Receiver Operating Characteristic Analysis and Confidence-Accuracy Characteristic Analysis in Investigations of System Variables and Estimator Variables that Affect Eyewitness Memory

Laura Mickes

Royal Holloway, University of London

Author Note

Laura Mickes
Department of Psychology
Royal Holloway, University of London
Egham TW20 0EX
United Kingdom
+44 1784 443711
laura.mickes@rhul.ac.uk

This work was supported by the National Science Foundation SES-1155248. The content is solely the responsibility of the author and does not necessarily reflect the views of the National Science Foundation. The author thanks John T. Wixted for his comments and suggestions; Travis M. Seale-Carlisle for programming the experiments; and Vivian Hwe for help with data collection.

Correspondence concerning this article should be addressed to Laura Mickes (laura.mickes@rhul.ac.uk).

Abstract

Two graphical techniques, receiver operating characteristic (ROC) analysis and what might be termed "confidence-accuracy characteristic" (CAC) analysis, are important tools for investigating variables that affect the accuracy of eyewitness identifications (e.g., type of lineup, exposure duration, same-race vs. other-race identifications, etc.). CAC analysis (a close relative of calibration analysis) consists of simply plotting suspect identification accuracy for each level of confidence. Two parties interested in the results of such investigations include (1) legal policymakers (e.g., state legislators and police chiefs) and (2) triers of guilt and innocence (e.g., judges and jurors). Which type of analysis is the most relevant to which party? The answer is largely a matter of whether the variable in question is a system variable or an estimator variable. ROC analysis, which measures discriminability, is critical for understanding system variables that affect eyewitness accuracy (e.g., the best lineup procedures). Thus, policymakers should be particularly attuned to the results of ROC analysis when making decisions about those variables. CAC analysis, which directly measures the confidence-accuracy relationship for suspect IDs, is critical for understanding the effect of estimator variables on eyewitness accuracy (e.g., exposure duration). Thus, triers of guilt and innocence should be particularly attuned to the results of CAC analysis. The utility of both analyses to system and estimator variables is illustrated by examining both types of analyses on previously published experiments and new experiments.

Keywords: Eyewitness Memory, Confidence and Accuracy, ROC Analysis, Calibration Analysis Recollection and Familiarity, System Variables and Estimator Variables

Two relatively new analyses have been recently recommended to elucidate certain issues in eyewitness identification (ID) research: receiver operating characteristic (ROC) analysis and calibration analysis. ROC analysis measures discriminability (i.e., the ability to discriminate innocent from guilty suspects), and was introduced to the field of eyewitness memory by Wixted and Mickes (2012). Calibration analysis measures the relationship between the subjective probability that an ID is correct (measured using a 100-point confidence scale) and the objective probability that it is correct. This method was introduced to the field of eyewitness memory by Juslin, Olsson, and Winman (1996). Calibration analysis is a specific example of a more general approach that I will refer to as *confidence-accuracy characteristic* (CAC) analysis. CAC analysis simply consists of plotting identification accuracy of suspect IDs (ignoring filler IDs) for each level of confidence regardless of the specific scale that is used (e.g., even if the scale amounts to nothing more than rating confidence as low, medium or high). The aim of this paper is to consider the utility of these two analyses for system variables and estimator variables that affect eyewitness memory (Wells, 1978). System variables can be controlled across criminal cases (e.g., lineup format, lineup size, etc.), whereas estimator variables cannot be controlled in particular criminal cases (e.g., exposure duration, presence or absence of a weapon, etc.). For reasons elaborated upon below, ROC analysis usually best informs decisions made by policymakers about system variables that influence eyewitness memory, whereas CAC analysis best informs decisions made by triers of fact about estimator variables that influence eyewitness memory.

The Meaning of Eyewitness "Accuracy"

The arguments presented in this paper have to do with variables that affect the overall accuracy of eyewitness memory. Thus, it is important to first clarify the term "accuracy", which is often used to refer to different aspects of eyewitness identification performance. Consider, for

example, how retention interval (an estimator variable) affects eyewitness performance. All else being equal, few would doubt that eyewitness memory generally weakens – and, therefore, that eyewitness accuracy generally decreases – as the retention interval increases. For example, as the retention interval increases, the correct ID rate might decrease and false ID rate might increase. The correct ID rate is the proportion of guilty suspects picked from a target-present lineup (i.e., lineups in which the perpetrator is present); and the false ID rate is the proportion of innocent suspects picked from a target-absent lineup (i.e., lineups in which the perpetrator is not present). When the correct and false ID rates are combined into an accuracy measure like percent correct, *d′* or partial area under the ROC curve, they will change in such a way as to reflect a reduced ability, on average, to distinguish between innocent and guilty suspects.

This type of accuracy could be referred to as "accuracy in the *d′* sense" or, more commonly, as "discriminability". Thus, the higher the percent correct or *d′*, or the greater the area under the ROC curve, the greater the accuracy (e.g., discriminability would be higher after short retention intervals relative to long retention intervals). My main claim about the relevance of ROC analysis and CAC analysis to policymakers and triers of fact pertains to variables that are thought to affect eyewitness memory in this sense (i.e., variables that affect the aggregate level of discriminability across a population of eyewitnesses). Such variables include not only retention interval but also exposure duration, same-race vs. other-race IDs, lineup type, number of foils in a lineup, presence vs. absence of a weapon, and so on.

A different use of the term "accuracy" applies to the performance of different subsets of eyewitnesses in a condition involving a single aggregate level of discriminability. For example, holding retention interval constant at 1 week, witnesses who express high confidence might provide many correct IDs and few false IDs (a high proportion correct), whereas witnesses who

express low confidence might provide as many false IDs as correct IDs (a low proportion correct). CAC analysis measures accuracy in this sense, and it can reveal how the relationship between confidence and accuracy changes (or not) across variables that generally affect eyewitness memory (i.e., that affect discriminability). Thus, for example, one can ask how the confidence-accuracy relationship changes as the retention interval increases from 1 day to 1 week. Critically, the relationship between confidence and accuracy can remain the same even if discriminability changes and vice versa. Thus, the two kinds of analyses do not convey the same information.

Whether or not the confidence-accuracy relationship changes as a function of discriminability, my argument will be that ROC analysis is most relevant to informing policymakers about system variables that affect discriminability, whereas CAC analysis is most relevant to informing triers of fact about estimator variables that affect discriminability. This point is important to emphasize because not all system variables affect discriminability, which means that not all variables fall within the scope of my claim. For example, the use of biased vs. unbiased instructions – a system variable – presumably affects response bias (i.e., inclination to choose someone from a lineup) rather than discriminability (Clark, 2005). Although ROC analysis would be useful for testing whether or not that is true, if it turned out to be true, the outcome of the ROC test would not directly indicate to policymakers which instruction ought to be used. That determination is a complex function of subjective values and presumed base rates (Clark, 2012; Mickes, Flowe, & Wixted, 2012). Similarly, the question of whether or not to use a confidence rating scale, the use of which presumably does not affect the ability of an eyewitness to distinguish between innocent and guilty suspects, is another system variable issue that is not directly informed by the outcome of ROC analysis. Calibration analyses provide useful

information about these system variables (e.g., in the complete absence of calibration, policymakers might choose to not take confidence ratings from eyewitnesses), but ROC analysis does not. The focus here is not on variables like these but is instead on system variables and estimator variables that affect eyewitness memory (i.e., variables that affect discriminability).

ROC Analysis vs. CAC Analysis of Estimator Variables

The main goal of ROC analysis is to measure discriminability. An ROC plot is a plot of the correct ID rate and the false ID rate pairs across different levels of response bias (typically measured across different levels of confidence). Although it has been assumed that accuracy for a system variable can be effectively assessed using a diagnosticity ratio (correct ID rate / false ID rate) based on a single correct and false ID rate pair (e.g., Steblay, Dysart, Fulero, & Lindsay, 2001; Steblay, Dysart, & Wells, 2011), this measure is flawed because it conflates response bias with discriminability (Gronlund, Wixted, & Mickes, 2014; National Research Council, 2014). ROC analysis does not. The greater the area under the curve of the ROC, the better eyewitnesses can distinguish between innocent and guilty suspects.[1] A condition that yields a higher ROC is therefore objectively superior to a condition that yields a lower ROC. If the variable in question is a system variable (e.g., simultaneous vs. sequential lineups), the condition that yields the higher ROC should be preferred because for any correct and false ID rate that can be achieved by the lower ROC condition, the higher ROC condition can yield both a higher correct ID rate and a lower false ID rate. This is why it would be sensible for policymakers to choose the procedure that yields the higher ROC.

In contrast to ROC analysis, the main goal of CAC analysis is to measure the relationship between confidence and accuracy across different eyewitnesses whose aggregate performance is

[1] Gronlund et al. (2014) provide a tutorial on conducting ROC analysis for lineup data.

associated with a given level of discriminability (e.g., the aggregate discriminability associated with a 1-week retention interval). This relationship has most often been measured using the point-biserial correlation coefficient, but Juslin et al. (1996) showed that this measure is flawed because its value can vary across a wide range even when the confidence-accuracy relationship exhibits perfect calibration. Perfect calibration exists when an eyewitness expresses a level of confidence that corresponds to the percentage of eyewitnesses who are correct when they express that level of confidence. Thus, eyewitnesses who express 50% confidence in an ID are 50% correct and eyewitnesses who express 90% confidence in an ID are 90% correct are examples of perfect calibration. Instead of using the point-biserial correlation coefficient, which can be both confusing and misleading, Juslin et al. (1996) recommended a calibration approach. This kind of information is what triers of fact would benefit from knowing. For example, whereas a policymaker would prefer a simultaneous lineup if it yields a higher ROC than a sequential lineup, to a juror, the type of lineup procedure that was used to identify the defendant is an estimator variable. Information that would be most informative to a judge or juror is whether a lineup ID that is made with high confidence means that that ID is likely to be accurate. ROC analysis does not provide that kind of information, but calibration analysis does. If a high-confidence ID is as accurate when coming from a low-discriminability lineup as a high-discriminability lineup, it would not matter to a judge or juror that the eyewitness was exposed to the low-discriminability lineup condition. The ID would be equally trustworthy either way. These considerations apply even if the confidence scale is something other than a 100-point subjective probability scale. Thus, for example, even if the confidence scale is nothing more than “high” vs. “low”, triers of fact are better informed by the CAC plot (i.e., a plot relating each level of confidence to accuracy) than by the ROC plot.

When examining the effect of estimator variables on memory performance (e.g., short exposure duration vs. long exposure duration), and when confidence ratings are also recorded, one can easily perform both ROC analysis and CAC analysis. Take, for example, the data from Experiment 1 in Palmer, Brewer, Weber, and Nagesh (2013). In this experiment, a research assistant approached individuals while a second research assistant appeared for 5 s or 90 s, and the participants were tested on their ability to identify the second research assistant from a lineup. A 100-point confidence scale was used in this experiment. Data from the 5 s and 90 s conditions were used to construct the ROC plots and calibration plots shown in Figures 1A and 1B, respectively.[2]

Over the false ID range of 0 to 0.32, the partial area under the curve (pAUC) for the 90 s condition (.128) was higher than the pAUC for the 5 s condition (.098), $D = 1.93$, $p = .053$. Thus, not surprisingly, memory was better (i.e., discriminability was higher) when exposure duration was longer, which is consistent with what eyewitness memory experts might testify to in a court of law. Indeed, the expert might argue that the trustworthiness of an ID made by an eyewitness who only had a brief exposure to the perpetrator is low, whereas the trustworthiness of an ID made by an eyewitness who had a longer exposure to the perpetrator would be higher. However,

[2]The "estimated" false ID rate is the false ID rate divided by the number of lineup members, and this is the typical practice when there is no innocent suspect designated in the target-absent lineups. When measuring pAUC, frequency counts of suspect IDs from target-present lineups are used in the analysis and all foil IDs from fair target-absent lineups are used in the analysis. That is, at this stage of the analysis, one need not divide by lineup size if both conditions have the same number of lineup members (e.g., a comparison of a 6-member simultaneous lineup vs. a 6-member sequential lineup). The reason is that the ROC plot will be visually identical whether the target-absent foil ID rate is plotted on the x-axis or the estimated target-absent foil ID rate is plotted on the x-axis. The only difference between the two plots would be the scale values shown on the x-axis, and changing those values does not change which condition yields the higher ROC. The exception to this rule is Experiment 2, because the comparison was a showup vs. a lineup, so the false ID for the lineup needed to be estimated from the outset.

the trustworthiness of an ID is not what ROC analysis measures. Thus, the fact that discriminability is lower when exposure time is brief might not be a relevant consideration for judges and jurors. The trustworthiness of an ID might be the same across conditions, depending on what the CAC analysis reveals. If so, the fact that the conditions differ in terms of discriminability would not be relevant.

CAC Analysis vs. Calibration Analysis. In calibration studies, calibration accuracy (*C*) is computed using the formula *C* = # correct IDs / (# correct IDs + # incorrect IDs). *A* is computed separately for IDs made with different levels of confidence made using a 100-point confidence scale (e.g., $C_{90\text{-}100}$ would be computed using correct and incorrect IDs made with confidence ratings of 90 to 100, $C_{70\text{-}90}$ would be computed using correct and incorrect IDs made with confidence ratings of 70 to 89, and so on). Although correct IDs always consist of suspect IDs made from target-present lineups, what counts as an incorrect ID varies from study to study (see Juslin, et al., 1996 and Palmer, et al., 2013 for two different examples). Using one approach, all of the errors are counted (including filler IDs, whether they come from target-present or target-absent lineups). Using a different approach, only the innocent suspect IDs (or an estimate of the innocent suspect IDs) are counted, which means ignoring filler IDs on target-present trials and dividing filler IDs by the number of lineup members on target-absent trials. The former approach probably makes the most sense if testing a psychological theory of calibration (because filler picks are errors in the mind of the participant, and the possibility of a filler pick likely informs the participant's confidence rating), whereas the latter approach makes the most sense if the information is to be used by judges and juries (because triers of fact are specifically concerned with *suspects* who have been identified).

As the phrase is used here, CAC analysis refers specifically to a plot of the relationship between confidence and accuracy for correct and incorrect suspect IDs made with varying degrees of confidence (regardless of the type of confidence scale that is used). That is, for each level of confidence, suspect ID accuracy (*A*) = # correct suspect IDs / (# correct suspect IDs + # incorrect suspect IDs). For example, if two levels of confidence are taken (High vs. Low), then A_{High} = # correct high-confidence suspect IDs / (# correct high-confidence suspect IDs + # incorrect high-confidence suspect IDs) and A_{Low} = # correct low-confidence suspect IDs / (# correct low-confidence suspect IDs + # incorrect low-confidence suspect IDs). If no innocent suspect is designated in target-absent lineups (in which case incorrect suspect IDs cannot be directly counted), and if the lineup is fair, then the number of target-absent filler IDs for each level of confidence would be divided by lineup size to estimate the number of innocent suspect IDs.

If the base rates of target-present and target-absent lineups are equal, as is typically true in experimentally controlled studies of eyewitness identification, *A* represents the posterior probability of guilt (i.e., the probability of guilt given that the suspect was identified). Similarly, instead of computing *A* for each level of confidence, one could compute a diagnosticity ratio (*DR*) separately for each level of confidence, where *DR* = # correct suspect IDs / # incorrect suspect IDs. *DR* represents the posterior odds of guilt (i.e., the odds of guilt given that the suspect was identified). Although either measure effectively captures the "information value" of an ID made with a particular level of confidence, the accuracy score seems generally preferable for CAC analyses (which are intended, in part, for consumption by triers of fact) because it corresponds to the more familiar and more intuitive measure "proportion correct." Still, either measure could be used. The situation is more complicated when base rates are unequal (because in that case the

posterior odds of guilt differ depending on what the base rate happens to be), but for the studies considered here, the base rates were equal or very close to being equal.

CAC Analysis and the Diagnosticity Ratio. Because we have argued against the utility of the diagnosticity ratio in the past (Wixted & Mickes, 2012), a few more words about that measure are in order. In the discussion above, the *DR* value was computed separately for each level of confidence. When the *DR* has been used in the past to assess system variables like simultaneous vs. sequential lineups, it has been computed based on correct and false ID rates collapsed across confidence. That is, *DR* = correct ID rate / false ID rate, where correct and false IDs are counted no matter the level of confidence. Just as the *DR* computed separately for each level of confidence provides useful information to triers of fact about the trustworthiness of an ID made with a particular level of confidence, the *DR* computed from correct and false ID rates (collapsed across confidence) does so as well. The problem with the diagnosticity ratio is not that it is uninformative per se; instead, it is uninformative when the goal is to differentially evaluate system variables such as simultaneous vs. sequential lineups. However, it does provide useful information for triers of fact. For example, if (ignoring confidence) sequential lineups typically yield a higher diagnosticity ratio than simultaneous lineups – as some claim (e.g., Steblay et al., 2012) but others dispute (Clark, 2012; Gronlund et al., 2009) – then, disregarding confidence, an ID made from a sequential lineup would be more trustworthy than one made from a simultaneous lineup. This is useful information to a jury (i.e., it is useful information when lineup type is an estimator variable), but it is a mistake to assume that this same information is what policymakers should use to decide which type of lineup to use.

Why is such information not useful for policymakers? Should they not prefer the procedure that yields more trustworthy IDs? Actually, policymakers should prefer the procedure

that yields the higher ROC. As noted earlier, the reason is that if Procedure A yields a higher ROC than Procedure B, then, then for any level of performance achieved using Procedure B (i.e., a certain correct ID rate, false ID rate, and diagnosticity ratio), a higher level of performance can be achieved by Procedure A (a higher correct ID rate, lower false ID rate, and higher diagnosticity ratio) simply by adjusting response bias. These same considerations explain why policymakers should prefer the procedure that yields a higher ROC even if CAC analysis shows that the procedure with a lower ROC is associated with a higher (and more desired) level of high-confidence accuracy. By encouraging witnesses to be more cautious about making a high-confidence ID, the procedure associated with a higher ROC can achieve even higher accuracy while simultaneously achieving a higher high-confidence correct ID rate and lower high-confidence false ID rate.

CAC Plots of Real Data. CAC plots for the participants in the exposure duration condition of Palmer et al. (2013) are shown in Figures 1B and 1C. Figure 1B retains the confidence levels binning as Palmer et al. reported[3] and Figure 1C shows the lower confidence levels collapsed further to reduce noise. This analysis focuses only on suspect IDs, which was accomplished by counting all suspect IDs from the target-present lineups and all foil choices divided by lineup size from the target-absent lineups, eight, because there was no designated innocent suspect. The CAC plots indicate that the participants appreciated the effect that exposure time would have on their memory and compensated for it by appropriately adjusting their confidence, particularly at the high-confidence end of the scale. In other words, a high-confidence ID made from the 5 s condition was as likely to be correct as a high-confidence ID

[3] The difference between the CAC curves in the current paper and the calibration curves Palmer et al. (2013) paper results from the fact that Palmer et al. did not divide by the number of lineup members in the target-absent lineups.

made from the 90 s condition. This is the key point. Even though the ROC shows lower discriminability in the short exposure condition, the CAC analysis suggests that while participants in that condition were less likely to make relatively high-confidence IDs, when they did, they were as accurate as the high-confidence IDs from the long exposure condition.

These considerations illustrate why, for an estimator variable, ROC analysis is generally less informative than CAC analysis. When confidence ratings are available, what is important for the legal system to know is how trustworthy an ID made with a particular level of confidence is, and its trustworthiness could be the same whether the memory conditions were good (higher ROC) or bad (lower ROC). As noted above, assuming equal base rates, the trustworthiness of an ID is equivalent to the posterior probability of guilt (i.e., the probability that the identified suspect is guilty), and that is precisely the dependent measure plotted on the y-axis of a CAC plot like those shown in Figures 1B and 1C.

Next, both ROC analysis and CAC analysis is used to investigate another estimator-like variable[4]: whether or not the witness recollects details associated with a suspect identified from a lineup.

Experiment 1

Identification decisions can be based on the familiarity of the suspect's face or may also involve the recollection of additional details (such as what the suspect was wearing). Because the legal system has no control over whether or not recollection occurs, recollection is like an estimator variable. Palmer, Brewer, McKinnon, and Weber (2010) used the Remember/Know procedure to determine whether IDs accompanied by the recollection of details (indicated by a "Remember" judgment) were more accurate than IDs that were not accompanied by the

[4] Recollection is like an estimator variable in the sense that it is a determinant of discriminability that is not under the control of the legal system.

recollection of any details (indicated by a "Know" judgment). As with many previous list-memory studies, they found that recollection-based IDs were more accurate than familiarity-based IDs, but the difference was no longer apparent once confidence was considered. However, they did not perform either ROC analysis or CAC analysis. Experiment 1 was designed to further illustrate how these analyses are used. Because participants can be confused by the terms “remember” and “know” to indicate recollection- or familiarity-based memories, respectively, without lengthy and detailed training, participants were simply required to answer whether they recollected details about the perpetrator presented during the study phase or not. Despite this methodological difference, the results reinforce the claims of Palmer et al. Although one might question the need for such a replication (since the methodological points presented above could be made using existing data), the field's current replicability crisis suggests that further illustrating a methodological point in the context of a replication study could be doubly useful.

Method

Participants

University of California, San Diego (UCSD) undergraduates (n = 307) participated online for course credit.

Materials

In a 30 s video of a mock carjacking crime, a victim approaches and sits in her car, and a perpetrator opens the door and pulls her out. The perpetrator’s face was shown for 10 s. One hundred and forty three filler images were selected from the supervised population in the Florida Offender Database (dc.state.fl.us) based on descriptions from 15 additional participants who watched the video and answered questions about the perpetrator’s appearance. That defined the following search terms: white male; age ranged from 18-29 years; weight ranged from 150-190

lbs.; height ranged from 5'9 – 6'0; hair color blond. All images were set to gray scale. All lineups were 6-person simultaneous lineups (in a 2x3 array) that either contained the perpetrator (target-present lineups) or contained six foils (target-absent lineups). No foils were designated as the innocent suspect. The filler images were randomly pulled from the large pool and displayed in random positions for each participant. The target image was also presented in any one of the six positions in the target-present lineups.

Procedure

Participants were randomly assigned to the target-present or target-absent condition. Participants were informed that they would watch a brief video and they should pay special attention because they would answer questions about the video later. They then watched the video and took part in a 5-minute distractor task (a game of Tetris). During the test phase, participants were informed that they would see six faces in a lineup, and the perpetrator from the video may or may not be in the lineup. They were instructed if they saw the perpetrator from the video in the lineup to select the button under the image, and if he was not present, select the "not present" button. They were also informed they would be asked to rate the level of confidence that they did or did not see the perpetrator from the video in the lineup and then answer several more questions about the video. If they chose a lineup member, they rated their confidence on a scale from 0-100% (0 = "Just Guessing" and 100% = "Absolutely Certain"). Also if they chose a lineup member, they then answered "yes" or "no" whether they recollected details about the person they selected (the specific question was "can you recollect details about this person?"). They then answered four multiple-choice questions about the video (two questions about his clothing, and one question about the weather) including a validation question (what crime did he commit?). After they answered the questions, they were debriefed.

Results and Discussion

Five participants incorrectly answered the validation question and were therefore not included in the analysis. The remaining 302 participants had been randomly assigned to the target-absent (n=150) or the target-present (n = 152) lineup condition. Of those, 166 made an ID, and 122 reported that they recollected details; and 136 did not make an ID (52 missed the target from a target-present lineup and 84 correctly rejected the suspect from a target-absent lineup). Response frequencies for false IDs, foil picks, misses, and correct IDs per level of confidence for those who recollected details and those who did not are displayed in the table in Appendix A.

To determine if ID decisions were more accurate when participants reported that they recollected details, ROC analysis was conducted in the manner described above. Figure 2A shows the ROC curves for the recollection and no-recollection responses (c.f., Slotnick, 2010). The pAUC was significantly higher for the recollection responses (0.029) than the no-recollection responses (0.010), $D = 3.22$, $p < 0.001$ (false ID range was 0 - 0.13). Thus, participants who recollected details about the event discriminated between the innocent suspects and the guilty suspect better than those who did not. Consequently, one might be tempted to regard an ID accompanied by the recollection of details to be more trustworthy than an ID that is not accompanied by recollection. However, whether or not that is true is addressed by CAC analysis, not ROC analysis.

To determine if those who claimed they could not recollect details about the event adjust their confidence to reflect their likely accuracy on the ID decision, CAC analysis was conducted in the manner described above. The CAC plot in Figure 2B shows that confidence and accuracy are related similarly for the responses associated with recollection of details and those without recollection of details. Because of the small number of responses in some bins, the low

confidence responses were collapsed to 0-60. Moreover, there are no apparent differences between the calibration curves for recollection and no-recollection responses. This result suggests that participants appreciate when their memories are not strong and *appropriately adjust their confidence to reflect that fact*. This is the key consideration. The practical implication is that even though overall memory performance is clearly worse when recollection does not occur, that is not a relevant consideration for triers of fact if confidence ratings are available. As with the exposure duration manipulation considered earlier, a moderately high-confidence ID[5] is similarly accurate, and therefore similarly trustworthy, whether memory conditions are good or bad.

The Utility of ROC Analysis in Examining System Variables

To demonstrate the utility of ROC analysis in measuring lineup discriminability, Mickes, et al. (2012) conducted experiments in which they manipulated the type of lineup procedure. Because the legal system determines the nature of the lineup procedure, this is considered to be a system variable. After watching a video of a mock crime, memory for the perpetrator was tested on a simultaneous or a sequential lineup procedure. In Figure 3A (from Mickes et al. Experiment 1A), the simultaneous ROC curve falls further from the chance line than the sequential ROC from Experiment 1a (i.e., the simultaneous lineup procedure yielded greater discriminability). This surprising finding goes against what had been repeatedly concluded in the past when the diagnosticity ratio was used (e.g., Lindsay & Wells, 1985; Steblay et al. 2001; Steblay et al. 2011). However, based on the small number of studies that have been conducted thus far, ROC

[5] There were only two high-confidence IDs (both correct) made by those who claimed they did not recollect details about the perpetrator. Because two IDs are too few to base any conclusions on, those ratings are collapsed with lower confidence ratings when conducting ROC analysis and CAC analysis.

analysis suggests that the simultaneous procedure reliably outperforms the sequential lineup procedure (Gronlund et al. 2012; Dobolyi & Dodson, 2014; Carlson & Carlson, 2014).

Figure 3B shows the CAC curves for Experiment 1a of Mickes et al. (2012). The data for the sequential condition are a bit variable, but it is clear that the two procedures perform similarly at the high end of the confidence scale. If the CAC curves for simultaneous and sequential lineups turned out to be identical, would it mean that the ROC data are irrelevant? That depends on who is using the information. Policymakers, for example, are in a position to implement either lineup procedure. Thus, from their perspective, lineup procedure is a system variable, and ROC analysis is highly relevant even if the CAC curves associated with simultaneous and sequential lineups are identical. The reason is that even if high-confidence IDs are equally trustworthy whether the ID was made from a simultaneous or a sequential lineup, the use of the lineup procedure associated with a higher ROC will yield a larger number of cases associated with IDs made with different levels of confidence. As an example, imagine that for both simultaneous and sequential lineups, a CAC analysis shows that suspect IDs made with 70% confidence or more are 90% accurate (as in Figure 3B). If simultaneous lineups yield a higher ROC than sequential lineups, then out of a set of 1000 eyewitnesses, there might be 300 high-confidence IDs that provide such reliable information, whereas if a sequential lineup is used there might be only 150 such IDs. In other words, if two conditions yield the same calibration curve and different ROCs, the procedure with the higher ROC would yield a greater number of suspect IDs associated with a given level of reliability (e.g., a greater number of IDs with >70% reliability). For that reason, the procedure that yields the higher ROC would be more useful to the legal system, so it would make sense for policymakers to mandate the use of that procedure.

Wells and colleagues have long argued that, as a system variable, the sequential

procedure is superior to the simultaneous procedure because the former yields a higher diagnosticity ratio than the latter (ignoring confidence). However, even if one accepts that empirical claim (and many do not; e.g., Ebbeson & Flowe, 2002; Malpass, Tredoux, & McQuiston-Surrety, 2009), the diagnosticity ratio is irrelevant to the system-level question of which procedure should be used by the legal system. The systems-level question is better addressed using ROC analysis (National Research Council, 2014). Although the diagnosticity ratio is not relevant to the system-variable question about which lineup procedure should be used as a matter of policy, it is relevant to the estimator-variable question of whether or not one lineup procedure yields a more trustworthy ID than the other. In its crudest form, the diagnosticity ratio can be computed for each procedure without regard for the level of confidence expressed by the eyewitness. In fact, this is how the diagnosticity ratio is usually computed. However, because the diagnosticity ratio is strongly related to confidence (e.g., Brewer & Wells, 2006), a much better approach (if one wished to use lineup procedure as an estimator variable) would be to compute the diagnosticity ratio for choosers separately for different levels of confidence. To make the results more understandable when base rates are equal, it would be better (in my view) to convert the ratio associated with each level of confidence into a proportion correct score for each level of confidence. In that case, one would have a CAC curve that represents the posterior probability of guilt separately for each level of confidence.

Experiment 2

Experiment 2 illustrates the use of ROC analysis and calibration analysis by conducting a study of what would usually be construed of as a system variable, namely, lineup size. Conceivably, lineup size affects discriminability. Four faces were presented to participants in an incidental learning task. Unbeknownst to them, one of those faces was designated as the

"perpetrator", and memory for that face was tested on a showup or on a 6-person simultaneous lineup.

Method

Participants

UCSD undergraduates (n = 500) participated online for course credit. Participants were randomly assigned to a target-absent 6-person lineup (n = 139), a target-present 6-person lineup (n = 99), a target-absent showup (n = 126), or a target-present showup (n = 136).

Materials

Stimuli were 350 photos of faces of young (born after 1983) White males taken from the Arkansas Department of Corrections database (adc.arkansas.gov). Three additional photos of "non-targets" were taken from the same database (a young White female, a young African American male, and an old White male). All images were set to gray scale and altered. The altering of the images involved rotating them horizontally, applying a 5% noise filter, adding 60 level of brightness, and applying a linear burn filter. The reason for doing this was to minimize concerns that participants were remembering the image itself, not the face, per se.

All lineups were 6-person simultaneous lineups (images were presented in a 2x3 array) that either contained the perpetrator (target-present lineups) or contained six foils (target-absent lineups). No foils were designated as the innocent suspect. The showup image was displayed in the center of the screen and was the same size as the images in the lineups. The same fillers were used for both lineups and the showups. The filler images were randomly pulled from the large pool and displayed in random positions for each participant. The target image was also presented in any one of the six positions in the target-present lineups.

Procedure

Participants were informed that their task was to count the number of faces that were presented because they would be asked to provide the number later. During the study phase, four faces appeared for 3s with a 250ms ISI (in random order for each participant). Three of the study faces were non-targets, and the target, or perpetrator, was a photo randomly drawn from the large pool of young White male images.

After a 10-minute distractor task (a game of Tetris), participants were randomly assigned to the showup or lineup condition (either target-absent or target-present). If assigned to the showup condition, participants were informed that they would see a face and that if it was one of the faces presented earlier, they should select it by clicking on the button below the face. If not, they should click on the "not shown" button. If assigned to the lineup condition, participants were informed that they would see six faces in a lineup, and if they could identify one of the faces that was presented earlier, they should select it by clicking the button below the face, and if not, they should select the “not shown” option. After making their decision, in the showup and lineup condition, participants indicated their confidence on the same scale used in Experiment 1. Next, they provided the number of faces presented during study (M = 4.33, SD = 1.01). Finally, they were debriefed.

Results and Discussion

Response frequencies for false IDs, foil picks (for the lineup condition), misses, correct rejections, and correct IDs per level of confidence for both conditions are displayed in the table in Appendix B. To conduct ROC analysis, the false ID rate was estimated by dividing by six. This must be done when comparing a lineup to a showup, or else the lineup is necessarily placed at an unfair advantage. Figure 4A shows the showup and lineup ROCs. Two trends are apparent: 1) the simultaneous lineup yields higher discriminability than the showup, and 2) the showup

yields more liberal responding than the lineup (i.e., the showup ROC points are shifted more to the right than the lineup ROC points).

With regard to the statement that the showup yields more liberal responding than the lineup, imagine that lineups and showups yield the same ROC curve and that the correct and false ID rates using a showup were deemed to be too high. To achieve more conservative responding, the administrator could psychologically change the witness's criterion (urging more caution before making an ID), or the administrator could achieve the same empirical result by using a lineup (effectively achieving more conservative responding without actually changing the witness's decision criterion). The lower correct and false ID rates would be obtained using a lineup because many of the IDs made by the witness would land on fillers instead of on the suspect (guilty or innocent).

The showup pAUC (0.037) was significantly less than the lineup pAUC (0.065), D = 2.44, p = 0.015 (false ID range was 0 - 0.12). These findings replicate those reported by Gronlund et al. (2012) and are consistent with much additional non-ROC evidence indicating that showups are diagnostically inferior to simultaneous lineups (e.g., Clark, 2012). With regard to the system variable issue of which ID procedure should be used when there is a choice between a showup and a 6-person simultaneous lineup, these ROC results show that the lineup should be preferred. For example, suppose it was determined that high confidence IDs from a showup would be used to further police investigations and criminal prosecutions. According to the showup ROC, the high-confidence (>90%, represented by the leftmost point on the showup ROC) correct ID rate is 0.49, so many guilty suspects would be further pursued. Unfortunately, the high-confidence false ID rate is 0.10, so quite a few innocent suspects would be further pursued as well. However, using a lineup, one could achieve a higher correct ID rate and a lower

false ID rate even if IDs made with lower levels of confidence were used. For example, for IDs made with 70% confidence or more (fourth point from the left on the simultaneous ROC), the correct ID rate is 0.57 and the false ID rate is 0.06.

Figure 4B shows the corresponding CAC analyses. Although it is theoretically possible that participants would appreciate the lower accuracy associated with a showup and adjust their confidence ratings accordingly, these data do not seem to bear that out. For high-confidence IDs (e.g., confidence >70%), showup accuracy is lower than a high-confidence ID made from a lineup. Indeed, this conclusion would seem to hold throughout a fairly wide range of confidence (except for low confidence ratings of 0-60%). Thus, with regard to the estimator variable issue of whether high-confidence IDs made from a lineup vs. a showup are equally trustworthy for a given level of confidence, the answer based on these data would appear to be no. A high-confidence ID made from a simultaneous lineup appears to be more trustworthy than a high-confidence ID made from a showup.

Does the fact that the showup procedure yielded data that were closer to "perfect" calibration make it better than the lineup procedure? I would argue that the answer is no because the legal system cares more about high accuracy than perfect calibration. An important point to consider is that even low confidence responses were higher in accuracy in the lineup procedure. This result may seem at odds with the literature, but in the Brewer & Wells (2006) paper, for example, the low confidence IDs had 3 to 1 probability of being correct (if limited to suspect IDs). When the matter makes it to the courtroom, judges and jurors should know that (at least according to the current data) even if the witness expressed a low-confidence ID when choosing from a lineup, that ID is likely to be moderately high in accuracy. How much weight to attach to an ID made with "moderately high" accuracy is a judgment call for the jurors to make.

General Discussion

Two main points were made in this paper: 1) ROC analysis of system variables that affect eyewitness memory (e.g., lineup size, lineup format, etc.) is most applicable for policymakers, and 2) CAC analysis of estimator variables that affect eyewitness memory (e.g., duration of exposure, retention interval, etc.) is most applicable to triers of guilt or innocence. A decision about a system variable can be best informed by ROC analysis because the procedure that yields a higher ROC can be used to achieve a higher correct ID rate and a lower false ID rate than the alternative procedure. A decision about an estimator variable is generally best informed by CAC analysis because, whether or not one condition yields a lower ROC (and, therefore, lower discriminability) than another, high-confidence IDs could be equally trustworthy from either condition (and a jury, for example, is interested in the trustworthiness of a high-confidence ID).

Because certain conditions adversely affect eyewitness memory (i.e., they adversely affect discriminability), a natural assumption is that IDs based on memories formed during adverse conditions are not as trustworthy as IDs made under better conditions. Based on the finding that the correlation between confidence and accuracy decreases when conditions are worse, Deffenbacher (1980) proposed the *optimality hypothesis*, which holds that more ideal conditions under which a crime is witnessed result in a better confidence-accuracy relationship. But this hypothesis fails to take into account the possibility that witnesses appreciate the effect poorer conditions have on memory and adjust their confidence accordingly. Indeed, using calibration analysis, Palmer et al. (2013) showed that participants appropriately adjusted confidence under a variety of poor memory conditions such that accuracy associated with different levels of confidence was the same as in the corresponding good memory conditions. As they pointed out, these findings are contrary to the optimality hypothesis. In cases like this, the

fact that an estimator variable affects eyewitness memory may not be as relevant as it is often assumed to be. Whether memory conditions are good or poor, an ID made with a particular level of confidence seems to be equally trustworthy for the estimator variables investigated by Palmer et al.

Why are people often good at appropriately adjusting confidence across conditions that differ in terms of discriminability? Mickes, Hwe, Wais, and Wixted (2011) offered the following explanation based on error feedback training that occurs during the course of life:

> "… experience may be what teaches a participant to express high confidence when memory is strong (and likely to be accurate) and to express low confidence when it is weak (and likely to be inaccurate). (p. 255).

In other words, because of prior learning based on error feedback, individuals are reasonably adept at assessing the probability of making errors based on the subjective state of their memory and, in turn, assign fitting confidence. For example, when there is a large chance of being wrong about a recognition decision (based on prior experience involving states of memory similar to the state that currently prevails), an adult will indicate they are guessing. If, on the other hand, an adult says they are very confident in a memory report, they are indicating that, given the prevailing state of memory, experience has taught that there is a small chance they are making an error about that recognition decision. Thus, in situations where their extensive past experience is relevant, adults appear to be reasonably expert at knowing the likelihood that the current state of their memory will lead to an accurate decision.

Although participants may appropriately adjust confidence when they have learned from experience that the prevailing state of memory is associated with low accuracy, certain conditions may adversely affect memory accuracy unbeknownst to the eyewitness, in which case confidence may not be as closely tied to accuracy. For example, due to limited experience,

individuals may be typically unaware that they are less able to discriminate people of different races (as in the own race bias), and that lack of awareness may result in overconfidence and a weakening of the confidence-accuracy relationship (Hourihan, Benjamin, & Liu, 2012). Similarly, participants in Experiment 2 did not adjust their confidence in such a way as to maintain high-confidence high-accuracy performance for showups. One possibility is that participants fail to fully appreciate that showups are more difficult than they seem because, for this memory test, even innocent suspects will match the description of the perpetrator and may therefore seem familiar. By contrast, simultaneous lineups immediately offer useful information to the participant about how difficult this memory test really is (Wixted & Mickes, 2014). Unlike their performance with respect to showups, participants in Experiment 1 seemed to appreciate that memory is worse when details are not recollected (something that ordinary life experience would probably teach) and adjusted their confidence criteria accordingly (e.g., becoming unwilling to make a high-confidence ID unless the recollection-free memory was very strong). Thus, from a judge's or juror's perspective, a high-confidence ID may be equally trustworthy regardless of whether that ID was accompanied by the recollection of details or not. The CAC data presented in this paper are examples of the kind of data that jurors and judges ought to know.

Practical Applications

Depending on the nature of the variable (system or estimator), questions regarding discriminability and the confidence-accuracy relationship can best be answered with ROC analysis and CAC analysis, respectively. Because policymakers are in the position to make systemic changes, they should be informed by results from ROC analysis to make decisions about which system variables that affect eyewitness memory to endorse. For example, when

deciding which of two lineup procedures to use (e.g., simultaneous versus sequential lineup procedures), the results from ROC analysis are most relevant. The endorsed procedure should be the one that yields the highest ROC curve because that is the procedure that can simultaneously maximize correct IDs while minimizing incorrect IDs.

Because triers of guilt or innocence are in the position to weigh evidence from various estimator variables that affect eyewitness memory, their decisions should be primarily informed by results from CAC analysis. This may also call for a re-thinking of the way that experts testify about certain matters in the courtroom. Experts are often called upon to testify on a range of factors that may or may not affect the reliability of a witness's ID. For example, an expert may testify that a witness's ID is unreliable because duration of exposure was short. In the language of ROC analysis, such an expert is essentially testifying to the fact that IDs made from low ROC conditions are less reliable compared to high ROC conditions. However, if confidence ratings associated with the initial eyewitness ID are available, the calibration results – not the ROC results – are relevant. If it is consistently shown that individuals can appreciate when certain conditions are poor and adjust their confidence accordingly, then this is what experts ought to focus the jurors and judges' attention on, not the fact that one condition tends to have lower accuracy (i.e., lower d') than another. Thus, instead of saying "Memory is poor when the witness only gets a brief look at the perpetrator", for example, a more appropriate statement would be along the following lines: "Though memory tends to be poorer when the witness only gets a brief look at the perpetrator, they typically appreciate and adjust their confidence to reflect that circumstance. So, a high confidence suspect ID is likely to be as accurate as a high-confidence response from a witness exposed to a longer look at the perpetrator."

References

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relation in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30.

Carlson, C. A. & Carlson, M. A. (2014). An Evaluation of Perpetrator Distinctiveness, Weapon Presence, and Lineup Presentation using ROC Analysis. *Journal of Applied Research in Memory and Cognition, 3,* 45-53.

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

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

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence. *Law and Human Behavior, 4,* 243-260.

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: a criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied, 19,* 345-357.

Ebbesen, E. B., & Flowe, H. D. (2002). Simultaneous v. sequential lineups: What do we really know? Retrieved from http://www2.le.ac.uk/departments/psychology/ppl/hf49/SimSeq%20Submit.pdf

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S., Wooten, A. & Graham, M. (2012). Showups Versus Lineups: An Evaluation Using ROC Analysis. *Journal of Applied Research in Memory and Cognition, 1,* 221-228.

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating Eyewitness Identification Procedures Using ROC Analysis. *Current Directions in Psychology Science, 23,* 3-10.

Hourihan, K.L., Benjamin, A.S., & Liu, X. (2012). A cross-race effect in metamemory: Predictions of face recognition are more accurate for members of our own race. *Journal of Applied Research in Memory and Cognition, 1* (3), 158-162.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1304-1316.

Keast, A., Brewer, N. & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology, 97*, 286–314.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556–564.

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D., (2009). Public policy and sequential lineups. *Legal and Criminological Psychology, 14,* 1-12.

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

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong memories are hard to scale. *Journal of Experimental Psychology: General, 140,* 239-257.

Palmer, M., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Pscyhologica, 133,* 137-145.

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.

Slotnick, S. D. (2010). "Remember" source memory ROCs indicate recollection is a continuous process. *Memory, 18,* 27-39.

Steblay, N. K., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25,* 459-473.

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

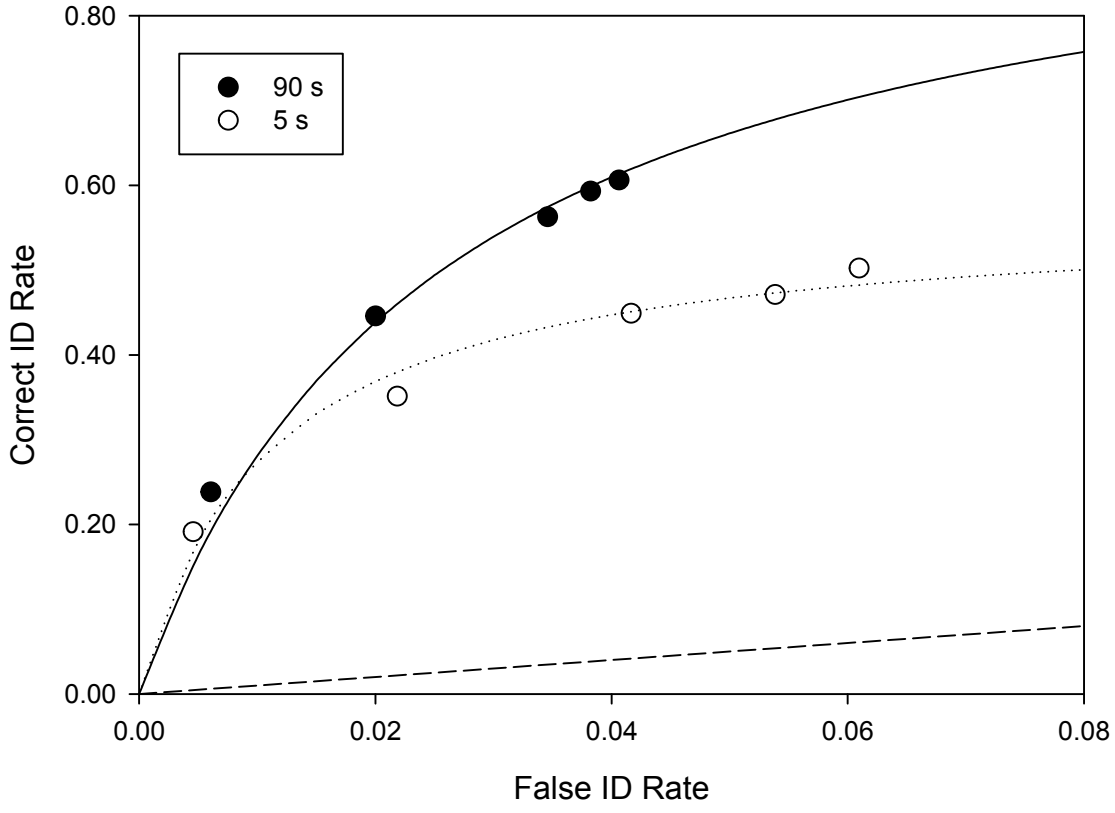
Wells, G. L. (1978). Applied eyewitness-testimony research: system variables and estimator variables. *Journal of Personality and Social Psychology, 12,* 1546-1557.

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.

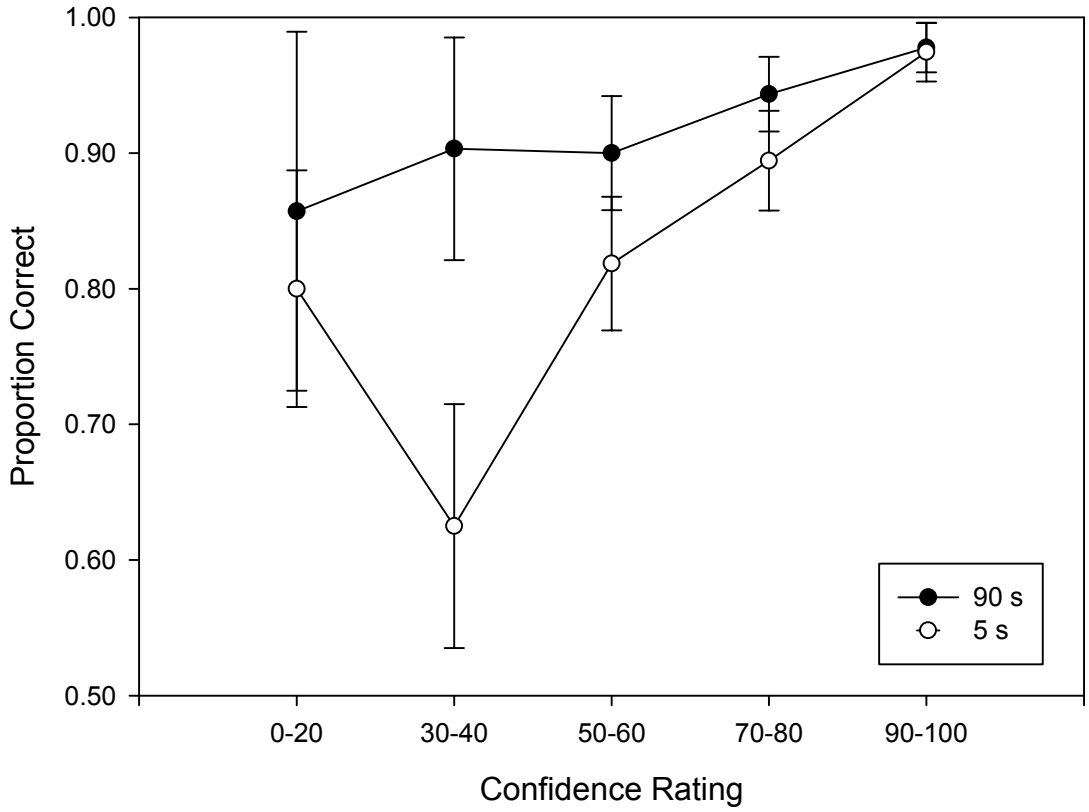
Wixted, J. T. & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121,* 262-276.

Figure 1. ROC curves (Figure 1A) and CAC curves (Figures 1B and 1C) for the exposure duration conditions in Experiment 1 of Palmer et al. (2013). Figure 1B retains the confidence levels binning as Palmer et al. reported and Figure 1C shows the lower confidence levels collapsed further. The dashed line in Figure 1A represents chance performance and the bars in Figures 1B and 1C represent standard error bars.

A.



B.


1.00
0.90
0.80
0.70
0.60
0.50
Proportion Correct
0-20
30-40
50-60
70-80
90-100
Confidence Rating
90 s
5 s

C.

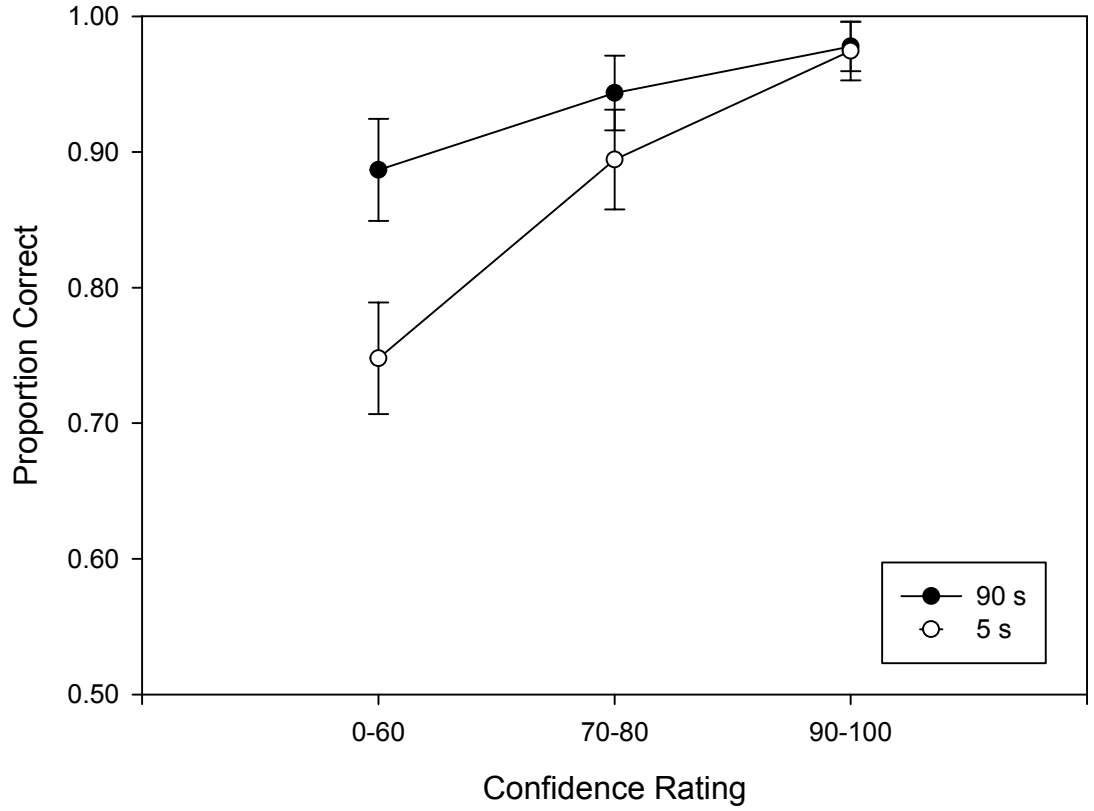
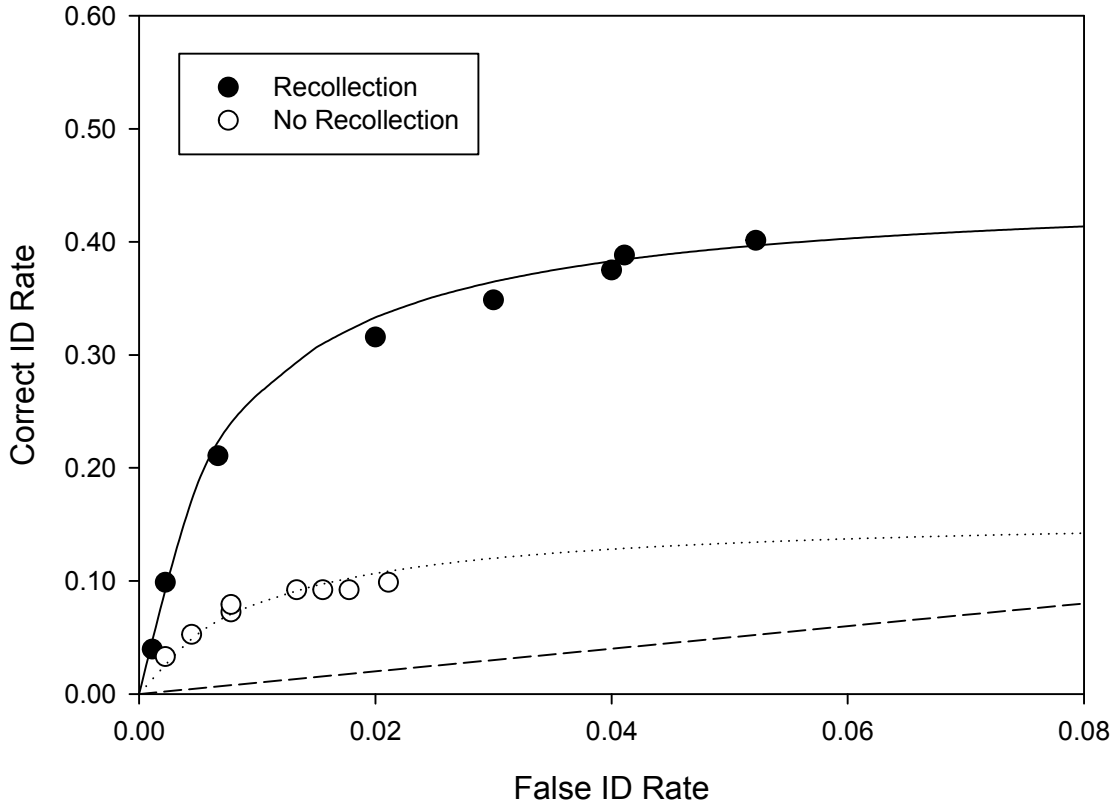

1.00
0.90
0.80
0.70
0.60
0.50
Proportion Correct
0-60
70-80
90-100
Confidence Rating
90 s
5 s

Figure 2. ROC curves (Figure2A) and CAC curves (Figure 2B) for the recollection and no recollection responses in Experiment 1. The dashed line in Figure 2A represents chance performance and the bars in Figure 2B represent standard error bars.

A.



B.

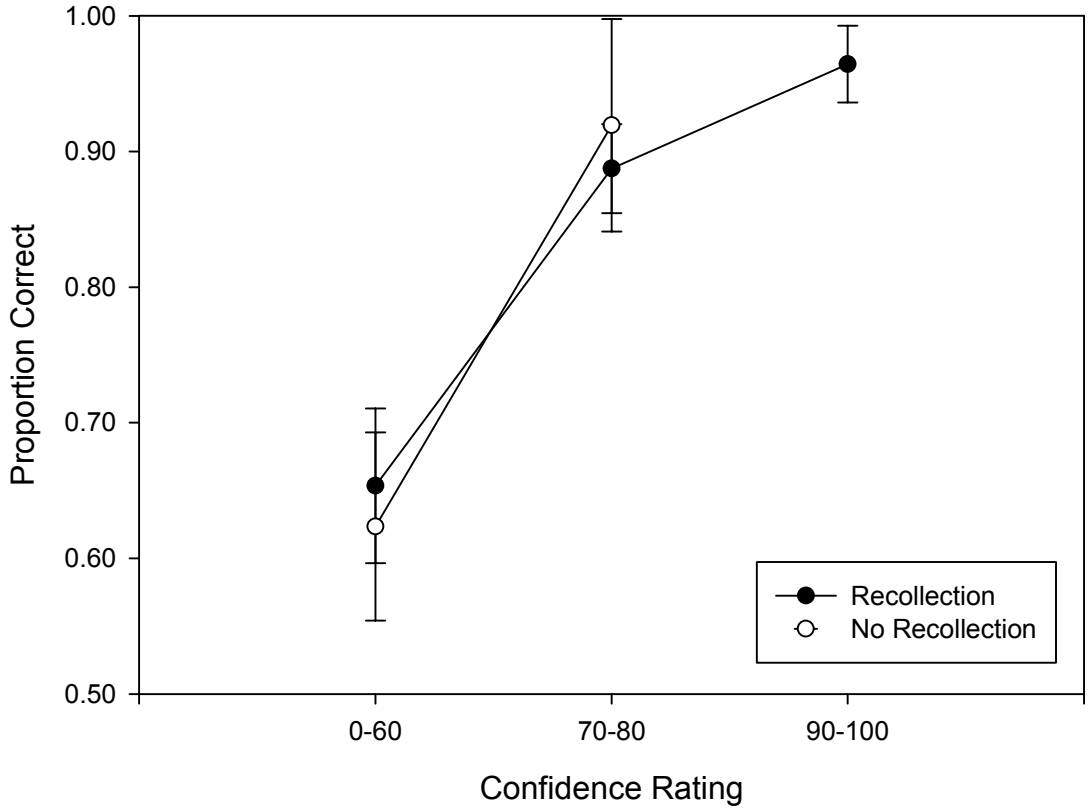
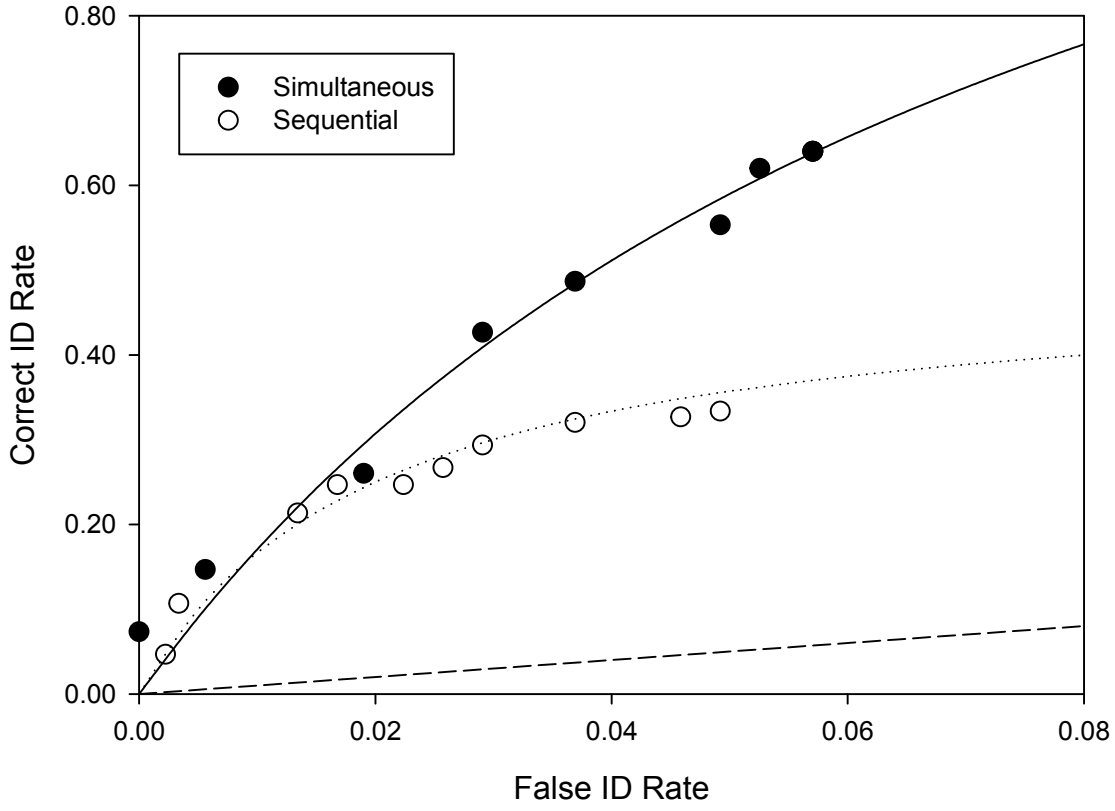

1.00
0.90
0.80
0.70
0.60
0.50
Proportion Correct
0-60
70-80
90-100
Confidence Rating
Recollection
No Recollection

Figure 3. ROC curves (Figure 3A) and CAC curves (Figure 3B) for simultaneous and sequential lineup procedures in Experiment 1a of Mickes, Flowe & Wixted (2012). The dashed line in Figure 3A represents chance performance and the bars in Figure 3B represent standard error bars.

A.



B.

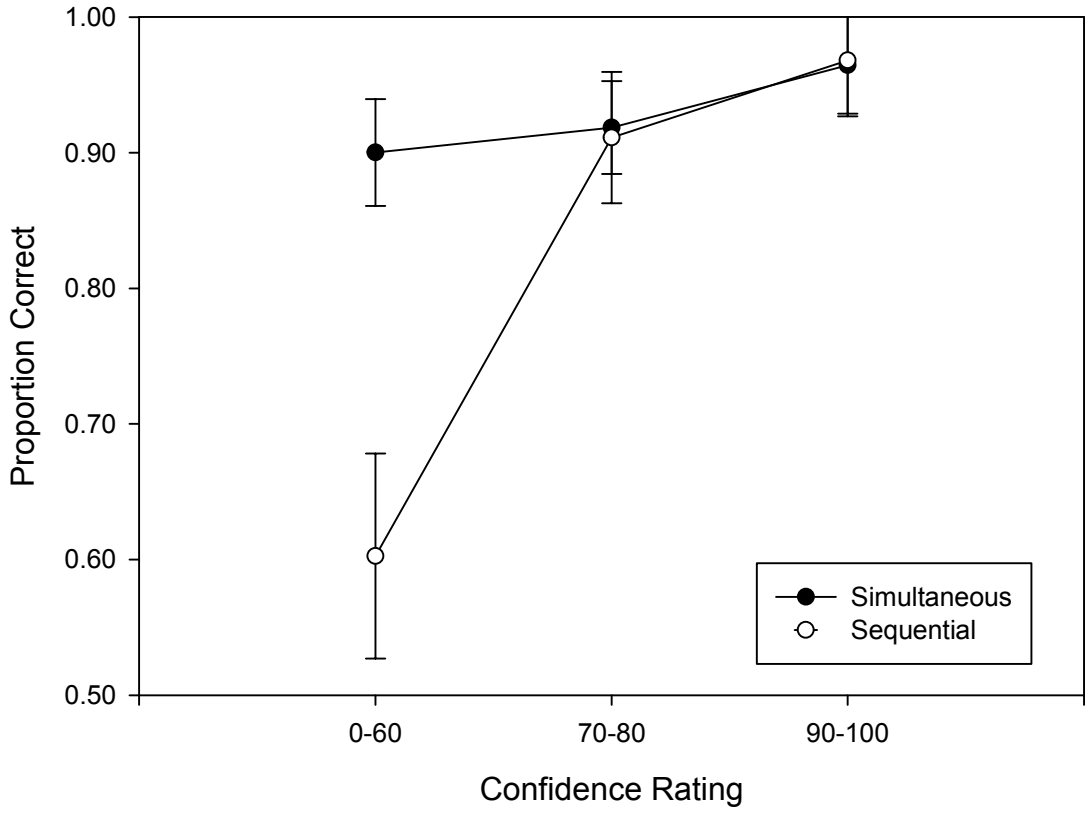
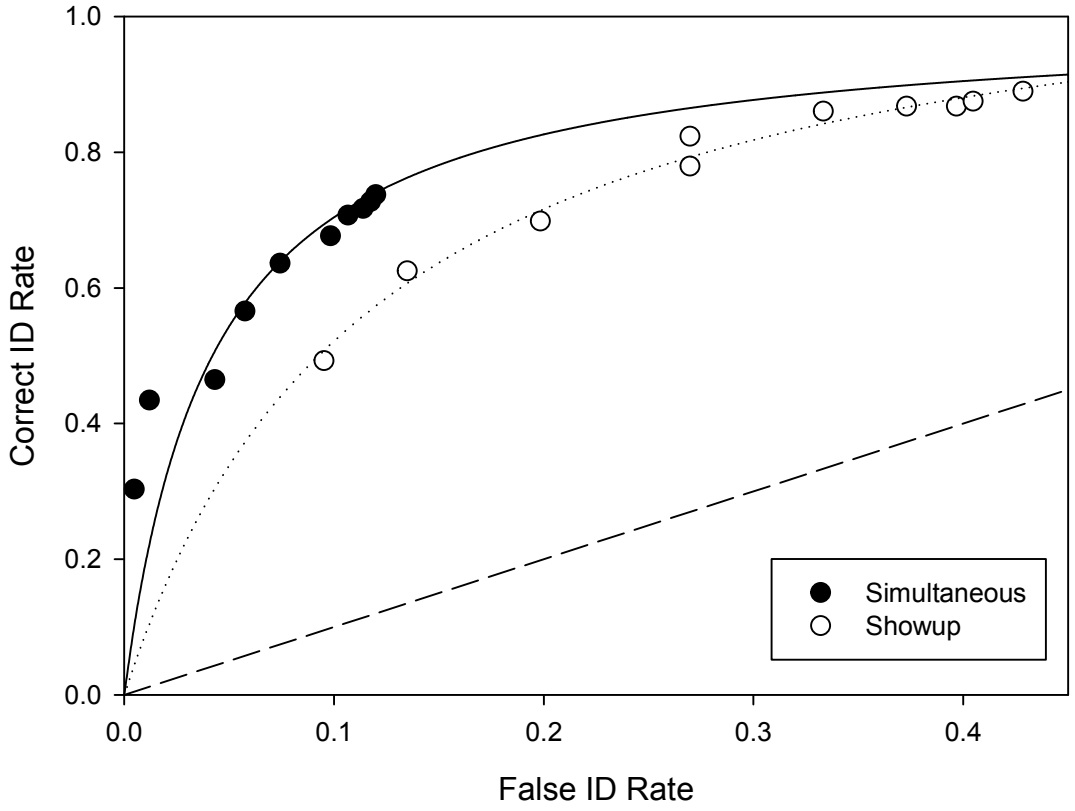
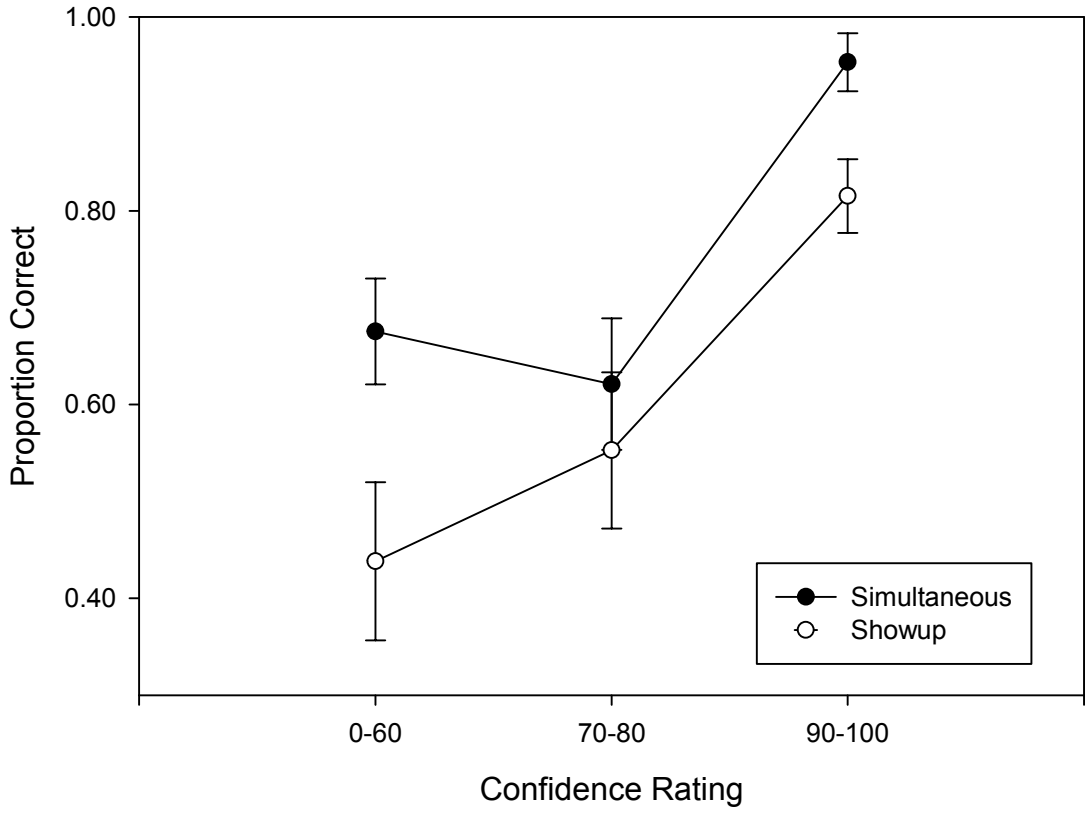

1.00
0.90
0.80
0.70
0.60
0.50
Proportion Correct
0-60
70-80
90-100
Confidence Rating
Simultaneous
Sequential

Figure 4. ROC curves (Figure 4A) and CAC curves (Figure 4B) for simultaneous lineup and showup procedures in Experiment 2. The dashed line in Figure 4A represents chance performance and the bars in Figure 4B represent standard error bars.

A.



B.


1.00
0.80
0.60
0.40
Proportion Correct
0-60
70-80
90-100
Confidence Rating
Simultaneous
Showup

Appendix A. False IDs, foil picks, and correct IDs for every level of confidence from Experiment 1.

Confidence	False IDs	Foil Picks	Correct IDs
	No Recollection		
0	3	1	1
10	2		
20	2	3	
30	5	2	2
40		1	1
50	3		3
60	2	2	3
70	2		1
80		1	2
90			2
100			
	Recollection		
0	1		
10	1		
20	2	1	
30	6		2
40	1	4	2
50	9		4
60	9	2	5
70	12	5	16
80	4	1	17
90	1		9
100	1	1	6

Appendix B. False IDs, foil picks, misses, correct rejections, and correct IDs for every level of confidence from Experiment 2.

Confidence	False IDs	Foil Picks	Misses	Correct Rejections	Correct IDs
		6-Person Lineup			
0			1		1
10	2	1			
20	3	1		1	1
30	6	3		4	1
40	7			4	3
50	20	5	1	6	4
60	14	4		2	7
70	12			6	10
80	26	4	1	4	3
90	6	4	1	5	13
100	4			7	30
		Show-up			
0	1				
10	2			2	2
20	1		2	1	1
30	3		1	2	
40	5		1	8	1
50	8		4	9	5
60				8	6
70	9		1	10	11
80	8		2	10	10
90	5		1	8	18
100	12		3	14	67