5-1-2020

The Effect of Viewing Distance on Empirical Discriminability and the Confidence–Accuracy Relationship for Eyewitness Identification

Robert F. Lockamyeir

Curt A. Carlson

Alyssa R. Jones

Maria A. Carlson

Dawn R. Weatherford

The effect of viewing distance on empirical discriminability and the confidence-accuracy relationship for eyewitness identification

Robert F. Lockamyeir[1], Curt A. Carlson[1], Alyssa R. Jones[1], Maria A. Carlson[1], & Dawn R. Weatherford[2]

[1]Texas A&M University – Commerce

[2]Texas A&M University – San Antonio

Corresponding author:
Curt A. Carlson
PO Box 3011
Commerce, TX 75429
903-468-8723
curt.carlson@tamuc.edu

The authors declare that they have no conflicts of interest.

Abstract

The distance from which an eyewitness views a perpetrator is a critical factor for eyewitness identification, but has received little research attention. We presented three mock-crime videos to participants, varying distance to three perpetrators (3, 10, or 20 m). Across two experiments, increased distance reduced empirical discriminability in the form of a mirror effect, such that correct identifications decreased while false identifications increased. Moreover, high confidence identifications were associated with high accuracy at 3 m (Experiment 1 and 2) and 10 m (Experiment 2), but not at 20 m. We conclude that eyewitnesses may be less likely to identify a perpetrator viewed at a distance, and also more likely to falsely identify an innocent suspect. Furthermore, there may be certain boundary conditions associated with distance and the impact it has on the confidence-accuracy relationship. More research is needed to elucidate the effect of estimator variable manipulations on the confidence-accuracy relationship.

Keywords: eyewitness identification, distance, empirical discriminability, confidence-accuracy relationship

The effect of viewing distance on empirical discriminability and the confidence-accuracy relationship for eyewitness identification

The number of individuals who have been exonerated based on DNA evidence continues to rise, approaching 400 in the US, and mistaken eyewitness identification (ID) was a factor in approximately 70% of cases (Innocence Project, 2020). In 1972, the United States Supreme Court implemented five criteria to guide judges and jurors in assessing the reliability of an eyewitness. These are known as the Biggers' criteria: (a) view, (b) attention, (c) detail, (d) time, and (e) certainty of a witness' report and suspect ID (Neil v. Biggers, 1972). Our study focused on the first criterion, in the form of the distance between the eyewitness and perpetrator at the time of a crime. Of the estimator variables that have been investigated (e.g., cross-race, weapon presence, exposure time), viewing distance has received surprisingly little research attention. We will first describe the most relevant studies on this issue, followed by how our study fills important gaps remaining in the eyewitness ID literature regarding this variable.

Effect of Distance on Face Recognition Accuracy and Response Bias

Loftus and Harley (2005) proposed the distance-as-filtering hypothesis, which is based on two observations of the human visual system: (a) the image processing mechanism becomes progressively poorer through higher spatial frequencies (i.e., fine-grained details), and (b) the image's spatial frequency spectrum is scaled down in a way that is inversely proportional to distance. In other words, with the loss of fine-grained details, face perception becomes more reliant upon coarse-grained details (i.e., low spatial frequency details). So, as viewing distance increases, the ability to accurately encode a face decreases. McKone (2009) found an optimal face recognition range of approximately 2-10 m.

Wagenaar and Van der Schrier (1996) were the first to investigate the effect of (simulated) distance on eyewitness ID by testing with lineups (target-present or -absent). Using variations in picture size, they simulated seven distances (3, 5, 7, 12, 20, 30, 40 m) along with differences in illumination. They found that recognition performance decreased as viewing distance increased, and from these data they suggested a diagnostic rule of thumb called the “rule of 15,” referring to a steep decline in performance once viewing distance exceeds 15 m.

However, this rule of 15 has not been consistently supported with additional research. Lindsay, Semmler, Weber, Brewer, and Lindsay (2008) and Lampinen, Erickson, Moore, and Hittson (2014) supported the general pattern of recognition accuracy being inversely related to viewing distance, coming in the form of a consistent decline in performance as distance increased. One important difference in the results of these two studies involves response bias rather than accuracy. Lindsay et al. concluded that the overall likelihood of choosing was not affected by distance, whereas Lampinen et al. found more liberal choosing at longer distances. The difference may be attributable to how these two studies assessed eyewitness performance, as Lindsay et al. used lineups and Lampinen et al. used a yes/no recognition test, but it is unclear why testing method would impact response bias. We will attempt to further this discussion in the literature by also analyzing choosing rate across distances with perpetrator-present (PP) and perpetrator-absent (PA) lineups.

In a more recent study, Nyman et al. (2019) explored the effect of distance on perpetrator IDs across various lighting conditions in a live encoding task. Replicating Wagenaar and Van der Schrier, they found that correct IDs declined as distance increased and less light was available. At 20 m and in the poorest lighting conditions, correct ID rate was a mere .11. In a related study, Nyman, Lampinen, Antfolk, Korkman, & Santtila (2019) sought to establish a distance threshold at which IDs would no longer be reliable. Overall accuracy suffered as distance increased from 5 to 110 m. Although accuracy was harmed by distance, high confidence still indicated high accuracy for IDs made at distances of 5-40 m. However, by collapsing their data across such a wide distance range, it remains unclear whether or not there could be a stronger relationship between confidence and accuracy for short (e.g., 3 or 10 m) relative to longer distances (e.g., 20 m). Also, this study did not have enough high confident IDs at distances greater than 40 m, so they could not determine if there was a distance at which high confidence was no longer a reliable indicator of accuracy.

Moreover, none of these studies (Lampinen et al., 2014; Lindsay et al., 2008; Nyman et al., 2019; Nyman, Lampinen et al., 2019; Wagenaar & Van der Schrier, 1996; see also Lampinen, Roush, Erickson, Moore, & Race, 2015) utilized a mock-crime eyewitness ID paradigm, which is the most common procedure in the literature, and an important test of estimator variable manipulations. Instead, all but Wagenaar and Van der Schrier (1996) presented live stationary targets in the field. Though this approach has benefits, there are also weaknesses,[1] which our study will address by being the first to manipulate distance to a perpetrator with mock-crime videos, followed by random assignment to PP or PA lineups. After each ID decision, we collected confidence judgments in order to conduct Receiver Operating Characteristic (ROC) analysis (described below; see also Lampinen et al., 2014) and also to address the confidence-accuracy (CA) relationship.

The Confidence-Accuracy Relationship

Within the context of previous literature, we have described two important measures of eyewitness ID performance that could be affected by distance: accuracy and response bias (i.e., likelihood of choosing). Despite the fact that the criminal justice system has no control over estimator variables, it is still important for judges and juries to know how to evaluate such evidence (Mickes, 2015). Although researchers have investigated the CA relationship for decades, the eyewitness literature has recently updated previous findings about the CA relationship by showing that confidence can be a good indicator of accuracy under certain circumstances (e.g., when recorded immediately after an ID; see Wixted & Wells, 2017). A consistent finding is that estimator variable manipulations largely do not negatively impact the overall CA relationship. Indeed, manipulations of weapon presence (Carlson, Dias, Weatherford, & Carlson, 2017), exposure time (Carlson et al., 2016; Palmer, Brewer, Weber, & Nagesh, 2013), retention interval (Sauer, Brewer, Zweck, & Weber, 2010), and race (Dodson & Dobolyi, 2016) have all shown that regardless of condition, accuracy increases with confidence, and highly confident eyewitnesses tend to be highly accurate.

---

[1] We will discuss this tradeoff in the General Discussion.

Semmler, Dunn, Mickes, and Wixted (2018) recently continued the argument for a strong CA relationship (particularly in the highest confidence range) regardless of estimator variable manipulations, and expanded it to include distance. They re-analyzed data from Lindsay et al. (2008) to create Confidence-Accuracy Characteristic (CAC) curves (Mickes, 2015), and found a strong CA relationship for both short (4-15 m) and long (20-50 m) distances. Furthermore, based on Signal Detection Theory (SDT; see Green & Swets, 1966; Macmillan & Creelman, 2005), they supported the argument that this pattern should be expected even for variables that have a substantial effect on discriminability. As mentioned previously, Nyman, Lampinen et al. (2019) also found support for a fairly strong CA relationship regardless of distance, but they collapsed across IDs made from eight distances ranging from 5-40 m. Therefore, a primary goal of this study is to address the strength of the CA relationship across specific distances ranging from short (3 m) to intermediate (10 m) to long (20 m).

The Present Research

We had three primary goals for the present experiments. First, no prior study has applied the mock-crime eyewitness ID paradigm to manipulations of distance, in order to yield results that may be more applicable to real-world eyewitness IDs. Second, in terms of response bias, we will address the mixed findings from previous studies. What will be the effect of distance on the willingness to choose from a lineup? Third, we will provide an examination of a priori predictions regarding the effect of distance on the CA relationship (see also Nyman, Lampinen et al., 2019). Again, Semmler et al. (2018) found with a re-analysis of data from Lindsay et al.

(2008) that the CA relationship may hold up well across distance. We will directly test this hypothesis.

Experiment 1

Method

Participants. Participants (*N* = 452) were recruited from psychology department subject pools across multiple universities. Those who incorrectly answered a manipulation check question (what was stolen in the mock-crime videos?) were excluded (*n* = 14), leaving 438 participants for analysis (77.2% female; mean age = 21.73, *SD* = 6.69). The majority of participants were Caucasian (38.4%), followed by Hispanic/Latino (35.4%) and African-American (19.4%).

Materials.

Mock-Crime Video. We used Qualtrics to create and administer the experiment. We filmed nine mock-crime videos using three different perpetrators (all young Caucasian men with dark hair and no distinctive features), each of which was filmed at three different distances (3, 10, 20 m; see Figure 1 for still images). These three distances were chosen to test short, medium, and long distances based on the literature. Lindsay et al. (2008) used distances of between 4-15 m (their short condition) and 20-50 m (their long condition), and Lampinen et al. (2014) used distances ranging from 5-40 yards (4.5 - 36.5 m). Each video was displayed in Qualtrics with a resolution of 400 x 650 pixels and was measured 8 in diagonally on the screen, resulting in 95.4 pixels per inch (PPI). The videos present a woman (the victim) sitting on a bench and reading a magazine, with her purse to her right. From the left of the screen, the perpetrator emerges from behind a large tree and approaches the bench. He looks at the camera as he sits on the bench next to the woman's purse. While sitting on the bench, the perpetrator looks to his right, left, and behind the bench. After looking at the camera a final time, he grabs the purse and runs away in the same direction he approached the bench, chased by the victim. We instructed our actors to adhere to these scripted behaviors in order to ensure that each video sequence took place in the same way for all three perpetrators and distances. Each video lasts 20-25 s with the perpetrator in view for 12-15 s.

Although we filmed each mock-crime at the three pre-determined distances, we acknowledge that viewing the videos on a computer screen diminishes the visual angle subtended, which could yield a different perceived distance for participants. To our knowledge, there is no obvious way to calculate the distance between a person or object and a camera based on the size (or PPI) of the person or object on a computer screen. However, it is still important to understand how far away participants think the perpetrator on the screen actually is (as if they are viewing the mock-crime through a window rather than their computer screen). We failed to obtain a distance estimate from participants in our first experiment, but afterward we presented an independent group of participants ($N = 14$) with three of our mock-crime videos at three different distances (i.e., each participant viewed just one of our three perpetrators, but at all three distances) and asked them to estimate the distance to the perpetrator as if watching the crime through a window. Participants' perceived distances were all underestimates of the actual distances at which the mock-crimes were filmed. They reported an average distance of 2 (*SD* = 1.3), 6.4 (3.9), and 16.3 (12.6) m for our 3, 10, and 20 m conditions, respectively. Lindsay et al. (2008) also found that participants underestimated shorter distances, but unlike our data they overestimated longer distances. Although our participants' perceived distance was underestimated, they were a fairly close approximation of the actual distance from our camera to our perpetrators. We will expand upon this issue in the General Discussion, as it is relevant to all eyewitness ID studies involving a filmed mock-crime.

Lineups. To construct lineups, we first downloaded mugshots from various online public offender databases (e.g., State of Arkansas Department of Corrections, State of Florida Department of Corrections) to serve as fillers. These images were then edited to remove any background content, leaving only the face cropped from the neck up, presented on a white background. We then assessed lineup fairness by presenting an independent group of participants (*N* = 32) with each lineup and asking them to choose the member that best matched a basic description of the perpetrator (e.g., white male, 20-30 years old, short brown hair). Their choices were used to calculate Tredoux's *E'* (Tredoux, 1998), which ranges from 1 (very biased) to 6 (very fair). For each of our three perpetrators, we selected a fair PP and fair PA lineup that had similar Tredoux's *E'* values: Perpetrator 1 (PP 4.57, PA 4.38), Perpetrator 2 (PP 4.34, PA 4.20), Perpetrator 3 (PP 4.34, PA 4.88). There was no designated innocent suspect for the PA lineups; they simply contained six fillers. We used a different-fillers design across our three different perpetrator lineups, as well as across PP and PA lineups. This was intended to increase ecological validity, as police often select fillers based on their similarity to the suspect, thus producing different fillers for lineups that contain a guilty versus innocent suspect (Clark & Tunnicliff, 2001). The perpetrator in PP lineups was always in either position 2 (top row middle) or 5 (bottom row middle).

Design and Procedure. We utilized a 3 (distance from camera to perpetrator in mock-crime: 3 vs. 10 vs. 20 m) x 2 (perpetrator presence in lineup: present vs. absent) within-subjects design. After providing informed consent, participants took part in three consecutive experimental blocks. As a reminder, we utilized three different actors to play the role of perpetrator, and each participant viewed all three perpetrators and all three distances across the three blocks, but the particular combination of perpetrator and distance was counterbalanced. In other words, across participants, Perpetrator 1 was presented at all three distances (but never within a given participant), and the same occurred for Perpetrator 2 and 3. For example, a participant could see Perpetrator 1 at 3 m, then Perpetrator 2 at 10 m, and then Perpetrator 3 at 20 m. Another participant could see Perpetrator 3 at 10 m, followed by Perpetrator 1 at 20 m, and then Perpetrator 2 at 3 m. Due to some Qualtrics difficulty, we did not assign each participant to a randomized order of distances and lineup types (PP versus PA) across the three blocks. Instead, we created four versions of the experiment, each with a different randomized order of distances combined with PP/PA lineup, and randomly assigned each participant to one of them: (a) Block 1: 3 m distance followed by PP lineup, Block 2: 10 m distance followed by PA lineup, Block 3: 20 m distance followed by PP lineup, (b) B1: 20 m followed by PP, B2: 3 m followed by PA, B3: 10 m followed by PP, (c) B1: 10 m followed by PP, B2: 20 m followed by PA, B3: 3 m followed by PA, and (d) B1: 20 m followed by PA, B2: 10 m followed by PA, B3: 3 m followed by PP. Block 2 always consisted of a PA lineup, which was an unintended consequence of randomization.

At the beginning of each block, participants viewed one of the three mock-crime videos (containing only one perpetrator), and were instructed to pay close attention to everything presented on the screen as they may be tested on it later. After each video, participants engaged in a 3 min distractor task (solving 15 anagrams of U.S. states), followed by the lineup ID task. A 6-member PP or PA simultaneous lineup (2 x 3 array) was presented, along with instructions that the perpetrator may or may not be present. After their lineup decision, participants entered their confidence on an 11-point scale, ranging from 0% (not at all confident) to 100% (extremely confident). Finally, following the completion of all three blocks, participants answered a manipulation check question (what was stolen in the videos?) and demographic questions (e.g., sex, age, race).

Results

The within-subjects design created non-independence in our data sets, which requires non-parametric analyses. Therefore, in the first section below, we describe several repeated-measures logistic regressions applied separately to PP correct ID rate, PP filler ID rate, PP rejection rate, and PA false ID rate. We utilized Bonferroni correction for multiple comparisons within each measure ($\alpha = .05/3 = .017$). See Table 1 for all cell sizes and proportions. In the second section, we describe the results of ROC analysis, which combines correct and false IDs with confidence data (e.g., Gronlund & Neuschatz, 2014) to provide a more complete understanding of the effect of distance on empirical discriminability versus response bias (see also Lampinen et al., 2014). Lastly, the final section describes the effect of our distance manipulation on the CA relationship with CAC analysis (Mickes, 2015; see also Nyman, Lampinen et al., 2019; Semmler et al., 2018).

Repeated-Measures Logistic Regression. Starting with PP lineups, a repeated-measures logistic regression confirmed that increased distance reduced the number of correct IDs, $\chi 2$ (2, *N* = 671) = 69.63, $p < .001$. The number of correct IDs dropped precipitously from 3 to 10 m ($\chi 2$ (1, *N* = 438) = 30.30, $p < .001$) and from 3 to 20 m ($\chi 2$ (1, *N* = 468) = 61.26, $p < .001$). In fact, increasing distance from 3 to 10 m cut the correct ID rate by half, from 50% to 24% (see Table 1). There was also a significant decline in correct IDs from 10 to 20 m ($\chi 2$ (1, *N* = 436) = 6.30, $p = .012$). PP filler IDs were also affected by distance, but in a non-linear fashion (see Table 1), $\chi 2$ (2, *N* = 671) = 18.74, $p < .001$. They increased from 3 to 10 m ($\chi 2$ (1, *N* = 438) = 18.65, $p = .001$) and marginally increased from 3 to 20 m ($\chi 2$ (1, *N* = 468) = 5.90, $p = .02$), but decreased from 10 to 20 m, although non-significantly ($\chi 2$ (1, *N* = 436) = 4.05, $p = .044$). Lastly, participants were generally more likely to reject a PP lineup with greater distance, $\chi 2$ (2, *N* = 671) = 38.81, $p < .001$. Though the number of rejections did not change significantly from 3 to 10 m ($\chi 2$ (1, *N* = 438) = 2.13, ns), they did increase from 3 to 20 m ($\chi 2$ (1, *N* = 468) = 33.37, $p < .001$), as well as from 10 to 20 m ($\chi 2$ (1, *N* = 436) = 18.23, $p < .001$).

Turning now to PA lineups, as a reminder we had no designated innocent suspect, so all filler IDs were treated as false IDs. This means that false ID rate and rejection rate are perfectly inversely related (i.e., as the number of false IDs increase, the number or rejections naturally decrease to the same degree). We found that the number of false IDs increased with distance, $\chi2$ (2, $N = 643$) = 38.81, $p < .001$. There were more false IDs as distance increased from 3 to 10 m ($\chi2$ (1, $N = 438$) = 9.56, $p = .002$), from 3 to 20 m (1, $\chi2$ ($N = 408$) = 38.77, $p < .001$), and from 10 to 20 m ($\chi2$ (1, $N = 440$) = 12.27, $p < .001$). Overall, the correct and false ID patterns reflect a mirror effect (Glanzer & Adams, 1985), such that the increase in distance resulted in fewer correct IDs and more false IDs (in replication of Lampinen et al., 2014, and Lindsay et al., 2008).

In terms of overall choosing (i.e., all IDs collapsed across PP and PA lineups), another repeated-measures logistic regression revealed no differences across distances: 3 vs. 10 m ($\chi2$ (1, $N = 876$) = 0.31, ns); 3 vs. 20 m ($\chi2$ (1, $N = 876$) = 0.18, ns); 10 vs. 20 m ($\chi2$ (1, $N = 876$) = 0.02, ns). This suggests that eyewitnesses' willingness to choose is unaffected by distance, which supports Lindsay et al. (2008) but contrasts with Lampinen et al. (2014). When separating PP and PA lineups, there was a significant drop in PP IDs (i.e., correct IDs plus filler IDs) from 3 to 20 m ($\chi2$ (1, $N = 438$) = 33.37, $p < .001$), and 10 to 20 m ($\chi2$ (1, $N = 438$) = 18.23, $p < .001$), but not 3 to 10 m ($\chi2$ (1, $N = 438$) = 2.13, ns). The opposite pattern was found for PA IDs (see above), such that IDs increased with distance. Again, these results are similar to Lindsay et al., who found (in their immediate judgment condition, which is closest to a standard eyewitness ID paradigm) a decrease in IDs for PP lineups (84.4% to 76.6%) and an increase in IDs for PA lineups (45.5% to 55.7%) when distance was long. The results are clear for PA lineups: IDs increase with distance. However, results for PP lineups are less clear: correct IDs decline linearly with distance, but filler IDs do not. If a perpetrator is far away, eyewitnesses should realize that this could harm the accuracy of their lineup decision, and therefore abstain from choosing. However, the fact that participants are still willing to choose at long distances (especially in PA lineups) may reflect poor metacognitive ability to appropriately determine whether or not they should make an ID based on distance. We will expand upon this issue below with an investigation of the confidence-accuracy relationship.

Receiver Operating Characteristic (ROC) Analysis. To better understand how our manipulation of distance affected empirical discriminability independently of response bias, ROC analysis is necessary[2] (e.g., Gronlund, Wixted, & Mickes, 2014; Rotello & Chen, 2016; Wixted & Mickes, 2012). Due to our within-subjects design, non-independence exists in our data, so we treated the ROC curves as correlated by using Equation 3[3] from Robin et al. (2011), which adjusts the variance estimate by taking covariance into account. As shown in Figure 2, each curve represents a particular distance based on correct and false ID rates across varying levels of confidence. As is commonly done in the literature when there is no designated innocent suspect (e.g., Carlson et al., 2019; Mickes, 2015), we divided the number of PA filler IDs by lineup size (6) to obtain a false ID rate that is comparable to the correct ID rate. For a given

---

[2] We acknowledge that there are some researchers opposed to the application of ROC analysis to lineup data (e.g., Lampinen, 2016; Smith, Lampinen, Wells, Smalarz, & Mackovichova, 2019; Wells, Smalarz, & Smith, 2015), however, there are many in favor (e.g., Gronlund et al., 2012, 2014; National Research Council, 2014; Rotello & Chen, 2016; Wixted & Mickes, 2012, 2018).

[3] $\text{var}(\theta_1 - \theta_2) = \text{var}(\theta_1) + \text{var}(\theta_2) - 2\text{cov}(\theta_1, \theta_2)$, where θ_1 and θ_2 are the two pAUCs

condition, the far-right point on the curve comprises all IDs regardless of level of confidence. As the curve extends to the left, participants who made IDs with low levels of confidence are dropped, thus the second point from the far right excludes IDs made with 0-20% certainty. This procedure continues for each point on the ROC curve until the far-left point, where only IDs made with the highest level of confidence (90-100%) are remaining.

Empirical discriminability is measured with partial Area Under the Curve (pAUC; Robin et al., 2011). A curve has greater empirical discriminability when it rises closest to the upper-left area in ROC space. The curves are compared using D = (pAUC1 - pAUC2)/s, where s is the standard error of the difference between two pAUCs after bootstrapping (see Gronlund et al., 2014, for a tutorial). As seen in Figure 2, empirical discriminability increased as distance decreased. Empirical discriminability was greater at 3 m (pAUC = .044 [.033 - .054]) compared to both 10 m (pAUC = .019 [.013 - .026], $D = 4.11$, $p < .001$) and 20 m (pAUC = .006 [.003 - .011], $D = 6.60$, $p < .001$); it was also greater at 10 m compared to 20 m ($D = 3.29$, $p < .001$).

Confidence-Accuracy Characteristic (CAC) Analysis. Based only on suspect IDs, CAC analysis is a valuable way of portraying the CA relationship (Mickes, 2015). Due to the lack of an innocent suspect in TA lineups, we again divided the total number of PA filler IDs by lineup size (6) to obtain a number of false IDs that can be compared to correct IDs. As shown in Figure 3, CAC analysis plots proportion correct (correct IDs/correct IDs + false IDs) on the y-axis and three bins representing low, medium, and high confidence (0-50%, 60-70%, 80-100%) on the x-axis. The value of confidence can be seen in two ways from a CAC plot: (a) whether or not the curves have a positive slope, indicating that as confidence increases, so to does accuracy; and (b) how high accuracy is at the highest level of confidence. Ideally, each condition's curve would converge at the upper right portion of the plot, indicating that regardless of their viewing condition, if witnesses report high confidence, they are likely to be highly accurate. We focus our interpretation on high confidence IDs because they are often the only IDs that are admitted into the courtroom, thus making it crucial for judges and jurors to know how to best evaluate such evidence (Mickes, 2015).

As depicted in Figure 3, at a distance of 3 m we found that high-confidence IDs were associated with high accuracy (and higher accuracy than low-confidence IDs, $z = 2.01$, $p = .022$), but this was not the case at 10 or 20 m. Indeed, IDs supported by 80-100% confidence after seeing a perpetrator at 3 m were significantly more accurate than IDs supported by 80-100% confidence after seeing a perpetrator at 10 or 20 m, $z = 2.71$, $p = .003$. These results contrast with Semmler et al. (2018) and Nyman, Lampinen et al. (2019), who found that the reliability of IDs made with high confidence was unaffected by distance (up to 40 m in Nyman, Lampinen et al.). One explanation for why high confidence IDs were not associated with high accuracy at 10 and 20 m comes from Sauer et al. (2010), who speculated that poor eyewitness ID performance would restrict the amount of variation in accuracy and consequently reduce the extent to which confidence can postdict accuracy. Given that we had almost chance performance at 10 m and chance performance at 20 m, this account could be appropriate. Another explanation is that we simply did not have enough data in the 10 and 20 m conditions (i.e., too few IDs, leading to the wide error bars in Figure 3) to draw conclusions regarding the association between high confidence and high accuracy.

However, a closer look at Figure 3 reveals that the error bars for 10 and 20 m conditions were very wide only above 50% confidence. In other words, participants viewing the perpetrator at these longer distances were generally unwilling to support an ID with medium or high levels of confidence at all. This could indicate a certain metacognitive awareness that distance will harm accuracy, and could be considered another measure of eyewitness metacognition beyond CAC curves. To evaluate this possibility, we conducted some additional analyses. After collapsing across medium and high confidence because so few IDs were made in this range at 10 and 20 m, chi-square tests revealed that participants in the 3 m condition were more likely to support an ID with above 50% confidence, compared to 10 m ($\chi2$ (1, N = 550) = 49.46, $p < .001$, $\phi = .30$) and 20 m ($\chi2$ (1, N = 548) = 97.54, $p < .001$, $\phi = .42$). Critically, a binomial test revealed that at 3 m, participants were more willing to support an ID with medium/high confidence (above 50%) compared to low confidence (below 50%), $p = .004$. However, at 10 and 20 m, the pattern reversed, such that participants were significantly *less* willing to assign above 50% confidence to their IDs ($p < .001$). This pattern may reflect a certain degree of metacognitive awareness that distance harms accuracy, which is not detected by CAC analysis alone because of its focus exclusively on suspect IDs. This is particularly problematic if there are very few suspect IDs, in which case CAC analysis is not as informative. We will address this issue with a high-powered between-subjects design in our second experiment.

Discussion

In replication of other eyewitness ID studies manipulating distance (e.g., Lampinen et al., 2014; Lindsay et al., 2008), increasing distance harmed overall face recognition ability, which ultimately resulted in a mirror effect (Glanzer & Adams, 1985). At longer distances, people are less likely to correctly identify the perpetrator and more likely to mistakenly identify someone when the perpetrator is not present in a lineup. Additionally, ROC analysis revealed that empirical discriminability was significantly reduced for longer distances – reaching chance level at the longest distance we manipulated (20 m). This suggests that distance is a powerful estimator variable that substantially affects encoding strength and may impair ultimate ID performance more than previously known.

Though estimator variables cannot be controlled by the criminal justice system, it is still important to extract some practical information from an ID made at a particular distance so that those in the courtroom can appropriately evaluate it. This practical information comes in the form of confidence, specifically whether or not confidence can be a postdictor of accuracy. We did not find support for the fairly robust finding that poor encoding conditions can still yield a strong CA relationship and that high confidence is indicative of high accuracy (see Wixted & Wells, 2017). Based on our results, eyewitness who viewed a perpetrator from just 10 m away were not highly accurate when reporting high confidence – indicating people may not be metacognitively aware how severe of an impact distance has on recognition ability. However, we also found that people were generally less willing to respond with high confidence at 20 m, which could imply people actually are metacognitively aware that recognition ability suffers greatly at longer distances.

This experiment has at least two limitations that are worth noting. First, it is now becoming common to have 500-1000 data points per condition for ROC analysis (e.g., Carlson et al., 2019; Colloff, Wade, & Strange, 2016; Seale-Carlisle, Wetmore, Flowe, & Mickes, 2019), which this study did not have. This lack of power is particularly evident with the wide error bars in our CAC graph, due to low performance at 10 and 20 m, which resulted in too few suspect IDs. This restricted our ability to draw strong conclusions regarding the CA relationship (Sauer et al., 2010). Another limitation of this study was the within-subjects design. Participants saw three consecutive blocks consisting of three similar mock-crime videos. Although a given participant never saw the same perpetrator or distance twice, they may have been able to learn the task at hand, which may have influenced the amount of attention paid to each perpetrator in the videos, as well as affecting their ID decision and confidence. In Experiment 2, we addressed these concerns by substantially increasing power and conducting a between-subjects design with only one mock-crime video per participant.

Experiment 2

Although the impact of distance on empirical discriminability appears to be clear, the relationship between confidence and accuracy across distance is less clear. In Experiment 2, we sought to reproduce the empirical discriminability advantage for shorter distances. More importantly, we also wanted to provide more conclusive evidence pertaining to how distance affects the CA relationship. To this end, we increased power by extending the experiment to a nationwide sample via SurveyMonkey and created a between-subjects design using the same stimuli from Experiment 1.

Method

Participants. Participants (*N* = 4134) were recruited from a nationwide sample via SurveyMonkey. After dropping those who did not complete the study or failed to answer the manipulation check multiple-choice question (what type of crime did you witness?) correctly (*N* = 372), we were left with a sample of 3762 for analysis (i.e., an average of over 600/cell). See Table 2 for demographics.

Materials. We used the same stimuli as Experiment 1, except instead of using three different targets we only used two.

Design and Procedure. This experiment featured a 3 (distance from camera to perpetrator in mock-crime: 3 vs. 10 vs. 20 m) x 2 (perpetrator presence in lineup: present vs. absent) between-subjects design. Following informed consent, participants were randomly assigned to one of six videos based on the two targets and three distances. Immediately following the video, they were asked to estimate the distance (in feet) between the people in the video and the camera. Then, they watched a 3-min nature video as a distractor task, followed by random assignment to a 6-member (2x3 array) PP or PA lineup (based on the appropriate target), and were instructed that the man from the video may or may not be present. Next, they were asked to enter a confidence rating regarding their lineup ID decision on a scale of 0-100%. Lastly, participants answered a manipulation check multiple-choice question (what type of crime did you witness?) and provided demographic information (e.g., sex, age, race).

Results

See Table 3 for ID and rejection frequencies and proportions across the two targets and three distances. There was no difference in responding between targets, so we collapsed across target for the following analyses. Again we start by conducting several logistic regressions applied separately to PP correct ID rate, PP filler ID rate, PP rejection rate, and PA false ID rate. A Bonferroni correction was applied to multiple comparisons within each measure ($\alpha = .05/3 = .017$). We then combined correct and false IDs with confidence data to apply ROC analysis (Gronlund & Neuschatz, 2014). Lastly, we constructed CAC curves (Mickes, 2015) to describe the effect of distance on the CA relationship.

Logistic Regression. Overall, the number of correct IDs declined with increased distance, $\chi 2$ (2, $N = 1944$) = 170.72, $p < .001$. Specifically, the number of correct IDs fell dramatically from 3 to 10 m ($\chi 2$ (1, $N = 1368$) = 75.67, $p < .001$), 3 to 20 m ($\chi 2$ (1, $N = 1299$) = 165.40, $p < .001$), and 10 to 20 m ($\chi 2$ (1, $N = 1221$) = 21.50, $p < .001$). PP filler IDs were also affected by distance, $\chi 2$ (2, $N = 1944$) = 23.71, $p < .001$, such that they increased from 3 to 10 m ($\chi 2$ (1, $N = 1368$) = 23.28, $p < .001$) and from 3 to 20 m ($\chi 2$ (1, $N = 1299$) = 9.12, $p = .003$), however there was no significant difference between 10 and 20 m ($\chi 2$ (1, $N = 1221$) = 2.65, ns). Finally, the tendency to reject a PP lineup was greater at longer distances, $\chi 2$ (2, $N = 1944$) = 111.66, $p < .001$. There were significantly more rejections at 10 ($\chi 2$ (1, $N = 1368$) = 24.18, $p <$

.001) and 20 m (χ2 (1, *N* = 1299) = 116.78, *p* < .001) relative to 3 m. There were also significantly more rejections at 20 m compared to 10 m (χ2 (1, *N* = 1221) = 35.52, *p* < .001).

For PA lineups, the number of false IDs increased with longer distances, χ2 (2, *N* = 1818) = 117.79, *p* < .001. There were fewer false IDs at 3 compared to both 10 m (χ2 (1, *N* = 1200) = 43.31, *p* < .001) and 20 m (χ2 (1, *N* = 1215) = 119.84, *p* < .001). There were also significantly fewer false IDs at 10 than 20 m (χ2 (1, *N* = 1221) = 20.01, *p* < .001). The patterns observed for correct and false IDs largely replicate Experiment 1; we found a mirror effect (Glanzer & Adams, 1985), such that longer distances yielded a decrease in correct IDs and an increase in false IDs.

To investigate overall choosing behavior, we analyzed all IDs collapsed across PP and PA lineups. Like in E1, we found that distance did not influence the tendency to make an ID overall: 3 vs. 10 m (χ2 (1, *N* = 2568) = 1.53, ns); 3 vs. 20 m (χ2 (1, *N* = 2514) = 0.29, ns); 10 vs. 20 m (χ2 (1, *N* = 2442) = 1.07, ns). When separating PP and PA lineups, we found that PP IDs (i.e., correct IDs plus filler IDs) declined greatly from 3 to 10 m (χ2 (1, *N* = 1368) = 24.18, *p* < .001), from 3 to 20 m (χ2 (1, *N* = 1299) = 116.78, *p* < .001), and from 10 to 20 m (χ2 (1, *N* = 1221) = 35.52, *p* < .001). In contrast, PA IDs (see above) increased with distance. These findings largely replicate E1 and findings by Lindsay et al. (2008).

ROC Analysis. We followed the same process as described in Experiment 1 for assessing empirical discriminability, however we did not need to treat the ROC curves as correlated due to the between-subjects design. As illustrated in Figure 4, we replicated

Experiment 1: empirical discriminability was greater at 3 m (pAUC = .054 [.048 - .060]) compared to both 10 m (pAUC = .024 [.020 - .029], $D = 7.77$, $p < .001$) and 20 m (pAUC = .011 [.008 - .015], $D = 12.23$, $p < .001$), and empirical discriminability at 10 m was also greater compared to 20 m ($D = 3.96$, $p < .001$).

CAC Analysis. As portrayed in Figure 5, we found a strong CA relationship at distances of 3 and 10 m in the sense that each curve increased monotonically. Within the 3 m condition, high-confidence IDs were associated with higher accuracy than low-confidence IDs, $z = 3.68$, $p < .001$. However, this difference was not significant at 10 m ($z = 1.65$, $p = .049$) or 20 m ($z = .17$, ns). In fact, people in the 20 m condition tended to be overconfident, implying that they were not able to metacognitively adjust their confidence appropriately given the poor viewing conditions and high likelihood of being inaccurate at a long distance. Of particular applied importance, high confidence evinced high accuracy at 3 and 10 m, but not at 20 m. For IDs made with 90-100% confidence, accuracy was marginally higher for 3 m relative to 10 m ($z = 1.9$, $p = .029$), but significantly higher compared to 20 m ($z = 5.64$, $p < .001$). Accuracy was also marginally higher for 10 m relative to 20 m, $z = 1.99$, $p = .023$. This suggests that the CA relationship may have some limitations under certain encoding conditions. These patterns largely track the pattern of results found in Experiment 1 with one exception: the 10 m condition in Experiment 2 showed a consistent positive slope rather than dropping off at 80-100% confidence like in Experiment 1. This is likely due to the larger number of suspect IDs made with high confidence, thus strengthening the interpretability of these CAC curves. Additionally, this experiment did not fall prey to the potential carryover effects from the within-subjects design of E1.

Distance Estimation. Participants provided an estimate (in feet) of how far away they perceived the perpetrator in the video was from the camera. Although all participants were prompted to answer this question, an error in SurveyMonkey originally allowed this question to be skipped. This was subsequently fixed so that all participants were required to answer before moving on. As a result, we were left with 3159 distance estimates to analyze. We first converted all distance estimates from feet to meters and then averaged them for each condition. As in Experiment 1, participants slightly underestimated the actual distance at which each mock-crime video was filmed. They reported an average distance of 3.19 (*SD* = 2.8), 7.25 (5.5), and 18.65 (29.3) m for our 3, 10, and 20 m conditions, respectively. We still acknowledge the concern regarding distance potentially being amplified when viewing a video on a computer screen, but these estimates from the same participants in the experiment itself (as opposed to the independent sample we used in Experiment 1) show that they have a reasonably accurate perception of how far away each perpetrator is from the camera.

Discussion

Again, in replication of previous literature (e.g., Lampinen et al., 2014; Lindsay et al., 2008) and Experiment 1, increasing distance harmed recognition accuracy. This function is in the form of a mirror effect (Glanzer & Adams, 1985), which was also found in Experiment 1. As ROC analysis illustrates, empirical discriminability declined as a consequence of increasing distance. These results demonstrate that people are at a significantly greater risk of mistaken identification after having viewed a perpetrator from a distance of several meters. Although performance overall increased somewhat from E1, a distance of just 10 m again caused considerable harm to eyewitness ID accuracy.

We could not draw strong conclusions regarding the CA relationship in Experiment 1 due to a lack of power and the potential that the within-subjects design influenced responding. However, with increased power and a between-subjects design in the present experiment, these results have more solid ground for interpretation. We found a strong CA relationship at distances of 3 and 10 m, such that confidence and accuracy increased monotonically and high confidence was associated with high accuracy. Although empirical discriminability dropped from 3 to 10 m, the CA relationship was resilient in postdicting accuracy for highly confident responders. This was not the case at a distance of 20 m, as we found a weak CA relationship, such that accuracy was no different between those who reported low (0-59%) and high (90-100%) confidence. From these data, we can conclude that there may be certain boundary conditions in estimator variable manipulations that do not translate to a strong CA relationship. For example, viewing a perpetrator at 20 m may be too poor of an encoding scenario for eyewitness confidence to have a meaningful relationship with accuracy.

General Discussion

We had three goals for our experiments. First, using a mock-crime eyewitness ID paradigm, we sought to extend upon prior studies that manipulated distance using a stationary target in the field (Lampinen et al., 2014; Lindsay et al., 2008; Nyman, Lampinen et al., 2019). Second, we investigated the effect that distance may have on response bias. Lastly, we addressed the impact of distance on the CA relationship. Our experiments yielded two primary findings: (a) increasing distance harmed empirical discriminability in the form of a mirror effect (decreasing correct IDs and increasing false IDs), and (b) the CA relationship was generally strong across distances, but high confidence did not indicate high accuracy at the longest distance we manipulated (20 m).

In agreement with prior findings (e.g., Lampinen et al., 2014; Lindsay et al., 2008), we found that accuracy declined with increased distance. Furthermore, we found a mirror effect (Glanzer & Adams, 1985), which is consistent with the patterns found in Lampinen et al. (2014) and Lindsay et al. (2008). One account of the mirror effect (i.e., the strength-based mirror effect) proposes that as memory strength decreases, overall recognition performance also decreases in the form of a decrease in correct IDs and an increase in false IDs. In terms of distance, there are more fine-grained facial details (i.e., internal facial features) available to be encoded at shorter distances, resulting in a more specific memory trace, which can ultimately aid in the subsequent recognition of a perpetrator. Having more fine-grained facial details to rely on during a recognition test (i.e., a lineup) increases the likelihood that the perpetrator will be chosen when present in a lineup (i.e., a correct ID), while also allowing for a lineup to be rejected when the perpetrator is not present. As distance increases, fine-grained facial details are lost, leading to individuals relying on coarse-grained facial details (i.e., external facial features) to inform their

ID decision, thereby resulting in fewer correct IDs and correct rejections. As seen in Tables 1 and 3, our data follow this pattern, which supports this strength-based account of the mirror effect, and cannot be easily explained by a criterion shift account.

Similarly, Clark and Godfrey (2009) investigated the effects of different eyewitness conditions on correct and false IDs in the extant literature. When a memory trace is non-specific regarding details of a target's face (e.g., at long distances), correct IDs decline. However, in PA lineups (regardless of whether there is a designated innocent suspect or not), Clark and Godfrey argued that when correct rejections decline, false IDs are evenly distributed across lineup fillers (assuming the lineup is fair). Although these erroneous selections minimally impact false ID rate for the innocent suspect (after dividing by 6), our data illustrate that false ID rate indeed increases (while correct ID rate decreases) with increased distance.

Notably, whereas we utilized mock-crime videos, others (e.g., Lampinen et al. 2014; Lindsay et al. 2008; Nyman, Lampinen et al., 2019) manipulated distance with live encoding conditions. We argue that there are pros and cons with both approaches. It could certainly be argued that experiments with live encoding conditions are more ecologically valid. However, both Lindsay et al. and Lampinen et al. had participants look directly at a stationary confederate for a full 10 s, and Nyman, Lampinen et al. used an encoding duration of 20 s. In contrast, for our mock-crime videos, participants were not instructed to focus on only the perpetrator for the duration of the crime. It is important to test estimator variables with a mock-crime video in order to assess how memory for a perpetrator is affected when a dynamic crime event is occurring and the participant does not know on whom or what they will ultimately be tested. Although we were intending to minimize any stimulus-specific effects (Wells & Windschitil, 1999) by choreographing all of our mock-crime videos to be as similar as possible, we acknowledge that this is a limitation when using a within-subjects design in Experiment 1. It is possible that participants became accustomed to the mock-crime scenario, therefore by the third block, the crime event was no longer surprising. They also could have determined whom they were supposed to identify, thus dedicating more time to studying only the perpetrator and not the surroundings. However, we conducted follow-up analyses that did not reveal any order effects as a result of our within-subjects design (see Appendix).

Another point of consideration is the nature of mock-crime videos in general. Although our mock-crimes were filmed at specific distances, the visual angle subtended when viewing a video on a computer screen is different (likely smaller) than viewing a live target in the field, therefore creating a distance that may be longer than anticipated. However, when we asked participants how far away they perceived the perpetrator to be, they reported distances quite close to our filmed distances, and actually underestimated these distances. We were satisfied with these results, but the relationship between the quality/size of mock crime videos and perceived distance should be a consideration by eyewitness researchers going forward. The distance at which a perpetrator is viewed in a particular mock-crime video is rarely reported in the literature, and these videos could be reflecting longer distances in the real world than is likely intended. Even when a mock-crime is filmed from a reasonable distance, performance can vary based on the perceived distance of the perpetrator. We suggest that more information about the distance of perpetrators in mock-crime videos be reported and results be interpreted with this taken into account. Moreover, we encourage researchers to add a question about perceived distance to the perpetrator from participants' point of view, as this is particularly relevant to their eyewitness ID performance and potentially their confidence.

It is also important to understand how distance affects response bias, as the literature has produced mixed results. Lindsay et al. (2008) found that distance did not affect choosing rates from lineups, but Lampinen et al. (2014) found with yes/no recognition tests that choosing increased with distance. Our results, though also a bit mixed, are more similar to those by Lindsay et al., which could be due to the use of lineups in both studies. Our PA lineup results are straightforward, such that IDs increase with distance. However, our PP lineup results are more complicated. Although rejection rates consistently increase with distance, we cannot claim that this is due to a decrease in choosing because, whereas correct IDs consistently decline, filler IDs follow a non-linear pattern as distance increases. It is unclear whether Lampinen et al.'s results are related to the use of yes/no recognition rather than lineups, but we do not have an explanation for the mixed results regarding choosing rates. Although we believe we have furthered the conclusions that can be made regarding the effect of distance on response bias, more research is needed to uncover a fuller picture of this relationship.

Lastly, we investigated how distance influences the CA relationship. The literature supports a reasonably strong CA relationship across estimator variable manipulations (Mickes, 2015; Wixted & Wells, 2017), including distance (Nyman, Lampinen et al., 2019; Semmler et al., 2018). In Experiment 1, we found that high-confidence IDs were indicative of high accuracy at 3 m, but not at 10 and 20 m. This could be due to the lack of suspect IDs that were supported by medium or high confidence at these longer distances, which led us to investigate participants' willingness to support any ID (of suspect or fillers) with low versus medium/high confidence. Although participants were more likely to support IDs with medium/high confidence compared to low confidence when viewing a perpetrator at 3 m, the opposite pattern arose at 10 and 20 m, with participants being largely unwilling to support IDs with above 50% confidence. This finding suggests that individuals may be metacognitively aware that long distances may negatively impact accuracy.

Experiment 2 proffered stronger evidence that the CA relationship may not be completely resilient to increased distance. Specifically, the CA relationship was strong at 3 and 10 m, but not at 20 m. Distance severely impacts recognition ability, and our data suggest that there could be a distance that is too far for eyewitnesses to appropriately assess their confidence based on their probable accuracy. Nyman, Lampinen et al. (2019) found a good CA relationship up to 40 m, however this was just the longest distance that they had enough data to analyze, so the exact boundary condition of this relationship is yet to be discovered. Wixted and Wells (2017) proposed that testing under "pristine" conditions would lead to a strong CA relationship regardless of encoding conditions, but there may be some encoding boundaries (contributing to poor performance; see Sauer et al., 2010) that will not produce a strong CA relationship even if tested with the following five “pristine” conditions: (a) witnesses only view one lineup, (b) the lineup is fair, (c) memory is uncontaminated, (d) there is no administrator influence, and (e) confidence is immediately recorded after a lineup ID decision. These are all system variables that the criminal justice system can control. Unfortunately, there may be uncontrollable aspects of a crime that restrict the ability of confidence to postdict accuracy, especially at high confidence, which is the type of evidence on which the criminal justice system largely relies.

Conclusions and Implications

Increasing distance between an eyewitness and a perpetrator harms eyewitness ID accuracy. Our manipulation of distance revealed a mirror pattern, such that distance not only reduced the chances of correctly identifying the perpetrator, but also increased the risk of selecting an innocent suspect. This finding is obviously detrimental for innocent suspects who are identified after the witness was far away from the perpetrator. These results have implications for jurors who may not be adequately educated on the true impact of distance on recognition memory for faces. Clearly more research is needed on this important and understudied estimator variable.

Finally, regarding the effect of distance on the CA relationship, our second experiment showed that eyewitnesses who are highly confident are likely to be highly accurate up to 10 m, but not at 20 m, where we found a weak CA relationship. This implies that eyewitnesses who view a perpetrator from a long distance may not be able to reconcile poor encoding conditions with providing meaningful confidence information. This is crucial for triers of fact to understand because when evaluating eyewitness evidence, it is important to know what can and cannot be reliable. We argue that more research is needed on the effect of estimator variables like distance on the CA relationship, specifically further investigating particular encoding conditions that may not produce a strong CA relationship, even under “pristine” testing conditions.

References

Carlson, C. A., Dias J. L., Weatherford, D. R., & Carlson, M. A. (2016). An investigation of the weapon focus effect and the confidence–accuracy relationship for eyewitness identification. *Journal of Applied Research in Memory and Cognition*, *6*, 82-92. doi:10.1016/j.jarmac.2016.04.001

Carlson, C. A., Jones, A. R., Whittington, J. E., Lockamyeir, R. F., Carlson, M. A., & Wooten, A. R. (2019). Lineup fairness: Propitious heterogeneity and the diagnostic feature-detection hypothesis. *Cognitive Research: Principles and Implications, 4*, 20–26. http://dx.doi.org/10.1186/s41235-019-0172-5

Carlson, C. A., Young, D. F., Weatherford, D. R., Carlson, M. A., Bednarz, J. E., & Jones, A. R. (2016). The influence of perpetrator exposure time and weapon presence/timing on eyewitness confidence and accuracy. *Applied Cognitive Psychology, 30*, 898–910. https://doi.org/10.1002/acp.3275.

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review*, *16*, 22-42. doi:10.3758/PBR.16.1.22

Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behavior*, *25*, 199–216. doi:10.1023/A:1010753809988

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups don't just make witnesses more willing to choose the suspect, they also make them more likely to confuse innocent and guilty suspects. *Psychological Science, 27*, 1227–1239. https://doi.org/10.1177/0956797616655789.

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and eyewitness identifications: The cross-race effect, decision time and accuracy. *Applied Cognitive Psychology*, *30*, 113-125. doi:10.1002/acp.3178

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics, a reprint, with corrections of the original 1966 ed.* Huntington: Robert E. Krieger Publishing Co.

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Showups versus lineups: an evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *1*, 221-228. doi:10.1016/j.jarmac.2012.09.003

Gronlund, S. D., & Neuschatz, J., S. (2014). Eyewitness identification discriminability: ROC analysis versus logistic regression. *Journal of Applied Research in Memory and Cognition*, *3*, 54-57. doi:10.1016/j.jarmac.2014.04.008

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science*, *23*, 3-10. doi:10.1177/0963721413498891

Innocence Project. (2019). *The Innocence Project*. Retrieved from http://www.innocenceproject.org

Lampinen, J. M. (2016). ROC analyses in eyewitness identification research. *Journal of Applied Research in Memory and Cognition*, *5*, 21–33. doi:10.1016/j.jarmac.2015.08.006

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

Lampinen, J. M., Roush, A., Erickson, W. B., Moore, K. N., & Race, B. (2015) The effects of simulated distance on recognition of same race and other race faces, *Visual Cognition, 23*, 678-698. doi: 10.1080/13506285.2015.1084071

Lindsay, R. C. L., Semmler, C., Weber, N., Brewer, N., & Lindsay, M. R. (2008). How variations in distance affect eyewitness reports and identification accuracy. *Law and Human Behavior*, *32*, 526-535. doi:10.1007/s10979-008-9128-x

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone closer than far away? *Psychonomic Bulletin and Review*, *12*(1), 43-65.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.).* New York, NY: Cambridge University Press.

McKone, E. (2009). Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. *Vision Research*, *49*, 268-283. doi:10.1016/j.visres.2008.10.020

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. doi:10.1016/j.jarmac.2015.01.003

National Research Council (2014). *Identifying the culprit: Assessing eyewitness identification*. Washington, DC: The National Academic Press.

Neil v. Biggers, 409 U.S. 188 (1972).

Nyman, T. J., Antfolk, J., Lampinen, J. M., Tuomisto, M., Kaakinen, J. K., Korkman, J., & Santtila, P. (2019). A stab in the dark: The distance threshold of target identification in low light. *Cogent Psychology*, *6*. https://doi.org/10.1080/23311908.2019.1632047

Nyman, T. J., Lampinen, J. M., Antfolk, J., Korkman, J., & Santtila, P. (2019). The distance threshold of reliable eyewitness identification. *Law and Human Behavior*. doi:10.1037/lhb0000342

Palmer, M., 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. doi:10.1037/a0031602

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.

Rotello, C. M., & Chen, T. (2016). ROC Curve analyses of eyewitness identification decisions: An analysis of the recent debate. *Cognitive Research: Principles and Implications, 1:10*.

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. doi:10.1007/s10979-009-9192-x

Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D., & Mickes, L. (2019, May 16). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied*. Advance online publication. doi:10.1037/xap0000222

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. doi:10.1037/xap0000157

Smith, A. M., Lampinen, J. M., Wells, G. L., Smalarz, L., & Mackovichova, S. (2019). Deviation from perfect performance measure 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. doi:10.1016/j.jarmac.2018.09.003

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

Wagenaar, W. A., & Van der Schrier, J. H. (1996). Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime and Law*, *2*, 2321-2332. doi:10.1080/10683169608409787

Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory*, *4*, 313–317. doi:10.1016/j.jarmac.2015.08.008

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

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science*, *7*, 275-278. doi:10.1177/1745691612442906

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

Wixted, J. T., & Mickes, L. (2018). Theoretical vs. empirical discriminability: The application of ROC methods to eyewitness identification. *Cognitive Research: Principles and Implications*, *3*. doi:10.1186/s41235-018-0093-8

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. doi:10.1177/1529100616686966

Table 1.

Identification and rejection rates by distance from Experiment 1

Distance	**Perpetrator-Present Lineups**			**Perpetrator-Absent Lineups**	
	Correct ID Rate	**Filler ID Rate**	**Rejection Rate**	**Filler ID Rate**	**Rejection Rate**
3 meters	.50 (118/235)	.34 (81/235)	.15 (36/235)	.35 (72/203)	.65 (131/203)
10 meters	.24 (49/203)	.55 (112/203)	.21 (42/203)	.50 (118/235)	.50 (117/235)
20 meters	.15 (34/233)	.45 (106/233)	.40 (93/233)	.67 (137/205)	.33 (68/205)

Note: ID = Identifications

Table 2.

Demographic Information from Experiment 2

	Nationwide SurveyMonkey Sample
Sex	
Male	1,674
Female	1,976
No Response	112
Age	
18 – 29	938
30 - 44	798
45 - 60	1,336
Over 60	578
No response	112
Ethnicity	
Black or African-American	238
White or Caucasian	2,570
Hispanic or Latino	290
Asian or Pacific Islander	236
American Indian or Alaskan Native	68
Other	80
Choose Not to Answer	168
No Response	112
N	3,762

Table 3.

Identification and rejection rates by distance and target from Experiment 2

Overall Distance	**Perpetrator Present Lineup**			**Perpetrator Absent Lineup**	
	Correct ID Rate	**Filler ID Rate**	**Rejection Rate**	**Filler ID Rate**	**Rejection Rate**
3 meters	.54 (387/723)	.32 (228/723)	.15 (108/723)	.41 (243/597)	.59 (354/597)
10 meters	.30 (195/645)	.44 (285/645)	.26 (165/645)	.60 (360/603)	.40 (243/603)
20 meters	.19 (108/576)	.40 (228/576)	.42 (240/576)	.72 (444/618)	.28 (174/618)

Target 1 Distance	**Perpetrator Present Lineup**			**Perpetrator Absent Lineup**	
	Correct ID Rate	**Filler ID Rate**	**Rejection Rate**	**Filler ID Rate**	**Rejection Rate**
3 meters	.54 (195/360)	.29 (105/360)	.17 (60/360)	.38 (114/300)	.62 (186/300)
10 meters	.33 (96/285)	.43 (123/285)	.23 (66/285)	.61 (180/295)	.39 (115/295)
20 meters	.17 (51/306)	.41 (126/306)	.42 (129/306)	.75 (237/318)	.25 (81/318)

Target 2 Distance	**Perpetrator Present Lineup**			**Perpetrator Absent Lineup**	
	Correct ID Rate	**Filler ID Rate**	**Rejection Rate**	**Filler ID Rate**	**Rejection Rate**
3 meters	.53 (192/365)	.34 (123/365)	.13 (48/365)	.43 (129/297)	.57 (168/297)
10 meters	.28 (99/360)	.45 (162/360)	.28 (99/360)	.58 (180/309)	.42 (129/309)
20 meters	.21 (57/270)	.38 (102/270)	.41 (111/270)	.69 (207/300)	.31 (93/300)

Note: ID = Identifications

Figure 1. Mock-crime video still images at each distance as they appeared to participants. Each video was filmed at a distance of, from top to bottom: 3 m, 10 m, 20 m.

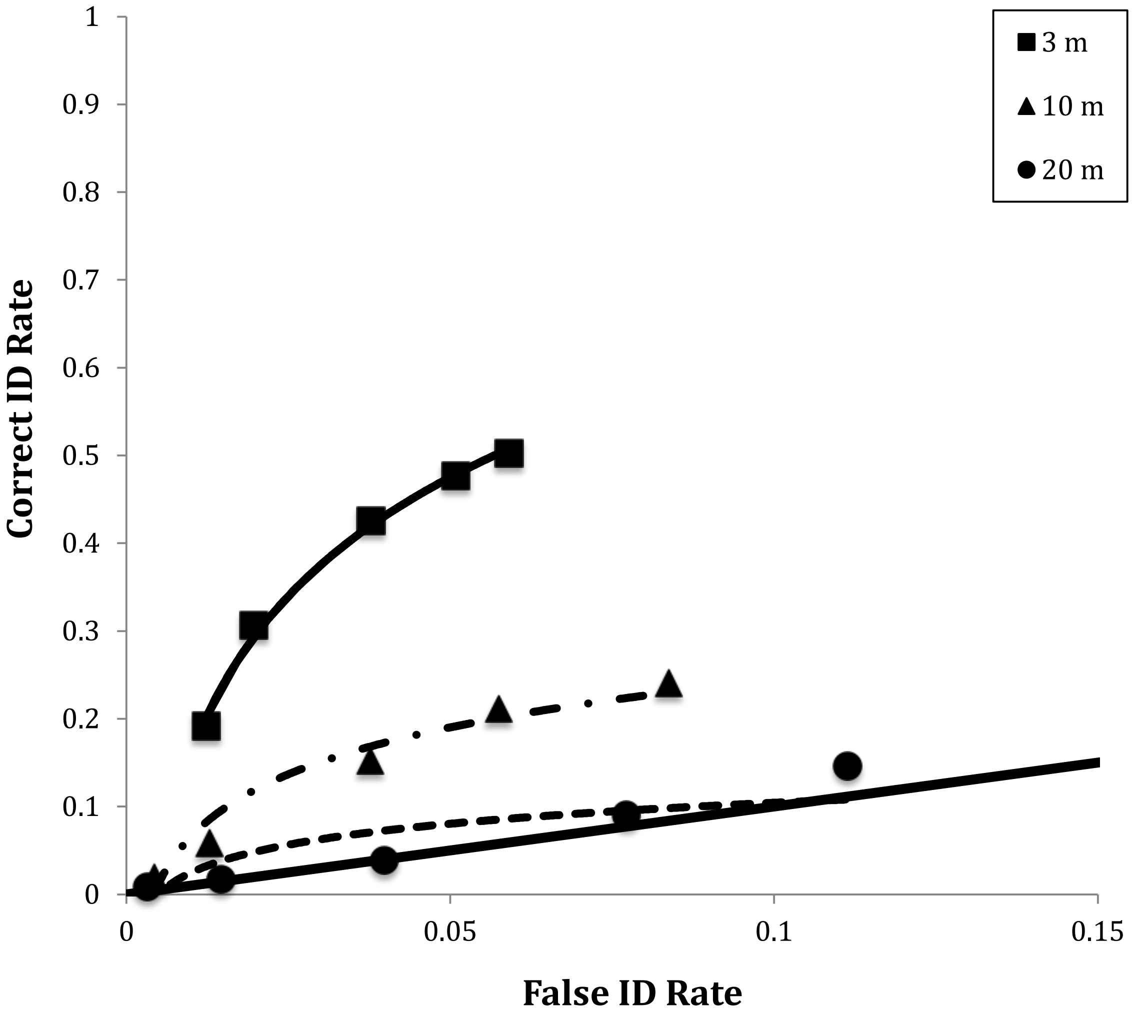

Figure 2. ROC curve for each distance condition in Experiment 1.

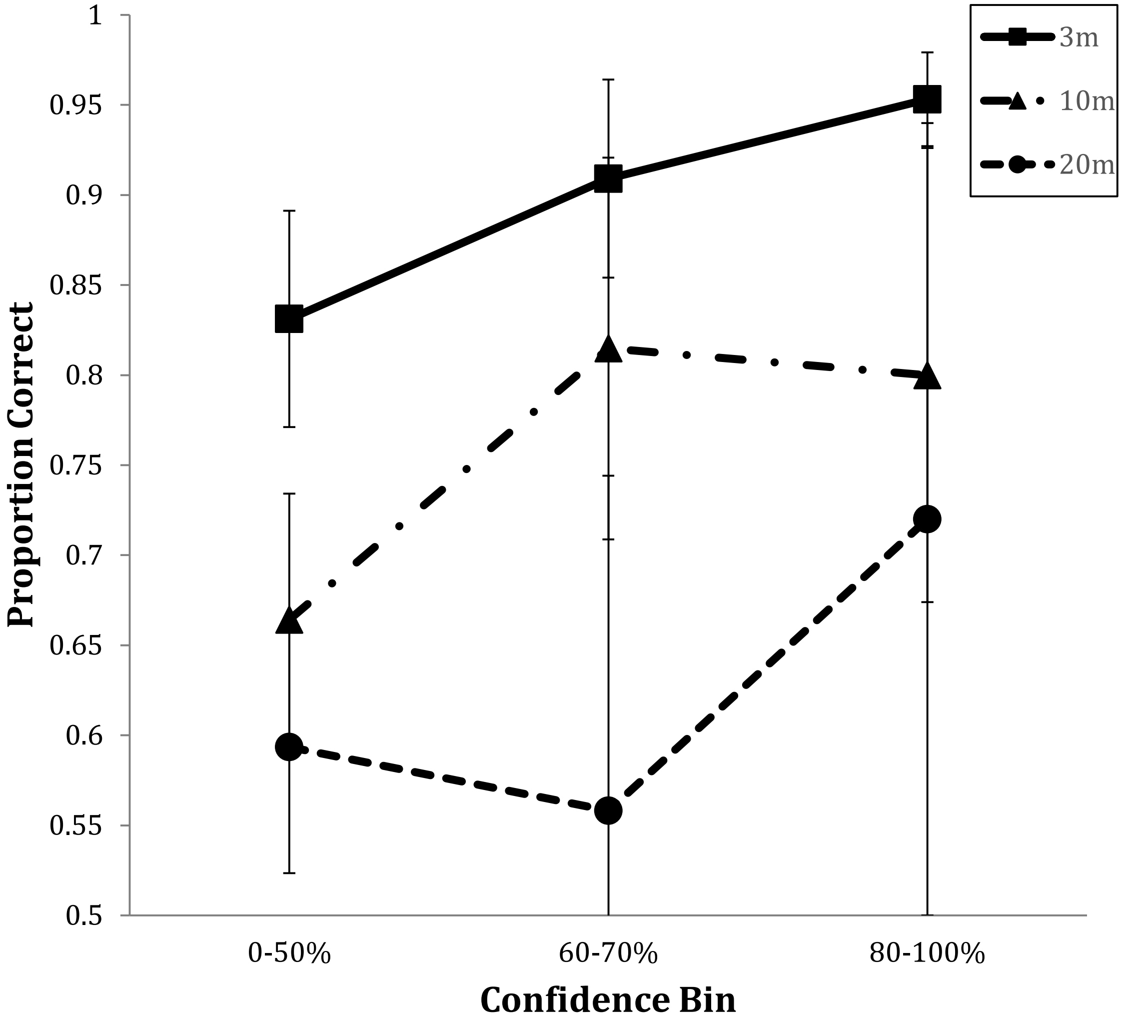

Figure 3. CAC curve for each distance condition in Experiment 1. Error bars represent standard error.

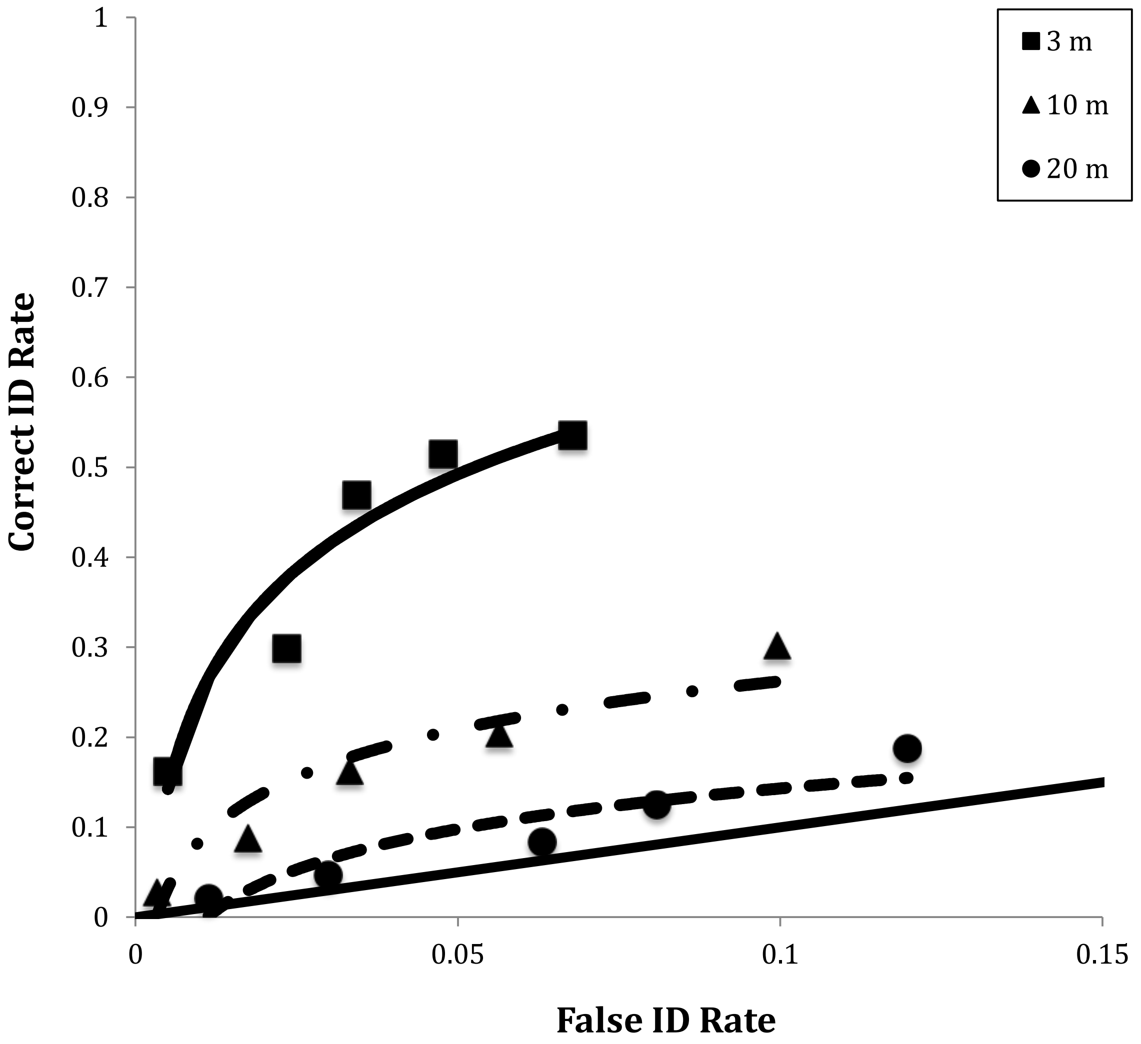

Figure 4. ROC curve for each distance condition in Experiment 2.

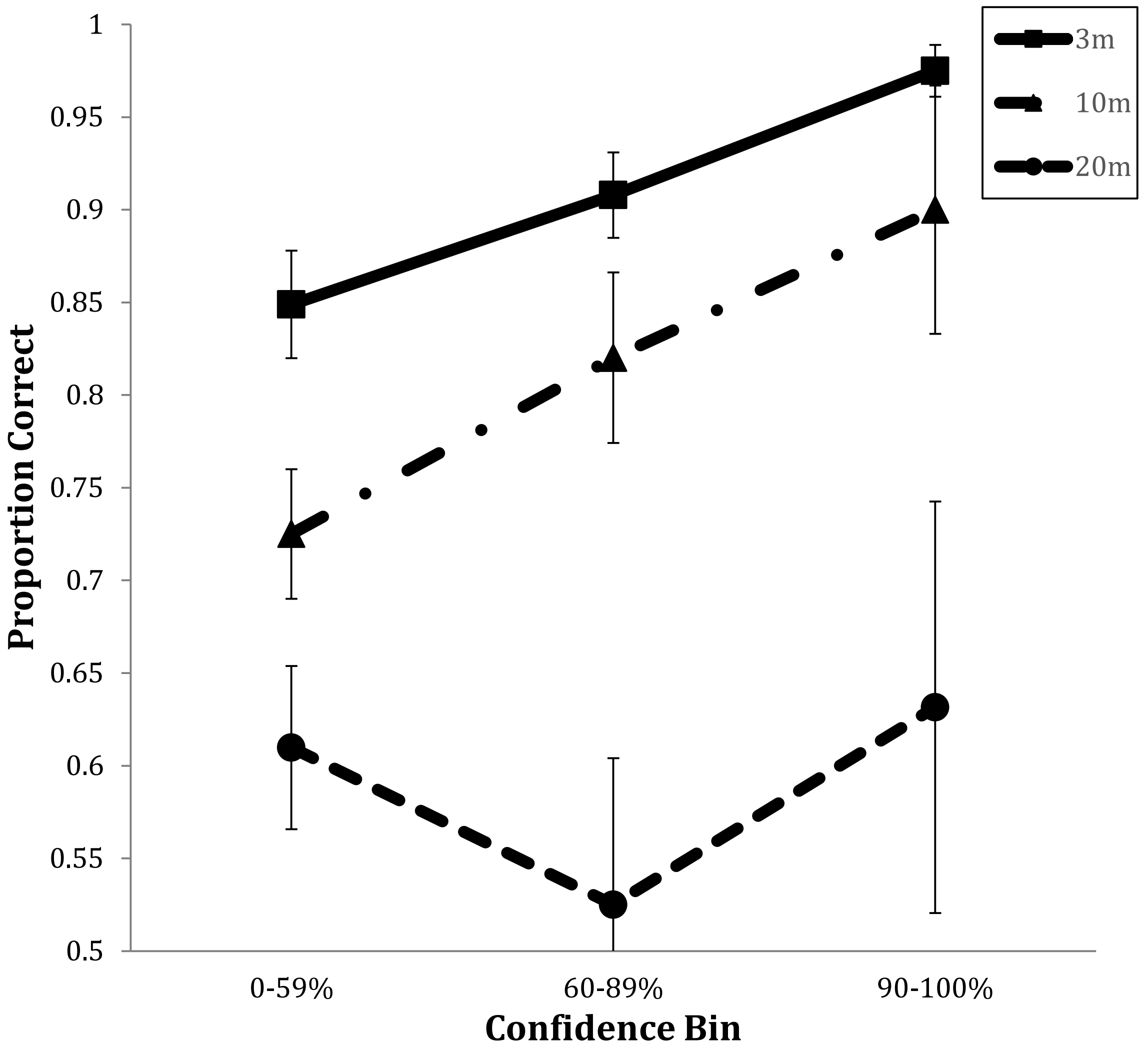

Figure 5. CAC curve for each distance condition in Experiment 2. Error bars represent standard error.

Appendix: Analyses of Lineup Decisions by Block in the Repeated-Measures Design of Experiment 1

Correct Identifications across Experimental Blocks

In order to address potential order effects due to our within-subjects design, we conducted several repeated-measures logistic regressions to ensure that our pattern of results was consistent across the three consecutive blocks of the experiment. However, being that the PA lineup was almost always presented in Block 2, we were unable to evaluate potential order effects for false ID rate. It is still important to show that our correct ID rate pattern (higher for distance of 3 m compared to 10 or 20 m) held up regardless of block. Indeed, a repeated-measures logistic regression revealed no effect of block on correct ID rate when collapsing across distance, with Block 1 correct ID rate = .29 and Block 3 correct ID rate = .31, $\chi 2$ (1, N = 671) = 0.47, p = .50. Analyzed a different way, we also found no effect of block on correct ID rate within each distance. For 3 m, Block 1 (correct ID rate = .47) did not differ from Block 3 (correct ID rate = .54), $\chi 2$ (1, N = 235) = 0.99, p = .36. For 10 m, Block 1 (correct ID rate = .22) did not differ from Block 3 (correct ID rate = .26), $\chi 2$ (1, N = 203) = 0.61, p = .51. Lastly, for 20 m, Block 1 (correct ID rate = .12) did not differ from Block 3 (correct ID rate = .17), $\chi 2$ (1, N = 233) = 1.16, p = .35.

Choosing by Position across Experimental Blocks

For all PP lineups, the perpetrator was placed in either position 2 (top middle) or 5 (bottom middle). It is unlikely that participants could learn where the perpetrator was placed because there were only three blocks, the second block always presented a PA lineup, and we had no designated innocent suspect in a particular position in the lineup. However, it is important to investigate whether participants still somehow picked up on the position of the perpetrator, which would appear as a greater choosing rate for positions 2 and 5 in Block 3 compared to Block 1. A repeated-measures logistic regression confirmed that participants did not learn where the perpetrator was placed in the lineup. Collapsed across PP and PA lineups, positions 2 and 5 received 162 identifications in both Block 1 and 162 identifications in Block 3. For PP lineups, there was no difference in position 2 and 5 responses in Block 1 compared to Block 3 (χ2 (1, *N* = 671) = 0.46, *p* = .50). There was also no difference in position 2 and 5 responses in Block 1 compared to Block 3 for PA lineups (χ2 (1, *N* = 205) = 3.11, *p* = .078).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.