Spring 5-16-2019

The Dud Effect: The Effect of Dissimilar Fillers in Eyewitness Lineups

Shannon Booth
CUNY John Jay College, boothshannon37@gmail.com

The Dud Effect: The Effect of Dissimilar Fillers in Eyewitness Lineups

Shannon Booth

John Jay College of Criminal Justice

New York, NY

Table of Contents

Abstract

Similarity and confidence are often analyzed in relation to eyewitness accuracy. Duds, highly dissimilar fillers, can impact an eyewitness' accuracy and confidence. This study analyzed the impact of duds, similarity and accuracy in the context of lineup bias. This study found that in target-absent conditions there is a significant positive relationship between lineup bias and confidence. This study also found that highly similar lineups results in a lower proportion of correct identification, even in highly confident witnesses, compared to medium similar and low similar lineups. Further research should look closer at the "sweet spot" of similarity in order to advise and put in place proper similarity scores when picking fillers in lineups.

Keywords: eyewitness, confidence, CAC, duds

Introduction

Seventy percent of DNA exonerees that were wrongfully convicted involved faulty eyewitness testimony (Innocence Project, 2018). Eyewitness unreliability has been highly studied for the past few decades with the hope of administering lineups in the least bias way. There have been substantial improvements in the administration of lineups due to the research, such as some states now require double blind administration, where the administrator of the lineup does not know who the suspect is (Malpass, 2006). Even with the improvements, there is still no consistent procedure for picking fillers for the lineup.

Fillers are very important because they can impact a witness' ability to correctly identify a perpetrator while limiting the identification of an innocent filler. In research there are two conditions, target-absent (i.e. perpetrator from video or scenario is not present) and target-present (i.e. perpetrator is present) lineups. This is due to the fact police do not know if the actual perpetrator is present in their lineup. Eyewitnesses' judgments are relative due to the fact that the eyewitnesses make comparisons among the choices (Wells, 2001). For this reason, fillers play a crucial role in the identification process and it is important to study fillers.

The focus of the current research is the relationship between filler similarity and confidence and accuracy of eyewitness identification. Filler selection and influences on filler selection such as filler similarity will be discussed first. Lineup bias and duds will then be discussed. The relationship between lineup bias and duds and confidence is a focus of this research. In addition, the relationship between filler similarity and confidence-accuracy relationship will be examined.

Choosing fillers and filler similarity

There is no consistent procedure on how fillers are chosen for lineups. Two ways that fillers are selected are match to description and match to appearance. Match to appearance strategy is most commonly used and involves fillers being chosen due to their similarity to the

suspect (Fitzgerald, Price, Oriet & Charman, 2013). Bias towards choosing the suspect arises if witnesses are choosing the suspect at a higher rate due to factors other than remembering the suspect in the videos. This bias has been found to depend on different criteria used when picking fillers (Navon, 1992; Clark, 2003). For example, if the suspect is identified due to the similarity to the description given by the eyewitness but then the fillers are chosen due to their similarity to the identified suspect, there will be bias to the suspect in the lineup if the fillers resemble the suspect instead of the initial description. To try and eliminate this bias it has been suggested that fillers should be chosen based on the eyewitness' description of the perpetrator with no regard to the suspect or the match to description strategy.

Match to description strategy theoretically should have witnesses rely on recognition because there is similarity among the lineup members. In addition, Wells, Rydell & Seelau (1993) found that the match to description strategy resulted in a low false foil identification (picking a filler instead of the suspect) and a higher accurate identification rate. This finding is not always the case with some researchers finding that neither strategy affects foil identifications or correct identification rates (Darling, Valentine, & Memon, 2008).

Facial similarity is also useful when picking fillers. There has been a long-held belief that having the foils look too similar to the description of the offender can put unreasonable demand on the witnesses' discrimination powers (Brewer & Williams, 2007). In addition, the most similar fillers available may adversely affect correct identification rates (Fitzgerald, Oriet, & Price, 2014; Oriet & Fitzgerald, 2018) and produce rejection rate increases in both target present and target absent conditions (Bruer, Fitzgerald, Therrien, & Price, 2014; Oriet & Fitzgerald, 2018). A similar pattern was also found with moderately high similarity lineups having higher correct identification rates than in very high similarity lineups (Fitzgerald, Oriet, & Price, 2014). Research is not entirely consistent with some findings suggesting that lineups

with moderate or high similarity fillers compared to lineups with low similarity fillers were far more likely to elicit suspect identifications (Fitzgerald, Price, Oriet, & Charman, 2013).

Lineup fairness

Lineup fairness refers to the extent to which the fillers resembled the suspect in a lineup (Malpass, Tredoux, & McQuiston-Surrett, 2007). In an unbiased lineup, the chance of picking the suspect is defined as 1 divided by the number of lineup members, in cases where the person picking did not witness the event (Malpass, Tredoux, & McQuiston-Surrett, 2007). Biased lineups were found to increase participants' identification of the suspect and impaired the subjects' ability to distinguish between innocent and guilty suspects (Colloff, Wade, & Strange, 2016).

Signal Detection Theory. Signal detection theory (SDT) has been used to measure eyewitness performance. Lee & Penrod (2019) developed a new SDT-based framework that takes into the account the role of fillers in discriminability of a guilty suspect from an innocent suspect. This research will be utilized in this study to analyze the influence of lineup bias due to the fact that this criterion is important for how witnesses make their decisions.

Confidence-Accuracy

Confidence of eyewitnesses has been frequently studied in regards to a confidence-accuracy relationship. The relationship between confidence and accuracy is especially important due to the fact that jurors rely on perceived confidence when evaluating eyewitness identification (Charman, Wells, & Joy, 2011; Wells, Lindsay, & Ferguson, 1979). Research suggests that there is actually a modest correlation between eyewitness confidence and identification accuracy (Sporer, Penrod, Read & Cutler, 1995).

As there have been changes in the administration of lineups (lineup instruction, lineup fillers, etc.) and lineup fairness has increased, there has been an increase in studies finding

confidence and accuracy being strongly related (Wixted & Wells, 2017). There have also been changes in the way researchers examine the confidence-accuracy relationship.

There is also a relationship between confidence-accuracy and target-filler similarity. Horry & Brewer (2016) found that when participants correctly identified the target, confidence increased but when target-filler similarity increased the proportion correct decreased. When the fillers do not match and are instead duds or implausible options, individuals became more confident (Charman, Wells, & Joy, 2011). Charman, et al. found that the inclusion of the duds in lineups causes the nondud alternatives to be perceived as more similar to the memory of the actual perpetrator (Charman, Wells, & Joy, 2011). Hanczkawoski, Zawadzka, & Higham (2014) also found that the inclusion of duds inflated confidence for the nondud alternatives. In these studies, the duds are picked prior to the lineup being administered and are picked by the researchers based on their similarity to the innocent or guilty suspect in the lineup. This study will utilize human similarity ratings and BetaFace, a computer-based face similarity tool, in order to assess if a filler is considered a dud and the impact the presence of the dud has on discriminability and confidence ratings. Fillers will be considered duds based on similarity rating scores and the number of times the fillers were picked by all the participants.

Recently, confidence-accuracy characteristics (CAC) analysis has started to be used to evaluate the confidence-accuracy relationship. Suspect identification accuracy is based on the number of guilty suspect ids divided by the number of guilty suspect ids plus the number of innocent suspect ids at the same confidence level (Wixted & Wells, 2017). CAC analysis has been used to examine confidence-accuracy relationship and most research has found that highly confidence witnesses are overconfident and are not more likely to be correct as moderately confident witnesses (Wixted & Wells, 2017).

Study Overview

To better understand the relation between filler similarity, specifically fillers that are duds, and accuracy, the present study will perform secondary analysis on a previous eyewitness study performed by Evelo, Lee, Modjadidi and Penrod (2018). One hypothesis of this study is that confidence will increase when the number of duds increases based on similarity ratings. Another hypothesis of this study is that as lineup bias increases so does the proportion of correct identifications and correct rejections in target-present and target-absent lineups, respectively.

Method

Research Design

The dud effect is analyzed using data from Evelo, Lee, Modjadidi, and Penrod (2018). The eyewitness study consisted of 2,000 participants.

The study used a 2 (video length: 15s & 45s) x 2 (lineup presentation: simultaneous & sequential) x 3 (base rate of target presence in lineups: no information & 33% & 66%)) x 2 (target race: black & white) x 2 (target presence: target-present & target-absent) x 2 (lineup bias: more biased & fairer) mixed factorial design. Of the independent variables, the video length, target race and target presence were within-subject variables. The other variables were between subject-variables.

Materials

Videos. There were four short videos containing two targets (a pair of one black man and one white man). There are two versions of every video- one is 15 seconds long and the other is 45 seconds in duration. Video 1 is a pair playing a card game. Video 2 is a pair talking at a kiosk. Video 3 is a pair talking at a table in a cafeteria. Video 4 is a pair talking outside on a bench.

Fillers. A photo pool of fillers for each target was created—the fillers were picked on match to descriptions of the targets. A preliminary Mturk experiment was designed and participants assigned a similarity score between fillers and the corresponding target. BetaFace,

an online software, was also utilized in giving a separate similarity score. Based on human rating scores, the top six fillers were selected for a fairer lineup, and the bottom six fillers were selected for a biased lineup. It should be noted that the fillers are highly similar to each other and to the innocent and guilty suspects.

Lineup Task. Each of the four videos had eight lineups. These lineups varied on three variables: target race, lineup bias and target presence, resulting in a 2 (target race: black and white targets) x 2 (lineup bias: high and low) x 2 (target presence: target-present & target-absent) design. In the target-present lineups, the perpetrator in the video was present in the lineup. In target-absent lineups, the perpetrator from the video was not in the lineup and instead there was another innocent foil or innocent suspect.

Confidence Ratings. Participants were asked to rank their confidence in their identification from 0% (not at all confident) to 100% (completely confident). The scale was in 10% increments.

Demographics Questionnaire. Participants were asked about their age, gender, race and formal education level.

Procedure

Participants volunteered for the study by clicking the link in the MTurk advertisement. A brief description of the study was posted on MTurk, where users view a description of the study and then choose to complete the study. At the beginning of the study, participants were presented with an online consent form, which described the study. The consent form was reviewed individually without the presence of any research staff but participants had the opportunity to email the PI with further questions if they did not understand the information presented. Participants were then able to click ‘agree’ to indicate that they had read and understood the consent form and wished to continue to the study. Participants then viewed four separate videos. After watching the four videos, participants participated in a 3-minute filler

task (a word puzzle game); and they were then asked to identify the targets from each of the 8 lineups. Participants watched all four videos before completing the corresponding lineups.

Results

Lineup bias was measured using the new framework of signal-detection- theory for lineups put forth in Lee & Penrod (2019). According to their model, the discriminability of a guilty suspect from an innocent suspect in target-present (TP) lineups is d'(GI). The discriminability of a guilty suspect from fillers in target-present (TP) lineups is d'(GF_P). The discriminability of an innocent suspect from fillers in target-absent (TA) lineups is d'(IF_a). The discriminability of fillers in TP lineups from fillers in TA lineups is d'(F_aF_p). Because d'(GF_p) and d'(IF_a) reflect the distance between memory-strength distributions of fillers and the guilty/innocent suspect, d'(GF_p) and d'(IF_a) are regarded as TP and TA lineup bias, respectively. However, TP lineup bias is more precisely calculated as d'(GF_p)-d'(RR), because d'(GF_p) is the product of TP lineup bias and the memory strength of the guilty suspect. Given that d'(RR), the discriminability of perpetrator-presence versus perpetrator-absence, reflects memory strength for the perpetrator, d'(GF_p)-d'(RR) measures TP lineup bias more accurately. d'(RR) was calculated by subtracting the z score of rejection rates in the target present conditions from the z score of rejection rates in the target absent conditions. Thus, getting the discriminability from perpetrator-presence versus perpetrator-absence. The higher the values of d'(GF_p)-d'(RR) and d'(IF_a) the more biased the TP and TA lineups are, respectively. Mean levels of d'(GF_p)-d'(RR), and confidence levels assigned to selections of each of the targets are presented in Table 1. Mean levels of d'(IFa) and confidence assigned to the selection of each filler are presented in Table 2. Although we report the conventional statistical test statistics, the correlations are based on grouped data resting on hundreds of observations so we will treat all results as reliable. Accuracy, confidence, and accuracy-confidence will be analyzed first in the terms of lineup bias and then by examining the influence of duds.

Accuracy

The relationship between accuracy and lineup bias for target presence was analyzed. In the target-present condition, the correlation between d'(GF)-d'(RR) (target-present lineup bias) and proportion correct was strongly positive, $r(16) = 0.704$ $p=.003$. There is a strongly negative relationship between d'(GF)-d'(RR) and proportion of false identifications, $r(16) = -0.821$ $p=.000$. In the target-absent condition, the correlation between d'(IF_a) (target-absent lineup bias) and the proportion that the filler was chosen has a strong positive relationship, $r(96) = .795$ $p=.000$.

In the target-present condition the relationship between the number of duds based on similarity ratings and accuracy was analyzed. In the target-present condition, the correlation between the number of duds based on human ratings and proportion correct was substantially negative, $r(16) = -.254$ $p=.342$. There is a substantial positive relationship between the number of duds based on human ratings and proportion of false identifications, $r(16) = .344$ $p=.192$. In the target-present condition, the correlation between the number of duds based on BetaFace ratings and proportion correct was substantially negative, $r(16) = -.377$ $p=.150$. There is a substantial negative relationship between the number of duds based on BetaFace ratings and proportion of false identifications, $r(16) = -.303$ $p=.254$.

In the target-absent condition the relationship between the number of duds based on similarity ratings and accuracy was analyzed. There is a substantial negative relationship between proportion of correct rejection and the number of duds based on human ratings, $r(16) = -.182$ $p=.501$. There is a slight positive relationship between proportion of false identifications and the number of duds based on human ratings, $r(16) = .053$ $p=.846$. In the target-absent condition, there is a substantial positive relationship between proportion of correct rejection and the number of duds based on BetaFace ratings, $r(16) = .443$ $p=.086$. There

is a substantial negative relationship between proportion of false identifications and the number of duds based on BetaFace ratings, $r(16) = -.400$ $p=.125$.

Confidence

First the relationship between lineup bias and confidence was analyzed for target-absent and target-present conditions, respectively. In the target-absent condition there is a modest negative relationship between d'(IF_a) (target-absent lineup bias) and confidence, $r(96) = -0.207$ $p=.043$. In the target-present condition, the correlation between d'(GF)-d'(RR) (target-present lineup bias) and TP confidence was strongly positive, $r(16) = 0.83$ $p=.761$.

For the purposes of this analysis if a filler was identified 20 times or less in each condition it was considered a dud. Table 3 is the number of duds and average confidence in the target-absent condition. There is a modest negative relationship between the number of duds and average confidence in the target-absent condition, $r(16) = -.382$, $p=.144$. Table 4 reports the number of duds and average confidence in the target-present condition. There is a substantial negative relationship between the number of duds and grand average confidence in the target-present condition, $r(16) = -.449$, $p=0.81$.

For the purposes of the next two analyses, duds were determined based on similarity ratings. There are two similarity ratings, the first is human rating and the second is BetaFace rating. Table 5 shows the average, standard deviation and number of duds for human ratings and BetaFace ratings. Due to the fact that they are a very different mechanism of rating similarity they have been treated as two separate conditions. There was a substantial positive relationship between the two similarity ratings, $r(16) = 0.502$, $p=.047$. A filler was considered a dud when the rating was one standard deviation below the average rating. There is a substantial positive relationship between the number of duds determined by the human rating and confidence, $r(16) = 0.49$, $p=.857$. There is a substantial negative relationship between the number of duds determined by the BetaFace rating and confidence, $r(16) = -.291$, $p=.274$.

Regression Analyses

Multiple regression analyses were conducted with confidence as the outcome variable; and lineup bias measures and other condition variables (race, video types, lineup type, exposure time, target presence, base rate and difficulty) as predictors to investigate the association of the lineup bias measures and eyewitness confidence, controlling for the other variables. The correlation matrices underlying these analyses are reported in the Appendices—for data at the individual level and at the array level (corresponding to each of the targets). All categorical variables were coded as dummy variables (for base rates_dummy1, none vs. 33%; for base rates_dummy2, none vs. 66%; for lineup bias based on similarity ratings, more biased=0 vs. fairer=1; for presentation mode, simultaneous=0 vs. sequential=1; for exposure time, short=0 vs. long=1; for race combination, same=0 vs. cross=1, for target presence, absence=0 vs. presence=1). As shown in Table 6, only exposure time was a significant predictor of confidence in TP arrays (data based on the 16 arrays), in Table 7 baserate (the three levels of that variable represented as two dummy variables), difficulty, sequential vs simultaneous and exposure time all influenced confidence judgments in TA arrays (again at the array level). Table 8 reports a regression analysis based on data at the individual level with confidence in both TP and TA arrays included in the analysis—in this instance baserate, difficulty, sequential vs simultaneous, exposure time and bias all influenced confidence judgments. When the base rate of target presence in lineups was provided, participants' confidence tended to increase; particularly the comparison of none vs. 33% was statistically significant (β=.06, p=.05). For the comparison of simultaneous vs. sequential lineups, sequential lineups led to an increase in the confidence (β=.17, p<.001). Longer exposure time was associated with higher confidence (β=.10, p<.001). Race combination did not predict participants' confidence (β=-.003, p=.90). The target-present condition yielded higher confidence than did the target-absent condition (β=.37, p<.001).

In this regression model, two types of lineup-bias predictors were included--lineup bias, which was manipulated based on human and Betaface ratings, and d' bias, which was estimated based on the discriminability between a suspect and targets. As expected, higher d' bias was associated with higher confidence (β=.06, p =.03). However, for the lineup bias manipulated with similarity ratings, fairer lineups were associated with higher confidence (β=.05, p=.04). In terms of effect size, bias displayed a small effect in TA and the overall analyses *partial r* = .02 and .03 respectively, but was of more substantial magnitude with respect to TP arrays (*partial r* =.15).

Confidence-Accuracy Characteristics (CAC) Analysis

For the purposes of the first analysis, there is no designated innocent suspect. Confidence-accuracy is based on (the number of guilty suspect ids) / (the number of guilty suspect ids + the number of total TA innocent suspect ids divided by the number of lineup members in the TA lineup) at the same confidence level. See Figure 1 for CAC curve for this analysis.

Lineup similarity was determined based on average BetaFace ratings. For these two CAC curves, it is under the assumption that there is no designated innocent suspect. The lineups were grouped into low similarity, medium similarity and high similarity based on equally spaced similarity ratings (Figure 2). Low similarity were lineups with an average BetaFace scores between 62 and 66. There were 5 lineups in this category. Medium similarity was average BetaFace scores between 66.1 -70.and there were 9 lineups in this category. High similarity was average BetaFace scores between 70.1 and 75 and there were 2 lineups in this category. A secondary CAC curve (Figure 3) was created with equal number of lineups in each of the low similarity, medium similarity and high similarity groups. Low similarity were 5 lineups with an average BetaFace score between 62 and 64.7. The medium similarity group was compromised of 5 lineups with an average BetaFace score between 65.1 and 67.8. The

high similarity group was compromised of 6 lineups with an average BetaFace score between 68.7 and 73.3.

Discussion

The current study examined whether lineup bias inflates eyewitness confidence; and influences eyewitnesses' Confidence-Accuracy relationship. Overall, lineup bias has some effect on confidence. In target-absent conditions, bias was negatively related to confidence, as lineups increased in bias, confidence decreased. In target-present conditions, there was a positive relationship between bias and confidence. Contrary to the original hypothesis, there was a negative relationship between the number of duds based on BetaFace similarity ratings and confidence. The CAC analysis reveals that the proportion correct increases as reported confidence increases. When the CAC analysis is done in respect to lineup similarity, highly similar lineups have a lower proportion of witnesses correctly picking the guilty suspect in the higher confidence intervals. In sum, the current study demonstrated that, as lineups became fairer, eyewitness confidence increased; and the CA relationship became weaker at the high confidence levels.

The purpose of this research was to examine the impact fillers that are considered duds on confidence levels. The examination of duds will then be used to advise police officers on how they pick their fillers based on similarity to suspect. Foil similarity is often examined in terms of its relationship with confidence. When suspects are surrounded by fillers that are duds or implausible options, individuals rate their confidence higher (Charman, Wells, & Joy, 2011). The current study had similar results, there was a positive association between the number of duds and confidence for human similarity ratings of determining duds. When BetaFace similarity ratings were used to determine duds there was (for reasons that are not obvious) a negative association. The correlations between the two types of similarity ratings and confidence have contradictory results. This is important due to the fact that most lineups

are based on human similarity ratings and picking fillers that resemble the suspect. Both of these correlations are important due to the fact that confidence of eyewitnesses impact jury decisions. It has been found that jurors rely on perceived confidence when evaluating eyewitness identification (Charman, Wells, & Joy, 2011; Wells, Lindsay, & Ferguson, 1979). Similarly, this current study found negative relationships between lineup bias, which was estimated by d' measures, and confidence.

This study is based on highly similar lineups. Due to the procedure of choosing fillers, the duds that were determined are still very similar to the other fillers. But, highly similar fillers can create difficulties for witnesses. Similar to this study, Brewer & Williams (2007) also found that too similar fillers put an unreasonable demand on the witnesses' discrimination powers. Brewer & Williams (2007) also found higher foil identification rates with highly similar lineups. This should serve as a caution when using fillers that are so highly similar to each other and to the suspect. The current study also found that medium similar lineups have higher correct identification rates than the highly similar lineups. The current study found that correct proportion rates lowered as similarity increased for each confidence interval. This result was previously seen in a study (Fitzgerald, Oriet & Price, 2014) but not seen in another study (Fitzgerald, Price, Oriet & Charman, 2013). This holds for cases of really high similarity where the fillers in the highly similar lineups are almost clones of the innocent or guilty suspect.

Extremely similar lineups can also be difficult for witnesses when it comes to choosing the correct guilty suspect. Biased lineups were found to increase participants' identification of the suspect and impaired the subjects' ability to distinguish between innocent and guilty suspects (Colloff, Wade, & Strange, 2016). The current study found that even when witnesses respond 100% confident, they are incorrectly identifying the guilty suspect when the lineup is more similar. This could be due to the fact that the high similarity lineups are so similar that the fillers are almost clones of each. Overall, the highly similarity might not impact

accuracy as most lineups traditionally done by police don't reach the level of similarity that was in this study.

There has been an increase in studies finding a significant association between confidence and accuracy (Wixted & Wells, 2017). The current study found an association between proportion correct and confidence in the overall CAC plot (see Figure 1). But, it is important to note that when the CAC analysis is performed as a function of lineup similarity, the proportion correct is lower for highly similar lineups then medium similar lineups. Low similarity has a higher proportion correction for each confidence interval. This is important because jurors consider witness confidence. If the witness looked at a highly similar lineup and responded 100% on the confidence interval, the witness is correct about 85% of the time.

Limitations and Future Research

There are limitations to the current study, which is why future research is important. One limitation is the fact that the research methodology was entirely done online. Some of the limitations of a completely online study is the potential impact on the internal validity of the study. There is limited control over external circumstance such as noise level or interruptions that could influence a subject's response. These circumstances are dramatically different from how real witnesses experience the event and then perform an identification procedure. Due to the nature of eyewitness events it is difficult to expose the participant to the same circumstances as a real witness.

Another limitation is the lack of random sampling. Witnesses occur at random in the environment but the advertisement for the study could potentially have self-selection. Participants with interest in witness tasks could have picked this study and it could have influenced how they performed.

In addition, it was impossible to extract the full dataset which includes all condition variables. This is a limitation of the secondary data analysis method. A majority of the analyses

did not control other condition variables (e.g., base rate, exposure time, lineup presentation, etc.), which could affect the association of lineup bias and confidence.

The study also lacks generalizability. Real lineup constructors are police officers, detectives or other criminal justice personnel. Instead these lineups were constructed by research assistants using similarity ratings typically not available to the police. All of the lineups have extremely high similarity ratings and this is not traditionally found in lineups done by the police. Overall, the lineups constructed by researchers may not be representative of the lineups constructed in real situations.

Despite these limitations, the current study uses an approach that can be expanded upon to strengthen policy implications. Future research should examine software that can further assess filler similarity and that can, in turn, be used for police officers to select fillers. Further research should also involve actual police officers picking fillers to look closer at their decision-making process when picking fillers and to make it more applicable to the real world since police officers are usually the ones constructing lineups. Future research should also look at using less similar lineups. Due to the fact that all the lineups were highly similar, the duds might not be true duds due to the fact they resemble the description that was given by the witness.

Conclusion

The current study examined the association between lineup bias, d' measures, and the number of duds and eyewitness confidence. The study found a positive relationship between the lineup bias, which was estimated by d' measures, and confidence for target-present lineups. In target-absent lineups there was a significant negative relationship between bias and confidence. It is important to note that although the conventional statistical test statistics were reported, the correlations are based on grouped data resting on hundreds of observations so we will treat all results as reliable. The study also found that highly similar lineups impact the

proportion of witnesses are correct even if they respond highly confident. Further research should look closer at the “sweet spot” of similarity in order to advise and put in place proper similarity scores when picking fillers in lineups.

References

Brewer, N., & Williams, K. (2007). *Psychology and law: An empirical perspective.* New York: Guilford.

Bruer, K., Fitzgerald, R., Therrien, N., & Price, H. (2014). Line-up member similarity influences the effectiveness of a salient rejection option for eyewitnesses. *Psychiatry, Psychology & Law, 22* (1), doi: 10.1080/13218719.2014.919688.

Charman, S., Wells, G., & Joy, S. (2011). The dud effect: adding highly dissimilar fillers increases confidence in lineup identifications. *Law Human Behavior, 35*, 479-500, doi: 10.1007/s10979-010-9261-1.

Clark, S. E. (2003). A memory and decision model for eyewitness identification. Applied Cognitive Psychology, 17, 629–654. doi:10.1002.acp.891

Colloff, M., Wade, K., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27* (9), 1227-1239, doi: 10.1177/095679616655789.

Darling, S., Valentine, T., & Memon, A. (2008). Selection of lineup foils in operational contexts. Applied Cognitive Psychology, 22, 159–169. doi:10.1002/acp.1366

Evelo, A., Lee, J., Modjadidi, K., & Penrod, S. D. (2018). The Role of Lineup Bias in Witness Accuracy, the Confidence-Accuracy Relationship and the Courtroom Value of Witness Confidence. Paper presentation at Annual Conference of the European Association of Psychology and Law, Turku, Finland.

Fitzgerald, R., Oriet, C., & Price, H. (2014). Suspect filler similarity in eyewitness lineups: a literature review and a novel methodology. *Law and Human Behavior, 39* (1), 62-74, doi: 10.1037/lhb0000095.

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19*(2), 151-164, doi: 10.1037/a0030618

Hanczakowski, M., Zawadzka, K. and Higham, P.A. (2014) The dud-alternative effect in memory for associations: Putting confidence into local context. *Psychonomic Bulletin & Review, 21*(2), doi: 10.3758/s13423-013-0497-x.

Horry, R., & Brewer, N. (2016). How target-lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology, 145* (12), 1615-1634, doi: 10.1037/xge0000227.

Innocence Project (2018). Retrieved from innocenceproject.org

Lee, J., & Penrod, S. (2019. New signal-detection-theory-based framework for eyewitness performance in lineups. Unpublished manuscript.

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

Malpass, R. S. (2006). A policy evaluation of simultaneous and sequential lineups. *Psychology, Public Policy, and Law, 12*(4), 394-418, doi:10.1037/1076-8971.12.4.394

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. (2007). Lineup construction and lineup fairness. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 155-178).

Mansour, J., Beaudry, J., Kalmet, N., Betrand, M., & Lindsay, R. (2017). Evaluating lineup fairness: Variations across methods and measures. *Law and Human Behavior, 41* (1), 103-115, doi: 10.1037/lnb0000203.

Navon, D. (1992). Selection of lineup foils by similarity to the suspect is likely to misfire, *Law and Human Behavior, 16* (5), 575-593, doi: 10.1007/BF01044624

Oriet, C., & Fitzgerald, R. (2018). The single lineup paradigm: a new way to manipulate target presence. *Law and Human Behavior, 42* (1), 1-12, doi: 10.1037/lhb0000272.

Smith, A., Wells, G., Lindsay, R., & Penrod, S. (2016). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior, 41* (2), 127-145, doi: 10.1037/lhb0000219.

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

Wells, G. L. (2001). Police lineups: Data, theory, and policy. *Psychology, Public Policy, and Law, 7*, 791-801. doi: 10.1037/1076-8971.7.4.791

Wells, G. L., Lindsay, R.C. L., & Ferguson, T.J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psycholgoy, 64,* 440-448.

Wells, G., Small, M., Penrod, S., Malpass, R., Fulero, S., & Brimacombe, E. (1998) Good practice recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603-647.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78*(5), 835-844, doi: 10.1037/0021-9010.78.5.835

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

Windschitl, P., & Chambers, J. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 30(1), 198-215, doi: 10.1037/0278-7393.20.1.198.

Zawadzka, K., Hingham, P.A., & Hanczakowski, M. (2017). Confidence in forced-choice recognition: what underlies the ratings? *Journal of Experimental Psychology: Learning, Memory and Cognition,* 43(3), 552-564.

Table 1

Target-present lineup bias and confidence

Lineup	d'(GF_p) - d'(RR)	TP Confidence	Proportion Correct	Proportion of False Identification
Video 1- Black Target- biased	-0.65	60%	.48	.34
Video 1- Black Target- fair	-0.77	55%	.31	.49
Video 1- White Target- biased	-0.32	54%	.38	.31
Video 1- White Target- fair	-0.69	58%	.25	.37
Video 2- Black Target- biased	0.07	53%	.50	.23
Video 2- Black Target- fair	-0.43	54%	.33	.41
Video 2- White Target- biased	-0.58	58%	.24	.28
Video 2- White Target- fair	-0.86	57%	.18	.38
Video 3- Black Target- biased	-0.80	45%	.22	.45
Video 3- Black Target- fair	-1.03	43%	.21	.53
Video 3- White Target- biased	-0.02	50%	.39	.24
Video 3- White Target- fair	-0.50	45%	.35	.36
Video 4- Black Target- biased	-0.49	59%	.39	.37
Video 4- Black Target- fair	-0.74	60%	.31	.46
Video 4- White Target- biased	-0.70	70%	.41	.34
Video 4- White Target- fair	-0.25	68%	.44	.30

Table 2

Target-absent bias and average confidence

Video	Lineup	Race of Target	Difficulty	Fillers	d'(IFa)	Average Confidence
1	1	Black	Biased	Filler 1	-1.307	41%
1	1	Black	Biased	Filler 2	-1.378	31%
1	1	Black	Biased	Filler 3	-1.040	34%
1	1	Black	Biased	Filler 4	-0.162	40%
1	1	Black	Biased	Filler 5	-1.739	35%
1	1	Black	Bia.49sed	IS	-1.632	38%
1	2	Black	Fair	Filler 1	-1.359	47%
1	2	Black	Fair	Filler 2	-1.054	40%
1	2	Black	Fair	Filler 3	-1.668	46%
1	2	Black	Fair	Filler 4	-0.481	48%
1	2	Black	Fair	Filler 5	-2.490	42%
1	2	Black	Fair	IS	-2.177	56%
1	3	White	Biased	Filler 1	-0.348	44%
1	3	White	Biased	Filler 2	-1.322	46%
1	3	White	Biased	Filler 3	-1.546	42%
1	3	White	Biased	Filler 4	-1.522	39%
1	3	White	Biased	Filler 5	-1.343	43%
1	3	White	Biased	IS	-1.263	40%
1	4	White	Fair	Filler 1	-0.668	38%
1	4	White	Fair	Filler 2	-1.572	42%
1	4	White	Fair	Filler 3	-0.726	38%
1	4	White	Fair	Filler 4	-2.447	53%
1	4	White	Fair	Filler 5	-1.331	32%
1	4	White	Fair	IS	-0.993	43%
2	5	Black	Biased	Filler 1	-1.515	45%
2	5	Black	Biased	Filler 2	-0.906	43%
2	5	Black	Biased	Filler 3	-1.489	34%
2	5	Black	Biased	Filler 4	-1.040	42%
2	5	Black	Biased	Filler 5	-1.417	37%
2	5	Black	Biased	IS	-0.736	42%
2	6	Black	Fair	Filler 1	-0.779	40%
2	6	Black	Fair	Filler 2	-1.356	39%
2	6	Black	Fair	Filler 3	-2.015	45%
2	6	Black	Fair	Filler 4	-0.918	45%
2	6	Black	Fair	Filler 5	-1.841	38%
2	6	Black	Fair	IS	-1.523	41%
2	7	White	Biased	Filler 1	-0.825	32%
2	7	White	Biased	Filler 2	-0.845	33%
2	7	White	Biased	Filler 3	-1.041	30%
2	7	White	Biased	Filler 4	-1.709	44%
2	7	White	Biased	Filler 5	-0.950	44%
2	7	White	Biased	IS	-0.972	35%
2	8	White	Fair	Filler 1	-1.732	44%
2	8	White	Fair	Filler 2	-1.385	47%

2	8	White	Fair	Filler 3	-1.293	35%
2	8	White	Fair	Filler 4	-0.487	35%
2	8	White	Fair	Filler 5	-1.072	38%
2	8	White	Fair	IS	-1.110	35%
3	9	Black	Biased	Filler 1	-1.837	46%
3	9	Black	Biased	Filler 2	-1.400	44%
3	9	Black	Biased	Filler 3	-0.868	42%
3	9	Black	Biased	Filler 4	-1.127	37%
3	9	Black	Biased	Filler 5	-1.219	39%
3	9	Black	Biased	IS	-1.299	42%
3	10	Black	Fair	Filler 1	-1.389	39%
3	10	Black	Fair	Filler 2	-1.373	36%
3	10	Black	Fair	Filler 3	-1.793	40%
3	10	Black	Fair	Filler 4	-0.983	39%
3	10	Black	Fair	Filler 5	-1.389	34%
3	10	Black	Fair	IS	-1.484	43%
3	11	White	Biased	Filler 1	-0.637	33%
3	11	White	Biased	Filler 2	-0.953	34%
3	11	White	Biased	Filler 3	-0.712	34%
3	11	White	Biased	Filler 4	-1.856	39%
3	11	White	Biased	Filler 5	-1.575	43%
3	11	White	Biased	IS	-1.546	42%
3	12	White	Fair	Filler 1	-1.754	40%
3	12	White	Fair	Filler 2	-0.504	42%
3	12	White	Fair	Filler 3	-1.395	43%
3	12	White	Fair	Filler 4	-0.717	33%
3	12	White	Fair	Filler 5	-1.784	37%
3	12	White	Fair	IS	-1.571	49%
4	13	Black	Biased	Filler 1	-1.173	41%
4	13	Black	Biased	Filler 2	-0.875	43%
4	13	Black	Biased	Filler 3	-1.693	38%
4	13	Black	Biased	Filler 4	-0.917	41%
4	13	Black	Biased	Filler 5	-1.324	43%
4	13	Black	Biased	IS	-1.718	37%
4	14	Black	Fair	Filler 1	-2.072	42%
4	14	Black	Fair	Filler 2	-1.370	38%
4	14	Black	Fair	Filler 3	-1.504	40%
4	14	Black	Fair	Filler 4	-0.839	47%
4	14	Black	Fair	Filler 5	-1.612	31%
4	14	Black	Fair	IS	-0.981	41%
4	15	White	Biased	Filler 1	-0.991	41%
4	15	White	Biased	Filler 2	-1.515	34%
4	15	White	Biased	Filler 3	-0.660	40%
4	15	White	Biased	Filler 4	-1.318	46%
4	15	White	Biased	Filler 5	-0.826	33%
4	15	White	Biased	IS	-1.425	32%
4	16	White	Fair	Filler 1	-1.014	45%
4	16	White	Fair	Filler 2	-1.507	48%
4	16	White	Fair	Filler 3	-0.891	38%

4	16	White	Fair	Filler 4	-1.048	37%
4	16	White	Fair	Filler 5	-1.631	44%
4	16	White	Fair	IS	-0.891	58%

Table 3

Number of duds and average confidence in target-absent lineups

Lineup- Target-Absent	# of Duds (ID <20)	Average Confidence
Video 1- Black Target- biased	2	37%
Video 1- Black Target- fair	1	46%
Video 1- White Target- biased	0	42%
Video 1- White Target- fair	1	41%
Video 2- Black Target- biased	0	40%
Video 2- Black Target- fair	0	41%
Video 2- White Target- biased	1	36%
Video 2- White Target- fair	1	39%
Video 3- Black Target- biased	1	39%
Video 3- Black Target- fair	0	37%
Video 3- White Target- biased	3	37%
Video 3- White Target- fair	2	39%
Video 4- Black Target- biased	0	41%
Video 4- Black Target- fair	2	40%
Video 4- White Target- biased	1	37%
Video 4- White Target- fair	0	43%

Table 4

Number of duds and average confidence in target-present lineups

Lineup- Target Present	# of Duds (ID <20)	Average Confidence
Video 1- Black Target- biased	0	60%
Video 1- Black Target- fair	1	55%
Video 1- White Target- biased	0	54%
Video 1- White Target- fair	2	58%
Video 2- Black Target- biased	3	53%
Video 2- Black Target- fair	1	54%
Video 2- White Target- biased	1	58%
Video 2- White Target- fair	1	57%
Video 3- Black Target- biased	2	45%
Video 3- Black Target- fair	1	43%
Video 3- White Target- biased	2	50%
Video 3- White Target- fair	3	45%
Video 4- Black Target- biased	1	59%
Video 4- Black Target- fair	0	60%
Video 4- White Target- biased	1	70%
Video 4- White Target- fair	1	68%

Table 5

Average, Standard Deviation and # of Duds for Human and BetaFace Similarity Rating

Lineup	Average Human Rating	SD of Human Rating	# of Duds- Human Ratings	Average BetaFace Rating	SD of BetaFace Rating	# of Duds- BetaFace
Video 1- Black Target-biased	2.5	0.02	0	63.9	1.6	0
Video 1- Black Target-fair	3.31	0.32	0	68.7	3.6	0
Video 1- White Target-biased	2.32	0.04	1	64.7	3.5	1
Video 1- White Target-fair	3.12	0.33	0	62.0	4.5	1
Video 2- Black Target-biased	2.88	0.04	1	69.0	3.6	1
Video 2- Black Target-fair	3.51	0.18	1	70.6	4.1	1
Video 2- White Target-biased	2.21	0.04	1	64.1	4.1	2
Video 2- White Target-fair	2.78	0.16	1	65.1	2.3	1
Video 3- Black Target-biased	2.62	0.08	1	69.8	5.1	1
Video 3- Black Target-fair	3.37	0.23	2	73.3	5.7	1
Video 3- White Target-biased	2.44	0.05	0	65.2	3.2	1
Video 3- White Target-fair	3.33	0.29	0	66.5	3.6	1
Video 4- Black Target-biased	2.78	0.02	1	67.8	4.2	1
Video 4- Black Target-fair	3.30	0.11	2	69.2	3.8	0
Video 4- White Target-biased	2.33	0.03	2	64.7	3.3	1
Video 4- White Target-fair	3.51	0.62	0	65.3	2.2	0

Table 6

Regression – Confidence in TP Arrays

	t	*p*	*β*
Basrerate_d1	.35	.73	.03
Baserate_d2	.61	.54	.05
Difficulty	-.88	.38	-.08
Mode	1.28	.04	.11
Exposure Time	2.06	.04	.04
Race	-1.45	.15	-.13
d'(GF)-d'(RR)	.58	.57	.07

Table 7

Regression – Confidence in TA Arrays

	t	*p*	*β*
Basrerate_d1	2.01	.05	.07
Baserate_d2	1.24	.21	.04
Difficulty	2.45	.02	.07
Mode	16.12	.00	.19
Exposure Time	3.25	.00	.10
Race	.30	.77	.01
d'(IF)	1.43	.15	.04

Table 8

Regression – Confidence in Combined TP & TA Arrays

	t	*p*	*β*
Basrerate_d1	1.99	.05	.06
Baserate_d2	1.36	.17	.04
Difficulty	2.08	.04	.05
Mode	6.31	.00	.17
Exposure Time	3.71	.00	.10
Race	-.13	.90	-.00
TPTA	14.37	.00	.37
Bias_Standardized	2.13	.03	.06

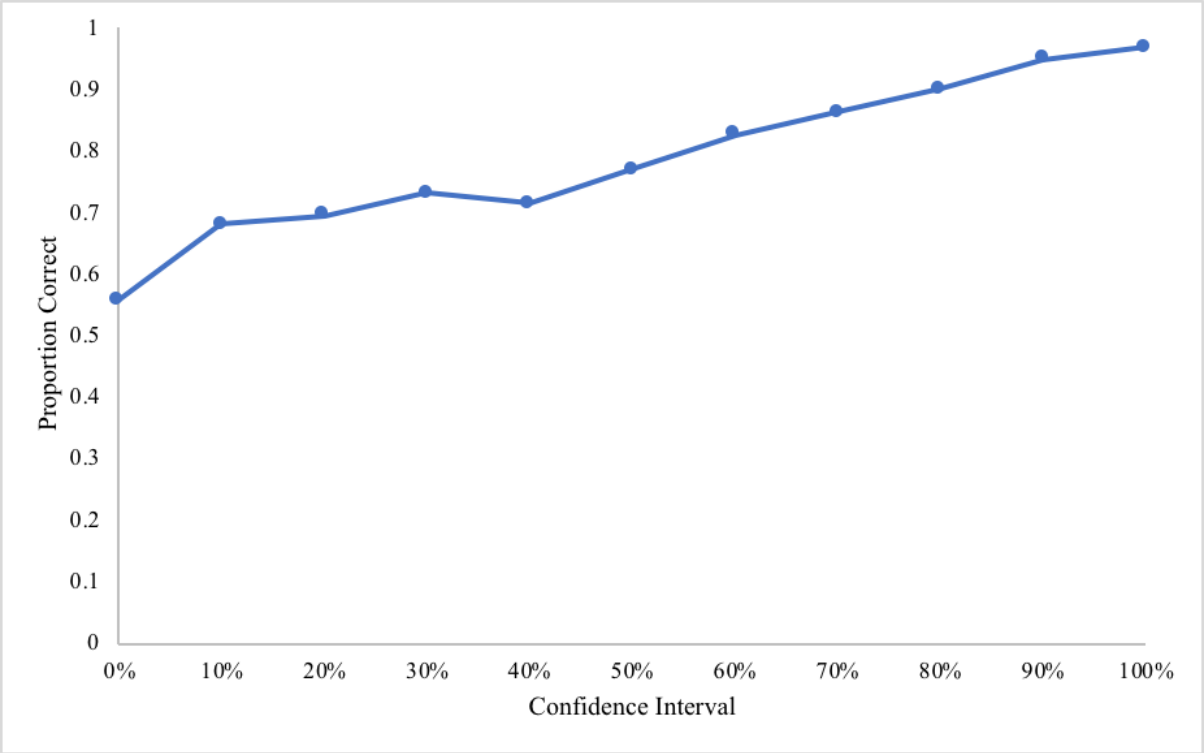

Figure 1. CAC curve without designated innocent suspect.

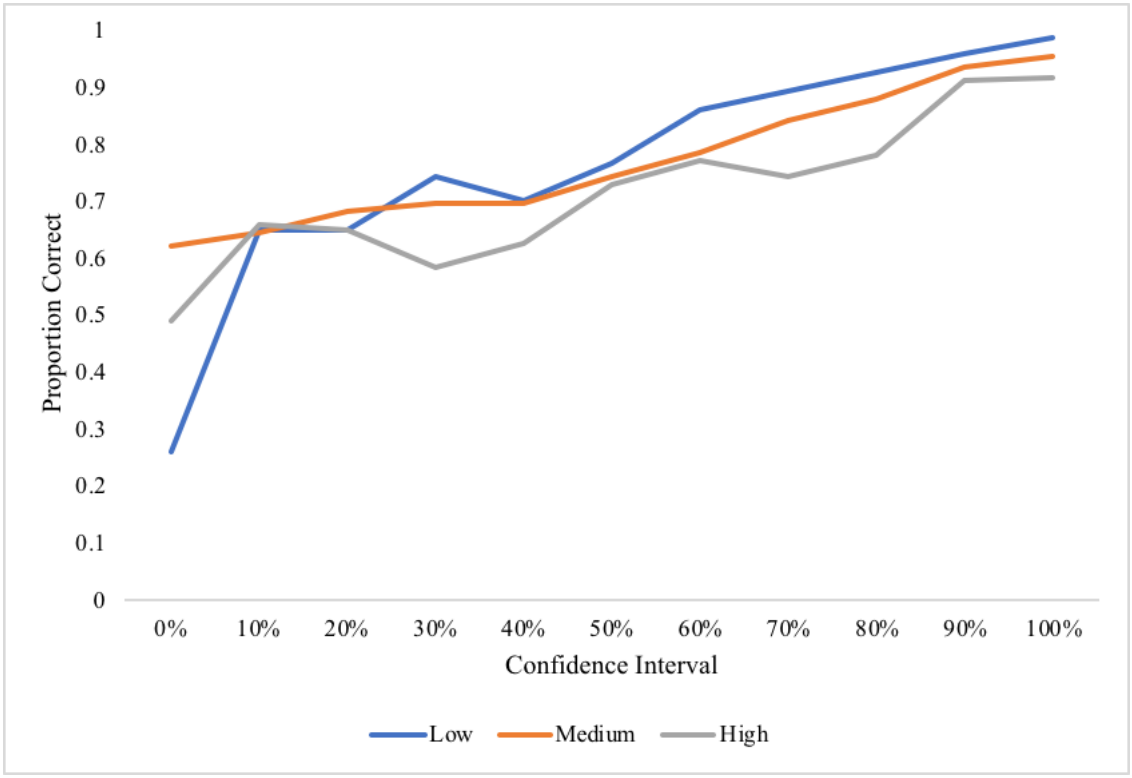

Figure 2. CAC curve for lineup similarity with unequal number of lineups in similarity levels.

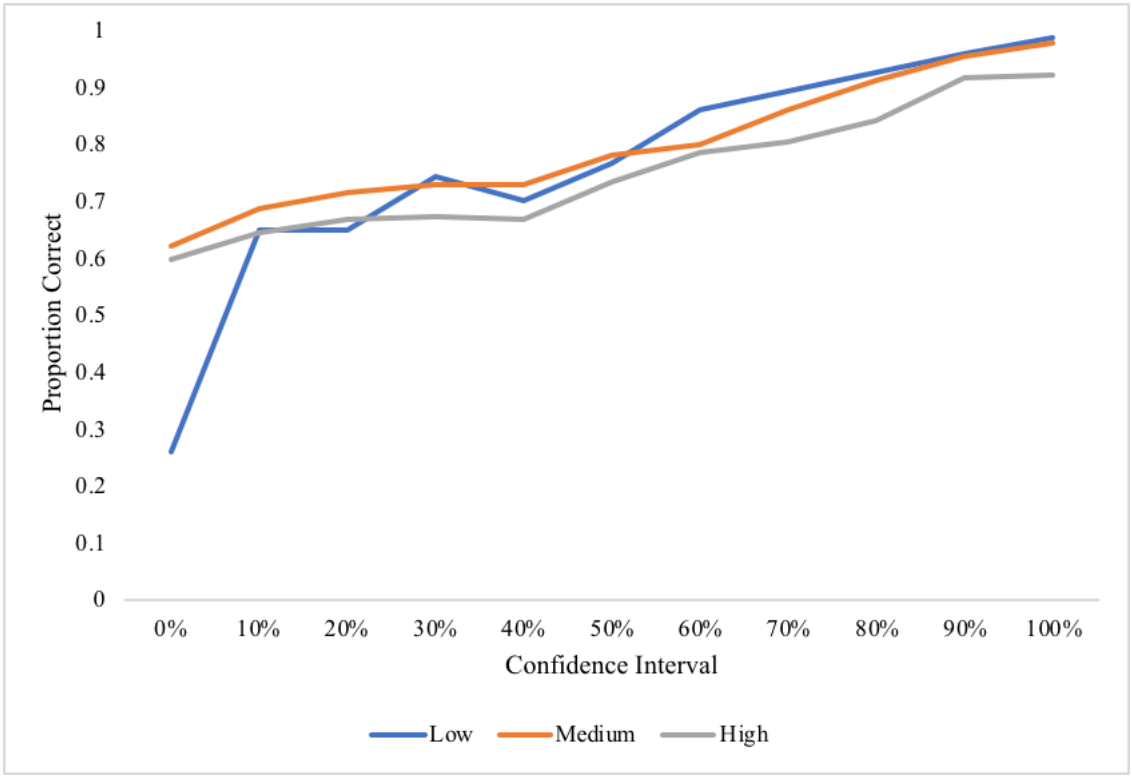

Figure 3. CAC curve for lineup similarity with equal number of lineups in similarity levels.

Appendix A
Correlations-TP Arrays

	1	2	3	4	5	6	7	8	9
1. GFpRR	-								
2. TPCon	.083	-							
3. Difficulty	-.382	-.077	-						
4. Target Race	.198	.266	.000	-					
5. Prop Correct	.704**	.403	-.402	-.074	-				
6. PropFalse IDTP	-.821**	-.297	.561*	-.518*	-.592*	-			
7. Dud Human Rating	-.364	.043	-.086	-.258	-.254	.344	-		
8. Dud Betaface Rating	.076	-.298	-.356	.356	-.377	-.303	.235	-	
9. TPDuds20	.365	-.449	.000	.139	.022	-.284	-.310	.362	-

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Appendix B

Correlations-TA Arrays

	1	2	3	4	5	6	7	8
1. Difficulty	-							
2. Target Race	.000	-						
3. Dud Human Rating	-.086	-.258	-					
4. Dud Betaface Rating	-.356	.356	.235	-				
5. Prop Correct Rejection	-.528*	.678**	-.182	.443	-			
6. PropFalse IDTA	.422	-.628**	.053	-.400	-.933**	-		
7. TA Con	.363	-.218	-.319	-.486	-.543*	.498*	-	
8. TPDuds20	-.070	.209	-.305	-.157	.213	-.116	-.382	-

*. Correlation is significant at the 0.05 level (2-tailed).

***. Correlation is significant at the 0.01 level (2-tailed).

Appendix C

Correlations- Combined TP & TA at Individual Level

	1	2	3	4	5	6	7	8	9
1. Baserate_d1	-								
2. Baserate_d2	-.500**	-							
3. Difficulty	.000	.000	-						
4. Mode	.000	.000	.00	-					
5. Exposure Time	.000	-.001	-.001	.001	-				
6. Race	.000	.001	.001	-.001	-.001	-			
7. TPTA	.000	.000	.000	.000	.000	.000	-		
8. Bias_ Standardized	.048	-.041	- .163**	-.156**	- .084**	- .215**	.000	-	
9. Confidence	.040	.007	.046	.154**	.101**	-.024	.370	.032	-

*. Correlation is significant at the 0.05 level (2-tailed).

***. Correlation is significant at the 0.01 level (2-tailed).

Appendix D

Correlations- TP at Individual Level

	1	2	3	4	5	6	7	8	9	10
1. Baserate_d1	-									
2. Baserate_d2	-.500**	-								
3. Difficulty	.000	.001	-							
4. Mode	.000	-.001	-.001	-						
5. Exposure Time	.000	.000	.000	.000	-					
6. Race	.000	.000	.000	.000	-.001	-				
7. D'(IF)	.038	-.034	-.114**	-.105**	.049	-.177**	-			
8. TA_Confidence	.051	-.001	.071*	.182**	.099**	-.003**	.019	-		
9. D'(GF)-d'(RR)	.102	-.077	-.435**	-.445**	.282**	-.426**	.410**	-.081	-	
10. TP_Confidence	.012	.055	-.106	.078	.189**	-.165*	.071	.042	.156*	-

*. Correlation is significant at the 0.05 level (2-tailed).

***. Correlation is significant at the 0.01 level (2-tailed).

Appendix E

Correlations- TA at Individual Level

	1	2	3	4	5	6	7	8
1. Difficulty	-							
2. Race	.000	-						
3. Dud_Human Rating	-.086	-.258	-					
4. Dud_BetaFace Rating	-.356	.356	.235	-				
5. Prop_ CorrectRej	-.528*	.678**	-.182	.443	-			
6. Prop_FalseID	.422	-.628**	.053	-.400	-.933**	-		
7. TA_Confidence	.363	-.218	-.319	-.486	-.543*	.498*	-	
8. TA_Duds20	-.070	.209	-.305	-.157	.213	-.116	-.382	-

*. Correlation is significant at the 0.05 level (2-tailed).

***. Correlation is significant at the 0.01 level (2-tailed).