitness identification. *Journal of Applied Research in Memory and Cognition*, *6*(1), 82-92.

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

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, *7*(3), 238-259.

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior*, *32*(3), 187-218.

Colloff, M. F., & Wixted, J. T. (2020). Why are lineups better than showups? A test of the filler siphoning and enhanced discriminability accounts. *Journal of Experimental Psychology: Applied, 26*(1), 124-143. https://doi.org/10.1037/xap0000218

Colloff, M. F., Wade, K. A., Strange, D., & Wixted, J. T. (2018). Filler-siphoning theory does not predict the effect of lineup fairness on the ability to discriminate innocent from guilty suspects: Reply to Smith, Wells, Smalarz, and Lampinen (2018). *Psychological Science*, *29*(9), 1552-1557.

Curran, T., DeBuse, C., & Leynes, P. A. (2007). Conflict and criterion setting in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 2-17. https://doi.org/10.1037/0278-7393.33.1.2

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification. *Law and Human Behavior*, *11*(3), 233-258.

de Keijser, J. W., De Lange, E. G. M., & Van Wilsem, J. A. (2014). Wrongful convictions and the Blackstone ratio: An empirical analysis of public attitudes. *Punishment & Society, 16*, 32-49.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*(6), 687-706.

Denker, M., & Woyczynski, W. A. (1998). *In introductory statistics and random phenomena: Uncertainty, complexity, and chaotic behavior in engineering and science*. Birkhäuser (Vol. 41). Springer Science+Business Media, LLC. https://doi.org/10.2307/1270584

Dodson, C. S., & Dobolyi, D. G. (2015). Misinterpreting eyewitness expressions of confidence: The featural justification effect. *Law and Human Behavior*, *39*(3), 266-280. https://doi.org/10.1037/lhb0000120

Duncan, M. (2006). *A signal detection model of compound decision tasks* (Technical Note DRDC TR 2006 –256). Ottawa, Ontario, Canada: Defense Research and Development Canada.

Erickson, W. B., Lampinen, J. M., & Leding, J. K. (2014). The weapon focus effect in target present and target absent lineups: The roles of threat, novelty and timing. *Applied Cognitive Psychology 28*(3), 349-359.

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2013). Of guns and geese: A meta-analytic review of the ‘weapon focus’ literature. *Psychology, Crime & Law*, *19*(1), 35-66.

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine Learning, 31*, 1–38

Findley, K. A. (2012). Judicial gatekeeping of suspect evidence: Due process and evidentiary rules in the age of innocence. *Georgia Law Review*, *47*, 723.

Fitzgerald, R. J., Rubínová, E., & Juncu, S. (2021). Eyewitness identification around the world. In *Methods, measures, and theories in eyewitness identification tasks* (pp. 294-322). Routledge.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443.

Giacona, A. M., & Lampinen, J. M. (2022). Diagnostic featural detection or filler siphoning: A red box study. Paper presented at the American Psychology Law Society (AP-LS), Denver, CO, March 17-20, 2022.

Giacona, A. M., Lampinen, J. M., & Anastasi, J. S. (2021). Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law and Human Behavior, 45*(3), 256-270. https://doi.org/10.1037/lhb0000381

Giacona, A. M., Aspenson, E. B., Schuetter, B. S., & Lampinen, J. M. (2020). Estimating estimator variables in exoneration cases. Paper presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021.

Grabman, J. H., Dobolyi, D. G., Berelovich, N. L., & Dodson, C. S. (2019). Predicting high confidence errors in eyewitness memory: The role of face recognition ability, decision-time, and justifications. *Journal of Applied Research in Memory and Cognition, 8*(2), 233-243.

Gronlund, S. D., & Benjamin, A. S. (2018). The new science of eyewitness memory. In *Psychology of Learning and Motivation* (Vol. 69, pp. 241-284). Academic Press.

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *1*(4), 221-228.

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In *Psychology of learning and motivation* (Vol. 63, pp. 1-43). Academic Press.

Goldstein, A. G., Chance, J. E., & Schneller, G. R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Bulletin of the Psychonomic Society*, *27*(1), 71-74.

Innocence Project. (2021). *Marvin Anderson celebrates 17 years of freedom today*. Retrieved from https://innocenceproject.org/cases/marvin-anderson/

Innocence Project. (2022). *DNA exonerations in the United States*. Retrieved from https://innocenceproject.org/dna-exonerations-in-the-united-states/

International Association of Chiefs of Police. (2010). *Showups, photographic identifications, and lineups: Model policy.*

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language, 30*(5), 513-541.

Jones, C. E. (2012). I am Ronald Cotton: Teaching wrongful convictions in a criminal law class. *Ohio St. J. Crim. L.*, *10*, 609.

Kabzińska, J. (2015). Evaluating the probative value of facial composites by laypeople and prosecutor apprentices. *Psychology, Crime & Law, 21*(8), 715-733.

Katzman, J., & Kovera, M. B. (2022). Evidence strength (insufficiently) affects police officers' decisions to place a suspect in a lineup. *Law and Human Behavior, 46*(1), 30–44. https://doi.org/10.1037/lhb0000476

Kovera, M. B., & Evelo, A. J. (2017). The case for double-blind lineup administration. *Psychology, Public Policy, and Law*, *23*(4), 421-437.

Kovera, M. B., & Evelo, A. J. (2020). Improving eyewitness-identification evidence through double-blind lineup administration. *Current Directions in Psychological Science, 29*(6), 563-568.

Kramer, G. M., Wolbransky, M., & Heilbrun, K. (2007). Plea bargaining recommendations by criminal defense attorneys: Evidence strength, potential sentence, and defendant preference. *Behavioral sciences & the law*, *25*(4), 573-585.

Kutateladze, B. L., Lawson, V. Z., & Andiloro, N. R. (2015). Does evidence really matter? An exploratory analysis of the role of evidence in plea bargaining in felony drug cases. *Law and Human Behavior*, *39*(5), 431-442. https://doi.org/10.1037/lhb0000142

Lampinen, J. M., Erickson, W. B., Moore, K. N., & Hittson, A. (2014). Effects of distance on face recognition: Implications for eyewitness identification. *Psychonomic bulletin & review*, *21*(6), 1489-1494. https://doi.org/10.3758/s13423-014-0641-2

Lampinen, J. M., Race, B., Provenzano, A., Gipson, N., & Giacona, A. (2021). Utility approaches and eyewitness identification reforms. In *Methods, Measures, and Theories in Eyewitness Identification Tasks* (pp. 257-293). Routledge.

Lampinen, J. M., Roush, A., Erickson, W. B., Moore, K. N. & Race, B. (2015). The effects of simulated distance on recognition of same race and other race faces. *Visual Cognition*, *23*, 678-698.

Lampinen, J. M., Smith, A. M., & Wells, G. L. (2019). Four utilities in eyewitness identification practice: Dissociations between receiver operating characteristic (ROC) analysis and expected utility analysis. *Law and Human Behavior*, *43*(1), 26-44. https://doi.org/10.1037/lhb0000309

Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior, 43*(5), 436–454. https://doi.org/10.1037/lhb0000343

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556-564. https://doi.org/10.1037/0021-9010.70.3.556

Lindsay, R. C., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations?. *Journal of Applied Psychology*, *66*(1), 79-89. https://doi.org/10.1037/0021-9010.66.1.79

Livingston, J. A. (2003). Metacognition: An Overview. *Psychology, 13*, 259-266.

Loftus, E. F. (1979). The malleability of human memory: Information introduced after we view an incident can transform memory. *American Scientist*, *67*(3), 312-320.

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

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological science*, *25*(9), 1663-1673.

Major Cities Chiefs (2015). Improving policing to ensure accuracy in arrests and convictions. Policy Statement Adopted by the Major Cities Chiefs. https://www.majorcitieschiefs.com/pdf/news/wrongful_convictions_policy_statement.pdf

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

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

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). Lineup construction and lineup fairness. In *The Handbook of Eyewitness Psychology: Volume II* (pp. 169-192). Psychology Press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological review, 87*(3), 252.

Mecklenburg, S. H., Bailey, P. J., & Larson, M. R. (2008). The Illinois Field Study: A significant contribution to understanding real world eyewitness identification issues. *Law and Human Behavior, 32,* 22–27. http://dx.doi.org/10.1007/s10979-007-9108-6

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*(1), 3-35. https://doi.org/10.1037/1076-8971.7.1.3

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *4*(2), 93-102.

Mickes, L., Seale-Carlisle, T. M., Chen, X., & Boogert, S. (2022, March 9). pyWitness 1.0: A Python eyewitness identification analysis toolkit. doi.org/10.31234/osf.io/5ruks

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*(4), 361-376. https://doi.org/10.1037/a0030609

National Academy of Sciences, Committee on Scientific Approaches to Understanding and Maximizing the Validity and Reliability of Eyewitness Identification in Law Enforcement and the Courts. (2014). *Identifying the culprit: Assessing eye-witness identification.* Washington, DC: National Academy of Sciences.

National Registry of Exonerations. (2022). *Summary view.* Retrieved from https://www.law.umich.edu/special/exoneration/Pages/detaillist.aspx

National Registry of Exonerations. (2019). *Ronald Cotton.* https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3124

Nguyen, T. B., Pezdek, K., & Wixted, J. T. (2017). Evidence for a confidence–accuracy relationship in memory for same-and cross-race faces. *Quarterly Journal of Experimental Psychology, 70*(12), 2518-2534.

Nyman, T. J., Antfolk, J., Lampinen, J. M., Tuomisto, M., Kaakinen, J. K., Korkman, J., & Santtila, P. (2019). A stab in the dark: The distance threshold of target identification in low light. *Cogent Psychology*, *6*(1), 1632047.

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, *19*(1), 55-71. https://doi.org/10.1037/a0031602

Pennycook, G. (2017). A perspective on the theoretical foundation of dual-process models. *Dual process theory, 2,* 34-57.

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology*, *84*(6), 940.

Pickel, K. (2007). Remembering and identifying menacing perpetrators: Exposure to violence and the weapon focus effect. In *The Handbook of Eyewitness Psychology: Volume II* (pp. 353-374). Psychology Press.

Pickel, K. L. (1998). Unusualness and threat as possible causes of "weapon focus". *Memory*, *6*(3), 277-295.

Police Executive Research Forum. (2013). A national survey of eyewitness identification procedures in law enforcement agencies. Retrieved from https://perf.memberclicks.net/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies% 202013.pdf

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological review*, *99*(3), 518-535. https://doi.org/10.1037/0033-295X.99.3.518

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1-75.

Rotello, C. M., & Chen, T. (2016). ROC curve analyses of eyewitness identification decisions: An analysis of the recent debate. *Cognitive Research: Principles and Implications*, *1*(1), 1-12.

Rotello, C. M., Heit, E., & Dubé, C. (2015). When more data steer us wrong: Replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review, 22*(4), 944-954.

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*(4), 337-347.

Schmechel, R. S., O'Toole, T. P., Easterly, C., & Loftus, E. F. (2005). Beyond the ken-Testing jurors' understanding of eyewitness reliability evidence. *Jurimetrics*, *46*, 177-214.

Seale-Carlisle, T.M., Quigley-McBride, A., Teitcher, J., Crozier, W.E., Dodson, C.S., & Garrett, B.L. (2022). New insights on expert opinion about eyewitness memory research [Preprint]. https://www.researchgate.net/profile/Travis-Seale-Carlisle/publication/357168008_New_insights_on_expert_opinion_about_eyewitness_memory_research/links/61bf9ea963bbd93242a6a500/New-insights-on-expert-opinion-about-eyewitness-memory-research.pdf

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied*, *24*(3), 400-415. https://doi.org/10.1037/xap0000157

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the US population. *PloS one*, *6*(8), e22757.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational behavior and human decision processes*, *91*(2), 296-309.

Smith, A. M., Lampinen, J. M., Wells, G. L., Smalarz, L., & Mackovichova, S. (2019). Deviation from perfect performance measures the diagnostic utility of eyewitness lineups but partial area under the ROC curve does not. *Journal of Applied Research in Memory and Cognition*, *8*(1), 50-59.

Smith, A. M., Smalarz, L., Wells, G. L., Lampinen, J. M., & Mackovichova, S. (2022). Fair lineups improve outside observers' discriminability, not eyewitnesses' discriminability: Evidence for differential filler-siphoning using empirical data and the WITNESS computer-simulation architecture. *Journal of Applied Research in Memory and*

Cognition. Preprint.

Smith, A. M., & Wells, G. L. (2023). Using reflector variables to determine whether the culprit is present in or absent from a police lineup. In D. DeMatteo & K. C. Scherr (Eds.), *The Oxford Handbook of Psychology and Law* (online edition). Oxford Academic. https://doi.org/10.1017/CBO9780511808098.026

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Myerson, T. (2018). Eyewitness identification performance on showups improves with an additional-opportunities instruction: Evidence for present–absent criteria discrepancy. *Law and Human Behavior*, *42*(3), 215.

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Penrod, S. D. (2016). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior, 41*(2), 127. http://dx.doi.org/10.1037/lhb0000219

Smith, A. M., Wells, G. L., Smalarz, L., & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves the applied value of lineups without improving memory performance: Commentary on Colloff, Wade, and Strange (2016). *Psychological Science*, *29*(9), 1548-1551.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*(1), 22.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, *7*(1), 36.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421–440). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.026

State Bar of Michigan: Eyewitness Identification Task Force. (2012). *Law enforcement and eyewitness identifications: A policy writing guide.* https://www.michigan.gov/-/media/Project/Websites/mcoles/TD-*Webpage/state_bar_eyewitness_law_enforcement_policy_writing_guide.pdf?rev=6e19807bd78447d2b54e6f9cc7ab186b.* [Updated December 22, 2015.]

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior*, *16*(4), 413-424.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior*, *21*(3), 283-297.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, *27*(5), 523-540.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*(1), 99-139.

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1397-1410. https://doi.org/10.1037/0278-7393.24.6.1397

Technical Working Group for Eyewitness Evidence. (1999). Eyewitness evidence: A guide for law enforcement. Washington, DC: U.S. Department of Justice.

Tekin, E., Lin, W., & Roediger, H. L. (2018). The relationship between confidence and accuracy with verbal and verbal+ numeric confidence scales. *Cognitive research: principles and implications*, *3*(1), 1-8.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *Ariel, 137*, 51-43.

Topp-Manriquez, L. D., McQuiston, D., & Malpass, R. S. (2016). Facial composites and the misinformation effect: How composites distort memory. *Legal and criminological psychology*, *21*(2), 372-389.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22*(2), 217–237.

Tredoux, C. G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology, 13*, S9–S26.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232.

Valentine, T., & Heaton, P. (1999). An evaluation of the fairness of police line-ups and video identifications. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *13*(S1), S59-S72.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *17*(8), 969-993.

Vishney, D. (2016). *Suppressing Criminal Evidence*. James Publishing.

Wagenaar, W. A., & Schrier, van der J. H. (1996). Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime & Law, 2,* 2321–2332.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, *36*(12), 1546-1557. https://doi.org/10.1037/0022-3514.36.12.1546

Wells, G. L. (2020). Psychological science on eyewitness identification and its impact on police practices and policies. *American Psychologist*, *75*(9), 1316.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*(3), 360-376. https://doi.org/10.1037/0021-9010.83.3.360

Wells, G. L., & Turtle, J. W. (1987). Eyewitness testimony research: Current knowledge and emergent controversies. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *19*(4), 363-388. https://doi.org/10.1037/h0080000

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior, 44*(1), 3–36. https://doi.org/10.1037/lhb0000359

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

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, *78*(5), 835-844.

Wells G. L., Memon A., Penrod S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7, 45–75.

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

Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory and Cognition*, *4*(4), 324-328.

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2012). Eyewitness identification reforms: Are suggestiveness-induced hits and guesses true hits? *Perspectives on Psychological Science, 7*, 264 –271.

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: an experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39*(1), 1-14. https://doi.org/10.1037/lhb0000096

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior*, *39*(2), 99-122. https://doi.org/10.1037/lhb0000125

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, *2*(4), 343-364. https://doi.org/10.1037/1076-898X.2.4.343

Wise, R. A., Safer, M. A., & Maro, C. M. (2011). What U.S. Law enforcement officers know and believe about eyewitness factors, eyewitness interviews and identification procedures. *Applied Cognitive Psychology, 25*, 488 –500.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*(4), 1025-1054.

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262-276. https://doi.org/10.1037/a0035940

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*(1), 10-65.

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger III, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist*, *70*(6), 515-526. https://doi.org/10.1037/a0039510

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences*, *113*(2), 304-309.

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science*, *13*(3), 324-335.

Wixted, J. T., Read, J. D., & Lindsay, D. S. (2016). The effect of retention interval on the eyewitness identification confidence–accuracy relationship. *Journal of Applied Research in Memory and Cognition*, *5*(2), 192-203.

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive psychology*, *105*, 81-114.

Wooten, A. R., Carlson, C. A., Lockamyeir, R. F., Carlson, M. A., Jones, A. R., Dias, J. L., & Hemby, J. A. (2020). The number of fillers may not matter as long as they all match the description: The effect of simultaneous lineup size on eyewitness identification. *Applied Cognitive Psychology*, *34*(3), 590-604.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law*, *7*(1), 119-133. https://doi.org/10.1037/1076-8971.7.1.119

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied Cognitive Psychology: The*

Official Journal of the Society for Applied Research in Memory and Cognition, *17*(3), 365-373.

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology*, *10*(1), 75-84.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in showups and lineups. *Law and Human Behavior*, *20*(4), 459-477.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310-1321.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*(3), 361-379. https://doi.org/10.1037/0096-3445.130.3.361

UNIVERSITY OF ARKANSAS

To:	Amber M Giacona
From:	Douglas J Adams, Chair IRB Expedited Review
Date:	03/02/2023
Action:	**Exemption Granted**
Action Date:	03/02/2023
Protocol #:	2208419622
Study Title:	Memory for Faces

The above-referenced protocol has been determined to be exempt.

If you wish to make any modifications in the approved protocol that may affect the level of risk to your participants, you must seek approval prior to implementing those changes. All modifications must provide sufficient detail to assess the impact of the change.

If you have any questions or need any assistance from the IRB, please contact the IRB Coordinator at 109 MLKG Building, 5-2208, or irb@uark.edu.

cc: James M Lampinen, Investigator

Amber M. Giacona

216 Memorial Hall, Fayetteville, AR 72701
agiacona@uark.edu | [redacted]

EDUCATION

Summer 2023 — **University of Arkansas**, Fayetteville, AR
Ph.D. in Experimental Psychology, Cognitive concentration
Dissertation Chair: Dr. James M. Lampinen
Dissertation: *How systematically increasing estimator variables affects the confidence-accuracy relationship.*

May 2018 — **Sam Houston State University**, Huntsville, TX
M.A. in General Psychology
Thesis Chair: Dr. Marsha Harman
Thesis: *Parenthood and media: Television influences college students' perceptions of parenthood*

May 2016 — **Sam Houston State University**, Huntsville, TX
B.S. in Psychology, Minor in Sociology

PUBLICATIONS

Gipson, N. I., **Giacona, A. M.,** & Lampinen, J. M. (Under review). Here comes a regular: Mistaken eyewitness identifications of familiar but innocent suspects. Submitted to *Applied Cognitive Psychology*.

Pennekamp, P., **Giacona, A. M.,** Kask, K., & Lampinen, J. M. (Under review). Boundary conditions to encoding specificity: Unmasked lineups lead to higher suspect identification accuracy for surgical masked perpetrators. Submitted to *Journal of Police and Criminal Psychology*.

Giacona, A. M., Lampinen, J. M., & Anastasi, J. S. (2021). Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law and Human Behavior, 45*(3), 256–270. https://doi.org/10.1037/lhb0000381

Book Chapters

Lampinen, J. M., Race, B., Provenzano, A. C., Gipson, N. I., & **Giacona, A. M.** (2021). Utility approaches and eyewitness identification reforms. *In* Toglia, M., Smith, A., & Lampinen, J. M. (Eds.) *Methods, Measures, and Theories in Forensic Facial-Recognition.* Taylor and Francis.

MANUSCRIPTS IN PREPARATION

Giacona, A. M. & Lampinen, J. M. (In prep). Estimating estimator variables: Poor witnessing conditions in exoneration cases.

Giacona, A. M., *Regier, G. O., & Lampinen, J. M. (In prep). Distinctive features.

Lampinen, J. M., Provenzano, A. C., & **Giacona, A. M.** (In prep). Eyewitness memory for distinctive features: Is the absence of evidence, evidence of absence?

CONFERENCE PRESENTATIONS

*Denotes mentored student

*Dranow, L., **Giacona, A. M.,** & Lampinen, J. M. (2023). National Registry of Exoneration Cases and Case Variables. Poster presented at the Southwestern Psychological Association (SWPA), Frisco, TX, March 31-April 2, 2023.

*Johnson, M. R., **Giacona, A. M.,** & Lampinen, J. M. (2023). National Registry of Exoneration Cases and Case Variables. Poster presented at the Southwestern Psychological Association (SWPA), Frisco, TX, March 31-April 2, 2023. Undergraduate Poster Finalist.

Giacona, A. M. (2023). Why more systematic research of estimator variables is needed. Theoretical talk presented at the Southwestern Psychological Association (SWPA), Frisco, TX, March 31-April 2, 2023.

Giacona, A. M., *Johnson, M. R., & Lampinen, J. M. (2023). Variable frequency in National Registry of Exoneration cases. Data blitz presented at the American Psychology Law Society (AP-LS), Philadelphia, PA, March 16-18, 2023.

Giacona, A. M., & Lampinen, J. M. (2023). Face masks and lineups: Do they hinder or help recognition? Data blitz presented at the American Psychology Law Society (AP-LS), Philadelphia, PA, March 16-18, 2023.

Giacona, A. M., & Lampinen, J. M. (2023). Mental filler siphoning is present in simultaneous showups. Symposium, “New Ideas in Eyewitness Research,” presented at Midwestern Psychological Association (MPA), April 20-22, 2023.

Lampinen, J. M., Regier, G. O., Provenzano, A. C., & **Giacona, A. M.** (2023). Eyewitness descriptions: Failure to mention distinctive features. Research presented at Midwestern Psychological Association (MPA), April 20-22, 2023.

Giacona, A. M., & Lampinen, J. M. (2022). COVID-19 masks at encoding: Do masked lineups increase accuracy? Research presented at Midwestern Psychological Association (MPA), April 21-23, 2022.

*Schuetter, B.S., **Giacona, A. M.,** Lampinen, J. M. (2022). Diagnostic feature detection or filler siphoning: A red box study. Research presented at Midwestern Psychological Association (MPA), April 21-23, 2022.

Gipson, N. I., **Giacona, A. M.,** & Lampinen, J. M., (2022). Mistaken eyewitness identifications of familiar but innocent suspects. Paper presented at Midwestern Psychological Association (MPA), April 21-23, 2022.

Giacona, A. M., & Lampinen, J. M. (2021). Diagnostic featural detection or filler siphoning: A red box study. Paper presented at the American Psychology Law Society (AP-LS), Denver, CO, March 17-20, 2022.

Giacona, A. M., & Lampinen, J. M. (2021). Should lineups have face masks if the suspect was masked? Research presented at the American Psychology Law Society (AP-LS), Denver, CO, March 17-20, 2022.

Giacona, A. M., *Aspenson, E. B., *Schuetter, B. S., & Lampinen, J. M. (2021). Estimating estimator variables in exoneration cases. Paper presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021.

Lampinen, J. M., Gipson, N. I., **Giacona, A. M.,** Provenzano, A. C., Smith, A. M., Wells, G. L., & Smalarz, L. (2021). Lineup / showup differences: filler siphoning vs. diagnostic feature detection. Paper presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021.

*Petersen, R., **Giacona, A. M.,** & Lampinen, J. M. (2021). Examining the effect of lineup administrator behavior on eyewitness identification. Research presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021.

*Regier, G. O., **Giacona, A. M.,** & Lampinen, J. M. (2021). Eyewitness descriptions of distinctive facial features. Research presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021.

*Siner, C. L., **Giacona, A. M.,** & Lampinen, J. M. (2021). Co-witness influence on police lineup identifications. Research presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021.

*Siner, C. L., **Giacona, A. M.,** & Lampinen, J. M. (2021). The effects of police bias and assumptions on photo lineup administrations. Research presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021. Psi Chi Award Winner.

*Schuetter, B., *Aspenson, E. B., **Giacona, A. M.,** Lampinen, J. M., & Anastasi, J. S. (2021). Viewing conditions negatively affect the confidence-accuracy relationship. Research presented at Midwestern Psychological Association (MPA), Virtual Conference, April 22-24, 2021.

*Forteith, H. D., *LaRoe, A. M., **Giacona, A. M.,** & Lampinen, J. M. (2020). Metacognition and facial recognition. Research presented at the 28th Annual Meeting of the Southwest Cognition and Cognitive Neuroscience Society (Armadillo), Virtual Conference, October 2-3, 2020.

*Aspenson, E. B., *Schuetter, B. S., **Giacona, A. M.,** & Lampinen, J. M. (2020). Frequency of estimator variables in exoneration cases. Research presented at the 28th Annual Meeting of the Southwest Cognition and Cognitive Neuroscience Society (Armadillo), Virtual Conference, October 2-3, 2020.

Giacona, A. M., Lampinen, J. M., Anastasi, J. S., & *Siner, C. L. (2020). Multiple estimator variables negatively affect the confidence-accuracy relationship. Research presented at Midwestern Psychological Association (MPA), Chicago, IL, April 23-25, 2020 (Cancelled due to COVID-19).

Lampinen, J. M., **Giacona, A. M.,** Gipson, N. I., Smith, A. M., Wells, G. L., & Smalarz, L. (2020). Lineup / showup differences: filler siphoning vs. diagnostic feature detection. Paper presented at Midwestern Psychological Association (MPA), Chicago, IL, April 23-25, 2020 (Cancelled due to COVID-19).

*Siner, C. L., **Giacona, A. M.,** & Lampinen, J. M. (2020). Co-witness influence on police lineup identifications. Research presented at Midwestern Psychological Association (MPA), Chicago, IL, April 23-25, 2020 (Conference cancelled due to COVID-19).

*Siner, C. L., **Giacona, A. M.,** Lampinen, J. M. & *Gross, M. A. (2020). Co-witness influence on police lineup identifications. Research presented at the Southwestern Psychological Association (SWPA), Frisco, TX, April 3-5, 2020 (Cancelled due to COVID-19).

*Smith, D. D., *Reiger, G. O., **Giacona, A. M.,** Lampinen, J. M., Anastasi, J. S. & *Siner, C. L. (2020). Multiple estimator variables negatively affect the confidence-accuracy relationship. Research presented at the Southwestern Psychological Association (SWPA), Frisco, TX, April 3-5, 2020 (Cancelled due to COVID-19).

Giacona, A. M., Lampinen, J. M., & Anastasi, J. S. (2020). Are pristine conditions enough?: Eyewitness accuracy and confidence in different viewing conditions. Paper presented at the American Psychology Law Society (AP-LS), New Orleans, LA, March 5-7, 2020.

Giacona, A. M. & Lampinen, J. M. (2019). Eyewitness accuracy and confidence in different viewing conditions. Research presented at the 27th Annual Meeting of the Southwest Cognition and Cognitive Neuroscience Society (Armadillo), San Antonio, TX, October 4-5, 2019.

*Siner, C. L., *Olvera, D. L., **Giacona, A. M.,** & Lampinen, J. M. (2019). Eyewitness accuracy and confidence in different viewing conditions. Research presented at the Psychological Science Research Night (PSRN), Fayetteville, AR, May 1, 2019.

Lampinen, J. M., Race, B., Provenzano, A. C., Gipson, N., & **Giacona, A. M.** (2019). Police identification policies in a southwestern state. Research presented at the American Psychology Law Society (AP-LS), Portland, OR, March 14-17, 2019.

Giacona, A. M., Anastasi, J. S., Lee, J. R., & Burke, J. L. (2018). The effects of photographs on the acceptance of product and scientific claims. Research presented at the Association of Psychological Sciences (APS), San Francisco, CA, May 24-27, 2018.

Shehadeh, K. K., Burke, J. L., **Giacona, A. M.,** Anastasi, J. S., & Lee, J. R. (2018). Social media vs. news articles: The influence of photographs on false memories. Research presented at the Association of Psychological Sciences, San Francisco, CA, May 24-27, 2018.

Horgan, M., **Giacona, A. M.**, &. Banks, C. S. (2018). Parent perceptions of bullying responses. Research presented at the Undergraduate Research Symposium, Huntsville, TX, April 28, 2018.

Giacona, A. M., Anastasi, J. S., Lee, J. R., & Burke, J. L. (2018). The effects of photographs on the acceptance of product and scientific claims. Research presented at the Annual Meeting of the Southwestern Psychological Association (SWPA), Houston, TX, April 13-15, 2018.

Giacona, A. M. (2018). Parenthood and media: Television influences college students' perceptions of parenthood. Research presented at the Annual Meeting of the Southwestern Psychological Association (SWPA), Houston, TX, April 13-15, 2018.

Burke, J. L., Shehadeh, K. K., **Giacona, A. M.,** Anastasi, J. S., & Lee, J. R. (2018). Social media vs. news articles: The influence of photographs on false memories. Research presented at the Annual Meeting of the Southwestern Psychological Association (SWPA), Houston, TX, April 13-15, 2018.

Giacona, A. M., Anastasi, J. S., Burke, J. L., & Lee, J. R. (2018). The effects of photographs on the acceptance of product and scientific claims. Paper presented at the Third Annual Student Scholars Forum, Shreveport, LA, March 2, 2018.

UNDERGRADUATE MENTORSHIP/SUPERVISION

Supervised Honor's Students/Theses (8):

[redacted] (2022-2023): *Effect of Exposure Duration and Familiarity on Identification*

[redacted] (2022-2023): *Simultaneous Showups: A Replication and Extension of the Red Box Study*

[redacted] (2022-2023): *Rigid and Non-Rigid Facial Images: Effects on Prospective Person Memory in Missing Persons Cases*

[redacted] (2022-2023): *Effect of Expert Testimony in Relation to Trust in Science*

[redacted] (2021-2022): *Red Box Study: A Replication and Extension*

[redacted] (2020-2021): *Examining the role double-blind lineup administrators have on eyewitness identification and confidence*

[redacted] (2020-2021): *Eyewitness descriptions of distinctive facial features*

[redacted] (2019-2020): *The effects of police bias and assumptions on photo lineup administrations*

Supervised Undergraduate Research Assistants (70):

[redacted]

WORK EXPERIENCE

Research Assistant, University of Arkansas, Fayetteville, AR — Spring 2023

Instructor of Record, University of Arkansas, Fayetteville, AR

- ***PSYC 4123: Perception*** — Fall 2022
- ***PSYC 2013: Introduction to Statistics*** — Summer I, 2022
- ***PSYC 3103: Cognitive Psychology***
  - *Face-to-Face with Remote Delivery Option* — Spring 2022
  - *Face-to-Face with Remote Delivery Option* — Fall 2021
- ***PSYC 2003: General Psychology***

Synchronous & Asynchronous Remote Delivery Fall 2020
Face-to-Face/Synchronous Remote Delivery Fall 2020

Teaching Assistant, University of Arkansas, Fayetteville, AR
Advanced Seminar: Affective Neuroscience Spring 2021
Introduction to Statistics: Drill Leader (5 sessions weekly) Spring 2020
Introduction to Statistics: Drill Leader (4 sessions weekly) Fall 2019
Advanced Seminar: Psychology and Law Spring 2019
Advanced Seminar: Ownership Spring 2019
Advanced Research: Face Perception Fall 2018
General Psychology (2) Fall 2018

Writing Tutor, University of Arkansas, Fayetteville, AR
Graduate Writing Consultant Jan. 2022- May 2023

Teaching Assistant, Sam Houston State University, Huntsville, TX
Intro to Research Methods Lab (2) Spring 2018
Intro to Research Methods Lab (2) Fall 2017
Intro to Research Methods Lab Summer II, 2017
Intro to Research Methods Lab Summer I, 2017

CERTIFICATIONS

University of Arkansas, Fayetteville, AR
CITI Research with Human Subjects Training
- Social/Behavioral Research Course Exp. Oct. 2023
- Social and Behavioral Responsible Conduct of Research Exp. Oct. 2024
- Conflicts of Interest Exp. Oct. 2024

University of Arkansas, Fayetteville, AR
College Reading and Learning Association Certification Level One Sept. 2022
- Certified with 25 hours of face-to-face tutoring, 10 hours of professional training, professional development, and leading/creating workshops.

Sam Houston State University, Huntsville, TX
College Reading and Learning Association Certification Level Three May 2018
- Certified with 75 hours of face-to-face tutoring, 30 hours of professional training, professional development, and leading/creating workshops.

MEMBERSHIPS

Southwestern Psychological Association (SWPA)	Since 2018
Midwestern Psychological Association (MPA)	Since 2019
American Psychology- Law Society (AP-LS; APA Division 41)	Since 2019
Cognition and Cognitive Neuroscience Southwest Regional Conference (ARMADILLO)	Since 2019
R Ladies- Fayetteville, AR Chapter	2019-2023

SERVICE/AWARDS

GPSC Travel Grant ($500)	2023
Regional Jr. High Quiz Bowl Judge	2023
GPSC Travel Grant ($500)	2022
President Graduate Psychology Program	2021-2022
Service Committee Member	2021-2022
GPSC Graduate Student Mentorship Award Recipient	2021
GPSC Travel Grant ($500)	2020
Experimental Training Committee (ETC) Student Representative	2019-2020
Law and Memory Processes (LAMP) Lab Web Coordinator	2019-Present
Diversity Committee Member	2018-2019
Social Committee Member	2018-2019

SOFTWARE PROFCIENCY

Experimental Design and Stimulus Development

Adobe Photoshop, Adobe Premiere Rush, Adobe Creative Cloud, E-Prime

Online Study Design and Deployment

SONA, Survey Monkey, Qualtrics, Amazon’s Mechanical Turk, Prolifics

Statistical Analysis

SPSS, R, Microsoft Excel

For Coursework/Virtual Learning

Blackboard/Blackboard Collaborate, Cisco Webex, Zoom, GatherTown

Other

Microsoft Word, Excel, PowerPoint