Do Evaluators of Eyewitness Evidence Believe Highly Confident Eyewitnesses are

Poorly Calibrated When They Experienced Suboptimal Witnessing Conditions?

by

Taylor Cameron Lebensfeld

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved August 2022 by the
Graduate Supervisory Committee:

Laura Smalarz, Chair
Jessica Salerno
Andrea Arndorfer

ARIZONA STATE UNIVERSITY

December 2022

ABSTRACT

It has recently been argued that high-confidence eyewitness identifications are highly likely to be accurate regardless of the quality of viewing conditions experienced by the witness. However, new evidence suggests that evaluators of eyewitness identification evidence (e.g., jurors) do not trust highly confident eyewitnesses who experienced poor witnessing conditions. In fact, contextual information about poor witnessing conditions decreases evaluators' belief of eyewitnesses to a greater extent for highly confident witnesses than for moderately confident witnesses. Why is the effect of witnessing-condition information greater for evaluations of high-confidence witnesses than for less confident witnesses? The current research tested the possibility that information about witnessing conditions influences evaluators' perceptions of how well-calibrated a witness's identification confidence is with the eyewitness's accuracy. Using a paradigm adapted from the confidence calibration literature, I conducted an experiment to test this calibration account of the finding that witnessing condition information has a stronger effect on perceptions of highly confident witnesses than moderately confident witnesses. Although the results replicated the differential effects of witnessing condition context on perceptions of highly and moderately confident eyewitnesses, they failed to yield support for the confidence calibration hypothesis, potentially because the confidence calibration manipulation was ineffective. Directions for future research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

For decades, researchers believed that eyewitness confidence in a lineup identification is informative of accuracy only to the extent that the eyewitness experienced "optimal" viewing conditions (i.e., the "optimality hypothesis;" Deffenbacher, 1980, 2008). Yet much of the time, eyewitnesses do not experience optimal viewing conditions when witnessing a crime; instead, witnesses often view crimes when it is dark, from a distance, under stress, when the perpetrator attempts to conceal their appearance, etc. In recent years, however, scientific understanding of the eyewitness confidence-accuracy relation has evolved. Researchers now generally agree that eyewitness confidence can be highly informative of accuracy as long as the eyewitness's confidence statement was collected immediately at the time of the identification using pristine procedures (Wixted & Wells, 2017) and in the absence of factors that bias the witness toward the suspect (Smalarz, 2021). In even starker contrast to the optimality hypothesis, some researchers have argued that highly confident eyewitnesses are highly likely to be accurate regardless of whether the witness experienced optimal viewing conditions (Semmler et al., 2018). According to this argument, someone who witnessed the crime from a far distance, at night, or under stress, but later identified the suspect from an unbiased, pristine lineup with high confidence is highly likely to have made an accurate identification.

Accordingly, research has begun to investigate how high-confidence witnesses are perceived by evaluators. A recent series of studies found that informing evaluators that a witness experienced poor witnessing conditions leads them to doubt the witness's

1

identification accuracy more if the witness made a high-confidence identification than if the witness made a moderate confidence identification (Lebensfeld & Smalarz, 2022). These findings directly oppose what Semmler and colleagues (2018) would argue evaluators should do. They would argue evaluators perceptions of highly confident witnesses should remain unchanged for highly confident witnesses but decrease for moderately confident witnesses. So, why was this not the case? More specifically, the current research asked, why does contextual information about poor witnessing conditions have a stronger effect on perceptions of highly confident witnesses than it does on perceptions of moderately confident witnesses?

Perceptions of Confidence and Poor Witnessing Conditions

Lebensfeld and Smalarz (2022) were the first to investigate the effects of witnessing condition information on evaluations of high- and moderate-confidence eyewitness identifications. Participants in their studies read a series of crime vignettes that manipulated the presence of contextual information about poor witnessing conditions (e.g., poor lighting, far viewing distance, witness intoxication, and culprit disguise) and whether the witness was highly confident or moderately confident in her identification. Contextual information about poor witnessing conditions decreased evaluators' perceptions of identification accuracy compared to when contextual information was not given. More importantly, as noted above, this effect was stronger for evaluations of highly confident eyewitnesses than for evaluations of moderately confident eyewitnesses. In fact, the witnessing condition information led evaluators to believe moderately confident and highly confident witnesses at similar rates, effectively eliminating their sensitivity to differences in eyewitness confidence. This effect was replicated across two

2

more studies by Lebensfeld and Smalarz and was robust to manipulations intended to reduce the effect. Specifically, jury instructions (Study 2) and expert witness testimony (Study 3) that attested to the reliability of highly confident witnesses regardless of whether they experienced poor witnessing conditions failed to eliminate the effect (though expert testimony reduced the strength of the effect).

The goal of the current research was to test a theoretical explanation for why contextual information about poor witnessing conditions has a stronger effect on perceptions of highly confident witnesses than on perceptions of moderately confident witnesses. In particular, I tested whether perceptions of a witness's confidence calibration—how well a witness's confidence aligns with the accuracy of their testimony and/or identification decision—may account for the effect. I theorized that when a witness experiences poor witnessing conditions, evaluators perceive a highly confident witness to be over-confident and perceive a less confident witness to be appropriately confident. In other words, a highly confident witness who experienced poor witnessing conditions is believed to have poorly calibrated confidence (i.e., their level of confidence is not aligned to their likelihood of making an accurate identification). As described next, poorly calibrated confidence leads evaluators to be skeptical of eyewitness testimony.

Perceptions of Confidence Calibration

In the first demonstration of the importance of confidence calibration in evaluations of perceived accuracy, participant-evaluators were presented with a trial summary containing the testimony of a witness to a breaking-and-entering (Tenney et al., 2007). The witness was either confident in his testimony ("yes, sir, absolutely. I'm certain of it.") or not confident in his testimony ("no, sir, I am not certain of it.") and the

3

witness's testimony either contained an error (the time the witness saw the culprit leave the room) or contained no errors. When the witness did not make an error, evaluators perceived the confident witness to be more credible than the unconfident witness. However, when the witness made an error in his testimony, evaluators perceived the unconfident witness to be more credible than the confident witness. The authors reasoned that evaluators perceived the highly confident witness as having poorly calibrated confidence when he made an error whereas they perceived the less confident witness as having well-calibrated confidence when he made an error.

The second study expanded on those findings by pitting the testimony of two witnesses to the same car accident against one another (Tenney et al., 2007). In their testimonies, each witness described several details central to witnessed event (e.g., how the accident occurred, and which vehicle was responsible for the accident) and two details peripheral to the event (the weather and what they had done earlier in the day). One witness was highly confident about every detail recalled in their testimony (both central and peripheral to witnessed event). The other witness was highly confident about all of the central details and one peripheral detail (the weather), but moderately confident about the other peripheral detail (what they had done earlier in the day) (Tenney et al., 2007). After reading the witnesses' testimonies, evaluators were told that each witness correctly recalled what the weather was on that day but that both had incorrectly recalled what they had done earlier in the day. This "calibration information" provided evaluators with information about the witness's confidence calibration: One witness was highly confident about all details (and therefore made an error with high confidence) while the other witness was moderately confident about the erroneous detail (and therefore made an

error with low confidence). This communicated that the former witness was poorly calibrated while the latter witness was well-calibrated.

Evaluators in this study rated the credibility of each witness before and after calibration information was presented. Prior to receiving the calibration information, evaluators perceived the highly confident witness to be more credible than the moderately confident witness. However, once the calibration information was presented, participants perceived the moderately and highly confident witnesses to be similarly credible (Tenney et al., 2007). Of direct relevance to the current research, the calibration information reduced evaluators' belief of the high-confidence witnesses to a greater extent than for the low-confidence witness—the same pattern observed in the witnessing condition context research (Lebensfeld & Smalarz, 2022).

The next study, study one from Tenney et al. (2008), was designed to eliminate the alternative explanation that the finding from the previous study could have been the result of confidence alone as opposed to confidence calibration. The study design was similar to study two from Tenney et al. (2007) but included additional conditions where the moderately confident witness was poorly calibrated, such that they incorrectly recalled a detail they were highly confident about and correctly recalled the detail they were less confident about. In this study, similarly to the first, participants perceived the well-calibrated, moderately confident witness to be significantly more credible than the poorly calibrated, highly confident witness after receiving calibration information (Tenney et al., 2008). This finding was then replicated in the fourth and final study, which used a similar paradigm to the second study, but with a different testimony error (their identification of one of the passengers involved in the car accident) and an addition

following the calibration information manipulation. Specifically, evaluators were provided a justification for the witness's testimony error—namely, that the mis-identified passenger had an identical twin who was likely to have been the true passenger. Upon receiving this justification information, the confidence calibration information no longer reduced perceptions of the eyewitness's credibility, likeability, or honesty and returned perceptions to the level they were pre-calibration. The justification demonstrated that the highly confident witness was not poorly calibrated because their error was justified. Overall, therefore, when a witness is known to have made an unjustified high-confidence error, that witness is perceived to be over-confident, and this over-confidence reduces perceived credibility.

Linking Perceptions of Confidence Calibration to the Eyewitness Identification Literature

An inspection of the eyewitness literature reveals that perceptions of confidence calibration might provide a parsimonious account of multiple findings in the literature that, to date, have not been conceptually linked. One such finding that might operate through perceptions of confidence calibration is the featural justification effect. Featural justification refers to when a witness makes an identification, states their confidence, and then justifies their decision by stating that they recognize a particular feature in the suspect (Dodson & Dobolyi, 2015, 2017). For example, a witness might say "I remember his eyes," or "I remember the shape of his nose" (Dodson & Dobolyi, 2015). In the first studies examining the effect of featural justifications on evaluations of eyewitnesses, Dodson and Dobolyi (2015) presented participants, who were instructed to pretend they were a police officer, with a series of lineups along with the verbal confidence statements

of the witnesses who viewed each lineup. The studies manipulated whether the eyewitness was highly confident or moderately confident in their identification decision as well as the presence of a featural justification statement. Based on the provided information, participants were asked to translate the verbal confidence statements to numeric confidence levels. The presence of a featural justification statement decreased perceptions of eyewitness confidence, an effect that was especially pronounced for highly confident compared to moderately confident witnesses (Dodson & Dobolyi, 2015, 2017). This pattern has been replicated multiple times, including in a study which measured perceptions of identification accuracy instead of perceptions of confidence (Dodson & Dobolyi, 2017). Thus, providing a featural justification for an identification reduces perceptions the eyewitness's confidence and accuracy, and it does so to a greater extent for highly confident than for moderately confident eyewitnesses.

In each of these featural justification studies, the researchers also included an unobservable justification condition (e.g., "He is very familiar," "He has the same expression as before") and found that providing *unobservable* justifications did not reduce evaluations of eyewitness accuracy (Dodson & Dobolyi, 2015, 2017). Essentially, the researchers theorized that it was only when evaluators could independently evaluate the validity of the witness's justification that the justification might change their belief of the eyewitness. Specifically, for an observable justification, it is possible for evaluators to disagree with the diagnosticity, or distinctiveness, of that feature (Dodson & Dobolyi, 2015, 2017). This may then lead evaluators to doubt the ability of the witness to accurately identify the culprit. This theory, known as the perceived diagnosticity account, was further supported by later research which found that when the feature described by

7

the witness was in fact a distinctive feature, then the negative effect of featural justification was greatly reduced for both highly and moderately confident witnesses (Cash & Lane, 2018).

However, the perceived diagnosticity account does not provide a satisfactory explanation for why the featural justification effect is stronger for highly confident witnesses than for moderately confident witnesses. I suggest that perceptions of confidence calibration may underlie the differential effects of featural justifications on high and moderately confident witnesses. Specifically, it is possible that when an evaluator perceives the feature described in a witness's justification to be non-diagnostic of the identified individual and the witness made a high-confidence identification, the evaluator may believe the witness is overly confident. Meanwhile, if the witness made a moderate-confidence identification with the same non-diagnostic justification, the evaluator may believe the witness is appropriately confident. In other words, the non-diagnostic feature would act in the same way testimony errors did in the confidence calibration research (Tenney et al., 2007; Tenney et al., 2008). This would produce an effect wherein a non-diagnostic featural justification is more detrimental to perceptions of a highly confident witness than a moderately confident witness (Cash & Lane, 2018).

Another phenomenon in the eyewitness literature that may be explained by perceptions of confidence calibration involves evaluations of eyewitnesses who make multiple lineup decisions. Specifically, research investigating the effects of presenting evidence of a prior lineup decision on perceptions of a subsequent lineup decision found that when an eyewitness makes an initial incorrect selection (e.g., identifies a known innocent filler) with high confidence, they are perceived as less accurate than if they had

8

done so with low confidence (Smalarz et al., 2020). The authors theorized that evaluators perceive those who make an initial high-confidence identification error as having poorly calibrated confidence, thus evaluators are less inclined to believe a subsequent high-confidence identification. Meanwhile, those who make an initial low-confidence identification error or an initial high-confidence (correct) rejection are perceived as well calibrated, thus evaluators are more inclined to believe a subsequent high-confidence identification. This suggests that perceptions of confidence calibration can influence evaluators' judgements of lineup accuracy, at least under certain circumstances.

Overall, therefore, perceptions of confidence calibration might underlie a variety of effects in eyewitness literature. In the current research, I investigated whether perceptions of confidence calibration specifically underlie the differential effects of contextual information about witnessing conditions on the perceived accuracy of high- and moderate-confidence identifications.

The Current Research

The current research was designed to test a theoretical mechanism to explain why contextual information about poor witnessing conditions has a stronger influence on perceptions of highly confident eyewitnesses than on perceptions of moderately confident eyewitnesses. I specifically investigated whether calibration information—which reveals that the witness is well-calibrated—moderates the effect of contextual information about poor witnessing conditions on evaluations.

To investigate this question, I used a paradigm similar to a paradigm used in the confidence calibration literature (Tenney et al., 2007; Tenney et al., 2008). Participants in the current study were first presented with a witness's testimony transcript in which the

9

witness reported having witnessed a crime and described what she saw, including the quality of the witnessing conditions, and that she identified the suspect from a lineup with either high or moderate confidence. I reasoned that participants would perceive a witness who experienced poor witnessing conditions but made a high-confidence identification as poorly calibrated (i.e., overconfident), whereas they would perceive a witness who experienced poor witnessing conditions but made a moderate-confidence identification as well-calibrated (i.e., appropriately confident).

In addition to her testimony about central details of the crime, the witness recalled some peripheral details about the crime with high confidence and one peripheral detail with moderate confidence—these details would be referenced later to manipulate participants' perceptions of the witness's confidence calibration. Specifically, after reading the witness's testimony, participants learned that further investigation confirmed that the witness was accurate about her high-confidence peripheral details and inaccurate about her moderate-confidence peripheral details. Thus, this information revealed that the witness's confidence was well-calibrated with her accuracy. The same calibration information was presented to all participants in the current study. Therefore, there was no condition in the current research in which the witness was revealed to be poorly calibrated, unlike in Tenney et al. (2007; 2008). If perceptions of confidence calibration drive the disproportionate effect of witnessing condition information on highly confident versus moderately confident witnesses, then providing evidence to evaluators that the witness is *well-calibrated* should counteract the effect.

In summary, I reasoned that if perceptions of confidence calibration do underlie the effect of contextual information about witnessing conditions—such that knowing the

10

witness experienced poor witnessing conditions leads evaluators to doubt the witness's ability to identify the culprit confidently and correctly (Lebensfeld & Smalarz)—then providing information that conveys the witness is well-calibrated should eliminate or attenuate the effect of contextual information. Therefore, I hypothesized that confidence calibration information would reduce the impact of information about poor witnessing conditions on evaluations of eyewitness accuracy and that it would do so to a greater extent for highly confident than for moderately confident eyewitnesses. More precisely, the hypothesized effect would be evidenced by a significant three-way interaction of witnessing condition × identification confidence level × calibration information such that the interaction of witnessing condition × identification confidence level would be weaker post-calibration than pre-calibration. Pre-calibration, the simple main effect of poor witnessing conditions on perceptions of identification accuracy would be greater for high-confidence witnesses than for moderate-confidence witnesses, replicating Lebensfeld & Smalarz. Post-calibration, the simple main effect of poor witnessing conditions would be similar for both high- and moderate-confidence witnesses. This would suggest that witnesses who make high-confidence identifications after experiencing poor witnessing conditions are perceived to have poor confidence calibration, similar to how a witness would be perceived if they made a high-confidence error (Tenney et al., 2007; Tenney et al., 2008), unless evaluators are given information confirming that they are well-calibrated.

CHAPTER 2

METHOD

All procedures, hypotheses, and analyses were preregistered and are available on the Open Science Framework https://tinyurl.com/tclosf.

Participants

A power analysis using a Monte Carlo simulation for a mixed factorial design (Lakens & Caldwell, 2019), determined that 740 total participants were required to detect a three-way interaction between witnessing conditions (good vs. poor), identification confidence level (high vs. moderate), and calibration information (absent vs. present) with at least 90% power and an alpha level of .05. A total of 862 online participants completed the study, 652 of whom were undergraduate students recruited from SONA and 210 of whom were recruited from Amazon Mechanical Turk (Mturk). As compensation for participation, student participants received course credit and Mturk participants each received $1.50 upon completion of the survey. After exclusions (discussed in the Results section) the final sample size was 752 participants.

Of these 752 participants, 416 identified as White or Caucasian, 166 identified as Hispanic or Latinx, 58 identified as Asian or Asian-American, 44 identified as Black or African American, 11 identified as Middle-Eastern, 10 identified as Native Hawaiian or other Pacific Islander, three identified as Native American or Alaska Native, 33 identified as other or multiracial, and 11 did not indicate their race/ethnicity. Additionally, 244 participants identified as male, 490 identified as female, 12 responded "other" and six did not indicate their gender. Participants ranged in age from 18 to 72 ($M = 28.09$, $SD = 10.80$) with an average age of 28.

Design

The study used a 2 (witnessing conditions: good vs. poor) × 2 (identification confidence level: high vs moderate) × 2 (calibration information: absent vs. present) mixed factorial design. Witnessing conditions and confidence level were manipulated between-subjects, and calibration information was manipulated within-subjects. The main dependent variable was participants' perceived accuracy of the witness's identification decision.

Measures

Perceptions of the Eyewitness

The study included five dependent measures that were collected at two times points. Participants rated the likelihood that the eyewitness accurately identified the perpetrator on an 11-point scale, from 0 (Definitely Inaccurate) to 10 (Definitely Accurate). They also rated how credible the eyewitness's testimony was, how trustworthy the eyewitness was, and how convincing the eyewitness's testimony was on 11-point scales ranging from 0 (Not at all Credible/Trustworthy/Convincing) to 10 (Extremely Credible/Trustworthy/Convincing). Participants were also asked "…how much additional evidence would you need to convict the defendant?" to which they responded using an 11-point scale from 0 (No Additional Evidence Needed) to 10 (Much More Evidence Needed). While perceived accuracy was the primary dependent measure, the measures of credibility, trustworthiness, convincingness, and additional evidence needed were included to help reduce demand characteristics and suspicion that may have arisen because of the inclusion of a manipulation check measure (explained below), as well as to potentially expand the findings.

13

Attention Check

An attention check asked participants what crime was described in the eyewitness's testimony with the following response options: purse theft (the correct answer), bank robbery, vandalism, and assault.

Manipulation Checks

The study included three manipulation checks, one for each manipulated variable. The confidence calibration manipulation check asked participants to judge how "appropriate" the eyewitness's confidence level was, using an 11-point scale, from 0 (Very Underconfident) to 10 (Very Overconfident) with 5 labeled as "Appropriately Confident." Participants also responded to an identification confidence manipulation check, which asked, "According to her testimony, approximately how confident was the eyewitness in her identification of the defendant?" with a 0 (Not at all Confident) to 100 (Absolutely Confident) response scale. Finally, participants responded to a witnessing-condition manipulation check, which asked, "According to her testimony, approximately how good was the eyewitness's view of the culprit during the crime?" with an 11-point scale, from 0 (Very Poor) to 10 (Very Good).

Materials

The materials for this study, which consisted of a transcript and calibration information, were adapted from Tenney et al. (2008) and Lebensfeld and Smalarz (2022). All materials can be found in Appendix A. The transcript included the testimony of an eyewitness to a crime, specifically a purse theft. In the testimony, the eyewitness began by recalling the date, weather conditions, where the crime took place, and what she did that evening before the crime occurred (she left work early to meet a friend for dinner).

14

As she recalled these details, the eyewitness expressed that she was highly confident about each one, except for the detail regarding what she did earlier in the evening (leaving work early to meet a friend for dinner), which she said she was "not sure of."

Between-subjects confidence and viewing conditions manipulations. The eyewitness went on to explain that she later identified the suspect (now defendant) from a photographic lineup. For the identification confidence manipulation, the attorney asked the witness "How sure are you that [the defendant] is the man you saw commit the robbery?" to which the witness responded with either moderate (e.g., "I'm moderately confident," "I'm kind of sure") or high confidence (e.g., "I'm very certain," "I'm positive").

For the witnessing condition quality manipulation, the attorney asked the eyewitness "How well were you able to see the face of the culprit as you witnessed this?" Half of the participants read that the witnessing conditions were good ("I saw his face very clearly"), while the other half read that the witnessing conditions were poor. For the poor witnessing conditions condition, the eyewitness would either describe that she viewed the crime from a far distance ("I was across the street, around 40 feet away, so I couldn't see his face very well"), that there was dark lighting ("It was dark out and the street was dimly lit, so I couldn't see his face very clearly"), or that the perpetrator was wearing a medical mask (i.e., a disguise; "The culprit was wearing a face mask at the time, so I could only see part of his face"). This was done for stimulus sampling purposes (Wells & Windschitl, 1999). The transcript ended with the attorney asking the eyewitness to repeat the detail that she was unsure about earlier in the testimony (that she left work early to meet a friend for dinner).

Within-subjects calibration manipulation. The calibration information informed participants that all verifiable details from the eyewitness's testimony were investigated, and all but one were found to be correct. Specifically, participants read that further investigation confirmed that the eyewitness had correctly recalled the weather on the day of the crime as well as the street corner on which the crime occurred, but the witness had incorrectly recalled what she had done earlier in the evening (leaving work early to meet a friend for dinner). Critically, this was the detail about which the eyewitness expressed low confidence in her testimony. In other words, the eyewitness had been highly confident about the details she remembered correctly, and less confident about the detail she remembered incorrectly. This was intended to indicate that the eyewitness's confidence was well-calibrated with her accuracy.

Procedure

After consenting to participate, participants read the following instructions about the study: "In this study, you will read testimony from an eyewitness to a purse theft. The theft occurred about a month before the witness provided her testimony. The witness was asked to describe the incident and the events leading up to the incident. For confidentiality purposes, the names of the defendant, eyewitness, and attorney have been removed from the transcript. As you are reading the testimony, pretend that you are a member of a jury. It is important that you read the testimony very carefully, as you will be asked questions about it. If your responses indicate that you did not read the testimony carefully, you risk not receiving compensation for participating."

After reading the instructions, participants advanced to the page containing the testimony transcript. After reading the transcript (and on the same page as the transcript),

participants responded to the pre-calibration measures. As a reminder, participants were asked about their perceptions of identification accuracy, the eyewitness's credibility, her trustworthiness, the convincingness of the eyewitness's testimony, perceptions of the eyewitness's confidence calibration, and how much additional evidence was needed in order to convict the defendant.

Participants were then presented with the calibration information. Immediately following the calibration information (on the same page as the calibration information), there was a copy of the same testimony transcript from the previous page with the post-calibration measures just below that. The post-calibration measures were identical to the pre-calibration measures. Again, participants were asked about perceptions of identification accuracy, the eyewitness's credibility, her trustworthiness, the convincingness of the eyewitness's testimony, how much additional evidence was needed in order to convict the defendant, and perceptions of the eyewitness's confidence calibration.

After completing the post-calibration measures, participants responded to the remaining attention and manipulation checks. The study procedure ended following some basic demographic questions (e.g., gender, age, race), a quick debrief, and compensation.

CHAPTER 3

RESULTS

Preliminary Analyses

Exclusions

Consistent with the pre-registration for this study, participants who failed the attention check ($n = 45$) were excluded from analyses. I had additionally planned to exclude data from participants who submitted the survey too quickly and therefore likely did not thoughtfully consider and respond to the questions. More specifically, I planned to exclude participants who submitted either the page containing the testimony and pre-calibration measures or the page containing the calibration information and post-calibration measures more than two standard deviations below the mean submission time. However, in the final data analysis, two standard deviations below the mean submission time of either page produced negative numbers for both the page containing the testimony and pre-calibration measures ($M = 178.25$, $SD = 210.61$) and the page containing the calibration information and post-calibration measures ($M = 95.60$, $SD = 63.20$). I then attempted to identify outliers using the interquartile range criterion, but it identified only outliers above the mean (i.e., participants who spent much longer than average), which were not a concern here. Therefore, I ultimately established a minimum submission time of 60 seconds for the first page and 40 seconds for the second page. Based on these criteria, which were determined prior to analyzing the data, a total of 65 participants (7.96%) were excluded.
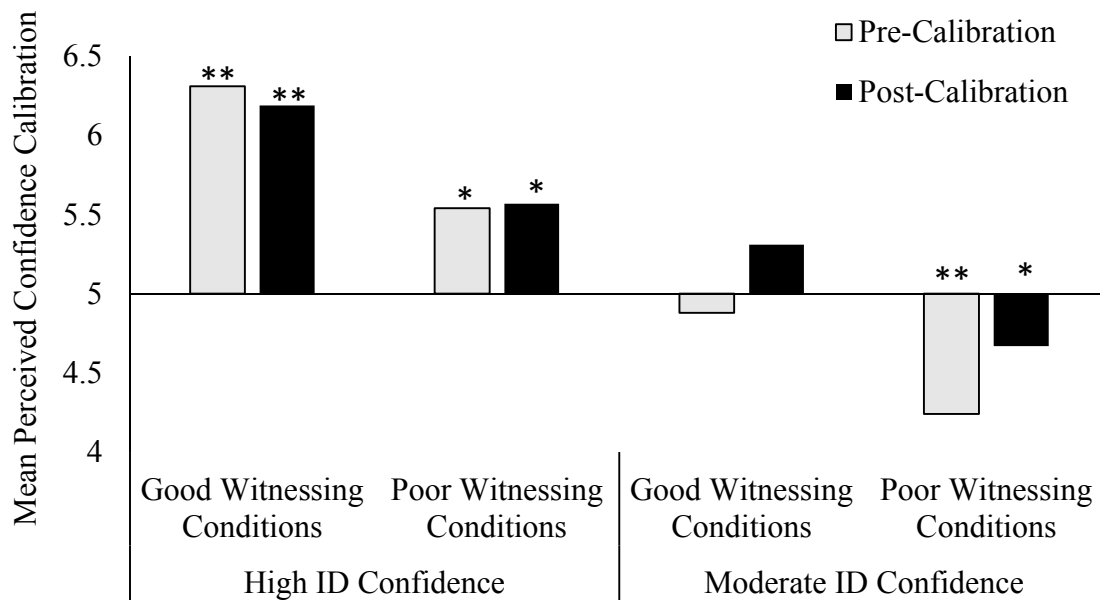
Manipulation Checks

I first assessed participants' responses to the confidence level manipulation check, which asked, "According to her testimony, approximately how confident was the eyewitness in her identification of the defendant?" Participants in the high-confidence condition reported that the witness was significantly more confident in her identification decision ($M = 89.75$, $SD = 15.74$), than did participants in the moderate-confidence condition ($M = 64.22$, $SD = 21.25$), $t(681.73) = 18.68$, $p < .001$. I then assessed participants' responses to the witnessing condition manipulation check, which asked, "According to her testimony, approximately how good was the eyewitness's view of the culprit during the crime?" Participants who read the good witnessing conditions testimony reported that the witness had a significantly better view of the culprit ($M = 8.72$, $SD = 1.71$), than did participants who read poor witnessing condition testimony ($M = 3.51$, $SD = 2.37$), $t(689.93) = 34.62$, $p < .001$. Thus, both between-subjects manipulations were effective.

Finally, I assessed participants' responses to the confidence-calibration manipulation check, which asked, "How appropriate is the eyewitness's level of confidence in her testimony?" and used an 11-point scale, from 0 (Very Underconfident) to 10 (Very Overconfident) with 5 labeled as "Appropriately Confident." Specifically, I conducted a series of one-sample t-tests for each of the eight groups in the factorial design to determine which of these groups were perceived as significantly poorly calibrated (either over- or under-confident) or whether they were perceived as well-calibrated (not significantly different from the scale midpoint). This series of statistical tests revealed that participants perceived each of the witnesses to be significantly poorly calibrated both pre-calibration and post-calibration information, except for the group that

read the testimony of the moderately confident witness who experienced good witnessing

conditions (see Table 1 and Figure 1). Specifically, participants perceived witnesses who

made high-confidence identifications as significantly overconfident both pre- and post-

calibration, regardless of witnessing conditions. Participants also perceived the witness

who made a moderately confident identification and experienced poor witnessing

conditions as underconfident both pre- and post- calibration. However, participants

perceived the moderately confident witness who experienced good witnessing conditions

to be appropriately confident, or not significantly greater or less than five on the scale,

both before and after the calibration information had been presented.

Figure 1

Perceived Confidence Calibration: Mean Differences from "Appropriately Confident"

Witnessing Condition Quality and Identification Confidence Level

* *p* < .05, ** *p* < .001

Taken together, these findings suggest that the confidence calibration

manipulation was ineffective. Had the manipulation worked as intended, participants

20

should have perceived all witnesses to be well-calibrated post-calibration, regardless of their perceptions pre-calibration. In addition, based on my theory about witnessing condition quality informing perceptions of a witness's confidence calibration, participants should have perceived the high-confidence witness who experienced good witnessing conditions and the moderate-confidence witness who experienced poor witnessing conditions as appropriately confident pre-calibration (i.e., well-calibrated). Furthermore, participants should have perceived the high-confidence witness who experienced poor witnessing conditions and the moderate-confidence witness who experienced good witnessing conditions as significantly poorly calibration (over- and under-confident, respectively).

Table 1

Confidence Calibration Manipulation Check Statistics

	Pre-Calibration			Post-Calibration		
	M	*t(df)*	*p*	*M*	*t(df)*	*p*
Good Witnessing Conditions & High Confidence	6.31	8.92(186)	< .001	6.19	8.87(186)	< .001
Good Witnessing Conditions & Moderate Confidence	4.88	-0.72(184)	.470	5.31	1.92(184)	.056
Poor Witnessing Conditions & High Confidence	5.54	2.86(193)	.004	5.57	3.27(193)	.001
Poor Witnessing Conditions & Moderate Confidence	4.24	-5.03(185)	< .001	4.67	-2.21(185)	.028

Because this confidence calibration manipulation check did not show the anticipated pattern, the pre-registered exploratory analyses involving the confidence calibration measure as a mediator were not performed. Instead, I performed other exploratory analyses in order to determine why the manipulation did not work as intended. Specifically, I used the same analysis used to examine the experimental measures: a linear mixed-effects model (lmer in R) to test the effects of the experimental manipulations (confidence level, witnessing conditions, and calibration information) on the perceived confidence calibration measure, with participants specified as a random effect. This analysis revealed a significant main effect of confidence level, such that participants believed witnesses who made high-confidence identifications were over-confident ($M = 5.90$, $SD = 2.27$) while witnesses who made moderate-confidence identifications were underconfident ($M = 4.77$, $SD = 2.16$), $F(1, 748) = 61.13$, $p < .001$, $d = 0.51$, 95% CI [0.38, 0.64]. Additionally, there was a significant main effect of witnessing condition quality such that participants believed witnesses who had experienced good witnessing conditions were over-confident ($M = 5.67$, $SD = 2.15$), while witnesses who had experienced poor witnessing conditions were seen as more appropriately confident ($M = 5.00$, $SD = 2.37$), $F(1, 748) = 21.51$, $p < .001$, $d = 0.30$, 95% CI [0.17, 0.44]. There was a significant main effect of confidence calibration information such that participants believed witnesses were more overconfident post-calibration information ($M = 5.44$, $SD = 2.20$) than pre-calibration information ($M = 5.24$, $SD = 2.37$), $F(1, 748) = 7.85$, $p = .01$, $d = 0.09$, 95% CI [0.03, 0.15].

There was also a significant confidence level $\times$ confidence calibration information interaction, $F(1, 748) = 11.73$, $p < .001$ (Figure 2). Simple main effects analyses revealed

that when a witness was highly confident, participants believed that the witness was similarly calibrated both pre-calibration information ($M = 5.11$, $SD = 2.36$) and post-calibration information ($M = 5.47$, $SD = 2.18$), $t(748) = 0.42$, $p = .67$. However, when a witness was moderately confident, participants believed that the witness was significantly better calibrated post-calibration information ($M = 4.99$, $SD = 2.13$) than pre-calibration information ($M = 4.56$, $SD = 2.17$), $t(748) = 4.40$, $p < .001$, $d = 0.20$, 95% CI [0.11, 0.28]. All other interactions were not significant (see Table 2).

Figure 2

Perceived Confidence Calibration: Identification Confidence Level × Confidence
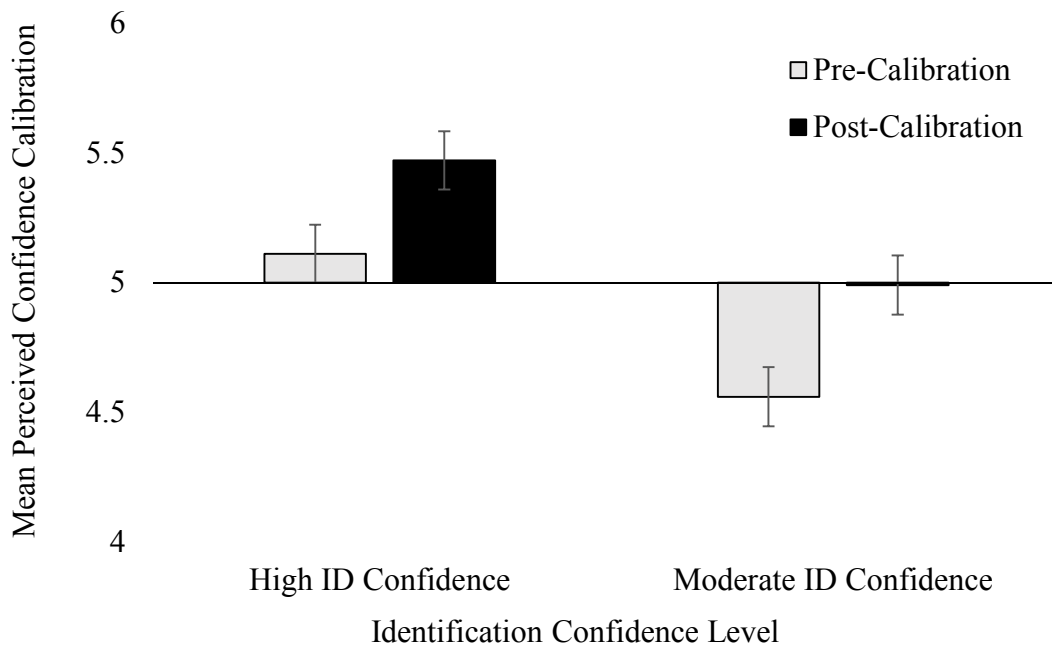
Calibration Information

Table 2

Perceived Confidence Calibration Model Statistics

	F	p	d	95% CI
Confidence Level	61.13	< .001	0.51	[0.38, 0.64]
Witnessing Condition	21.51	< .001	0.30	[0.17, 0.44]
Calibration Information	7.85	0.01	0.09	[0.03, 0.15]
Confidence Level × Witnessing Condition	0.04	0.84	--	--
Confidence Level × Calibration Information	11.73	< .001	--	--
Witnessing Condition × Calibration Information	0.36	0.55	--	--
Confidence Level × Witnessing Condition × Calibration Information	0.28	0.60	--	--

Primary Analyses

I used a linear mixed-effects model (lmer in R) to test the effects of the

experimental manipulations (confidence level, witnessing conditions, and calibration

information) on each of the dependent measures, with participants specified as a random

effect. I tested the effects of the manipulations on perceptions of the eyewitness's

identification accuracy along with the other dependent measures (credibility,

trustworthiness, convincingness, and additional evidence) (see Table 3 for a summary of

all marginal means and standard deviations from these analyses). Additionally, because

the different dependent measures were expected to align with one another, a correlation

matrix using Pearson's R correlations is displayed in Table 4.

24

Table 3

Marginal Means and Standard Deviations for All Primary Analyses Models

Measure		Pre-Calibration				Post-Calibration			
		High Confidence		Moderate Confidence		High Confidence		Moderate Confidence	
		Good Conds.	Poor Conds.	Good Conds.	Poor Conds.	Good Conds.	Poor Conds.	Good Conds.	Poor Conds.
Confidence Calibration	M	6.31	5.54	4.88	4.24	6.19	5.57	5.31	4.67
	SD	2.01	2.61	2.23	2.07	1.84	2.44	2.18	2.03
Accuracy	M	6.41	4.03	5.45	3.79	6.56	4.38	5.98	4.24
	SD	1.89	2.17	2.02	2.05	2.11	2.38	2.33	2.28
Credibility	M	6.19	4.39	5.25	4.10	6.47	4.86	5.99	4.85
	SD	2.21	2.23	2.31	2.24	2.30	2.41	2.45	2.40
Trustworth-iness	M	6.51	5.26	5.83	5.26	6.55	5.51	6.24	5.44
	SD	2.14	2.33	2.16	2.42	2.28	2.44	2.42	2.48
Convinc-ingness	M	6.01	4.21	4.90	3.89	6.30	4.63	5.67	4.51
	SD	2.35	2.42	2.50	2.53	2.25	2.51	2.30	2.45
Evidence	M	7.13	8.05	7.57	8.28	6.55	7.70	6.84	7.57
	SD	1.87	1.60	1.91	1.50	2.24	1.79	2.25	2.07

Table 4

Correlation Matrix for Dependent Measures

	Accuracy	Trustworthiness	Convincingness	Credibility	Evidence
Accuracy	1	--	--	--	--
Trust.	0.7**	1	--	--	--
Convinc.	0.77**	0.76**	1	--	--
Credibility	0.79**	0.82**	0.82**	1	--
Evidence	-0.52**	-0.44**	-0.53**	-0.51**	1

$**p < .001$

Perceived Identification Accuracy

As a reminder, participants rated identification accuracy using an 11-point scale, from 0 (Definitely Inaccurate) to 10 (Definitely Accurate). There was a significant main effect of confidence level such that participants rated high-confidence identifications as more likely to be accurate ($M = 5.34, SD = 2.43$) than moderate-confidence identifications ($M = 4.86, SD = 2.34$), $F(1, 748) = 11.11, p < .001, d = 0.21$, 95% CI [0.07, 0.36]. Additionally, there was a significant main effect of witnessing condition quality such that participants perceived a witness's identification to be more accurate when the witness experienced good witnessing conditions ($M = 6.10, SD = 2.13$) than when the witness experienced poor witnessing conditions ($M = 4.11, SD = 2.23$), $F(1, 748) = 196.59, p < .001, d = 0.92$, 95% CI [0.79, 1.05]. Lastly, there was a significant main effect of confidence calibration information such that participants rated identification accuracy higher post-calibration ($M = 5.29, SD = 2.49$) than they did pre-calibration ($M = 4.92, SD = 2.30$), $F(1, 748) = 196.59, p < .001, d = 0.17$, 95% CI [0.11, 0.23].

The only significant interaction for the perceived identification accuracy measure was the confidence level × witnessing condition quality interaction[1], $F(1, 748) = 4.15, p = 0.04$ (Figure 3). Simple main effects analyses revealed that when the witness experienced good witnessing conditions, participants believed that a highly confident witness was significantly more likely to be accurate ($M = 6.48, SD = 2.00$) than a moderately confident witness ($M = 5.72, SD = 2.20$), $t(748) = 3.79, p < .001, d = 0.36$,

---

[1] When participants who submitted either page too quickly were included in analyses, the confidence level × witnessing condition quality interaction became non-significant, $F(1, 813) = 3.34, p = 0.07$.

95% CI [0.17, 0.54]. However, when the witness had experienced poor witnessing conditions, participants believed that a highly confident witness was similarly accurate (*M* = 4.20, *SD* = 2.28) to a moderately confident witness (*M* = 4.01, *SD* = 2.18), *t*(748) = 0.94, *p* = 0.35. Note that this replicated the same pattern found by Lebensfeld and Smalarz (2022).

Figure 3

Perceived Identification Accuracy: Identification Confidence Level × Witnessing Condition Quality
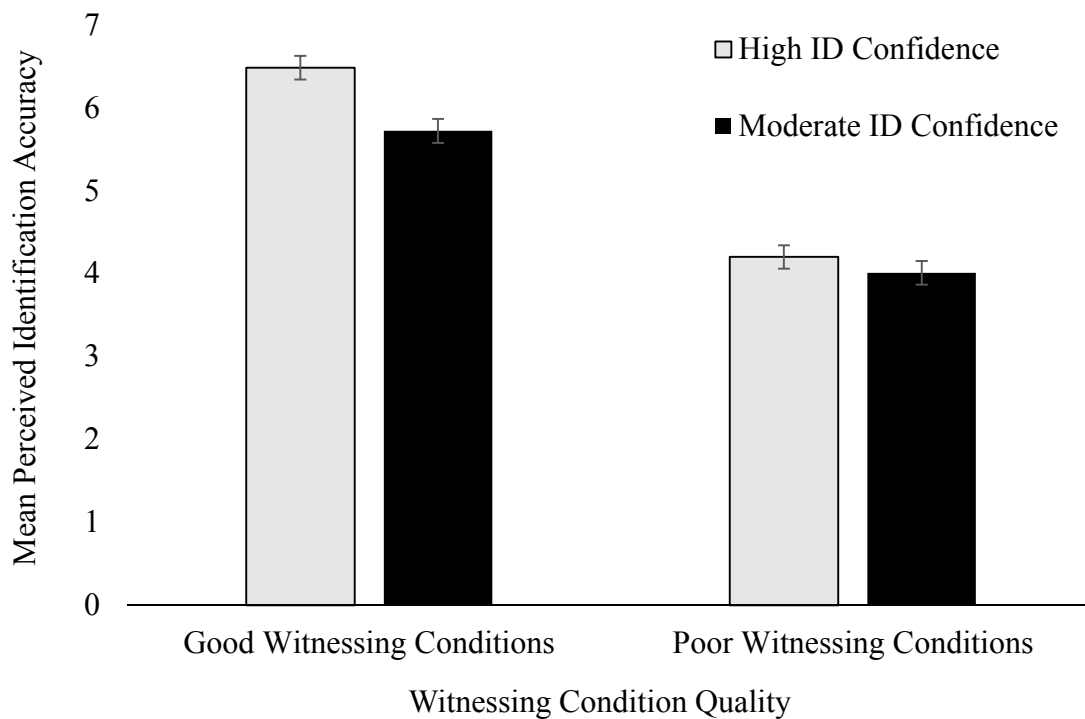
Contrary to the hypothesis, the confidence level × witnessing condition × calibration information interaction was not significant, *F*(1, 748) = 1.10, *p* = 0.29. All other interactions were also non-significant (Table 5).

Table 5

Perceived Identification Accuracy Model Statistics

|	*F*	*p*	*d*	*95% CI*
Confidence Level	11.11	< .001	0.21	[0.07, 0.36]
Witnessing Condition	196.59	< .001	0.92	[0.79, 1.05]
Calibration Information	29.21	< .001	0.17	[0.11, 0.23]
Confidence Level × Witnessing Condition	4.15	0.04	--	--
Confidence Level × Calibration Information	3.05	0.08	--	--
Witnessing Condition × Calibration Information	0.24	0.63	--	--
Confidence Level × Witnessing Condition × Calibration Information	1.10	0.29	--	--

Credibility

As a reminder, participants rated credibility using an 11-point scale, from 0 (Not at all Credible) to 10 (Extremely Credible). Participants' perceptions of the eyewitness's credibility paralleled their perceptions of eyewitness accuracy. There was a significant main effect of confidence level such that participants believed witnesses who made high-confidence identifications were more credible ($M = 5.48$, $SD = 2.44$) than witnesses who made moderate-confidence identifications ($M = 5.05$, $SD = 2.44$), $F(1, 748) = 8.16$, $p = .004$, $d = 0.18$, 95% CI [0.04, 0.31]. Additionally, there was a significant main effect of witnessing condition quality such that participants believed witnesses who had experienced good witnessing conditions were more credible ($M = 5.98$, $SD = 2.36$) than

28

witnesses who had experienced poor witnessing conditions ($M = 4.55$, $SD = 2.34$), $F(1,$
748) = 93.13, $p < .001$, $d = 0.62$, 95% CI [0.49, 0.74]. There was also a significant main
effect of confidence calibration information such that participants rated the witness's
credibility higher post-calibration ($M = 5.54$, $SD = 2.38$) than they did pre-calibration (M
= 4.99, $SD = 2.49$), $F(1, 748) = 46.45$, $p < .001$, $d = 0.24$, 95% CI [0.17, 0.31].

Figure 4

Perceived Credibility: Identification Confidence Level × Confidence Calibration
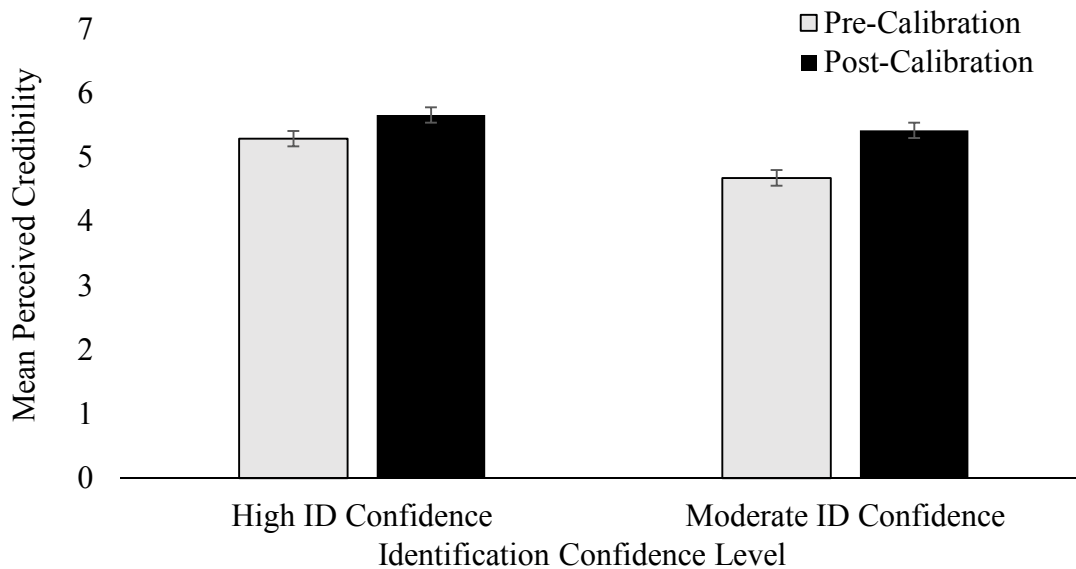
Information

Unlike with accuracy, there was not a significant confidence level × witnessing
condition quality interaction, $F(1, 748) = 3.52$, $p = 0.06$. However, there was a significant
confidence level × confidence calibration information interaction, $F(1, 748) = 5.30$, $p =$
0.02 (Figure 4). Simple main effects analyses revealed that when the witness was highly
confident, participants believed that the witness was significantly more credible post-
calibration ($M = 5.66$, $SD = 2.49$) than they were pre-calibration ($M = 5.29$, $SD = 2.39$),

$t(748) = 3.22$, $p = .001$, $d = 0.16$, 95% CI [0.06, 0.26]. Likewise, when the witness was

moderately confident, participants believed that the witness was significantly more

credible post-calibration ($M = 5.42$, $SD = 2.49$) than they were pre-calibration ($M = 4.68$,

$SD = 2.34$), $t(748) = 6.42$, $p < .001$, $d = 0.32$, 95% CI [0.22, 0.42].Additionally, note that

participants were more sensitive to differences in confidence level before receiving

calibration information, but not after. All other interactions were non-significant (see

Table 6).

Table 6

Credibility Model Statistics

	F	p	d	95% CI
Confidence Level	8.16	.004	0.18	[0.04, 0.31]
Witnessing Condition	93.13	< .001	0.62	[0.49, 0.74]
Calibration Information	46.45	< .001	0.24	[0.17, 0.31]
Confidence Level × Witnessing Condition	3.52	0.06	--	--
Confidence Level × Calibration Information	5.30	0.02	--	--
Witnessing Condition × Calibration Information	0.38	0.54	--	--
Confidence Level × Witnessing Condition × Calibration Information	0.32	0.57	--	--

Trustworthiness

As a reminder, participants rated their perceptions of the witness's trustworthiness

using an 11-point scale, from 0 (Not at all Trustworthy) to 10 (Extremely Trustworthy).

Unlike with accuracy and credibility, there was no significant main effect of confidence

level for the trustworthiness measure[2], $F(1, 748) = 2.92$, $p = .09$. There was a significant main effect of witnessing condition quality such that participants believed witnesses who had experienced good witnessing conditions were more trustworthy ($M = 6.28$, $SD = 2.26$) than witnesses who had experienced poor witnessing conditions ($M = 5.37$, $SD = 2.42$), $F(1, 748) = 34.92$, $p < .001$, $d = 0.39$, 95% CI [0.26, 0.52]. There was also a significant main effect of confidence calibration information such that participants rated the witness's trustworthiness higher post-calibration ($M = 5.93$, $SD = 2.32$) than they did pre-calibration ($M = 5.71$, $SD = 2.45$), $F(1, 748) = 9.46$, $p = .002$, $d = 0.09$, 95% CI [0.03, 0.15]. All other effects were non-significant (see Table 7).

Table 7

Trustworthiness Model Statistics

	F	*p*	*d*	*95% CI*
Confidence Level	2.92	0.09	--	--
Witnessing Condition	34.92	< .001	0.39	[0.26, 0.52]
Calibration Information	9.46	.002	0.09	[0.03, 0.15]
Confidence Level × Witnessing Condition	2.15	0.14	--	--
Confidence Level × Calibration Information	1.11	0.29	--	--
Witnessing Condition × Calibration Information	0.004	0.95	--	--
Confidence Level × Witnessing Condition × Calibration Information	2.44	0.12	--	--

---

[2] When participants who submitted either page too quickly were included in analyses, the main effect of confidence level was significant, such that highly confident witnesses were perceived as more trustworthy ($M = 5.99$, $SD = 2.34$) than moderately confident witnesses ($M = 5.60$, $SD = 2.41$), $F(1, 813) = 6.71$, $p = 0.01$, $d = 0.16$, 95% CI [0.04, 0.29].

Convincingness

As a reminder, participants rated convincingness using on an 11-point scale, from 0 (Not at all Convincing) to 10 (Extremely Convincing). There was significant main effect of confidence level such that participants believed witnesses who made high-confidence identifications ($M = 5.29$, $SD = 2.54$) were more convincing than witnesses who made moderate-confidence identifications ($M = 4.74$, $SD = 2.53$), $F(1, 748) = 11.79$, $p < .001$, $d = 0.22$, 95% CI [0.09, 0.36]. Additionally, there was a significant main effect of witnessing condition quality such that participants believed witnesses who had experienced good witnessing conditions were more convincing ($M = 5.72$, $SD = 2.41$) than witnesses who had experienced poor witnessing conditions ($M = 4.31$, $SD = 2.49$), $F(1, 748) = 79.47$, $p < .001$, $d = 0.58$, 95% CI: [0.45, 0.71]. Lastly, there was a significant main effect of confidence calibration information such that participants believed a witness was more convincing post-calibration ($M = 5.28$, $SD = 2.49$) than pre-calibration ($M = 4.75$, $SD = 2.58$), $F(1, 748) = 46.21$, $p < .001$, $d = 0.22$, 95% CI [0.15, 0.28].

Like with the accuracy measure, there was also a significant confidence level × witnessing condition quality interaction, $F(1, 748) = 4.15$, $p = 0.04$ (Figure 5). Simple main effects analyses revealed that when a witness experienced good witnessing conditions, participants believed that the highly confident witness was significantly more convincing ($M = 6.15$, $SD = 2.30$) than a moderately confident witness ($M = 5.28$, $SD = 2.43$), $t(748) = 3.86$, $p < .001$, $d = 0.36$, 95% CI [0.18, 0.54]. However, when a witness experienced poor witnessing conditions, participants believed that the highly confident witness ($M = 4.42$, $SD = 2.47$) and the moderately confident witness ($M = 4.20$, $SD = 2.51$) were both similarly convincing, $t(748) = 1.02$, $p = .31$.

Figure 5

Convincingness: Identification Confidence Level × Witnessing Condition Quality
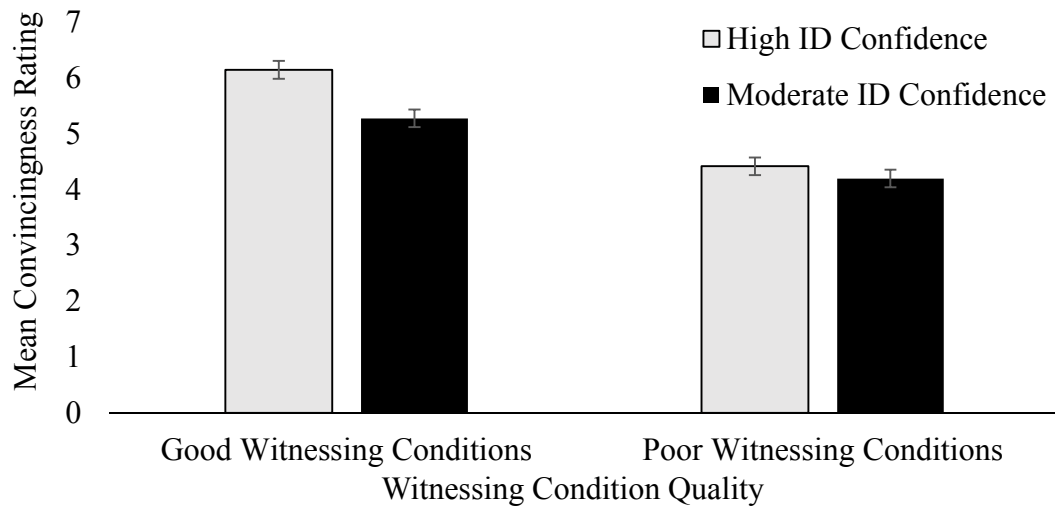
Figure 6

Convincingness: Identification Confidence Level × Confidence Calibration Information
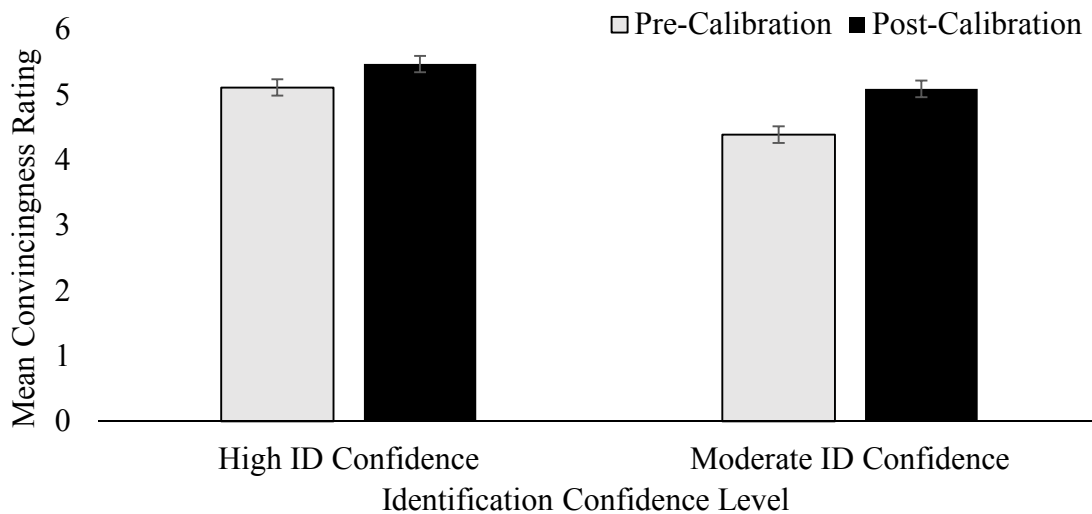
Lastly, similar to the credibility measure, there was a significant interaction of confidence level × confidence calibration information, $F(1, 748) = 5.30$, $p = 0.02$ (Figure 6). Simple main effects analyses revealed that when a witness was highly confident, participants believed that the witness was significantly more convincing post-calibration

($M = 5.47$, $SD = 2.53$) than they were pre-calibration ($M = 5.11$, $SD = 2.55$), $t(748) =$
3.30, $p = .001$, $d = 0.15$, 95% CI [0.06, 0.24]. Similarly, when a witness was moderately
confident, participants believed that the witness was significantly more convincing post-
calibration ($M = 5.09$, $SD = 2.44$), than they were pre-calibration ($M = 4.39$, $SD = 2.56$)
$t(748) = 6.32$, $p < .001$, $d = 0.29$, 95% CI [0.20, 0.38]. All other interactions not discussed
above were not significant (see Table 8).

Table 8

Convincingness Model Statistics

	F	p	d	95% CI
Confidence Level	11.79	< .001	0.22	[0.09, 0.36]
Witnessing Condition	79.47	< .001	0.58	[0.45, 0.71]
Calibration Information	46.21	< .001	0.22	[0.15, 0.28]
Confidence Level × Witnessing Condition	4.10	0.04	--	--
Confidence Level × Calibration Information	4.72	0.03	--	--
Witnessing Condition × Calibration Information	0.005	0.94	--	--
Confidence Level × Witnessing Condition × Calibration Information	0.84	0.36	--	--

Additional Evidence Needed

Lastly, participants indicated how much additional evidence was needed to make
a guilt decision using an 11-point scale from 0 (No Additional Evidence Needed) to 10
(Much More Evidence Needed). Like with trustworthiness, there was no significant main
effect of confidence level for the evidence measure, $F(1, 748) = 2.83$, $p = .09$. However,

there was a significant main effect of witnessing condition quality such that participants who read the testimony of a witness who had experienced good witnessing conditions indicated that they needed less additional evidence ($M = 7.02$, $SD = 2.10$) than those who read the testimony of a witness who had experienced poor witnessing conditions ($M = 7.90$, $SD = 1.77$), $F(1, 748) = 51.46$, $p < .001$, $d = 0.46$, 95% CI [0.33, 0.58]. Lastly, there was a significant main effect of confidence calibration information such that participants needed more additional evidence pre-calibration ($M = 7.76$, $SD = 1.78$) than they did post-calibration ($M = 7.16$, $SD = 2.14$), $F(1, 748) = 78.40$, $p < .001$, $d = 0.31$, 95% CI [0.24, 0.38]. All other effects[3] were non-significant (see Table 9).

Table 9

Additional Evidence Model Statistics

	F	p	d	95% CI
Confidence Level	2.83	0.09	--	--
Witnessing Condition	51.46	< .001	0.46	[0.33, 0.58]
Calibration Information	78.40	< .001	0.31	[0.24, 0.38]
Confidence Level × Witnessing Condition	1.59	0.21	--	--
Confidence Level × Calibration Information	3.69	0.06	--	--
Witnessing Condition × Calibration Information	0.80	0.37	--	--
Confidence Level × Witnessing Condition × Calibration Information	0.60	0.44	--	--

[3] When participants who submitted either page too quickly were included in analyses, the confidence level × confidence calibration information interaction became significant, $F(1, 813) = 75.91$, $p < .001$. There was a significant simple effect such when a witness was highly confident, significantly less evidence was needed post-calibration ($M = 7.13$, $SD = 2.10$) than pre-calibration ($M = 7.56$, $SD = 1.81$), $t(813) = 4.76$, $p < .001$, $d = 0.22$, 95% CI [0.13, 0.31]. When a witness was moderately confident, significantly less evidence was needed when the witness was post-calibration ($M = 7.21$, $SD = 2.19$) than pre-calibration ($M = 7.91$, $SD = 1.78$), $t(813) = 7.59$, $p < .001$, $d = 0.36$, 95% CI [0.27, 0.45].

CHAPTER 4

DISCUSSION

The goal of this study was to test a perceived calibration account of the finding that contextual information about poor witnessing conditions decreases evaluators' sensitivity to differences in confidence level in their judgments of identification accuracy (Lebensfeld & Smalarz, 2022). Specifically, I tested whether demonstrating that a witness is well-calibrated reduces the impact of poor witnessing conditions on evaluations of eyewitness accuracy, particularly for highly confident witnesses. I had predicted that it would reduce the impact based on the theory that information about witnessing conditions is used by evaluators in the same way testimony errors are used by evaluators—to assess the witness's confidence calibration (Tenney et al., 2007; Tenney et al., 2008). When the witnessing conditions were poor, evaluators interpret a witness's high confidence as over-confident, which I hypothesized would occur unless evaluators were given information demonstrating that the witness's confidence was well-calibrated. Contrary to my prediction, there was no significant three-way interaction between confidence level, witnessing conditions, and confidence calibration information for any measure.

My study did, however, successfully replicate the finding that contextual information about poor witnessing conditions reduces evaluators' belief of highly confident witnesses to a greater extent that it reduces belief of moderate-confidence witnesses (Lebensfeld & Smalarz, 2022). Furthermore, my findings not only replicated those of Lebensfeld and Smalarz, but the effect was arguably stronger in the current study. Whereas evaluators in Lebensfeld and Smalarz maintained some sensitivity to

36

differences in eyewitness confidence, in this study, there was no significant difference in perceived accuracy between high- and moderate-confidence witnesses in the context of poor witnessing conditions. One possible explanation that could account for the stronger effect observed here is that Lebensfeld and Smalarz used a within-subjects design while the current study used a between-subjects design. The within-subjects design used previously had evaluators read about multiple witnesses who experienced poor viewing conditions, which may have created the perception that poor witnessing conditions are common, thereby dampening their effect. In the current between-subjects design, however, evaluators read about only one eyewitness's testimony and may have been especially sensitive to information about poor witnessing conditions. Regardless, because the between-subjects design of the current studied more closely resembles real world scenarios in which jurors consider one case at a time (as opposed to making judgements about witnesses from several different cases), the current study expands the generalizability of the finding that contextual information about poor witnessing conditions reduces evaluators' sensitivity to differences in identification confidence.

Implications of the Findings for the Perceived Calibration Account

There are multiple possible reasons why the study failed to find support for the perceived calibration account of the differential effects of witnessing condition information on perceptions of high- and moderate-confidence witnesses. One possibility is that the perceived calibration account simply is not true. If that is the case, it will be necessary, for the purpose of better understanding legal decision-making, to develop an alternative explanation for why contextual information influences evaluators judgments more when judging the identification accuracy of a highly confident witness than a

moderately confident witness. However, because this is the first study to test this confidence calibration theory, it is far too early to draw the conclusion that the theory is false.

A more likely explanation at this point is that my hypothesis was not supported due to the ineffectiveness of the confidence calibration manipulation. Participants perceived that only the witness who experienced good witnessing conditions and expressed moderate confidence was well-calibrated after the calibration information was presented. They perceived the rest of the witnesses to be significantly poorly calibrated. Recall that the calibration information was meant to inform participants that the witness was well-calibrated. Therefore, it should have caused evaluators to perceive the witness as well-calibrated, or at least better calibrated than she was pre-calibration. However, on average, confidence calibration was perceived to be *worse* post-calibration information than pre-calibration information: Witnesses were believed to be more over-confident after confidence calibration information was presented than before it. Had the manipulation in this study worked as expected, the perceived confidence calibration should have been better following the calibration manipulation than before it.

A closer look at this manipulation check data revealed that participants believed highly confident witnesses were consistently overconfident, regardless of calibration information, whereas participants believed moderately confidence witnesses were more appropriately confident post-calibration than pre-calibration. Therefore, only moderately confident witnesses benefitted from the confidence calibration information; this same information made little difference when the witness was highly confident. One potential explanation for why this occurred is that evaluators penalized highly confident witnesses

for making any error. Indeed, some research has shown that any error in a witness's testimony negatively influences evaluators' perceptions of the witness (Lavis & Brewer, 2017). Perhaps evaluators think that an eyewitness should be highly confident only if they never make an error. Because the eyewitness in my study was always revealed to have made a (low-confidence) error, evaluators may have interpreted the witness's high identification confidence as being inappropriately high. However, because there was a main effect of confidence calibration on identification accuracy—such that the calibration information increased perceptions of identification accuracy—it seems unlikely that witnesses were penalized simply for making an error.

Alternatively, the calibration manipulation may have been ineffective in this study because the calibration information only informed participants about the accuracy of peripheral details that would be unrelated to the witness's identification accuracy and identification confidence. For example, being correct or incorrect about details unrelated to the crime (i.e., what the witness did early in the day in the current study) might not be perceived as informative of the witness's confidence calibration for her identification. I chose to use these details in the calibration information because they resemble the details used in previous studies (Tenney et al., 2007; Tenney et al., 2008). However, the previous research, which focused on perceptions of the witness's overall credibility, did not specifically test whether testimony information would influence perceptions of an *identification*. Therefore, the calibration information in this study should have shown how well-calibrated the witness was when recalling central details. This might have demonstrated her encoding capabilities at the time of the witnessed event and her ability to recall the event details. This information could then have been used evaluate her

confidence calibration for her identification decision. While I predicted that each of the five dependent measures (accuracy, credibility, trustworthiness, convincingness, and evidence needed) would produce convergent results, this was not always the case. There were some parallels: For example, the main effects of calibration information and witnessing conditions were consistent across all of the dependent measures.

Notwithstanding the issues with the confidence calibration manipulation, the current findings may provide some insight into evaluators' expectations regarding eyewitness confidence. I found that, in the absence of calibration information, evaluators perceived a highly confident witness as overconfident but perceived a moderately confident witness as underconfident. This contrasts with Tenney et al. (2007; 2008) who reasoned that evaluators assume people are well-calibrated unless they receive evidence that indicates otherwise.

It is also notable in the current findings that highly confident witnesses were perceived as being especially overconfident when they experienced good witnessing conditions while moderately confident witnesses who experienced good witnessing conditions were perceived as appropriately confident. As such, it appears that the witness's good-view statement in the current study may have been interpreted similarly to a statement of high confidence. Therefore, participants may have perceived the witness who made a moderate-confidence identification but reported having had a good view as more measured in her testimony as opposed to a witness who made a high-confidence identification and reported having a good view. According to this line of reasoning, evaluators might perceive people whose confidence is more variable (e.g., sometimes

high, sometimes low) as having better confidence calibration than people who report a consistent confidence level about all aspects of their testimony.

Differences Across Measures

Contextual information about poor witnessing conditions had stronger effects on perceptions of high-confidence witnesses than moderate-confidence witnesses for only two of the measures: judgements of identification accuracy and perceptions of the eyewitness's convincingness. On the one hand, the findings for these two measures replicate and extend the finding from Lebensfeld and Smalarz (2022). On the other hand, it is somewhat surprising that the confidence × witnessing conditions interaction was not significant for evaluators' judgments of the witness's trustworthiness, credibility, or additional evidence needed. These divergent patterns suggest that judgments of an eyewitness's trustworthiness, credibility, and additional evidence needed may be driven by different factors than judgments of an eyewitness's identification accuracy and convincingness. Indeed, the tendency for high-confidence identifications to lead to more favorable perceptions than moderate-confidence identifications was found for all dependent measures except for trustworthiness and perceptions of additional evidence needed, further suggesting that different mechanisms may underlie evaluators' perceptions of these judgments. With regard to trustworthiness, it is possible that trustworthiness judgements were more dispositional while other judgments were based on more situational information and, therefore, more prone to variation as a result of contextual manipulations. Future research will need to investigate different aspects of evaluators' perceptions of eyewitness testimony more systematically (e.g., dispositional vs. situational factors, trustworthiness vs. credibility, etc.).

41

Lastly, I found that calibration information increased participants' perceptions of both credibility and convincingness for highly and moderately confident witnesses. In particular, the moderately confident witness's credibility and convincingness improved post-calibration more than highly confident witness's credibility and convincingness. In addition, though the interaction was not significant for the perceived identification accuracy measure, the mean identification accuracy ratings formed the same pattern. These findings would seem to indicate that the calibration manipulation was more effective for moderately confident witnesses than for highly confident witnesses. Perhaps evaluators are skeptical of witnesses who make high-confidence identifications due to strong perceptions of over-confidence, and therefore it is more difficult to shift their perceptions even when it is confirmed that the witness is well-calibrated than it is for perceptions of witnesses who make moderate-confidence identifications.

Limitations

The generalizability of this study was at least partially limited by the use a majority student sample. Students in this study were sampled from a pool which contains students studying forensic psychology and similar topics which could influence their perceptions of eyewitnesses in ways that differ from the average juror. For example, these participants may have been taught that eyewitness confidence is not a reliable predictor of accuracy regardless of witnessing conditions (Deffenbacher, 1980, 2008) and misidentifications has lead to numerous wrongful convictions (Innocence Project). Furthermore, I did not screen for participants' prior knowledge or skepticism about eyewitnesses and eyewitness memory in general. Participants with specialized knowledge or skepticism regarding eyewitness memory might have different expectations regarding

42

eyewitness confidence and perceive eyewitness testimony differently than the average juror. Future research that measures participants' prior knowledge of eyewitness issues could help researchers quantify and control for these effects.

In addition, it is possible that the manipulation of good witnessing conditions in the current study could have been viewed as an expression of high confidence on its own, thus influencing perceptions of confidence calibration pre-calibration. While witnesses who experienced poor witnessing conditions always specified why they had a poor view (e.g., because it was dark out, because they were far away), witnesses who experienced good conditions always simply stated that they had a good view of the culprit (i.e., "I saw his face very clearly"). Because participants weren't given information that corroborated the witness's good view, they might have perceived witnesses' "good view" testimony as posturing. This would explain why participants perceived witnesses who experienced good witnessing conditions as more overconfident than witnesses who experienced poor witnessing conditions. Future research should have the witness provide an explanation for her good witnessing conditions (e.g., "The street was very well lit, so I could see his face very clearly," "He stood right in front of me, so I could see his face very clearly") and/or a manipulation in which the appropriate attorney provides corroborating evidence of the witness's claim that she had a good view of the culprit, so that this claim is not mistaken for overconfidence.

CHAPTER 5

CONCLUSION

Even though I failed to find support for my hypothesis, it is still too early to dismiss the possibility that perceptions of confidence calibration underlie the differential effects of contextual witnessing condition information on evaluations of high- and moderate-confidence eyewitnesses. Instead, my analyses suggest that the current null effects of confidence calibration were likely due to an ineffective confidence calibration manipulation. Therefore, future research, which I plan to pursue, should aim to design a confidence calibration manipulation that is more focused on perceptions of identification decisions. Potential improvements include providing calibration information which focuses on more central details (as opposed to peripheral details) and/or details which could logically be affected by poor witnessing conditions (e.g., the witness's recollection of clothing, colors, description of the culprit, etc.). Future research should also further investigate whether evaluators have expectations for a witness's confidence or if this finding was an artifact of the current research and the participants sampled.

While this research did not support its primary hypothesis, it replicated and expanded the previous finding that the influence of contextual information about poor witnessing conditions is especially strong when the witness is highly confident in an identification decision. This finding has now been replicated multiple times, using different paradigms. Researchers should continue to investigate the true effects poor witnessing conditions have on eyewitnesses' identification decisions. This line of research will aid the legal system in determining when poor witnessing conditions may (or may not) affect eyewitness reliability, how evaluators such as police, attorneys,

judges, and jurors perceive those witnesses and why they perceive them that way. Finally,

this knowledge can be used to determine what, if any, intervention (e.g., jury instructions)

or strategies (e.g., expert testimony) may be necessary to ensure that evaluators'

judgements of eyewitnesses align with the eyewitness's true reliability.

REFERENCES

Cash, D. K., & Lane, S. M. (2017). Context influences interpretation of eyewitness confidence statements. *Law and Human Behavior*, *41*(2), 180–190. https://doi.org/10.1037/lhb0000216

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior*, *4*(4), 243–260. https://doi.org/10.1007/BF01040617

Deffenbacher, K. A. (2008). Estimating the impact of estimator variables on eyewitness identification: A fruitful marriage of practical problem solving and psychological theorizing. *Applied Cognitive Psychology*, *22*(6), 815–826. https://doi.org/10.1002/acp.1485

Dodson, C. S., & Dobolyi, D. G. (2015). Misinterpreting eyewitness expressions of confidence: The featural justification effect. *Law and Human Behavior, 39*(3), 266.

Dodson, C. S., & Dobolyi, D. G. (2017). Judging guilt and accuracy: highly confident eyewitnesses are discounted when they provide featural justifications. *Psychology, Crime & Law, 23*(5), 487-508

Innocence Project. (n.d.). *Eyewitness Identification Reform.* https://innocenceproject.org/eyewitness-identification-reform/

Lakens, D., & Caldwell, A. R. (2019). Simulation-based power-analysis for factorial ANOVA designs. 1–11. https://doi.org/10.31234/osf.io/baxsf

Lavis, T., & Brewer, N. (2017). Effects of a proven error on evaluations of witness testimony. *Law and Human Behavior, 41*(3), 314-323. http://dx.doi.org/10.1037/lhb0000217

Lebensfeld, T. C., & Smalarz, L. (2022, March). *Estimator Variables Harm Eyewitness Credibility, Especially When the Eyewitness is Highly Confident.* [Paper presentation]. American Psychology-Law Society Annual Conference, Denver, CO.

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

Smalarz, L. (2021). Suspect bias: A neglected threat to the reliability of eyewitness identification evidence. *Journal of Applied Research in Memory and Cognition*, *10*(3), 356–362. https://doi.org/10.1016/j.jarmac.2021.06.005

46

Smalarz, L., Douglass, A. B., & Chang, A. (2020). Eyewitness-identification decisions as Brady material: Disclosing information about prior decisions affects evaluations of eyewitnesses. *Psychology, Public Policy, and Law, 26*(1), 10-21. http://dx.doi.org/10.1037/law0000213

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science*, *18*(1), 46–50. https://doi.org/10.1111/j.1467-9280.2007.01847.x

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, *44*(5), 1368–1375. https://doi.org/10.1016/j.jesp.2008.04.006

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125. https://doi.org/10.1177/01461672992512005

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10-65.

APPENDIX A

TESTIMONY TRANSCRIPT AND CALIBRATION INFORMATION

Testimony

Attorney: [Witness] please tell me everything you remember from the night of October
7th.

Witness: I remember what happened that night very well, but I remember some things
better than others. For example, I know for a fact that it was especially cold that
night, and that it had rained earlier in the evening, because I was wearing my
warm rain coat at the time. I'm extremely sure of that. What I'm not sure of is if
that was the evening I had left work an hour early to have dinner with a friend
who was visiting from out of town. I think we met up that day, but it's possible
I'm wrong. Anyway, I am certain that the incident occurred as I was standing at
the crosswalk on the corner of 32nd Street and 7th Avenue on my way home.
There was a man and an elderly woman who were also waiting to cross the
street. As the elderly woman was about to begin crossing the street, the man
grabbed the purse out of her hand and ran off down the street with it. After
making sure the woman was unharmed, I called the police and told them what
happened.

Attorney: How well were you able to see the face of the culprit as you witnessed this?

Witness: I saw his face very clearly.

I was across the street, around 40 feet away, so I couldn't see his face very well.
It was dark out and the street was dimly lit, so I couldn't see his face very
clearly.

The culprit was wearing a face mask at the time, so I could only see part of his
face.

Attorney: And you were contacted a week later by the detective investigating the case, is

that right?

Witness: Yes, he wanted to show me a photo lineup of six men and asked if I recognize

anyone as being the thief.

Attorney: And did you recognize anyone?

Witness: Yes, I recognized [the defendant].

Attorney: How sure are you that [the defendant] is the man you saw commit the robbery?

Witness: I'm positive.

I'm very certain.

I'm moderately confident.

I'm kind of sure.

Attorney: You said you weren't sure whether the event took place the night you left work

early to meet a friend for dinner, is that right?

Witness: Yes, as I said, I remember some things better than others. I'm not sure if that

was the day I left work early.

Attorney: Thank you. No more questions.

--

Calibration Information

After the witness gave her testimony, further investigation was conducted. From

interviewing other sources, it turns out that all but one of the details in the witness's

testimony were correct.

Rainfall and cold temperatures were recorded on October 7th, the day of the crime, and the incident occurred on the corner of 32nd Street and 7th Avenue. All of these facts were consistent with the witness's testimony.

However, October 7th was not the evening the witness left work early to meet up with her friend. According to records obtained from the witness's boss, October 5th was the evening the witness left work early. The witness's friend also provided a credit card receipt from the dinner, which confirmed that the dinner had taken place on October 5th, not on the night of the crime.

APPENDIX B

IRB APPROVAL

EXEMPTION GRANTED

Laura Smalarz
NCIAS: Social and Behavioral Sciences, School of (SSBS)
-
Laura.Smalarz@asu.edu

Dear Laura Smalarz:

On 5/4/2022 the ASU IRB reviewed the following protocol:

Type of Review:	Modification / Update
Title:	Evaluations of verbal and numerical eyewitness confidence in the context of estimator variables
Investigator:	Laura Smalarz
IRB ID:	STUDY00012728
Funding:	Name: (Unspecified)
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Confidence in Context IRB Protocol, Category: IRB Protocol; • Consent_Form_ConfidenceInContext, Category: Consent Form; • Updated Recruitment Material, Category: Recruitment Materials;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 5/4/2022.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

REMINDER - - Effective January 12, 2022, in-person interactions with human subjects require adherence to all current policies for ASU faculty, staff, students and visitors. Up-to-date information regarding ASU's COVID-19 Management

Strategy can be found here. IRB approval is related to the research activity involving human subjects, all other protocols related to COVID-19 management including face coverings, health checks, facility access, etc. are governed by current ASU policy.

Sincerely,

IRB Administrator cc:

Taylor Lebensfeld