Can Decision-Making be Improved by Allowing Eyewitnesses to Opt-Out?: Examining the Utility of a 'Not Sure' Option With Showups

By

Shaela T. Jalava

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Psychology

Carleton University

Ottawa, Ontario

Abstract

I examined the impact of an explicit opt-out option on eyewitness identification performance. I predicted that an opt-out option would decrease innocent-suspect identifications more than culprit identifications, and that this effect would be more pronounced when viewing conditions were worse. I randomly assigned participants (*N* = 2003) to watch either a clear or degraded simulated-crime video. After a brief filler task, participants viewed either a culprit-present or culprit-absent showup and responded either "Yes" or "No". Half of the participants were randomly assigned to have an additional option to respond, "Not Sure". Contrary to my prediction, the not-sure option decreased both culprit (44% to 36%) and innocent-suspect (19% to 14%) identifications; this effect was unaffected by viewing condition quality. Despite empirical evidence and theoretical rationale indicating an opt-out option would improve the culprit and innocent-suspect identification tradeoff, the present results suggest otherwise.

Acknowledgements

Throughout the process of writing this thesis I have received tremendous support and assistance. I would first like to acknowledge the Natural Science and Engineering Research Council for funding this project.

Next, I would like to extend my sincerest thanks and gratitude to my supervisor, Dr. Andrew Smith, without whom this would not have been possible. His expertise, advice, and camaraderie were invaluable, and he has put forward a great deal of his time to help me grow as a burgeoning researcher.

I would also like to thank my lab colleagues, Simona Mackovichova and Daniella Filoso, for their help with the ethics application, proofreading, and their emotional support.

Finally, I would like to thank my friends and family, especially my partner Sebastian Baglole, for lending a sympathetic ear throughout this process and providing much-needed breaks from my research.

Table of Contents

List of Tables

List of Figures

List of Appendices

Does Encouraging Strategic Regulation of Memory Improve Eyewitness Decision-Making When Memory is Weak?: Examining The Utility of a 'Not-Sure' Option with Showups

On a July evening in 1984, a man broke into the apartment of a then 22-year-old college student, Jennifer Thompson. He severed her phone lines, rifled through her belongings, and raped her at knifepoint. During the brutal attack, Jennifer tried to commit every detail of her assailant to memory. She understood that the more features she could remember, the more likely he was to be caught and punished for what he had done. When the police brought her in to identify her attacker, she studied a photo-array of six black men for several minutes. She eventually picked out two photos and spent some time deciding between them. Jennifer finally picked out a 22-year-old man by the name of Ronald Cotton. "I think this is the guy", Jennifer claimed. After some prompting from the officers, she confirmed that she was certain about her decision. She asked the detectives "Did I do OK?" and they answered, "You did great, Ms. Thompson." (Garrett, 2011).

One week later the police brought Jennifer back into the station and showed her a live lineup. Cotton was the only member of this lineup who was also present in the photographic lineup presented to Jennifer a week prior. She again spent several minutes wavering between Cotton and another lineup member. Eventually, she identified Cotton for a second time. After repeating the lineup procedure once more, she claimed that Cotton "looks the most like [the attacker]", (State v. Cotton, 1987). The detectives responded, "We thought that might be the guy. It's the same person you picked from the photos." At that point, she felt completely confident that she had identified the correct man. For Jennifer, it was a huge relief to have succeeded in catching who she thought was responsible. Her immediate ordeal, if not the lingering effects of the traumatic incident, was over (Frontline, 1997).

Ronald Cotton was put on trial and convicted of rape and burglary in January 1985. He received a sentence of life in prison plus 54 years. At the trial, the jury saw Jennifer point out – with absolute certainty – Cotton as the man who had viciously raped her. After many years behind bars, Ronald Cotton was exonerated by DNA evidence. DNA linked the crime to another man named Bobby Poole, who had been boasting about it from prison. Cotton was cleared of all charges and released on June 30, 1995.

The wrongful conviction of Ronald Cotton is one of many miscarriages of justice that can be attributed to mistaken eyewitness identification. Mistaken eyewitness identification has played a role in more than 70% of DNA-based exonerations in the United States (Innocence Project, 2017). By the time Jennifer reached the witness stand, she was absolutely certain that Ronald Cotton was the man who raped her. But, on the initial lineup procedure, she was not at all certain that Ronald Cotton was the man who raped her. Indeed, she spent several minutes wavering between Cotton and another innocent lineup member. That is not the behaviour of a confident eyewitness. Indeed, Jennifer was not at all sure that Cotton was the culprit as is evident from her statement, "I think this is the guy." What if, on the initial photographic lineup, Jennifer, who had the best intentions but lacked initial confidence in her decision, had been given an option to respond, "I'm not sure"? If Jennifer were able to express her uncertainty by opting out of making a definitive judgment, perhaps Ronald Cotton would never have been convicted. Maybe he would not have spent over 10 years in prison for a crime he did not commit.

The applied potential for an explicit option to withhold an identification can be illustrated through cases like this and other similar cases in which the witness lacked initial confidence in his or her decision (eg., Bain v. State, 2007; State v. Booker, 1985). My proposed thesis aimed to test whether the inclusion of an explicit 'not-sure' option for an eyewitness showup identification procedure (a one-person lineup) decreases innocent-suspect identifications to a greater extent than culprit identifications. I crossed the explicit not-sure option with the presence or absence of the culprit and with the quality of viewing conditions (a proxy for memory strength). As I will explain below, there are theoretical reasons to expect that a not-sure option will prove most important when participants are given a poor view of the culprit. Before explaining the theoretical rationale behind my experiment, I will discuss the roles of relevant system and estimator variables in the study of eyewitness identification performance (Wells, 1978). Then I will describe a previous experiment that utilized an explicit option to withhold identification decisions. Finally, I will outline the features of my thesis including the hypotheses, methodology, results, and discussion.

Eyewitness Identification Procedures

When an investigation leads law-enforcement personnel to suspect that a particular individual perpetrated the crime, they will often place this suspect in a lineup identification procedure. A lineup is a procedure in which the suspect is embedded amongst some number of known-innocent persons commonly referred to as fillers. Importantly, law-enforcement personnel do not know whether the suspect is guilty or innocent. If law-enforcement personnel knew the guilt status of the suspect, there would be no need for a lineup procedure. Indeed, despite potential misconceptions amongst lay audiences, the objective of a lineup procedure is not to test the ability of the eyewitness to identify the culprit from the lineup. Instead, a lineup procedure aims to test law enforcement's hypothesis that the suspect is the culprit (Charman & Wells, 2007). Sometimes the suspect in a lineup is the culprit, but other times the suspect is an innocent person and how the witness responds to the lineup procedure informs law enforcement on the probability that the suspect is guilty. Law enforcement are using the witness' memory as a tool to make an inference about the probability of the suspect's guilt. The lineup model is depicted in Table 1 (this is a revised version of the model depicted in Wells, 1993).

Table 1

Possible Outcomes from an Eyewitness Lineup Procedure

	Witness Decision		
Lineup Status	ID Suspect	ID Filler	Rejection
Culprit Present	Correct ID	Known False ID	False Rejection
Culprit Absent	Dangerous False ID	Known False ID	Correct Rejection

Note. In signal detection terms correct IDs are referred to as hits, false rejections are referred to as misses, and false identifications (of either a suspect or a filler) are referred to as false alarms.

Law enforcement are looking to see if the eyewitness will identify the suspect from the lineup. There are two possible lineup statuses when a witness is presented with a lineup: the culprit is present in the lineup or the culprit is not present in the lineup (in which case the suspect is innocent). Regardless of whether or not the culprit is present, there are three possible outcomes: the witness could identify the suspect (a correct identification in the culprit-present case, or a dangerous false identification in the culprit-absent case), the witness could identify an innocent filler (a known false identification in both cases), or the witness could reject the lineup (an incorrect rejection in the culprit-present case, or a correct rejection in the culprit-absent case).

It is important to distinguish between the two types of false identifications that can occur in a lineup: the false identification of an innocent suspect and the false identification of known-innocent filler. On the one hand, the false identification of an innocent suspect is considered dangerous because law enforcement personnel do not know that this witness has made an error. In fact, they think that the suspect is guilty and that is why they have placed him or her in a lineup. The false identification provides support for the hypothesis that the suspect is guilty, which could eventually lead to a false conviction. On the other hand, a known-innocent filler identification is not considered dangerous because law enforcement know that the witness has made an error. It is still an incorrect identification, but it is not regarded as dangerous in an applied manner. In fact, the purpose of including fillers in a lineup is to create the possibility for a known error – a response that law enforcement personnel can classify as an error without uncertainty. The problem with filler picks is not that they lead to the arrest and potential conviction of innocent persons, but rather that they destroy the witnesses' credibility. Indeed, a witness who picks a filler out of a lineup has impeached herself and would likely not be of further use to law enforcement should the investigation continue.

What is interesting is that each eyewitness behaviour is informative to a police investigation. Because suspect identifications are more common when the suspect is guilty than when the suspect is innocent, suspect identifications increase the probability that the suspect is the culprit (Wells, Yang, & Smalarz, 2015). To the contrary, because filler identifications and rejections are more likely when the suspect is innocent than when the suspect is guilty, these witness behaviors decrease the probability that the suspect is the culprit (Wells et al., 2015). This brings me back to the central purpose of surrounding a suspect with known-innocent fillers. Known-innocent fillers serve to protect innocent suspects from mistaken identification. Indeed, when the suspect is innocent, the majority of false-positive identifications do not land on the innocent suspect, but rather are "siphoned" off by known-innocent fillers. In the absence of known-innocent fillers (i.e., when law enforcement use a showup), suspects are at much greater risk of mistaken identification (e.g., Smith, Wells, Lindsay, & Penrod, 2017; Smith, Wells, Smalarz, & Lampinen, 2018).

Another commonly used identification procedure is the showup procedure. An eyewitness showup is essentially a one-person lineup in which only a lone suspect (and no fillers) is presented to the eyewitness for an identification attempt. As with a lineup, the suspect in a showup could be guilty or innocent. When the suspect is guilty, the witness could correctly identify the culprit (a hit) or could mistakenly reject the showup (a miss). When the suspect is innocent, the witness could correctly reject the innocent suspect or could mistakenly identify the innocent suspect (false alarm). There are no known-innocent fillers to protect against a false identification in a showup. Table 2 displays the four possible outcomes.

Table 2

Possible Outcomes from an Eyewitness Showup Procedure

	Witness Decision	
Showup Status	ID Suspect	Rejection
Culprit Present	Correct ID	False Rejection
Culprit Absent	False ID	Correct Rejection

Note. In signal detection terms correct IDs are referred to as hits, false rejections are referred to as misses, and false identifications are referred to as false alarms.

Showups are one of the most common identification procedures, with past estimates of their use ranging from 30% to 77% of all identification procedures (Gonzalez, Ellsworth, & Pembroke, 1993; McQuiston & Malpass, 2001). Showups offer some advantages and disadvantages when compared to lineups. I have already alluded to one potential disadvantage–without the presence of known-innocent fillers, a potentially innocent suspect is not offered any protection from a witness with a weak memory who decides to pick someone from the procedure. Because there are no fillers to siphon off false picks, all false identifications fall on the suspect, inevitably leading to more innocent-suspect identifications.

There is one primary advantage to a showup that becomes apparent when taking into consideration the feasibility of lineups in a real case. When law enforcement can locate someone fitting the general description of the culprit near the scene of the crime (in terms of both time and distance), that is not considered probable cause for an arrest, but it is cause for a brief investigative detention (Smith et al., 2014). Such detentions do not offer law enforcement sufficient time to construct a lineup, but they can use showups to avoid letting a potentially dangerous suspect go. So, showups fill an important gap in the investigative process: they have the potential to lead to the swift arrest of the guilty and to quickly exculpate the innocent. Note that I will examine performance from showup procedures in the present thesis. However, because most eyewitness identification research focuses on lineups, I will focus primarily on lineups for my literature review.

Eyewitness Confidence

In addition to asking witnesses to make an identification decision, law enforcement also asks them to qualify that decision with a confidence statement. This is a part of the complete identification procedure, and it applies to both lineups and showups (Wells et al., forthcoming). What is the purpose of collecting a confidence statement? Confidence can be a potential marker of the accuracy of an identification decision and can be used to help law enforcement distinguish between accurate and inaccurate suspect identifications.

Historically, many concluded that confidence is, in fact, an unreliable predictor of eyewitness accuracy (Kassin, Ellsworth, & Smith, 1989; Luus & Wells, 1994; Wells, Ferguson, & Lindsay, 1981). This notion came about partly because it was discovered that confidence could be easily influenced by external factors. Some of these external factors are beyond the control of law enforcement (e.g., individual differences); however, some external factors influencing a witness' confidence are under the direct control of law enforcement. For example, the lineup administrator might provide confirmatory feedback to the witness after she has identified the suspect, which tends to result in an increase in confidence (Wells & Seelau, 1995). Consequently, confidence is malleable, as information confirming the witnesses' belief is likely to produce an increase in confidence that may be unrelated to accuracy (Smalarz & Wells, 2013).

That is not to say that confidence and accuracy are always wholly unrelated. In fact, it is now widely believed that under at least some conditions, confidence is strongly associated with accuracy. A watershed paper by Wixted and Wells (2017) went so far as to argue that when lineups are conducted under what they refer to as "pristine conditions", the confidence-accuracy relationship will be strong and high-confidence identifications will imply high accuracy. The pristine conditions outlined by Wixted and Wells (2017) are as follows: include only one suspect per lineup, the suspect should not stand out in the lineup, caution that the culprit may or may not be present, use double-blind administration, and collect a confidence statement immediately after the identification. Wixted and Wells (2017) argued that when conditions are pristine, high confidence implies high accuracy and low confidence implies low accuracy; however, if conditions are not pristine, all levels of confidence are associated with low levels of accuracy.

It also seems that the strength of the relationship between confidence and accuracy is limited to eyewitnesses who make an affirmative identification. In a meta-analysis of the confidence-accuracy relation for eyewitnesses, Sporer, Penrod, Read, and Cutler (1995) concluded that choosing moderated the relationship. When only including those witnesses who made an identification, the correlation between confidence and accuracy was $r = .41$. In other words, witnesses who correctly identified the target tended to be more confident than witnesses who identified an innocent person. For non-choosers (those who rejected the lineup), however, the confidence-accuracy correlation was a much more modest $r = .12$. Confidence bore little relation to the accuracy of eyewitness rejections.

One important note of relevance to my experiment– Smalarz and Wells (2013) point out an important methodological confound in the research on the relationship of confidence and accuracy for non-choosers. They note that non-choosers are comprised of two sub-types of witnesses, those who reject the lineup because they do not believe the culprit is not present and those who are uncertain either way. Failing to have a mechanism to remove these uncertain witnesses may be obfuscating what might otherwise be a strong relationship between confidence and accuracy for non-choosers. If these two sub-types were parsed out, a stronger association between confidence and accuracy might appear for witnesses who reject the lineup because they do not believe the culprit is present. To this end, an explicit not-sure option for eyewitnesses would provide researchers and law enforcement alike with more direct information about what type of non-choosing eyewitness they are dealing with. Moreover, confidence would potentially have more utility in discriminating between correct and incorrect rejection decisions.

Variables Affecting Eyewitness Identification Performance

Keeping in mind the matrix of possible identification outcomes, it is important to consider how factors such as viewing conditions at the time of the crime and the response options available to witnesses during the procedure affect the quality of identification performance. Some 40 years ago, Wells (1978) sought to increase the impact of eyewitness identification research by distinguishing between two categories of eyewitness variables: system variables and estimator variables. System variables are those that are under the control of the criminal justice system. Wells (1798) identified four broad types of system variables: The method of presentation, the instructions that the eyewitness is given during the procedure, the content of the lineup itself, and behavioural influence of the lineup administrator, including how and when confidence is measured (Wells & Olson, 2003). For example, law enforcement has control over how the identification procedure is explained to the witness. They could choose to inform the witness that the culprit may-or-may-not be present in the lineup, and/or let them know that they have the option to opt out of making an identification decision altogether.

To the contrary, estimator variables are those variables that are not under the control of the criminal justice system. Although law enforcement cannot control these variables, the effects of these variables can be estimated, hence the name, ‘estimator variable’. And, like system variables, estimator variables have an appreciable impact on the quality of identification performance (Lindsay & Wells, 1985). Wells (1978) divided estimator variables into three categories: characteristics of the criminal event, characteristics of the defendant, and characteristics of the witness. Factors such as quality of viewing conditions and exposure duration fall under the category of characteristics of the criminal event.

Wells’ (1978) distinction between system and estimator variables has had a long-lasting impact on the field of eyewitness science and it continues to guide research to this day. This distinction has also given rise to several significant advances in the reduction of false identifications, which is critical because they can lead to the wrongful conviction of innocent persons, and have dire consequences for the integrity of the criminal justice system and society in general (Wells et al., 1998). One damaging aspect of false identification is that not only does it potentially ruin the life of an innocent person, but it also allows the guilty party to go free, without atoning for the crime, and to possibly commit more crimes in the future. With the goal of reducing false identifications in mind, my thesis examined an interaction between a system and estimator variable. Namely, the response options available to the witness and viewing conditions, which I will review in the following sections.

Viewing Condition Variables

The effects of race, crime seriousness, and witness age are a few of the estimator variables that have been researched historically (Kassin, 1984; Meissner & Brigham, 2001; Pozzulo & Lindsay, 1998). However, because my experiment focuses on quality of viewing conditions, I focus solely on variables that fall within this category. I will review the relevant research on exposure duration and distance and describe how manipulations of viewing conditions generally lead to a decrease in correct identifications and an increase in false identifications.

Exposure duration. Exposure duration, or time spent viewing an event, is considered an estimator variable because it occurs outside of the control of the criminal justice system. Although the effects of exposure duration on facial recognition in the general memory literature have been well-documented (Bornstein, Deffenbacher, Penrod, & McGorty, 2012; Shapiro & Penrod, 1986), a review of the eyewitness literature reveals a strikingly limited number of studies that explicitly manipulated the amount of time the witness was exposed to the culprit's face. Nevertheless, a small number of researchers have experimentally manipulated exposure duration, theorizing that less exposure to a criminal event should decrease eyewitness accuracy (Memon, Hope, & Bull, 2003; Read, 1995).

Memon et al. (2003) examined the effect of exposure duration (12s versus 45s) for a video of a simulated robbery on eyewitness accuracy. They collected lineup identifications from 84 young-adult participants (aged 17-24) and 80 older-adult participants (aged 59-81) 40 minutes after viewing the event. They found an interaction between exposure duration and suspect identifications– shorter exposures decreased culprit identifications and increased innocent-suspect identifications. Thus, the idea that a shorter exposure duration during the witnessing of a crime would lead to a worse identification outcome was supported.

Viewing distance. How is eyewitness accuracy affected by the distance from which the witness observed the crime? Knowledge about the human visual and perceptual system implies that faces viewed at a distance would be associated with compromised recognition memory. This is because as distance increases, the quality and quantity of information available to an observer decreases (Loftus & Harley, 2005). Phenomenologically, as distance increases, facial information becomes increasingly "blurred". As faces become increasingly "blurred", there would be less information available to an eyewitness, which means that from the outset, the eyewitness would likely develop a weaker mental representation of the culprit. To the extent that an eyewitness develops a weaker mental representation for the culprit at encoding, one would expect this eyewitness to have a weaker memory for the culprit during a later identification procedure. Hence, intuition suggests that the accuracy of eyewitness identifications should be inversely related to viewing distance.

In one of the few eyewitness studies directly examining the effect of viewing distance on accuracy, Lampinen, Erickson, Moore, and Hittson (2014) showed 185 undergraduate participants eight target individuals at one of eight viewing distances. They viewed the targets live for 10s each, during daylight, on the sidewalk of a college campus. The participants then viewed 16 photographs presented in a random order, eight of which depicted one of the target individuals and the other eight depicting innocent persons. Consistent with the impact of shorter exposure durations, Lampinen et al. (2014) found that increasing encoding distance also led to a reduction in hits and an increase in false alarms.

Both of the studies reviewed showed the same pattern of results. In fact, although there are a limited number of experiments dedicated to manipulating viewing conditions in the eyewitness context, they show similar patterns– that when memory strength is decreased through manipulation of viewing condition quality, there is both a reduction in culprit identifications and an increase in false identifications (Smith, Wilford, Quigley-McBride, & Wells, 2019). The present experiment also contains a viewing condition manipulation. Therefore, I expected to observe a reduction in hits and an increase in false alarms for the group of participants that received a relatively worse view of the culprit. My thesis also manipulated the response options given to the witness. In the following section I will explore how the idea of manipulating the available response options followed, in part, from some research on eyewitness instructions.

Instruction effects

One of the most important ways law enforcement can influence the accuracy of identification performance is how the procedure is explained to the witness, i.e., the instructions given to them. Instructions can be used to help guide the witness into making appropriate decisions and understanding the implications of a wrong one. For example, law enforcement can inform the witness that the culprit may not actually be present in the lineup. This instruction is intended to reduce false identifications by curtailing the common preconception that the purpose of an identification procedure is to pick the culprit from amongst the fillers (Wells et al., 1998). In the following section, I will review some of the instructions that have been developed to improve identification outcomes, and how the inclusion of a not-sure response option extends these ideas.

May-or-may-not instruction. Witnesses often make the assumption that if they are being brought in to view a lineup, then the culprit must be present (Memon, Gabbert & Hope, 2013). However, the base rate of culprit presence in the real world is likely quite low (around 30%, see Wixted et al., 2015), so this mistaken notion is likely to lead to a high rate of false identifications. Thus, there is a need to disabuse witnesses of this dangerous misconception, such as an instruction that takes a neutral position towards the presence of the culprit. Instructions that fall within that category are deemed unbiased instructions. Instructions that bias the witness towards making an identification – or provided no information about the presence of the culprit at all – are deemed biased instructions (Clark, 2005; Wells et al., forthcoming). The most prevalent unbiased instruction that is used currently is telling the witness that the culprit may-or-may-not be present in the identification procedure (Clark, 2005).

Comparisons between identification procedures with unbiased compared to biased instructions have revealed both positive and negative outcomes. Two meta-analyses on instruction effects concluded that unbiased instructions tend to lead to a decrease in both culprit and innocent-suspect identifications (Clark, 2005; Steblay, 1997). Whereas a decrease in innocent-suspect identifications in certainly a desired outcome, a decrease in culprit identifications is not. Providing witnesses with the instruction that the culprit may-or-may-not be present affects their willingness to make a positive identification. Despite mixed results, informing the witness that the culprit may-or-may-not be present in the lineup has become commonplace in research on eyewitness identification performance as well as in the field (Wogalter, Malpass, & McQuiston, 2004).

Additional-opportunities instruction. Smith, Wells, Lindsay, and Myerson (2018) reasoned that many innocent-suspect identifications result from witnesses being concerned that this is their one and only opportunity to make an identification and wanting to avoid missing that opportunity. Accordingly, they developed a simple instruction intended to disabuse witnesses of this misconception. They told half of their participants that if they did not think that the culprit was present in the showup procedure, they would have an additional opportunity to view another suspect later. The rationale behind the additional-opportunities instruction is that more witnesses encountering innocent suspects will have a weak match-to-memory experience compared to witnesses encountering guilty suspects. Witnesses with strong memory experiences are not likely to be influenced by scarcity concerns because they simply see a good match and make an identification. But, witnesses with weak memory experiences are especially prone to making identifications due to scarcity concerns. Thus, more innocent-suspect identifications (compared to guilty suspect identifications) might be caused by these scarcity concerns and if so, disabusing witnesses of the idea that his is their only opportunity to make an identification might reduce innocent-suspect identifications more than guilty suspect identifications. As predicted, the additional-opportunities instruction decreased the innocent-suspect identifications to a greater extent than culprit identifications (Smith et al., 2018), and this same pattern of results was replicated by Eisen, Smith, Olaguez, and Skerritt-Perta, (2017).

Opt-out option. In theory, the standard admonition informing the witness that the culprit may-or-may-not be present in the lineup can help increase the saliency of the potential response options a witness has available to them. After hearing the admonition, the witness should be aware that she can choose to make an identification if she believes the culprit is present, or that she is free to reject the lineup if she does not believe the culprit is present. But what about the subset of witnesses who may not be sure one way or the other? If witnesses who do not have enough memory information to make a definitive judgment are forced to make one, those witnesses are relatively likely to be incorrect (Brewer & Wells, 2006).

The standard admonition does not seem to be effective at alerting the witness to a possible third option – answering “I’m not sure” or “I don’t know”. Indeed, without being provided an explicit option to “opt out” of making a decision, few participants seem to do so (Warnick & Sanders, 1980; Weber & Perfect, 2012), and many eyewitness experiments do not give participants any ability to “opt out” from making a definitive decision despite providing them with the standard admonition. However, when provided with the option to withhold a decision, a substantial minority of participants responded, “I don’t know” (approximately 19%, Weber & Perfect, 2012). These results suggest that in the absence of an explicit opt-out option, participants who might be having trouble making an identification decision are being forced to do so. More importantly, in the real world, witnesses are often not given an explicit opt-out option. Of course, they can opt out in theory, but if they are not explicitly told by law enforcement that “I’m not sure” is a valid response, then their likelihood of spontaneously opting out is extremely low.

The idea that providing an explicit opt-out option would improve eyewitness identification performance stemmed from research demonstrating that superior accuracy of recalled items can be achieved by emphasizing the strategic regulation and monitoring of memory (e.g., Higham, 2007; Kelley & Sahakyan, 2003; Koriat & Goldsmith, 1996). These experiments demonstrated that permitting participants to withhold information – via the use of an opt-out option – increased the accuracy of information these participants provided. According to Koriat and Goldsmith (1996), the ability to withhold information or opt out can enhance accuracy to the extent that (a) witnesses wish to make accurate decisions, and (b) witnesses can effectively monitor the strength of their memories. Presumably all witnesses want to make correct decisions and there is an established confidence-accuracy association (Brewer & Wells, 2006), which suggests that witnesses can accurately monitor their memories. Weber and Perfect (2012) took inspiration from the research on memory monitoring and conducted one of the only experiments to date studying the effectiveness of an opt-out option for an eyewitness identification task, revealing some promising results.

Weber and Perfect (2012) tested the effect of a “don’t-know” option for a showup after either a three-minute delay or after a three-week delay. The experiment was conducted in-lab, where participants viewed a staged attempted car theft on a computer. After viewing the video, they saw a photograph of either the person from the video, displayed from the shoulders up, or a similar-looking innocent replacement, also displayed from the shoulders up. The authors expected that participants in the delayed showup condition would use the “don’t-know” option more frequently than the immediate condition. Research suggests that a delay in retrieval has a detrimental effect on memory quality (Shapiro & Penrod, 1986), thus they expected that participants who experienced a delay would be sensitive to how this delay impacted their memory for the culprit and would therefore be more likely to opt out from making a definitive decision. Surprisingly, the propensity to use the don’t-know option did not differ as a function of delay length (immediate delay: 20% versus short delay: 19%); however, providing participants with a don’t-know option decreased innocent-suspect identifications (from 27% to 13%) but had virtually no effect on culprit identification (37% vs. 36%). This effect is promising as it suggests that an opt-out option can improve the tradeoff between culprit and innocent-suspect identifications. The primary goal of the current thesis was to further evaluate the efficacy of providing witnesses with an option to opt out from making a definitive identification decision.

Moreover, I also examined whether the propensity to opt out from making an affirmative identification decision was more prevalent when witnesses are given a poor view of the culprit compared to when witnesses are given a clear view of the culprit.

Theoretical Rationale

Before elaborating on the rationale behind my proposed thesis, a short introduction to Signal Detection Theory and explanations of two related phenomena called the strength-based mirror effect (Glanzer & Adams, 1985) and present-absent criteria discrepancy (Smith et al., 2018) are warranted. Furthermore, an understanding of these concepts will help justify my prediction that the not-sure option would be most useful for witnesses who encounter a target-absent showup, compared to those who encounter a target-present showup.

Signal Detection Assumptions

Signal Detection Theory (SDT) provides a framework to describe decisions made under uncertain circumstances (Wickens, 2010). In a typical signal detection memory experiment, participants study a series of stimuli. They are then presented with a large number of randomly ordered target-present (signal) trials and target-absent (noise) trials and attempt to respond ‘old’ to studied stimuli (targets) and ‘new’ to non-studied stimuli (noise). Participants are trying to discriminate between the presence and absence of signal in a noisy environment. The problem with noise is that it can make it difficult to detect signal when present (leading to false negatives or misses) and it can masquerade as signal when there is no signal to detect (leading to false positives or false alarms). Because environments are noisy, memory performance will typically be less than perfect.

Beyond providing a conceptual explanation for why persons sometimes fail to recognize previously-seen items or falsely recognize non-studied items, Signal Detection Theory is also useful in that it provides measures of discriminability that are independent of response bias (Macmillan & Creelman, 2005). Discriminability refers to the ability of a decision-maker to distinguish between previously-seen and non-studied items and is commonly measured using either the area under the Receiver Operator Characteristic (ROC) curve or its theoretical stand-in, the discriminability index (d'). Independent of discriminability is the decision-maker's decision criterion, which refers to the amount of evidence the decision-maker requires to make an affirmative response (i.e., to classify a studied item as studied). Whereas the decision criterion is a latent construct, researchers typically make inferences about decision criteria on the basis of a decision-maker's response bias, which is simply her tendency to favour an affirmative response (e.g., studied) over a negative response (e.g., not studied). In the eyewitness literature, researchers most commonly use c as their measure of response bias, with negative values indicating a bias towards an affirmative response and positive values indicating a bias towards a negative response. Panel A of Figure 1 displays the signal and noise distributions and decision criterion that conform to the traditional equal-variance signal detection model.

Panel A of Figure 1 has two probability distributions. The black distribution represents the possible match-to-memory values for witnesses who encounter a culprit-present procedure, and the grey distribution represents the possible match-to-memory values for witnesses who encounter a culprit-absent procedure. The black curve exceeds the grey curve because the culprit, on average, provides a greater match-to-memory than does the innocent suspect. Match-to-memory for the culprit (black distribution) and match-to-memory for innocent persons (grey distribution) are represented as distributions because they reflect the fact that there is variation in match-to-memory from witness to witness. Discriminability is represented by the degree of overlap between the two distributions. Panel A has a stationary decision criterion that does not vary as a function of culprit (or target) presence, as is customary in almost all signal detection models.

Classic Signal Detection Theory assumes that criterion placement is a noise-free and stationary process. When applied to inanimate decision-makers, such as in a machine-learning context, this assumption is likely a safe bet. But, in the context of basic recognition research, there is evidence that where a participant places his or her decision criterion might vary from trial-to-trial (Benjamin, Diaz, & Wee, 2009). Moreover, in the context of eyewitness identification experiments, there is little debate that there must be variability in criterion placement. Indeed, in this context each witness makes only a single decision and researchers estimate decision criterion for each between-participant condition by averaging across the participants in that condition. Because different participants will require different amounts of evidence to make an affirmative identification, there must be variability around the decision criterion in the context of eyewitness identification (Smith et al., 2017). Hence, it would probably be more accurate to think of decision criteria not as a line as it is depicted in Figure 1A, but as a distribution, similar to how the match-to-memory distributions are conceptualized.

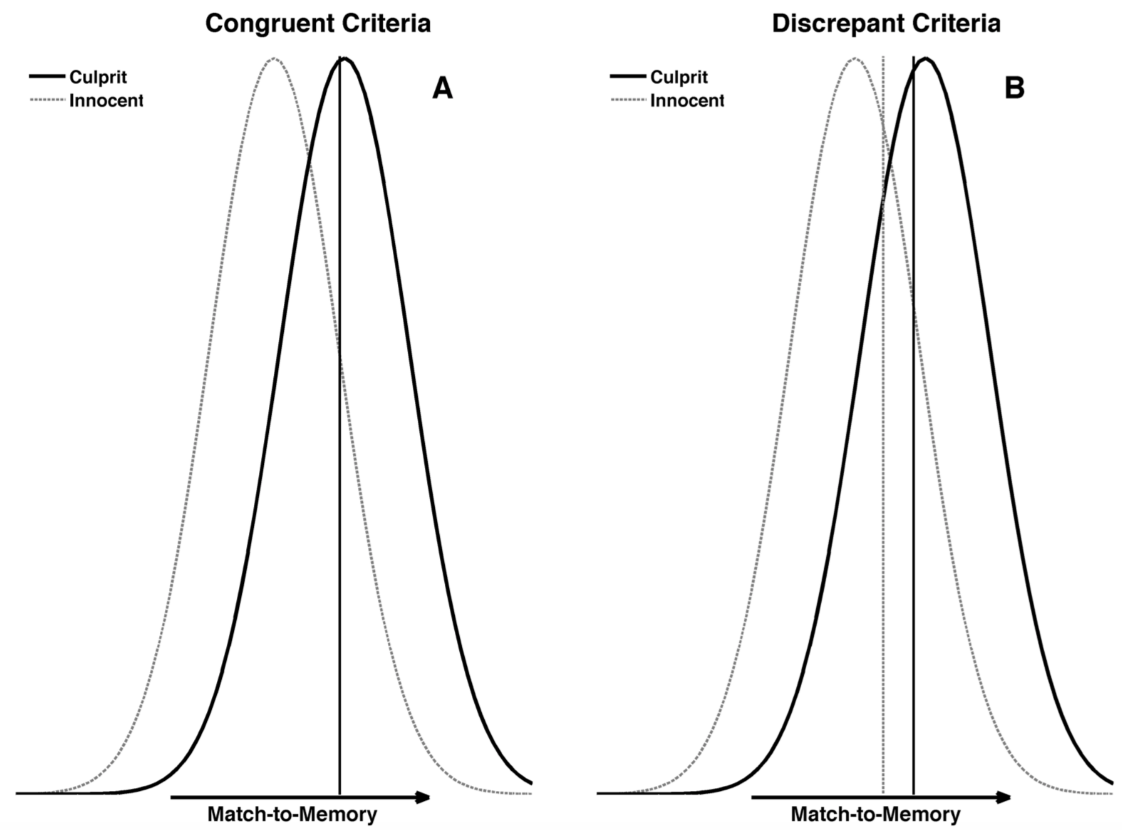

Figure 1. The figure compares the traditional SDT model (Panel A) to the model described by Smith et. al., (2018) (Panel B). Panel A has a stationary decision criterion that does not vary as a function of culprit (or target) presence, where Panel B has two different criterion placements (a more lenient decision criterion for trials where the culprit (or target) is absent). Adapted from "Eyewitness Identification Performance on Showups Improves with an Additional-Opportunities Instruction: Evidence for Present-Absent Criteria Discrepancy", by A. M. Smith, G. L. Wells, R. C. L. Lindsay, and T. Myerson, 2018, Law and Human Behaviour, 42. Copyright 2018 by the American Psychological Association. Adapted with Permission.

Strength-Based Mirror Effect

In almost all eyewitness experiments that manipulate quality of witnessing conditions, poorer views led to both a reduction in culprit identifications and an increase in false identifications (Lampinen et al., 2014; Memon et al., 2003). Characteristically speaking, the same pattern is also found nearly any time memory strength is weakened in the basic recognition memory literature (Glanzer, Adams, Iverson, & Kim, 1993). These findings typify a classic basic memory phenomenon referred to as the strength-based mirror effect (see Stretch & Wixted, 1998). The strength-based mirror effect describes a situation in which the weaker of two memory conditions is associated with both fewer hits and more false alarms when compared to the stronger of the two memory conditions.

Recall that both the study manipulating exposure duration (Memon et al., 2003) and the study manipulating viewing distance (Lampinen et al., 2014) produced a strength-based mirror effect. When witnesses had relatively worse witnessing conditions, they made fewer correct identifications and more false identifications. It is not at all surprising that poor witnessing conditions lead to a decrease in hits. Indeed, as the quality of encoding conditions decreases, it should be more difficult for witnesses to encode target persons' facial features, which in turn, should make it more difficult for witnesses to later recognize these persons. But it is surprising that poor encoding conditions lead to an increase in false alarms. Indeed, it is not as though a weak mental representation for target persons somehow made novel faces look more like target persons. So, what did cause this increase in false alarms? Researchers examining the strength-based mirror effect have theorized that false alarms increase because participants with weaker memories for studied items require less evidence to judge items as having been studied because they do not want to miss an opportunity to get a hit (Smith et al., 2019; Wixted & Gaitan, 2002).

Thus, according to this conceptualization, weaker memory conditions lead to a reduction in hits because it becomes more difficult to recognize target persons and lead to an increase in false alarms because people respond to these more difficult recognition conditions by lowering their criteria for making an affirmative response.

In the current experiment, I manipulated quality of viewing conditions, expecting to find the strength-based mirror effect for the poor-view condition. Based on the existing evidence from the recognition memory literature, correct identifications are likely to be lower and false identifications higher for those witnesses who had a poor view of the culprit. However, the presence of an opt-out option should help to mitigate the size of the mirror effect by affording witnesses with poor memory an opportunity to respond ‘not sure’ instead of forcing them to make an identification or outright rejection. Therefore, the opt-out option should reduce the number of false identification errors, especially for those witnesses who were in the poor-view condition.

Present-Absent Criteria Discrepancy

As elaborated on above, there is ample evidence that when memory is weak, witnesses lower their criterion for making an affirmative identification decision (Lampinen et al., 2014; Memon et al., 2003). Recent research suggests that it does not matter whether match-to-memory is weakened at encoding or at retrieval; in both cases, witnesses respond to the weak match-to-memory by lowering their criteria for making an affirmative identification (Kent, Lamberts, & Patton, 2018; Smith et al., 2019).

Of course, one way to make it more difficult to detect signal at the time of retrieval is to present the witness with a culprit-absent lineup. Clearly, a witness who is presented with a culprit-absent procedure will experience more difficulty in signal detecting than a witness presented with a culprit-present procedure, because there is no signal to detect. On average, innocent suspects will tend to provide a worse match-to-memory than will guilty suspects. Hence, when researchers randomly assign witness-participants to culprit-present and culprit-absent identification procedures they are also assigning them to strong (culprit present) and weak (culprit absent) match-to-memory conditions (Smith et al., 2018; Smith et al., 2019).

From this perspective it seems reasonable to suspect that witnesses who encounter innocent suspects would respond to the relatively weak match-to-memory experience the same way witnesses respond to any other weak match-to-memory experience – by lowering their criteria for an affirmative identification relative to those who encounter guilty suspects (and thus have a relatively strong match-to-memory experience). In essence, this present-absent criteria discrepancy hypothesis is a special case of the strength-based mirror effect in which the lowering of decision criteria results from the fact that innocent-suspects provide a weak match-to-memory. Although this hypothesis is consistent with a large body of research on the strength-based mirror effect, previous applications of Signal Detection Theory to the eyewitness identification literature have universally assumed that witnesses who encounter innocent suspects adopt the same decision criteria as witnesses who encounter guilty suspects (e.g., Clark, 2003; Wixted & Mickes, 2014).

But, based on past research demonstrating that witnesses lower their criteria when match-to-memory is weak, one would expect witnesses viewing innocent suspects to use more lenient criteria compared to witnesses viewing guilty suspects (Smith et al., 2018; Smith et al., 2019). To make a correct decision in a culprit-absent case, a witness must be able to attribute her weak recognition experience to the culprit's absence, and not to other factors such as poor encoding conditions. But, there is some evidence that witnesses who view innocent suspects misattribute their weak recognition experience, not to the absence of the culprit, but to factors such as having had a poor view. Bradfield, Wells, and Olson (2002) gave all participants the same initial encoding experience and then randomly assigned them to a target-present or target-absent lineup condition. When compared to participants in the target-present condition, participants who saw a target-absent lineup indicated that they were given a worse view and that they had paid less attention to the simulated-crime video.

Clearly, the balance of the evidence suggests that eyewitness-participants who are randomly assigned to culprit-absent conditions tend to adopt more lenient decision criteria than do eyewitness-participants who are randomly to culprit-present conditions (potentially because they misattribute their weak memory experience to factors other than the absence of the culprit). The propensity for witnesses who encounter target-absent identification procedures to use more lenient decision criteria compared to witnesses who encounter target-present identification procedures is called present-absent criteria discrepancy (Smith et al., 2018). Present-absent criteria discrepancy can be likened to a special case of the strength-based mirror effect, where the mirror effect results not from poor view or from degraded recognition, but from assigning people to either a strong (culprit present) or weak (culprit absent) recognition condition. Panel B of Figure 1 illustrates Smith and colleagues' (2018) hypothetical decision criterion placements for culprit-present and culprit-absent procedures.

The problem with present-absent criteria discrepancy is that it seriously undermines identification performance (Smith et al., 2018). Using a lower decision criterion for culprit-absent than for culprit-present conditions will produce worse performance than will using the same criterion for both cases, irrespective of whether that criterion is strict or lenient. Consider the distributions in Figure 1. Discriminability, represented the degree of overlap between the distributions, is the same in both panels. The black vertical line in Panel A represents a witness applying the same decision criterion to both culprit present and absent cases (if the suspect's match value exceeds the black line, the witness makes an identification). If, however, witnesses use a different criterion in culprit-absent cases, there needs to be a second more lenient criterion, as shown by the grey vertical line in Panel B of Figure 1. The outcome of using this criterion when the culprit is absent is an increase in false identifications without an accompanying increase in culprit identifications. Hence, present-absent criteria discrepancy undermines performance because it leads to an increase in innocent-suspect identifications without a concomitant increase in culprit identifications.

The effect of Smith et al.'s (2018) additional-opportunities instruction (AOI) provides support for present-absent criteria discrepancy. The authors attempted to create an instruction that would cause witnesses to raise their decision criteria when the culprit was absent but not when the culprit was present. If present-absent criteria discrepancy is correct and witnesses tend to set a lower criterion when the culprit is absent, the result of such an instruction would be an improved tradeoff between culprit and innocent-suspect identifications. They devised an instruction that targets witnesses who are more likely to have a weaker match-to-memory experience but will identify anyways out of fear that they might not have another opportunity to get a hit. Telling witnesses that they may have an opportunity to make an identification later on did in fact decrease innocent-suspect identifications to a greater extent than culprit identifications (Smith et al., 2018).

Within the traditional SDT framework, the only plausible explanation for the effect of the AOI is that it somehow increased discriminability, or the distance between the culprit and innocent-suspect distributions. This is because SDT assumes that people do not have different decision criteria for signal and noise trials. According to SDT, an instruction that affects the criteria for signal trials must have an equal effect for noise trials. If that were the case, Smith et al. (2018) would have observed an equal reduction in culprit identifications and innocent-suspect identifications. If the AOI impacted discriminability, then it could have improved the tradeoff between culprit and innocent-suspect identifications. However, SDT itself asserts that conservative instructions should only affect criteria placement and not discriminability (Green & Swets, 1966; Macmillan & Creelman, 2005). Furthermore, there is no recognized psychological process by which the instruction could decrease the distance between the culprit and innocent-suspect distribution. Thus, the impression that the AOI improved discriminability can be better explained by present-absent criteria discrepancy.

Overview of the Current Study

There are two lingering questions ensuing from the research by Weber and Perfect (2012) on the efficacy of an opt-out option that my proposed thesis attempted to answer. First, from a theoretical perspective, it was surprising that participants were equally likely to use the don't-know response under both short and long retention intervals (Weber & Perfect, 2012). This is in spite of the fact that length of delay was strongly associated with memory performance (Short-delay $d' = 0.80$; Long-delay $d' = 0.11$). This suggests that participants were not meta-cognitively aware of the impact delay had on memory performance. Although a three-week delay was apparently sufficient to impair memory performance, it might be the case that there is something about a delay that causes people to be insensitive to that fact.

Indeed, others have found that retention interval leads witnesses to a state of overconfidence (Sauer, Brewer, Zweck, & Weber, 2010). To be clear, longer retention intervals do not lead witnesses to become more confident in their identification decisions. Rather, this overconfidence results from the fact that even though longer retention intervals lead to less accurate identification performance, participants do not sufficiently lower their confidence. Perhaps witness-participants would have used the “don’t-know” option more frequently had they been more metacognitively aware of the negative impact of the delay on their memories. Indeed, Koriat, Bjork, Sheffer, and Bar (2004) found that participants who expected to receive a memory test a week or a year later predicted that they would recall as much information as those who expected to be tested immediately. When participants are sensitive to a factor that is likely to influence their memory, then they should be more likely to use an opt-out option. Perhaps a memory manipulation that is implemented at the time the witness-participant views the criminal event would be more salient.

Second, the authors claimed that a primary benefit of the don’t-know option was that it eliminates low-confidence responses (which are likely to be inaccurate). Although this claim fits with intuition, unfortunately the authors did not analyze eyewitness confidence, so the relationship between eyewitness confidence and the opt-out option could not be established in their experiment. My intention with the proposed experiment was to look specifically at how the opt-out option interacts with confidence by conducting confidence-accuracy calibration analysis and through examining how the not-sure option impacts the ROC curve. My expectation was that the not-sure option will decrease the tendency for witness-participants to make low-confidence decisions. Because low-confidence decisions are more likely to be incorrect than are high-confidence decisions, my expectation was that this would lead to a better tradeoff between culprit and innocent-suspect identifications.

The current study examined the effect of an opt-out option across two different levels of memory strength. The strength-based mirror effect provides evidence to suspect that when memory is weakened by a degraded witnessing experience, eyewitnesses are less likely to make culprit identifications and most at risk of making innocent-suspect identifications. Therefore, I expected an opt-out option to have a greater impact on the tradeoff between culprit and innocent-suspect identifications when memory was relatively weak compared to when memory was relatively strong. Because eyewitness showups do not include fillers and only a single suspect, they offer the clearest and most direct test of how an opt-out option might impact identification performance.

Participants were randomly assigned to either a target-present or target-absent showup. Both scenarios are possibilities in the real world, so it was important to examine how the showup procedure is affected under both conditions. Specifically, does the not-sure option decrease innocent-suspect identifications to a greater extent than culprit identifications and was this effect moderated by the quality of viewing conditions (i.e., witness memory) or eyewitness confidence?

As an alternative to manipulating the retention interval between encoding the culprit and being presented with a showup procedure (Weber and Perfect, 2012), I manipulated quality of view. Surprisingly, Weber and Perfect (2012) did not find that participants used the opt-out option more frequently when memory was weak (in the long retention interval) compared to when memory was strong (in the short retention interval). One possible explanation is that participants were insensitive to the impact that the delay had on their memories. Quality of view manipulations such as decreased lighting, increased distance, and shorter exposure durations all have the same effect – weaker memory. I reasoned that participants might be more sensitive to the impact that the poor-quality view had on their memories and therefore would be more likely to use the opt-out option if given a poor view relative to if given a strong view. In other words, it would provide a better test of the hypothesis that the not-sure option would be more effective under relatively weak memory conditions.

To degrade viewing conditions, I overexposed the image of the simulated-crime video. At first, I experimented with darkening the video to mimic nighttime conditions. However, my own tests of this method revealed that it was inconsistent across different lighting environments and because of the online nature of the experiment, I would be not able to control this for the participants. Accordingly, I created a quality of view manipulation that would be less affected by variations in the lighting environment. And, the goal of my thesis was not to examine some specific way of weakening memory, but to examine the utility of a not-sure option when memory was weakened through a quality of view manipulation. So, the overexposed video served this purpose to the extent that it greatly decreased the visibility of the features of the culprit's face.

Additionally, I examined whether the not-sure option decreased low-confidence decisions to a greater extent than high-confidence decisions. In the absence of a not-sure option, participants who are not sure would likely make a low-confidence decision. Hence, the not-sure option should primarily reduce low-confidence decisions. But, because Weber and Perfect (2012) did not examine confidence, this hypothesis has not yet been formally tested. The most straightforward way to determine this was to compare the average confidence of the not-sure option condition to the average confidence of the control condition. If the not-sure option primarily reduces low-confidence identification decisions than those participants who received the not-sure option and did not elect to use it should be more confident, on average, than participants who did not receive the option to respond "not sure".

Finally, for those participants that responded “not sure”, I asked them to indicate whether they would have made an identification or rejected the showup had they not opted out. I also asked them to indicate their anticipated confidence for that hypothetical decision. Analysis of those responses should provide ancillary information about what kind of witnesses the not-sure option is siphoning out– would they have in fact made an incorrect decision, and are they low-confidence witnesses as predicted? This analysis was exploratory and is included in the supplemental materials.

Hypotheses

1) Participants will use the not-sure option more frequently after receiving a poor view of the culprit compared to after receiving a clear view of the culprit.

2) Participants who view a culprit-absent showup will use the not-sure option more frequently than participants who view a culprit-present showup.

3) The not-sure option will decrease false identifications to a greater extent than correct identifications. In other words, relative to the control condition, the not-sure option will decrease affirmative identification decisions to a greater extent when the culprit is absent compared to when the culprit is present.

4) The effect size for the interaction predicted in hypothesis 3 will be moderated by memory strength. Specifically, the not-sure option will decrease false identifications to a greater extent than correct identifications, and the effect will be greater for participants who viewed the poor-quality video.

5) The not-sure option will reduce low-confidence decisions to a greater extent than high-confidence decisions.

6) The not-sure option will improve the confidence-accuracy relationship under degraded-viewing conditions to a greater extent than under clear-viewing conditions.

Methods

Participants

Participants (N = 2139) were recruited from Amazon's Mechanical Turk system. This sample consists of individuals who are over 18 and reside in the United States. See Appendix A for the recruitment notice. They were required to have an 80% or higher Human Intelligence Task approval rating to ensure that they were likely to be good quality participants. Participants attempting to complete the study with a mobile phone were screened out and asked to complete the study using either a tablet or computer. I expected that approximately 10% of participants would produce unusable data due to technical difficulties, not completing the study, and other factors, so I collected data until there were approximately 2000 usable data points and then halted data collection. Participants' data were deemed useable if they (1) passed an attention check by responding to a multiple-choice question about the video correctly, and (2) indicated that they did not experience technical difficulties viewing the video or showup photo. Forty-eight participants were excluded due to failure to pass the attention check, and 88 were excluded due to technical difficulties. The final number of participants was $N = 2003$. The average participant age was 36.90 ($SD = 11.67$) and was 57.86% female. 72.94% of participants indicated that they were Caucasian.

Design

My experiment conformed to a 2 (quality of view: clear or degraded) x 2 (not-sure option: yes or no) x 2 (target: present or absent) between-participants design. I used two target individuals (both males) and two sets of filler photos to achieve some minimal degree of stimulus sampling (Wells & Windschitl, 1999). Each set of filler photos consisted of three similar-looking innocent replacements for the corresponding target. I selected the innocent-replacements from a stimulus set that were created prior to this experiment. Each of the fillers matched the description of the target. The photos were front-facing and showed the individual from the neck up. No clothing was visible. As individual analyses of the targets replicated the patterns of inferential tests that were found when conducted together, and because it would not change any conclusions, the targets were collapsed across for the final analysis. The dependent variables of interest were identification decision (identify suspect, reject suspect, or opt out) and eyewitness level of confidence.

Materials

Informed consent form. Participants were required to give informed consent before participating in the study. See Appendix B.

Crime video. The simulated-crime video used in the study depicted a man facing the camera holding a gun. The gun was pointed towards the camera, and it obscured the man's face slightly. The man's face was in view for approximately 5 seconds. The degraded-view condition used exactly the same video, with the brightness and contrast altered to approximate viewing the crime in overexposed lighting. The specific level of exposure needed was determined by pilot testing for a different experiment that used the same stimuli (Jalava, Smith, & Wells, under review). I proceeded with the full experiment when I found a level of exposure that impaired performance while leaving some ability to discriminate between the culprit and innocent-suspect intact.

Demographic information. After viewing the video, participants were asked to provide their age, gender, and ethnicity.

Anagrams task. As a distractor task between viewing the video and the showup procedure, participants solved a series of anagrams for 10 minutes.

Showup task. Participants viewed a single photo displaying either the target from the video or an innocent suspect. I asked participants if they believed that the man in the photo was the man from the video. All of the participants received the generic instruction “In a moment you will see a photo of a man who may or may not be the culprit from the video you saw at the beginning of the experimental session. If you believe that the man in the photo is the man from the video, you should identify him (press “YES”). If you do not believe the man in the photo is the man from the video, you should reject the photo (press “NO”).” Those participants in the not-sure option condition were given the additional instruction “If you are not sure whether the man in the photo is the man from the video, you should indicate that you are not sure (press “NOT SURE”).” There was also an accompanying option labeled “NOT SURE” for those participants. All participants had to check a box indicating that they had read and understood the instructions before moving on.

Procedure

After obtaining ethics approval for the experiment, a recruitment notice (Appendix A) was posted on Amazon’s Mechanical Turk (MTurk). Participants who signed up for the study were given an anonymous link to Qualtrics, where the study was hosted. They digitally signed an informed consent form (Appendix B) and then were randomly assigned to watch either a clear or degraded version of video of a staged crime (see Appendix C for a description). They were told to pay attention to the video as their memory may be tested later. Following the video, participants filled out demographic information, and then solved a series of anagrams for 10 minutes. This served as a filler task before the identification procedure.

After completing the filler task, participants were instructed that they would be shown a single photograph and asked to indicate whether they believe the man displayed in the photo is the same man from the video. All photos were front facing with only the face and neck visible. Participants were randomly assigned to either a target-present or target-absent condition. In the target-present condition, the man in the photo was the same as the man from the video. In the target-absent condition, the man in the photo was someone the participant has never seen before but who generally resembles the target person. All participants were given the standard admonition ("The man in the photo may or may not be the man in the video"). They were randomly assigned to either the control or not-sure option condition. In the control condition, participants pressed a button labelled either YES or NO to make their identification decision. In the not-sure option condition, in addition to YES or NO, they had the option to press a button labelled NOT SURE. After making their decision, I asked participants to report their level of confidence from 0-100 in 10-pt increments on a sliding scale (except those participants who responded NOT SURE). For those that responded NOT SURE, I asked them to indicate which response option they would have selected had they not had the option to respond NOT SURE, and their hypothetical confidence in that decision.

Following the identification task, participants completed a retrospective questionnaire (Appendix D) about their experience, including a manipulation check. Finally, participants were asked to read a debriefing form (Appendix E) which explained more about the research goals and gave contact information for resources if they had questions or concerns about the study. The study took, on average, 13.06 minutes to complete. I paid participants $1.00 USD for completing the study.

Results

I calculated what percentage of participants elected to opt out in order to verify that participants used the response option. Out of the 1004 participants who were in the not-sure option condition, 307 opted-out (30.4%). So, around one third of participants opted-out of making an identification decision when given the opportunity to do so.

Logistic Regression

I used a series of hierarchical binary logistic regression analyses to test each of my hypotheses. To begin, I tested my first two hypotheses that the use of the not-sure option would be dependent on the quality of viewing conditions and on the presence of the culprit. Next, I tested my third hypothesis that the not-sure option would produce a superior tradeoff between culprit and innocent-suspect identifications than the control condition. Finally, I tested my fourth hypothesis that the tradeoff effect would be larger in magnitude when viewing conditions were poor.

Hypotheses 1 & 2. I used hierarchical binary logistic regression models to determine if quality of viewing conditions and culprit presence produced different frequencies of not-sure option responses. To test the hypotheses that the not-sure option would be used most frequently when participants had a poor view of the culprit, and when they were faced with a culprit-absent showup (Hypotheses 1 and 2), I conducted a 2 (quality of view: good vs poor) x 2 (culprit: present or absent) hierarchical binary logistic regression (dependent variable: response type, 'not sure' or other) for those participants who had the option to respond 'not sure'.

Participants were significantly more likely to use the not-sure option when given a degraded view of the culprit, $b = 1.05$, $SE = 0.15$, Wald's $\chi^2(1) = 49.68$, $p < .001$, $e^b = 2.86$, (95% CI [2.14, 3.83]), providing support for Hypothesis 1. However, participants were more likely to use the not-sure option when faced with a culprit-present showup, $b = .65$, $SE = 0.15$, Wald's $\chi^2(1) = 19.11$, $p < .001$, $e^b = 1.91$, (95% CI [1.43, 2.56]), contrary to what I predicted in Hypothesis 2. The effect was not qualified by a significant interaction between quality of view and culprit presence, b = -0.12, SE = 0.30, Wald's $\chi^2(1)$ = .16, p = .70, e^b = 0.89, (95% CI [0.49, 1.61]). Table 3 displays the frequency of responding for each response option in percentages for all conditions.

Table 3

Response Percentages as a Function of View Quality, Response Options, and Culprit Presence

		View Quality			
		Clear		Degraded	
Response Option		Culprit Present	Culprit Absent	Culprit Present	Culprit Absent
Control	Yes	58% (143)	15% (36)	31% (77)	23% (58)
	No	42% (102)	85% (208)	69% (171)	77% (199)
Not Sure	Yes	48% (128)	14% (37)	15% (37)	14% (33)
	No	29% (77)	70% (182)	40% (100)	46% (108)
	Not Sure	23% (61)	16% (42)	45% (111)	40% (93)

Note. Values in parentheses are the number of participants in that condition.

Hypotheses 3 & 4. To examine whether the not-sure option produced a better tradeoff between culprit and innocent suspect identifications (Hypothesis 3), I fit a 2 (quality of view: good vs. poor) x 2 (culprit: present vs. absent) x 2 (not-sure option: yes, no) hierarchical binary logistic regression model including all main effects and higher-order interaction terms. Because the intention of this analysis was to examine the impact of the predictors on suspect identifications, I coded suspect identifications as case outcomes (suspect identifications = 1) and all other responses as control outcomes ('not sure' and rejection responses = 0). Critically, I predicted a significant two-way interaction between culprit presence and the not-sure option that resulted from the not-sure option decreasing innocent-suspect identifications to a greater extent than culprit identifications. I also used this model to test my fourth hypothesis that the interaction between culprit presence and the not-sure option would have a larger effect when viewing conditions were relatively poor. The three-way interaction in this model directly tests this hypothesis.

On the first block, participants were significantly less likely to identify the suspect when viewing conditions were degraded, b = -0.73, SE = 0.11, Wald's $\chi^2(1)$ = 47.13, p < .001, e^b = 0.48, (95% CI [0.39, 0.59). They were also less likely to identify the suspect if they had the option to respond 'not sure', b = -0.50, SE = 0.11, Wald's $\chi^2(1)$ = 21.83, p < .001, e^b = 0.61, (95% CI [0.50, 0.75]).

On the second block, there were two significant two-way interactions. The interaction between quality of view and response-option condition was significant, b = -0.53, SE = 0.22, Wald's $\chi^2(1)$ = 5.69, p = .02, e^b = 0.59, (95% CI [0.38, 0.91]). Participants in the not-sure option condition were equally likely to identify the suspect when viewing conditions were clear, b = -0.24, SE = 0.13, Wald's $\chi^2(1)$ = 3.17, p = .08, e^b = 0.79, (95% CI [0.61, 1.02]), but less likely to identify the suspect when viewing conditions were degraded, b = -0.76, SE = 0.16, Wald's $\chi^2(1)$ = 21.78, p < .001, e^b = 0.47, (95% CI [0.34, 0.64]).

The interaction between quality of view and culprit presence was also significant, b = -1.66, SE = 0.22, Wald's $\chi^2(1)$ = 54.59, p < .001, e^b = 0.19, (95% CI [0.12, 0.30]). Relative to participants in the clear-view condition, participants in the degraded-view condition were less likely to identify the culprit, b = -1.33, SE = 0.14, Wald's $\chi^2(1)$ = 92.01, p < .001, e^b = 0.26, (95% CI [0.20, 0.35]), but equally likely to identify the innocent suspect, b = 0.30, SE = 0.17, Wald's $\chi^2(1)$ = 3.00, p = .08, e^b = 1.35, (95% CI [0.96, 1.89]). Thus, the viewing condition manipulation did not produce a clear-cut strength-based mirror effect, as degraded-viewing conditions did not lead to a statistically significant increase in innocent-suspect identifications. However, the effect was certainly trending towards significance with a marginal *p*-value of .08. So, at least qualitatively, these results fit with the well-grounded strength-based mirror effect.

Of relevance to my third hypothesis, the two-way interaction between culprit presence and response option did not reach statistical significance, $b = -0.37$, $SE = 0.23$, Wald's $\chi^2(1) = 2.66$, $p = .10$, $e^b = 0.69$, (95% CI [0.59, 1.08]). I wanted to probe the interaction to see if the not-sure option actually made the tradeoff between culprit and innocent identifications worse. Relative to participants in the control response-option condition, participants who were in the not-sure option condition were less likely to identify the innocent suspect, although the effect was only on the cusp of significance, $b = -0.34$, $SE = 0.17$, Wald's $\chi^2(1) = 3.84$, $p = .05$, $e^b = 0.71$, (95% CI [0.51, 1.00]), and also less likely to identify the culprit, $b = -0.53$, $SE = 0.13$, Wald's $\chi^2(1) = 16.60$, $p < .001$, $e^b = 0.59$, (95% CI [0.45, 0.76]). My third hypothesis was not supported. The not-sure option reduced choosing both when the culprit was present and absent.

On the fourth block, the three-way interaction term between quality of view, culprit presence and the not-sure option was not significant, $b = -0.003$, $SE = 0.45$, Wald's $\chi^2(1) = 0.00004$, $p = .995$, $e^b = 0.997$, (95% CI [0.41, 2.42]). Thus, my fourth hypothesis that any tradeoff benefit produced by the not-sure option would be greater when viewing conditions are relatively poor was not supported.

ROC Analysis

I augmented my binary logistic regression analyses with Receiver Operator Characteristic (ROC) analysis. All ROC curves were constructed using the pROC statistical package (Robin et al., 2011). Specifically, I compared all four ROC curves that were produced by crossing the not-sure manipulation with the quality of view manipulation using standard Area Under the Curve measures (AUC). In addition, I compared the 'not-sure' ROC curve to the control ROC curve (collapsing across quality of view). I also examined the portions of the ROC curves that included only those eyewitnesses who affirmatively identified a suspect with a novel Deviation from Perfect Performance measure (Smith, Lampinen, Wells, Smalarz, & Mackovichova, 2019), which can be found in the supplemental materials. Note that these analyses are largely replicating what was tested with my logistic regression analyses, but different groups of researchers have different preferences, so I analyzed the data in multiple ways.

Because I used a showup procedure, I was able to construct ROC curves that extended the full range of the ROC space (e.g., Smith et al., 2018). ROC curves for lineups can never extend the full range of the ROC space because fillers prevent the innocent-suspect identification rate from ever reaching 1.00. Due to the presence of fillers, researchers wishing to use ROC analysis to analyze lineup data have been forced to examine only partial ROC curves that are restricted to only the far left of the ROC space. When one procedure decreases innocent-suspect identifications and culprit identifications to a greater extent than another procedure (i.e., the difference between the two procedures is defined by a tradeoff), pAUC will not always inform on which procedure is superior (Lampinen, Smith, & Wells, 2019; Smith et al., 2019). So, it makes more sense to use an alternate, utility-based measure to assess which lineup procedure is superior, like Deviation from Perfect Performance (DPP) analysis (Smith et al., 2019). However, because I used a showup procedure, I can focus on the full AUC as is routine in the basic recognition literature (e.g., Verde & Rotello, 2007). I made use of the DPP analysis to examine the ROC curves restricted to choosers only. Because some eyewitness researchers focus only on choosers for their analyses, I did this as well, but using a measure that actually informs on which procedure has superior utility. This is something that pAUC analysis cannot do (Lampinen et al. 2019; Smith et al., 2019). This analysis is included in the supplementary materials as the results were largely consistent with my full AUC analysis.

To create curves that extend the full ROC space, I plotted culprit identifications against innocent-suspect identifications at cumulating levels of confidence (in 10-pt increments). The leftmost point on the curve represents the culprit and innocent-suspect identifications that were made with 100% confidence. Moving from left to right, the second point on the curve represents the culprit and innocent-suspect identifications made with 90% confidence plus all IDs made with 100% confidence. Construction of the curves continues in this fashion until all suspect identifications, correct or incorrect, and even those witnesses who made an identification with 0% confidence, are reflected in a single point on the ROC curve. This is where the ROC curve terminates with lineup procedures (due to the presence of fillers). With showups, the ROC curve can continue to be extended by plotting rejection decisions in a similar fashion to identification decisions. The next point represents all choosers plus those who rejected the showup with 0% confidence. The point after that includes all choosers, those who rejected the showup with 0% confidence, and those who rejected the showup with 10% confidence. Again, points are plotted in this fashion until all participants are reflected in a single point in the top right corner of the ROC space (e.g., Smith et al., 2018).

A critical question that has not been addressed in the eyewitness literature is how to include participants who used the not-sure option in the ROC curve. Theoretically, participants who select the not-sure option are saying that they are neither confident enough to make an affirmative identification nor are they confident enough to make a rejection decision. Hence, I included these participants at the point immediately following identifications made with 0%

confidence and immediately preceding rejections made with 0% confidence. This is consistent with how not-sure decisions are incorporated into signal-detection-based models of eyewitness identification (e.g., Clark, 2003).

After constructing the ROC curves, I conducted pairwise comparisons among them. Figure 2 displays ROC curves for the following conditions: control condition with a clear view, not-sure option with a clear view, control condition with a degraded view, and not-sure option with a degraded view. When viewing conditions were clear, there was no significant difference between the control (AUC = .77) and not-sure option (AUC = .77) conditions, $D = .07$, $p = .95$. Likewise, when viewing conditions were degraded, there was no significant difference between the control (AUC = .57) and not-sure option (AUC = .52) conditions, $D = 1.24$, $p = .21$. Better viewing conditions resulted in a better tradeoff between culprit and innocent-suspect identifications, both for the control condition (AUC = .77 vs. AUC = .57), $D = 6.21$, $p < .001$, and the not-sure option condition (AUC = .77 vs. AUC = .52), $D = 7.64$, $p < .001$. So, consistent with the logistic regression analysis, clear viewing conditions led to a better tradeoff between culprit and innocent-suspect identifications than did degraded viewing conditions, but the not-sure option did not lead to a better tradeoff than did the control condition.

Figure 3 displays the ROC curves for the response option conditions collapsing across quality of view. A similar pattern was found for the pairwise comparison of the AUC for these curves. There was no significant difference between the control (AUC = .68) and not-sure option (AUC = .66) conditions, $D = 0.76$, $p = .45$.

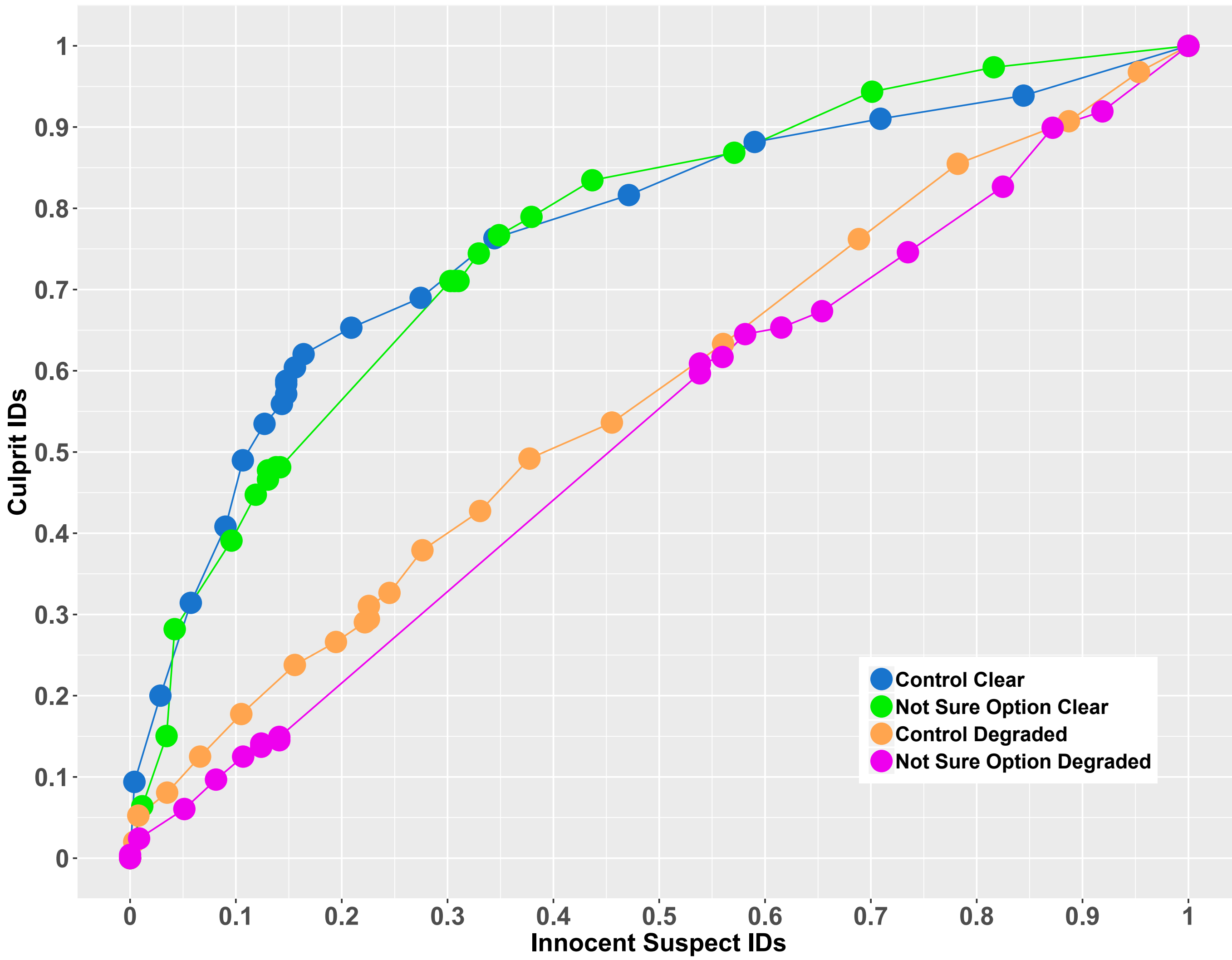

Figure 2. A graph displaying the ROC curves across all four conditions.

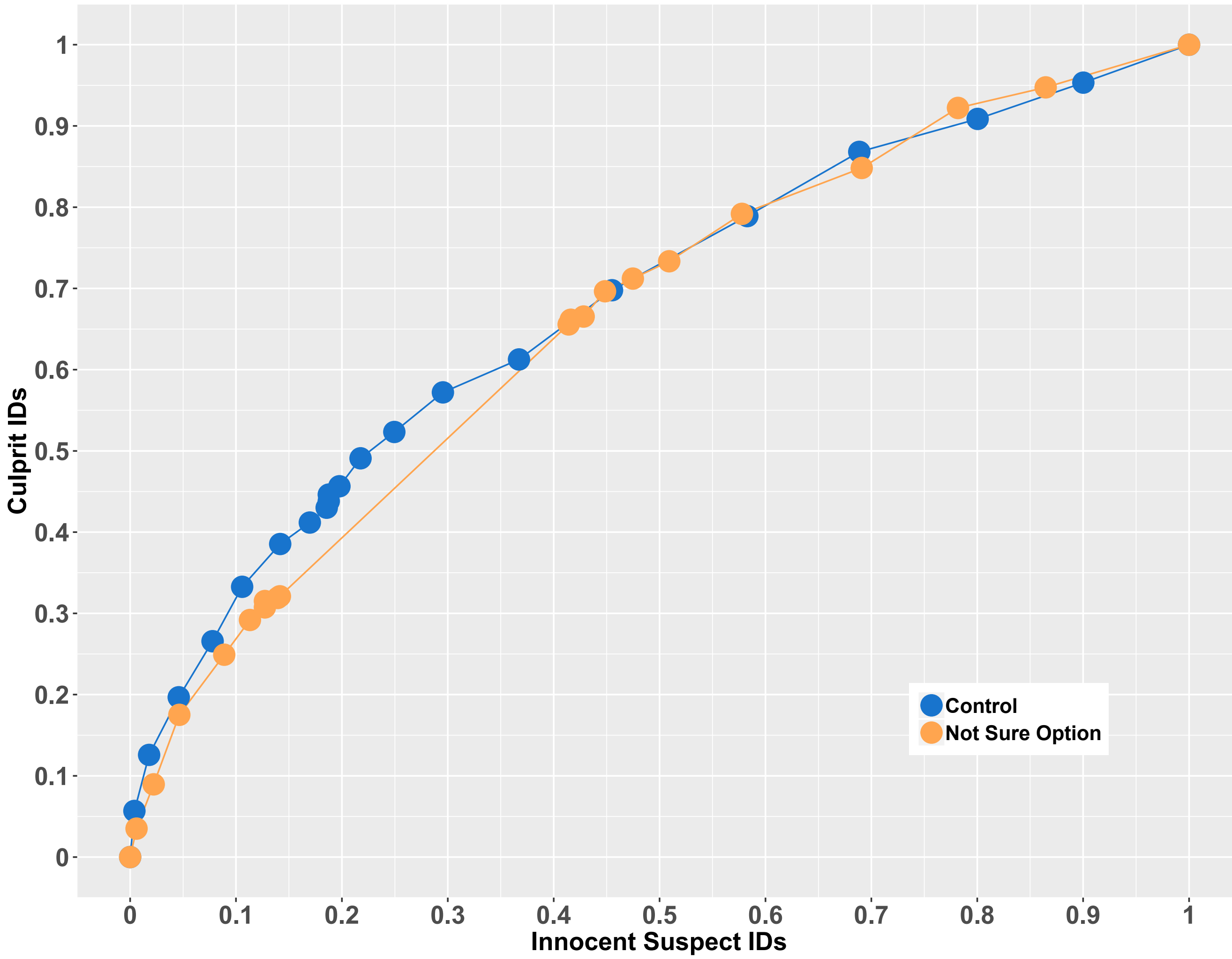

Figure 3. A graph displaying the ROC curves for the response option conditions, collapsing across quality of view.

Confidence-Accuracy Analysis

Next, I tested my fifth hypothesis that the not-sure option would decrease low-confidence decisions to a greater extent than high-confidence decisions by conducting an independent-samples t-test on the mean confidence for those who had the option to respond ‘not sure’ and those who did not. This analysis did not include participants who chose to opt out, as they did not provide an initial confidence judgment. Participants in the control condition reported significantly lower confidence in their decisions (M = 62.94, SD = 24.56) than those who had the option to respond ‘not sure’, (M = 72.78, SD = 20.38), $t(1694) = -8.71$, $p < .001$, d = 0.44, (95% CI around the difference [-12.10, -7.62]). Thus, my fifth hypothesis was supported.

Finally, I tested my sixth hypothesis that the not-sure option would improve the confidence-accuracy relationship to a greater extent under poor-viewing conditions than good-viewing conditions by regressing culprit presence on choosing, quality of view, response-option condition, confidence, and all higher-order interactions. The three-way interaction in the model provides a direct test of this hypothesis. This analysis also did not include participants who chose to opt out, as they did not provide an initial confidence judgment.

On the first block, the probability that the suspect was guilty decreased by 0.4% for every 10-pt increase in confidence, although the effect was only marginally significant, b = -0.004, SE = 0.002, Wald’s $\chi^2(1) = 3.84$, $p = .05$, $e^b = 0.996$, (95% CI [0.99, 1.00]). However, the main effect of confidence was qualified by a significant two-way interaction between confidence and choosing, b = 0.03, SE = 0.01, Wald’s $\chi^2(1) = 23.61$, $p < .001$, $e^b = 1.03$, (95% CI [1.02, 1.04]). When the participant identified the suspect, the odds that the suspect was guilty increased by 1.02 times for every 10-pt increase in confidence, b = 0.02, SE = 0.01, Wald’s $\chi^2(1) = 21.50$, $p < .001$, $e^b = 1.02$, (95% CI [1.01, 1.03]). When the participant rejected the suspect or responded ‘not sure’, the probability that the suspect was guilty decreased by 1.2% for every 10-pt increase in confidence, b = -0.01, SE = 0.002, Wald’s $\chi^2(1) = 25.25$, $p < .001$, $e^b = 0.988$, (95% CI [0.98, 0.99]). Thus, there was a positive relationship between confidence and suspect guilt for those who made an identification but a negative relationship for those who did not make an identification. This is consistent with previous literature showing a stronger relationship between confidence and accuracy for choosers compared to non-choosers (Sporer et al., 1995). The two-way interaction between confidence and response option condition was not significant, b = 0.003, SE = 0.01, Wald's $\chi^2(1) = 0.33$, $p = .57$, $e^b = 1.00$, (95% CI [0.99, 1.01]).

On the third block, the three-way interaction between quality of view, choosing, and confidence was not significant, $b = 0.01$, $SE = 0.01$, Wald's $\chi^2(1) = 1.17$, $p = .28$, $e^b = 1.01$, (95% CI [0.99, 1.04]). The three-way interaction between quality of view, response option condition, and confidence was also not significant, $b = 0.01$, $SE = 0.01$, Wald's $\chi^2(1) = 1.43$, $p = .23$, $e^b = 1.01$, (95% CI [0.99, 1.03]), indicating that the not-sure option did not improve the confidence-accuracy relationship to a greater extent under poor-viewing conditions than under good-viewing conditions. Thus, my sixth hypothesis was not supported.

On the fourth block, the four-way hypothesis was significant, $b = -0.05$, $SE = 0.02$, Wald's $\chi^2(1) = 4.82$, $p < .05$, $e^b = 0.96$, (95% CI [0.92, 0.99]). To break this interaction down, I examined the simple slope of confidence at every combination of the predictors. When the participant was in the control response-option condition, given a clear view, and identified the suspect, the odds that the suspect was guilty increased by 1.02 times for every 10-pt increase in confidence, although the effect was only marginally significant, $b = 0.02$, $SE = 0.01$, Wald's $\chi^2(1) = 3.19$, $p = .07$, $e^b = 1.02$, (95% CI [1.00, 1.04]). When the participant was in the control response-option condition, given a clear view, but did *not* identify the suspect, the probability that the suspect was guilty decreased by 1.8% for every 10-pt increase in confidence, $b = -0.02$, $SE = 0.01$, Wald's $\chi^2(1) = 12.73$, $p < .001$, $e^b = 0.982$, (95% CI [0.97, 0.99]).

When the participant was given the option to respond 'not sure', given a clear view, and identified the suspect, the odds that the suspect was guilty increased by 1.03 times for every 10-pt increase in confidence, $b = 0.03$, $SE = 0.01$, Wald's $\chi^2(1) = 6.38$, $p < .05$, $e^b = 1.03$, (95% CI [1.01, 1.05]). When the participant was given the option to respond 'not sure', given a clear view, but did *not* identify the suspect, the probability that the suspect was guilty decreased by 2.7% for every 10-pt increase in confidence, b = -0.03, SE = 0.01, Wald's $\chi^2(1) = 15.99$, $p <$ .001, e^b = 0.97, (95% CI [0.96, 0.99]).

When the participant was in the control response-option condition, given a degraded view, and identified the suspect, confidence did not discriminate between suspect guilt and innocence, b = 0.01, SE = 0.01, Wald's $\chi^2(1)$= 1.68, p = .20, e^b = 1.01, (95% CI [0.99, 1.03]). Nor did confidence discriminate between suspect guilt and innocence when the participant was in the control response-option condition, given a degraded view, but did *not* identify the suspect, b = -0.01, SE = 0.004, Wald's $\chi^2(1) = 1.73$, p = .19, e^b = 1.00, (95% CI [0.99, 1.00]).

When the participant was given the option to respond 'not sure', given a degraded view, and identified the suspect, confidence did not discriminate between suspect guilt and innocence, b = 0.01, SE = 0.01, Wald's $\chi^2(1) = 0.95$, p = .33, e^b = 1.01, (95% CI [0.99, 1.04]). Nor did confidence discriminate between suspect guilt and innocent when the participant was given the option to respond 'not sure', given a degraded view, but did *not* identify the suspect, b = 0.002, SE = 0.01, Wald's $\chi^2(1) = 0.16$, p = .69, e^b = 1.00, (95% CI [0.99, 1.01]). In summary, the four-way interaction revealed that both the magnitude and direction of the confidence-accuracy relationship was dependent on whether the participant made an identification, the quality of viewing conditions, and whether the witness had the option to respond, 'not sure'.

Next, I plotted confidence-accuracy characteristic (CAC) curves for choosers (Figure 4) and calculated the relevant statistics associated with those curves (the calibration statistic, the over/under statistic, and the Normalized Resolution Index), displayed in Table 4. Following previous research, I collapsed the confidence ratings into three bins, (low: < 70%, medium: 70-89%, and high: 90-100%) (Semmler, Dunn, Mickes, & Wixted, 2018). Collapsing into bins also ensures that a larger number of data points are captured in each confidence category, which is important for making stable estimates. The calibration or *C* statistic can range from 0 to 1, with values closer to 0 representing superior calibration. Perfect calibration occurs when 100% of decisions made with 100% confidence are correct, 90% of decisions made with 90% confidence are correct, and so on and so forth. The over/underconfidence (*O/U*) statistic can range from -1 to +1, with values less than 0 representing underconfidence and values greater than 0 representing overconfidence (Brewer & Wells, 2006). Confidence is said to be "underconfident" when the probability that a decision is correct exceeds the reported confidence level, and "overconfident" when the reported confidence level exceeds the probability that a decision is correct. The Normalized Resolution Index (NRI), can range from 0 to 1 and indicates how well confidence discriminated between correct and incorrect decisions (0 = no discrimination, 1 = perfect discrimination) (Baranski & Petrusic, 1994; Sauer et al., 2010). I also plotted the confidence-accuracy characteristic curves for participants who rejected the showup (Figure 5), and calculated the relevant statistics (*C*, *O/U,* NRI), displayed in Table 5.

Visual inspection of the CAC curves for choosers revealed that, in general, participants confidence was well-calibrated with identification accuracy. Because I binned confidence into low, medium, and high levels, as is becoming regular in analysis of the relationship between eyewitness confidence and accuracy (Mickes, 2015; Mickes, Flowe, & Wixted, 2012; Semmler et al., 2018; Wixted et al., 2015), it makes sense to examine calibration at those three levels. Low confidence, in general, was associated with relatively low accuracy. The percent correct averaging across all conditions at the lowest confidence range (<70%) was approximately 60%. Medium confidence, in general, was associated with relatively moderate accuracy. The average percent correct at the medium confidence range (70-89%) was approximately 68%. This average, although falling below the medium confidence range, is still higher than the average percent correct at the low confidence range. And high confidence, in general, was associated with relatively high accuracy. The average percent correct at the high confidence range (90-100%) was approximately 83%. Again, although falling below the high confidence range, this average is higher than the average percent correct at the medium confidence range.

One point of note is that in the high-confidence range, confidence was more indicative of accuracy than at the low or medium confidence range, for all four conditions. This observation is most clearly illustrated in the degraded view conditions. For low and medium confidence – which spans the range of 0-89% confidence – accuracy stayed below 60%. However, in the high confidence range, accuracy reached 82% when collapsing across response-option condition. This is consistent with the idea that confidence correctly discriminates between accurate and inaccurate suspect identifications (Brewer & Wells, 2006). However, 82% accuracy falls quite far below the confidence reported by participants who were in the high-confidence range, which encapsulates confidence of 90-100%. And, it is lower than what Wixted and Wells (2017) meant by 'high confidence implies high accuracy', which implied an identification accuracy rate of over 90%.

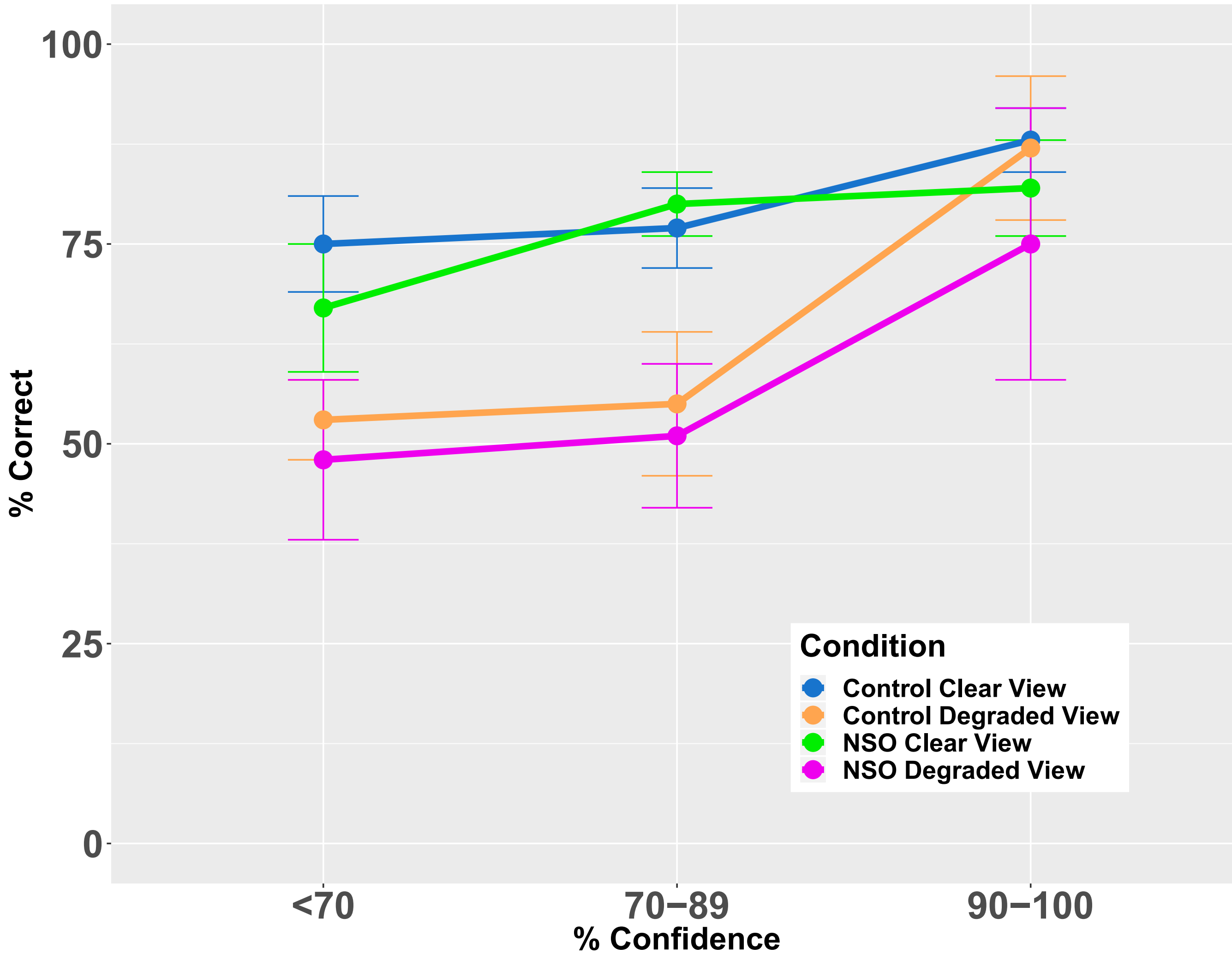

Figure 4. A graph displaying the CAC curves for choosers across all four conditions.

The observed differences among the statistics displayed in Table 4 regarding calibration, over/underconfidence, and the Normalized Resolution Index for choosers seem to be sporadic rather than systematic. For example, the not-sure option condition seems to show some significant overconfidence, but only for those who had a degraded view. The control condition shows a small deviation from perfect calibration, but only for those who had a clear view. Furthermore, the effects obtained were small, and the 95% confidence intervals for each condition's set of statistics show quite a bit of overlap with the 95% confidence intervals for the other conditions. The differences do not seem to fit an intuitive and consistent pattern, and they might be attributable simply to chance. So, rather than conducting a series of pairwise comparisons among the statistics, I will focus on describing macro-level patterns.

The *C* statistic is close to zero (ranging between .01 and .03) across all conditions, indicating good calibration. There is also not a great degree of over or underconfidence for any particular condition, save for the not-sure option condition with a degraded view. This overconfidence can be explained by inspection of the CAC curve for that condition. Identification accuracy was around 50% for both the low and medium confidence ranges, which reflects all identifications made with 0-89% confidence. A confidence judgment of over 50% given by a participant who was in the not-sure option, degraded-view condition was likely to be overconfident, given the average accuracy rate for that condition. The NRI ranges from .02-.07 across all conditions and all confidence intervals contain zero. Thus, no particular condition stands out as having meaningfully better discrimination than any other condition. In general, discrimination appears to be quite poor, as values closer to 1 indicate superior discrimination.

Table 4

Calibration Statistics by Condition for Choosers

	View Quality			
	Clear		Degraded	
Statistic	Control	Not Sure Option	Control	Not Sure Option
C	.02	.01	.01	.03
95% CI	.003, .04	-.003, .02	-.01, .03	-.01, .07
O/U	-.07	-.02	.01	.14
95% CI	-.13, -.01	-.08, .04	-.08, .09	.02, .26
NRI	.02	.02	.07	.04
95% CI	-.02, .03	-.05, .04	-.03, .18	-.09, .17

Note. C = calibration statistic, O/U = over/underconfidence, NRI = normalized resolution index, 95% CI = 95% confidence intervals.

Historically, research has shown a limited relationship between confidence and accuracy for those who make rejections from identification procedures (Brewer & Wells, 2006; Sporer et al., 1995). I predicted that the presence of a not-sure option would improve the relationship for non-choosers as it would parse out those witnesses who reject the procedure because they think the culprit is not present, and those witnesses who reject the procedure because they are uncertain and do not want to make a false identification, (Smalarz & Wells, 2013). Inspection of the CAC curves for non-choosers does not seem to provide much support for this hypothesis. However, confidence and accuracy seem to be generally well calibrated, at least for the clear view conditions.

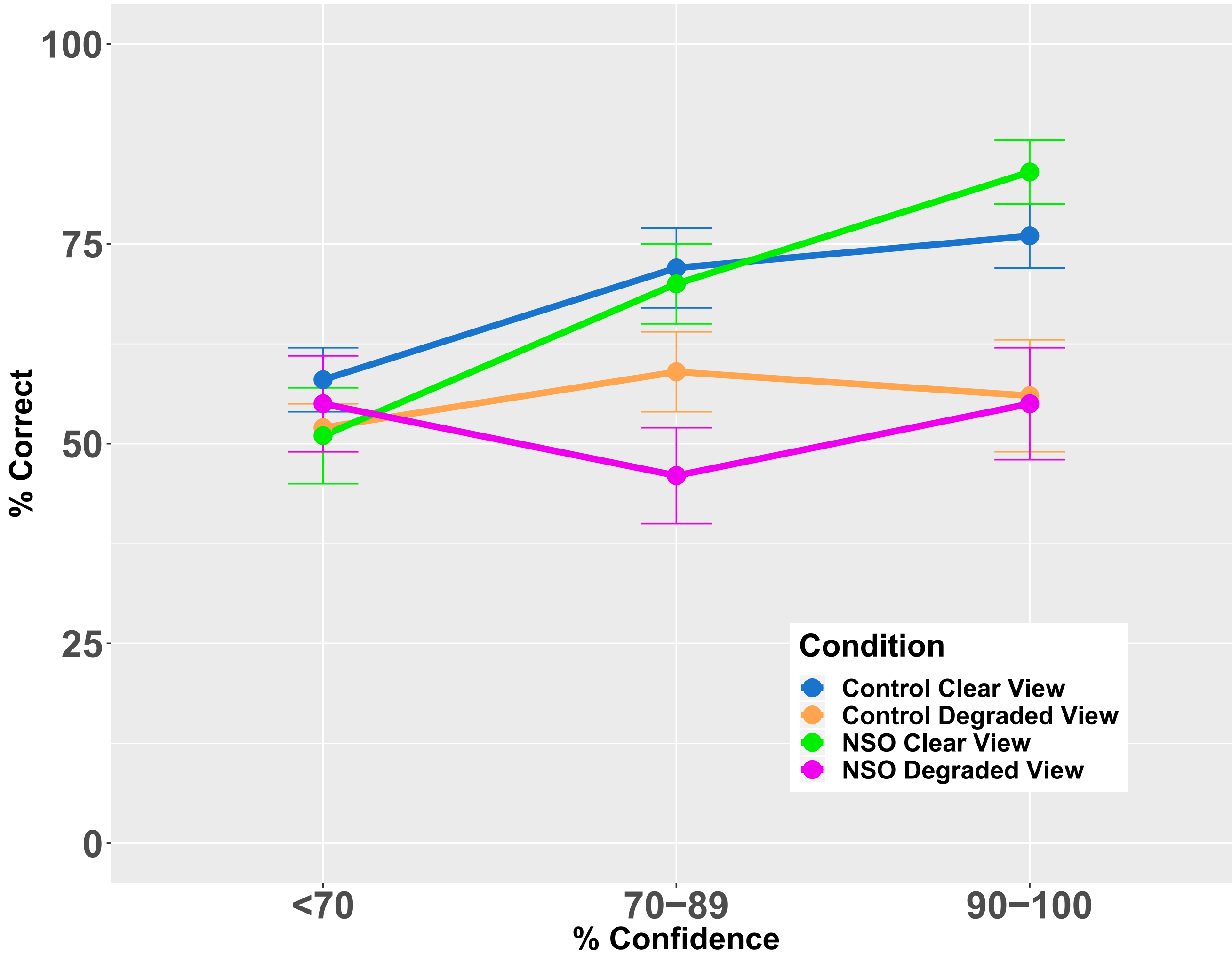

Figure 5. A graph displaying the CAC curves for non-choosers across all four conditions.

As with choosers, the differences observed amongst the calibration statistics seem to be sporadic rather than systematic. So, rather than focusing on pairwise differences, I will focus on what the statistics suggest at the aggregate. In general, the *C* statistic is close to zero for all four conditions, indicating that calibration was quite good for non-choosers. However, there does not appear to be a significant benefit to any of the calibration statistics for those who had the not-sure option compared to those in the control condition. This implies that not-sure option did seem not improve the relationship between confidence and accuracy for non-choosers, contrary to what I predicted.

There is some overconfidence for those participants who had the not-sure option and were given a degraded view, similar to the *O/U* for choosers in that condition. The same account given for choosers is also valid for non-choosers. Because the accuracy rate for that condition remained at around 50% regardless of confidence range, participants were likely to be overconfident if they reported confidence greater than 50%. Indeed, n = 39 participants in that condition reported 100% confidence, even though the accuracy rate for the highest confidence bin was only 55%. Discrimination as represented by the NRI was also quite poor here as it was for choosers. The only condition that differs significantly from zero as indicated by the confidence interval was the clear view, not-sure option condition.

Table 5

Calibration Statistics by Condition for Non-Choosers

	View Quality			
	Clear		Degraded	
Statistic	Control	Not Sure Option	Control	Not Sure Option
C	.02	.01	.04	.08
95% CI	.01, .03	-.001, .01	.02, .05	.04, .12
O/U	.01	.05	.02	.17
95% CI	-.05, .06	-.004, .11	-.04, .08	.10, .25
NRI	.03	.08	.01	.01
95% CI	-.01, .07	.01, .14	-.01, .02	-.04, .01

Note. C = calibration statistic, O/U = over/underconfidence, NRI = normalized resolution index, 95% CI = 95% confidence intervals.

Discussion

Contrary to past research showing that a not-sure option leads to improved identification performance (Steblay & Phillips, 2011; Weber & Perfect, 2012), I found no such evidence. Providing participants with a not-sure option failed to improve the tradeoff between culprit and innocent-suspect identifications and also failed to increase the accuracy of suspect identifications and this was true under both strong and weak memory conditions. The present pattern of results is inconsistent with both past research and with a theoretical framework explaining the strategic regulation of memory. According to Koriat and Goldsmith (1996), promoting the strategic regulation of memory by providing participants with the option to opt-out (i.e., say "I don't know" or "I'm not sure") from making an affirmative decision should improve the accuracy of information provided (Koriat & Goldsmith, 1996).

The purpose of this study was to examine if giving participants an explicit option to opt out of making an identification decision from an eyewitness showup led to an improved tradeoff between culprit and innocent identifications. I also assessed whether a tradeoff improvement was more pronounced under poor witnessing conditions. In fact, focusing only on those participants that had a clear view, culprit identifications decreased by 10% with the inclusion of an opt-out option (from 58% to 48%), but false identifications decreased by only 1% (from 15% to 14%). And, for those who had a degraded view, there was a promising 9% drop in false identifications (from 23% to 14%) but there was also a remarkable 16% decrease in correct identifications (from 31% to 15%). Thus, the main hypothesis of this experiment was not supported. Strangely, participants chose to opt out more frequently when faced with a showup that contained the culprit, which was contrary to my second hypothesis. Intuition suggests that witnesses should have greater uncertainty when viewing a showup containing an innocent suspect to when viewing a showup containing the culprit. To the extent that this is true, one would expect witnesses viewing innocent suspects to use the not-sure option more frequently than witnesses viewing guilty suspects. However, this is merely based on intuition, and will take further empirical research to assess when eyewitnesses generally tend to select opt-out options. As for the confidence-accuracy relationship, the not-sure option removed more low-confidence identifications than high-confidence identifications, as predicted. However, the not-sure option did not improve the confidence-accuracy relationship to a greater extent under poor viewing conditions than under good viewing conditions. Moreover, the accuracy of identifications made at any given level of confidence also failed to increase.

Approximately 30% of witnesses chose to opt out when given the option, indicating that it did indeed capture the state of mind of a large minority of witnesses. The opt-out option was also used more frequently when participants were given a degraded view of the culprit, which indicates that they were aware of the negative impact of poor witnessing conditions on their memories. And, primarily low-confidence identifications, which are likely to be incorrect, were removed by the not-sure option. Despite these positives, the present research is an instance where including an opt-out option produced arguably worse performance than *not* including an opt-out option.

These findings are surprising given the encouraging results of some previous research on the efficacy of an opt out option for eyewitnesses. Weber and Perfect (2012) found that giving participants an option to respond “I don’t know” to a lineup resulted in a significant decrease in innocent-suspect identifications without a significant decrease in culprit identifications. It is also surprising given the strong theoretical foundations for suspecting that such an option would allow witnesses to strategically monitor and regulate their memory, and therefore make better identification decisions (Koriat & Goldsmith, 1996).

Perhaps the not-sure option does not always lead to a better tradeoff between culprit and innocent-suspect identifications. Or, perhaps the results from the present experiment are a false negative. Maybe there are important, as of yet, unidentified moderators of the effectiveness of the not-sure option. I will present a suggestion as to why the current experiment failed to replicate past research showing the benefit of an opt-out option for eyewitnesses, followed by a discussion of the not-sure option's impact on confidence-accuracy relationship, and finally some limitations of the experiment and future directions for research.

Low Innocent-Suspect Identification Rate

There is little-to-no evidence from the present experiment that the presence of an explicit not-sure option improves eyewitness identification performance. Beyond the possibility that the not-sure effect does not typically improve the tradeoff between culprit and innocent-suspect identifications, there are potential moderators that might explain the discrepancy between the current results and those of Weber and Perfect's (2012). One such potential moderator is the innocent-suspect identification rate. In order to observe a substantial reduction in the innocent-suspect identification rate, there has to be a sufficiently high rate of innocent-suspect picks to begin with. Weber and Perfect (2012) obtained an innocent-suspect identification rate of around 30% for their control condition, which is quite typical for a showup procedure (Eisen et al., 2017; Smith et al., 2017). To the contrary, in the present experiment, the control clear-view condition had an innocent-suspect identification rate of only 15%. It would be difficult to reduce the rate of innocent-suspect identifications much more than that. And, in the degraded-view condition, there was some reduction in the innocent-suspect identification rate (from 23% to 14%), but that still might not have been high enough to start with to get the type of improvement that Weber and Perfect (2012) obtained. So, the relatively low innocent-suspect identification rate in the present study is one possible explanation for the discrepancy between the present results and those of Weber and Perfect (2012). I am not implying that if the innocent-suspect identification rate had been high enough, the not-sure option would have certainly been effective at reducing it. However, if any benefit of the not-sure option is to be seen, the innocent-suspect rate must have room to decrease.

Confidence-Accuracy Association

The presence of an opt-out option reduced low-confidence identifications to a greater extent than high-confidence identifications, as predicted. This makes intuitive sense because witnesses who are highly confident that the suspect is the culprit would not gravitate towards a response option that represents uncertainty. However, there is some evidence that high-confidence identifications are highly indicative of accuracy (Brewer & Wells, 2006; Wixted & Wells, 2017). So, it is surprising that although the presence of an opt-out option removed low-confidence identifications, identification performance did not improve. If the opt-out option was siphoning out low-confidence witnesses, who are likely to be incorrect, then what should remain is a higher proportion of accurate identifications. So, it is still somewhat unclear what effect an opt-out option has on the confidence-accuracy relationship (if any), and how that relationship relates to performance.

Inspection of the CAC curves for choosers revealed that high-confidence identifications from degraded encoding conditions were nearly as accurate as high-confidence identifications from clear encoding conditions. This fits with some recent research that found that poor viewing conditions negatively impact identification performance but not the reliability of high-confidence identifications (Palmer, Brewer, Weber, & Nagesh, 2013; Semmler et al., 2018). Nevertheless, high-confidence identifications were only moderately accurate, relative to what researchers typically mean by high accuracy (over 90%) under both degraded and clear encoding conditions.

Degraded-view high-confidence accuracy only reached 83%, and clear-view high-confidence accuracy reached only 85%.

Limitations and Future Directions

One limitation of the present study is that it was conducted entirely online, with no direct experimenter oversight. Despite the fact that many eyewitness studies – and psychology studies in general – are conducted in this manner, there is no way to completely verify that participants were paying attention and internalizing the instructions. There is also no way to completely ensure that they were invested in making correct decisions. Given that the premise of this experiment relies on the theory that people are able, willing, and motivated to strategically regulate and monitor their memories, it is important that the conditions of this experiment mirror the real-world experience of eyewitnesses as much as possible. However, I employed several techniques to ensure the data was of good quality, such as an attention check, as well as requiring participants to acknowledge their understanding of the instructions and agree to follow them before moving forward with each phase of the experiment. Future studies could replicate this experiment in a laboratory context to alleviate some of these concerns.

Another potential limitation of the present study is that no instructions that directed the participant on how best to use the not-sure option were provided, besides informing participants that if they were not sure then they should select the option. It could be that eyewitnesses are not able to use the opt-out option most effectively when left to their own devices. Despite strong theoretical and empirical reasons to expect good memory monitoring and control (Koriat & Goldsmith, 1996; Weber & Perfect, 2012), perhaps eyewitnesses need more guidance on how to best examine their memory. Because research on an opt-out option is relatively nascent in the eyewitness context, I wanted to see how participants would use the not-sure option in the absence of any explicit instructions. A large portion of participants did opt out, (~30%), suggesting that a not-sure option does indeed capture the state of mind of a large minority of eyewitnesses that was previously neglected.

However, future research should consider providing instructions about how to use the not-sure option appropriately. To help avoid a rise in the decision criterion for witnesses who view a culprit-present identification procedure, perhaps a statement could be included such as, “if you have a strong, immediate feeling of recognition when viewing the person, you should select YES. If it takes you a long time or if you are having a lot of difficulty determining whether the person is the culprit, select NO or NOT SURE”. This type of instruction may give witnesses the best chance to make correct identifications while still alerting them that they have the option to back out or make a rejection.

There are a multitude of instruction options that could be included depending on how conservative the procedure is meant to be. For example, telling participants to select the not-sure option if they are less than 80% or 90% confident in their identification or rejection decision. This would most likely result in a low culprit identification rate, but also a very low innocent-suspect identification rate, producing a good tradeoff. There is some convincing evidence that high-confidence identifications are likely to be highly reliable, so this instruction in combination with an opt-out option could help ensure that only the best-quality identifications are being made (Wixted & Wells, 2017).

Another future direction for research is to examine how an opt-out option interacts with biased and unbiased instructions. Biased instructions tend to guide the witness towards making an identification from both culprit-present and culprit-absent lineups (Clark, 2005; Steblay, 1997). However, a not-sure option may reduce the high innocent-suspect identification rate caused by biased instructions by giving witnesses an “out”. For instance, if the witness does not have a strong match-to-memory experience due to being in a culprit-absent condition, but feels pressure to identify the suspect due to biased instructions, the not-sure option may provide an appealing alternative to an outright rejection. There are also likely to be more innocent-suspect picks from biased instructions, which would give the not-sure option enough room to reduce them. For these reasons, the not-sure option may prove effective at improving the tradeoff between culprit and innocent-suspect identifications when the procedure also has biased instructions.

In any case, combining an opt-out option with different procedure instructions could help clarify the conditions under which such an option is most advantageous. Future research should try and determine why witnesses seem to opt out more frequently in situations where they should be making identifications, like when the culprit is present. Instructions around the proper use of an opt-out option could be useful to help improve the strategic regulation of eyewitness memory. More research should also be conducted to generalize the effect of an opt-out option across more estimator variables and identification procedure formats.

Conclusion

The present experiment demonstrates that the inclusion of an explicit opt-out option for an eyewitness showup does not necessarily improve the tradeoff between culprit and innocent-suspect identifications. Regardless of whether viewing conditions were clear or degraded, the presence of an option to respond ‘not sure’ decreased both culprit and innocent-suspect identifications. This was despite some evidence from the basic recognition memory literature and the eyewitness context that supports the inclusion of an explicit opt-out option (Koriat & Goldsmith, 1996; Weber & Perfect, 2012). Unmistakeably, there is now evidence from the present research that the implementation of an opt-out option may have potential drawbacks. Nevertheless, the inclusion of an explicit opt-out option is being promoted in the most recent white paper commissioned by the American Psychology-Law Society on the implementation of eyewitness procedures (Wells et al., forthcoming). Given the present results, it is unclear how beneficial explicit opt-out options will ultimately prove to be.

However, there may be a silver lining to be gleaned from the current experiment that justifies the recommendation, or at least future research on opt-out options. Participants opted out more frequently when they had worse witnessing conditions, indicating some metamemorial awareness of the impact of that view on their memory quality. Also, the fact that the not-sure option decreased primarily low-confidence identifications is quite promising. Recall Jennifer Thompson's mistaken identification of Ronald Cotton. Jennifer was a low-confidence witness– she took a long time to make her decision, wavered between lineup members, and was susceptible to bias from the lineup administrator. Low-confidence witnesses are associated with lower levels of decision accuracy, so an opt-out option should tend to increase accuracy (Brewer & Wells, 2006). It is not clear why this did not happen in the current study. However, the fact that the not-sure option functioned as predicted by siphoning out low-confidence responses does justify further research that fully explores the impact of opt-out options for eyewitnesses.

References

Bain v. State BT - So. 2d (2007). Fla: Dist. Court of Appeals, 2nd Dist.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, *55*, 412-428. https://doi.org/10.3758/BF03205299

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review, 116,* 84–115. https://doi.org/10.1037/a0014351

Bornstein, B. H., Deffenbacher, K. H., Penrod, S. D., McGorty, E. K. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime & Law, 18*, 473–490. doi: https://doi.org/10.1080/1068316X.2010.508458

Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *The Journal of Applied Psychology, 87,* 112–120. https://doi.org/10.1037/0021-9010.87.1.112

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

Charman, S. D., & Wells, G. L. (2007). Eyewitness lineups: Is the appearance-change instruction a good idea? *Law and Human Behavior, 31*, 3-22. https://doi.org/10.1007/s10979-006-9006-3

Clark, S. E. (2003). A memory and decision model for eyewitness identification. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in *Memory and Cognition, 17*, 629–654. https://doi.org/10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575-604. https://doi.org/10.1007/s10979-005-7121-1

Eisen, M. L., Smith, A. M., Olaguez, A. P., & Skerritt-Perta, A. S. (2017). An examination of showups conducted by law enforcement using a field-simulation paradigm. *Psychology, Public Policy, and Law, 23*, 1-22. https://doi.org/10.1037/law0000115

Frontline. (1997, February). What Jennifer Saw. *PBS Frontline.* Retrieved from: https://www.pbs.org/wgbh/pages/frontline/shows/dna/etc/script.html

Garrett, B. (2011, April). Getting It Wrong: Convicting the Innocent. *Slate Magazine*. Retrieved from http://www.slate.com/articles/news_and_politics/jurisprudence/features/2011/getting_it_wrong_convicting_the_innocent/how_eyewitnesses_can_send_innocents_to_jail.html

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8-20. https://doi.org/10.3758/BF03198438

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546-567. https://doi.org/10.1037/0033-295X.100.3.546

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology, 64*, 525-537. https://doi.org/10.1037/0022-3514.64.4.525

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General, 136,* 1–22.

https://doi.org/10.1037/0096-3445.136.1.1

Innocence Project. (2017). Eyewitness Misidentification. *Innocence Project.org.* Retrieved from: https://www.innocenceproject.org/causes/eyewitness-misidentification/

Kassin, S. M. (1984). Eyewitness identification: Victims versus bystanders. J*ournal of Applied Social Psychology, 14,* 519–529. https://doi.org/10.1111/j.1559-1816.1984.tb02257.x

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The "general acceptance" of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44*, 1089-1098. 10.1037/0003-066X.44.8.1089

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48,* 704-721. https://doi.org/10.1016/S0749-596X(02)00504-1

Kent, C., Lamberts, K., & Patton, R. (2018). Cue quality and criterion setting in recognition memory. *Memory and Cognition, 46,* 757-769. https://doi.org/10.3758/s13421-018-0796-6

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643-656. https://doi.org/10.1037/0096-3445.133.4.643

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the stategic regulation of memory accuracy, *Psychological Review, 103,* 490–517. http://dx.doi.org/10.1037/0033-295X.103.3.490

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*, 1489–1494. https://doi.org/10.3758/s13423-014-0641-2

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*, 26-44. http://dx.doi.org/10.1037/lhb0000309

Lindsay, R. C., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior*, *4*, 303-313. http://dx.doi.org/10.1007/BF01040622

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

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin and Review, 12*, 43-65. https://doi.org/10.3758/BF03196348

Luus, C. A., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology*, *79*, 714-723. http://dx.doi.org/10.1037/0021-9010.79.5.714

Macmillan, N. A., & Creelman, C. D. (2005). Classification Designs for Discrimination. *In Detection Theory: A User's Guide*. New York, NY: Psychology Press. https://doi.org/10.1017/CBO9781107415324.004

McQuiston, D., & Malpass, R. (2001). Eyewitness identifications in criminal cases: An archival study. *In Fourth biennial meeting of the Society for Applied Research in Memory and Cognition, Kingston, Ontario, Canada.*

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7,* 3–35. https://doi.org/10.1037//1076-8971.7.1.3

Memon, A., Gabbert, F., & Hope, L. (2013). The ageing eyewitness. In J.R. Adler (Ed.), *Forensic Psychology: Concepts, debates, and practice* (2nd ed., pp. 96-113). Cullompton: Willan.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339–354. https://doi.org/10.1348/000712603767876262

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*, 93-102. https://doi.org/10.1016/j.jarmac.2015.01.003

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*, 361-376. http://dx.doi.org/10.1037/a0030609

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*, 55-71. http://dx.doi.org/10.1037/a0031602

Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law and Human Behavior*, *22*, 549-570. https://doi.org/10.1023/A:1025739514042

Read, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology, 9,* 91–

121. https://doi.org/10.1002/acp.2350090202

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*, 77-88. https://doi.org/10.1186/1471-2105-12-77

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*, 337–347. https://doi.org/10.1007/s10979-009-9192-x

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

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139–156. https://doi.org/10.1037/0033-2909.100.2.139

Smalarz, L., & Wells, G. L. (2013). Eyewitness certainty as a system variable. Reform of Eyewitness Identification Procedures. American Psychological Association. https://doi.org/10.1037/14094-008

Smith, A. M., Bertrand, M., Lindsay, R. C. L., Kalmet, N., Grossman, D., & Provenzano, D. (2014). The impact of multiple show-ups on eyewitness decision-making and innocence risk. *Journal of Experimental Psychology: Applied, 20*, 247-259. https://doi.org/10.1037/xap0000018

Smith, A. M., Lampinen, J. M., Wells, G. L., Smalarz, L., & Mackovichova, S. (2018). 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,* 50-59. https://doi.org/10.1016/j.jarmac.2018.09.003

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 Behaviour*, *42*, 215-226. http://dx.doi.org/10.1037/lhb0000284.

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Penrod, S. D. (2017). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior, 41*, 127-145. https://doi.org/10.1037/lhb0000219

Smith, A. M., Wells, G. L., Smalarz, L., Lampinen, J. M. (2018). Increasing the similarity of fillers improves the applied value of lineups without improving memory performance: Commentary on Colloff, Wade and Strange (2016). *Psychological Science, 29*, 1-4. https://doi.org/10.1177/0956797617698528

Smith, A. M., Wilford, M., Quigley-McBride, A., & Wells, G. L. (2019). Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior*. DOI: 10.13140/RG.2.2.23975.19361

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327. https://doi.org/10.1037/0033-2909.118.3.315

State v. Booker BT - P. 2d (1985). Utah: Supreme Court.

State v. Cotton BT - SE 2d (1987). NC: Supreme Court.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior*, *21*, 283-297. https://doi.org/10.1023/A:1024890732059

Steblay, N. K., & Phillips, J. D. (2011). The not-sure response option in sequential lineup practice. *Applied Cognitive Psychology*, *25*, 768-774. https://doi.org/10.1002/acp.1755

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. J*ournal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1379–1396. https://doi.org/10.1037/0278-7393.24.6.1379

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition*, *35*, 254-262. https://doi.org/10.3758/BF03193446

Warnick, D. H., & Sanders, G. S. (1980). Why do eyewitnesses make so many mistakes? *Journal of Applied Social Psychology, 10*, 362–366. https://doi.org/10.1111/j.1559-1816.1980.tb00716.x

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36*, 28–36. https://doi.org/10.1037/h0093976

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

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

Wells, G. L., Ferguson, T. J., & Lindsay, R. C. (1981). The tractability of eyewitness confidence and its implications for triers of fact. *Journal of Applied Psychology, 66*, 688-696. https://doi.org/10.1037/0021-9010.66.6.688

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, *88,* 776-784. http://dx.doi.org/10.1037/0033-2909.88.3.776

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, and Law, 1,* 765-791. https://doi.org/10.1037/1076-8971.1.4.765

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603–647. https://doi.org/10.1023/A:1025750605807

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,* 324–328. https://doi.org/10.1016/j.jarmac.2015.08.010

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115–1125. https://doi.org/10.1177/01461672992512005

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*, 99–122. https://doi.org/10.1037/lhb0000125

Wickens, T. D. (2010). *Elementary Signal Detection Theory.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195092509.001.0001

Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning and Behavior, 30*, 289-305. https://doi.org/10.3758/BF03195955

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

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*, 515-526. http://dx.doi.org/10.1037/a0039510

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18,* 10–65. https://doi.org/10.1177/1529100616686966

Wogalter, M. S., Malpass, R. S., & Mcquiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime, and Law*, *10*, 69-82. https://doi.org/10.1080/10683160410001641873

Appendix A

Recruitment Notice

Study Name: Decision-Making About New Faces

Description: In the present experiment, you will be asked to make judgments about faces and unscramble some words. We are examining memory and decision-making within this context. You can expect to view a short video, view photos of faces, complete some short questionnaires and a word unscrambling task. Your small but valuable contribution of time will advance the science in this area. We are looking for participants to take part in a study conducted on Qualtrics that will take less than 15 minutes to complete.

Eligibility Requirements: We are looking for English speakers with good or corrected vision (with soft contact lenses). The study must be completed on a computer or tablet (not a mobile device). **You must have access to an internet browser other than Google Chrome or Safari in order to participate in this study. If you use either of these browsers, you will not be able to complete the study.**

Risks: You may view a video where there is a gun present and pointed in the direction of the camera. No shots are fired, and no other forms of violence are depicted. If you think this may be uncomfortable for you, please do not participate in this study. There are no physical risks associated with this study.

Duration and Locale: The study will take place online via Qualtrics online survey software.

Compensation: In exchange for your participation in this experiment, you will receive $1 as compensation deposited into your MTurk account.

Researchers: Andrew Smith (Principal Investigator); Shaela Jalava (MA Student)

Email: AndrewMSmith@cmail.carleton.ca, shaelajalava@cmail.carleton.ca

This study has received clearance by the Carleton University Research Ethics Board (B) (Clearance #109732).
If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca). For all other questions about the study, please contact the researchers.

Appendix B

Informed Consent Form

Name and Contact Information of Researchers:

Dr. Andrew Smith, Carleton University, Department of Psychology
Email: andrewm.smith@cmail.carleton.ca
Shaela Jalava, Carleton University, Department of Psychology
Email: shaelajalava@cmail.carleton.ca

Project Title

Decision-Making About New Faces

Project Sponsor and Funder (if any)

This study is funded by the Natural Sciences and Engineering Research Council.

Carleton University Project Clearance

Clearance: 109732 Date of Clearance: November 12, 2018

Invitation

You are invited to take part in a research project because you are registered for Amazon's Mechanical Turk. The information in this form is intended to help you understand what we are asking of you so that you can decide whether you agree to participate in this study. Your participation in this study is voluntary, and a decision not to participate will not be used against you in any way. As you read this form, and decide whether to participate, please ask all the questions you might have, take whatever time you need, and consult with others as you wish.

What is the purpose of the study?

The purpose of this study is to examine how people make judgments of new faces they haven't seen before.

What will I be asked to do?

If you agree to take part in the study, we will ask you to:

- Watch a short video
- Answer some demographic questions
- Do a word-unscrambling task
- Make decisions about some faces
- The study will last approximately 15 minutes, ending with a debriefing form

Risks and Inconveniences

You may view a video where there is a gun present and pointed in the direction of the camera. No shots are fired, and no other forms of violence are depicted. If you think this may be uncomfortable for you, please do not participate in this study. There are no physical risks associated with this study.

Possible Benefits
You may not receive any direct benefit from your participation in this study. However, your participation may allow researchers to better understand the processes underlying human decision-making.

Compensation/Incentives
In exchange for your participation in this experiment, you will receive $1 as compensation deposited into your MTurk account.

No waiver of your rights
By signing this form, you are not waiving any rights or releasing the researchers from any liability.

Withdrawing from the study
If you wish to withdraw from this study, you must withdraw your consent *during* the course of the study by closing the browser window. After the study is completed, there will be no way to tie your identity to your responses, and so you will not be able to withdraw from the study after it is completed. All information collected from you before your withdrawal will be discarded if and only if you withdraw before completion. You will not receive compensation if you wish to withdraw early.

Confidentiality
The data collected in this study will be anonymous at the outset because compensation is granted automatically through a unique code that you will receive upon completion. The data will be stored and protected by Qualtrics in a server in Toronto, ON, Canada, but may be disclosed via court order or data breach. The data from this study may eventually be published in an academic outlet and will be posted in an online repository for other researchers who may also wish to analyze these data. However, nobody will be able to determine who participated in this experiment nor will anybody be able to connect participants to their responses.

We will treat your personal information as confidential, although absolute privacy cannot be guaranteed. No information that discloses your identity will be released or published without your specific consent. Research records may be accessed by the Carleton University Research Ethics Board in order to ensure continuing ethics compliance. All data, including coded information, will be kept in a password-protected file on a secure computer. We will password protect any research data that we store or transfer.

Data Retention
After the study is completed, your de-identified data will be retained for future research use.

New information during the study
In the event that any changes could affect your decision to continue participating in this study, you will be promptly informed.

Ethics review

This project was reviewed and cleared by the Carleton University Research Ethics Board [B]. If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board (by phone at 613-520-2600 ext. 4085 or by email at ethics@carleton.ca).

Appendix C

Demographic Questions

General demographic information

1. Please indicate your ethnic background
   a. American Indian or Alaska Native
   b. Asian
   c. Black or African American
   d. Native Hawaiian or Other Pacific Islander
   e. Hispanic or Latino/Latina
   f. White or Caucasian

Other (Specify)

2. Please indicate your age using the slider below *(NB. Participants responded by using the slider provided on the computer screen to indicate their age.)*

3. Please indicate your gender
   a. Female
   b. Male
   c. Other

Video Description

The simulated crime video used in the study depicts a man facing the camera holding a gun. The gun is pointed towards the camera, and it obscures the man's face slightly. The man then states, "It's only money, hand it over and no one has to get hurt". The man's face is in view for approximately 5 seconds. The poor view condition used exactly the same video, with the brightness and contrast altered to approximate viewing the crime in overexposed lighting.

Appendix D

Retrospective Questionnaire

1. How good of a view did you get of the culprit? 1 (very poor) to 7 (very good)

2. How long would you estimate that the culprit's face was in view during the encoding video? 1 (very little time) to 7 (quite a bit of time)

3. How well were you able to make out specific features of the culprit's face from the video? 1 (not at all) to 7 (very well)

4. How far away was the culprit? 1 (not far) to 7 (very far)

5. How good was the lighting? 1 (not good) to 7 (very good)

6. How much attention were you paying to the culprit's face while viewing the video? 1 (none) to 7 (my total attention)

7. How easy or difficult was it for you to figure out whether the culprit was present in the photo? 1 (I needed almost no time to figure this out) to 7 (I had to look at the photo for a very long time to figure this out)

8. Generally, how good is your recognition memory for the faces of strangers you have encountered on only one prior occasion? 1 (very poor) to 7 (excellent)

9. How clear is the image you have in your memory of the culprit you saw in the video? 1 (not at all clear) to 7 (very clear)

Appendix E

Debriefing Form

Name and Contact Information of Researchers:
Dr. Andrew Smith, Carleton University, Department of Psychology
Email: andrewm.smith@cmail.carleton.ca
Shaela Jalava, Carleton University, Department of Psychology
Email: shaelajalava@cmail.carleton.ca

Project Title
Decision-Making About New Faces

Project Sponsor and Funder (if any)
This study is funded by the Natural Sciences and Engineering Research Council

Carleton University Project Clearance
Clearance #: 109732 Date of Clearance: November 12, 2018

What are we trying to learn in this research?

This research examines whether the inclusion of an option to opt out of making a decision about an eyewitness identification task (by responding "not sure") can help reduce the rate of false identifications. You viewed a video of a staged crime, receiving either a good or a poor view of the culprit. We are interested in whether those who saw a poor view of the culprit will use the not-sure option more often than those who received a good view of the culprit. If participants who were likely to make a false identification instead used the not-sure option, then this research will help make eyewitness identification procedures more accurate. The word-unscrambling task was merely a distractor and is not important to the goals of the experiment.

Why is this important to scientists or the general public?

Eyewitness identifications are a crucial component of a criminal investigation. Eyewitnesses usually think of an identification task as having two response options– choose to identify a suspect or choose to not identify a suspect. However, research shows that giving eyewitnesses the option to opt out of responding altogether can decrease the rate of false identifications without a cost to correct identifications. This research can help improve the accuracy eyewitness procedures and therefore safeguard the integrity of the criminal justice system.

What are our hypotheses and predictions?

We predict that those participants who were given a poor view of the culprit are more likely to falsely identify a suspect due to having a relatively worse memory of the culprit than those who were given a good view. Therefore, the not-sure option should be more useful for those who were given a poor view, and therefore be used more often by those participants.

Where can I learn more?

Is there anything I can do if I found this experiment to be emotionally upsetting? Yes. If you feel any distress or anxiety after participating in this study, please feel free to contact the Carleton University Health and Counseling Services at: 613-520-6674, or the Distress Centre of Ottawa and Region at 613-238-3311 (http://www.dcottawa.on.ca).

What if I have questions later?

If you have any remaining concerns, questions, or comments about the experiment, please feel free to contact Dr. Andrew Smith (Principal Investigator), at: andrewM.smith@cmail.carleton.ca (613-520-2600), or Shaela Jalava (MA Student), at: shaelajalava@cmail.carleton.ca (613-520-2600).

If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

Thank you for participating in this research!

Supplemental Material

Hypothetical Responses

I asked those participants who opted out of making an identification from the showup what response option they would have selected had they not been given the option to opt out. I also asked them to indicate their confidence in that hypothetical decision. The average confidence level was 45.08%. The response percentages of the hypothetical decisions separated by quality of view and culprit presence are displayed in Table 6. It seems that the not-sure option siphoned away primarily low-confidence responses and roughly equal percentages of correct and incorrect decisions. The fact that many correct decisions were removed by the not-sure option potentially contributed to its failure to improve the tradeoff between culprit and innocent-suspect identifications. However, these responses were purely hypothetical; there can be no guarantee that these participants would have made these choices had they been in the control condition. Nonetheless, these responses give some indication as to what types of responses the not-sure option removed.

Table 6

Hypothetical Response Percentages as a Function of View Quality and Culprit Presence

	View Quality			
	Clear		Degraded	
Response Option	Culprit Present	Culprit Absent	Culprit Present	Culprit Absent
Yes	56% (61)	45% (42)	41% (111)	37% (93)
No	44% (61)	54% (42)	59% (111)	63% (93)

Note. Hypothetical responses from those who initially opted out of the showup, $N = 307$.

Deviation from Perfect Performance Analysis

I also wanted to examine the ROC curves for only eyewitnesses who identified the suspect. Specifically, I wanted to examine the impact of the not-sure option collapsed over the two viewing conditions. Accordingly, I followed the recommendations of Smith et al. (2018) and used Deviation from Perfect Performance (DPP) to make these comparisons.

The equation for determining DPP is defined as:

$$\text{DPP}(c) = [1 - \text{Culprit}(c)] + [\text{Innocent}(c)]$$

where Culprit(c) is equal to the culprit identification rate at a given point on the ROC curve, and Innocent(c) is equal to the innocent-suspect identification rate at the same point on the ROC curve. An index of deviation from perfect performance for a given procedure– that is independent of expressed level of witness confidence– can be computed by calculating DPP for every ROC point and taking the average. To the extent that the DPP value is greater than 0, the procedure deviates from perfect performance. DPP is conceptually similar to pAUC, but it is not constrained to any particular region of the ROC space. Much like comparing the difference between two pAUC values to determine which procedure has better utility, comparisons can be made between two DPP values. Like all utility-based analyses, DPP requires researchers to explicate their assumptions about the relative costs of different decision errors and underlying base rates. I examined DPP for all combinations of three different cost ratios (innocent suspect identifications are 1x, 2x, or 10x more costly than culprit identifications) and three different underlying base rates of culprit presence (.25, .50, .75).

Table 7

Deviation from Perfect Performance as a Function of Base Rates and Cost Ratios for Response Option Condition

		Response Option		95% CI	
Base Rate	Cost Ratio	Control	Not Sure	LL	UL
0.25	1:1	.53	.57	-.004	.08
	2:1	**.98**	**1.07**	**.01**	**.19**
	10:1	**4.57**	**5.12**	**.11**	**.96**
0.50	1:1	.45	.46	-.03	.04
	2:1	.75	.80	-.02	.10
	10:1	**3.15**	**3.49**	**.05**	**.63**
0.75	1:1	.38	.35	-.06	.003
	2:1	.53	.52	-.05	.03
	10:1	1.73	1.87	-.01	.28

Note. The base rate refers to the probability that the culprit is present. Cost ratio has the same meaning as subjective utilities and in this case refers to how much more costly innocent suspect identifications are relative to missed culprit identifications (1:1 = equally as costly, 2:1 = twice as costly, 10:1 = 10 times more costly). The 95% CI refers to the 95% confidence interval of the difference between two DPP values. A significant difference between two DPP values is evidenced by a confidence interval that does not overlap zero. Bolded rows indicate significant differences between DPP values.

Table 4 displays DPP for response option condition by three different cost ratios and three different base rates of culprit presence. When the base rate is .75, the control and not-sure option conditions would produce roughly the same identification performance. When the base rate is .50, the control and not-sure response option conditions would produce roughly the same identification performance except when the cost of an innocent-suspect identification is 10 times more costly than a missed culprit identification. When the cost ratio is 10:1, the not-sure option would produce worse performance than the control condition. Finally, when the base rate is .25 and the cost of an innocent-suspect identification is two or 10 times more costly than a missed culprit identification, the not-sure option would produce worse performance than the control condition. Taken together, the findings suggest that when the base rate of culprit presence is .50

or lower and when the cost of an innocent-suspect identification is considered to be more costly than a missed culprit identification, including an opt-out option may actually be more detrimental to identification performance than not including an opt-out option.

Retrospective Questionnaire

One of the most interesting findings to come out of the research on the post-identification feedback effect is that witnesses' perceptions about their encoding experience can be influenced by external factors. Wells and Bradfield (1998; 1999) demonstrated that participants who were given confirmatory feedback after making an identification from a lineup reported that they had paid more attention to the simulated crime, had a better view of the culprit, and could better make out the details of the culprit's face – even though all participants had viewed the same video. Relatedly, Bradfield et al. (2002) showed that the distortive effect of confirmatory feedback on witnesses' retrospective experience is stronger for those who made an identification from a culprit-absent lineup compared to those who made an identification from a culprit-present lineup.

I wanted to see if the same pattern emerged from the responses to our retrospective questionnaire, even in the absence of confirmatory feedback. Although this analysis is exploratory, there is some *a priori* reason to suspect that participants who falsely identify will misattribute the difficulty they experience at the time of testing to having had worse conditions at the time of encoding, when compared to participants who do not falsely identify. Indeed, if participants who falsely identified had correctly attributed the difficulty they experienced to the fact that the culprit was not present in the lineup/showup, then they would have correctly rejected.

To test this hypothesis, I ran a series of independent-samples t-tests comparing the mean response for each question for participants who were randomly assigned to a culprit-present showup to the mean responses for participants who were randomly assigned to a culprit-absent showup. I conducted separate comparisons for choosers and non-choosers. When examining choosers only, compared to participants in a culprit-present condition, participants in a culprit-absent condition reported that (1) they had had a worse view of the culprit at encoding, (2) they were less able to make out specific features of the culprit's face, (3) the lighting quality was worse, (4) it was harder to make an identification decision, and (5) they had a less clear image of the culprit in their head. No significant differences were found between the mean responses on the questions regarding how long the culprit's face was in view, the distance of the culprit, how much attention was paid to the video, or how good participants' memory is for faces they have seen on one prior occasion. These comparisons confirm that participants who made an identification when they should have rejected the showup report a relatively worse encoding experience and more task difficulty than those who made a correct identification.

Table 8

Comparisons of Mean Responses on the Retrospective Questionnaire for Choosers

		Measure								
Culprit Presence		View of culprit	Time face visible	Details of face	Distance from culprit(i)	Lighting quality	Attention to video	Ease of decision(i)	Memory for faces	Image in memory
Present	*M*	**5.08**	4.33	**4.42**	2.28	**4.60**	5.76	**3.59**	4.56	**4.69**
	SD	1.65	1.91	1.61	1.19	1.98	1.22	1.73	1.47	1.49
Absent	*M*	**4.18**	3.99	**3.79**	2.43	**3.49**	5.65	**4.46**	4.33	**4.23**
	SD	1.65	1.92	1.46	1.25	2.25	1.39	1.42	1.38	1.49

Note. Choosers. Higher scores on a measure indicate relatively more ideal conditions (e.g., the culprit's face was in view for a longer amount of time). The subscript (i) on a measure indicates the reverse interpretation, with lower scores indicating more ideal conditions, (e.g., the culprit was a shorter distance from the camera). Columns in which the values are bolded indicate significant mean differences between the culprit-present and culprit-absent conditions.

Next, I wanted to examine non-choosers responses to see if the same pattern would emerge for incorrect rejecters. That is, for participants who should have made an identification but instead incorrectly rejected the showup, did they also report having had a worse encoding experience and increased task difficulty? When examining non-choosers only, compared to participants in a culprit-absent condition, participants in a culprit-present condition reported that (1) they had had a worse view of the culprit at encoding, (2) the culprit's face was in view for less time, (3) they were less able to make out specific features of the culprit's face, (4) the lighting quality was worse, (5) they paid less attention to the video, (6) it was harder to make an identification decision, and (7) they had a worse image of the culprit in their memory. No significant differences were found for the questions regarding the distance of the culprit, and participants' memory for faces in general. Taken together, these results suggest that participants who made incorrect decisions have a more distorted view of the witnessing experience than did participants who made correct decisions. And, it seems that this distorted impression of the witnessing experience is still present even in the absence of confirmatory feedback. It must be noted, however, that these analyses were purely exploratory.

Table 9

Comparisons of Mean Responses on the Retrospective Questionnaire for Non-Choosers

		Measure								
Culprit Presence		View of culprit	Time face visible	Details of face	Distance from culprit(i)	Lighting quality	Attention to video	Ease of decision(i)	Memory for faces	Image in memory
Present	*M*	**3.60**	**3.81**	**3.17**	2.39	**3.14**	**5.39**	**4.42**	4.03	**3.64**
	SD	1.90	1.98	1.56	1.17	2.15	1.55	1.75	1.63	1.66
Absent	*M*	**4.19**	**4.12**	**3.71**	2.35	**3.92**	**5.60**	**3.93**	4.06	**4.01**
	SD	2.01	1.91	1.67	1.15	2.28	1.41	1.79	1.56	1.65

Note. Non-choosers (rejected or opted out). Higher scores on a measure indicate relatively more ideal conditions (e.g., the culprit's face was in view for a longer amount of time). The subscript (i) on a measure indicates the reverse interpretation, with lower scores indicating more ideal conditions, (e.g., the culprit was a shorter distance from the camera). Columns in which the values are bolded indicate significant mean differences between the culprit-present and culprit-absent conditions.