Predicting Lineup Identifications

by

Mario J. Baldassari
Bachelor of Arts, Lake Forest College, 2011

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

Supervisory Committee

Predicting Lineup Identifications

by

Mario J. Baldassari
Bachelor of Arts, Lake Forest College, 2011

Supervisory Committee

Dr. D. Stephen Lindsay, (Department of Psychology)
Co-Supervisor

Dr. C. A. Elizabeth Brimacombe, (Department of Psychology)
Co-Supervisor

Dr. Michael E. J. Masson, (Department of Psychology)
Departmental Member

Dr. James W. Tanaka, (Department of Psychology)
Departmental Member

Gerard A. Ferguson, LLM, (Department of Law)
Outside Member

Abstract

Supervisory Committee
Dr. D. Stephen Lindsay, (Department of Psychology)
Co-Supervisor
Dr. C. A. Elizabeth Brimacombe, (Department of Psychology)
Co-Supervisor
Dr. Michael E. J. Masson, (Department of Psychology)
Departmental Member
Dr. James W. Tanaka, (Department of Psychology)
Departmental Member
Gerard A. Ferguson, LLM, (Department of Law)
Outside Member

Witnesses sometimes mistakenly identify innocent suspects in lineups from which the real culprit is absent, and those errors can have tragic consequences. Can we estimate in advance a witness's susceptibility to making false identifications in culprit-absent lineups? Kantner and Lindsay (2012) found that response criterion on a standard test of old/new recognition (of faces or words) correlated with the likelihood of making lineup identifications. Four experiments tested the predictive utility of a two-alternative forced choice facial recognition test that included trials in which neither face had been studied. Through Experiment 3 we observed several weak predictive relationships, including confidence on the facial recognition test with confidence on the lineup test, but not the hypothesized relationship: that the rate of false alarms on the TA face recognition trials would predict false alarm rates on the target-absent lineup trials. Experiment 4 implemented a substantial increase in the number of face recognition trials displaying two non-studied faces (from 4 trials to 30) and the originally hypothesized relationship was found (r=.45). Implications for future research aimed at developing measures with real-world utility are discussed.

Table of Contents

List of Tables

List of Figures

Acknowledgments

The author would like to acknowledge Steve Lindsay, Justin Kantner, and Joseph Sheppard for their origination of this research line and their subsequent input throughout the process. Special thanks to Steve, of course, for the support and resources with which I conducted this research.

Dedication

The author would like to dedicate this thesis to all those who supported me through this process: Dimitra, Mom, Dad, Gab, and the local cohort. Thanks for helping me maintain sanity and perspective.

Introduction

The psychological literature surrounding eyewitness suspect identification processes experienced a strong influx of interest and publication in the last 15 years. Through the course of many debates, reanalyses, retesting of old questions, and development of new questions the one consistent finding presented in each study was that eyewitness identifications are very unreliable evidence regarding the guilt of a suspect (Valentine, 2013). Variables such as exposure time, interview method, and lineup administration method were altered by many with only partial (if any) success in reducing false positive and false negative responses, and some psychologists moved to investigation of the differences among eyewitnesses rather than attempts to gain similar performance from all witnesses. Eyewitness quality also seemed to be very changeable in that some subjects may produce a false positive identification from a lineup in which the perpetrator is not present while others may be able to successfully reject such a lineup even when the subjects all have identical encoding situations (Darling, Martin, Hellman, & Memon, 2009; Valentine, Pickering, & Darling, 2003). Despite many attempts to develop a reliable predictor or postdictor of eyewitness accuracy, it may be the case that a combination of many measures will be the best way to reliably predict performance (e.g., Bindemann, Brown, Koyas, & Russ, 2012; Charman & Cahill, 2012; Darling et al., 2009; Smith, Lindsay, & Pryke, 2000). Some investigators searched for methods that assist or manipulate eyewitness performance, (e.g., Brewer, Weber, Clark, & Wells, 2008; Emmett, Clifford, & Gwyer, 2003; Macrae & Lewis, 2002), but most of these attempts have yielded weak results that again may work best when combined. This paper reviews

the previous study of prediction and manipulation of eyewitness accuracy and presents a new study of eyewitness skill prediction.

Since the late 1970's, many predictors (and postdictors) have been tested as measures of eyewitness accuracy. These predictors included personality measures, witness confidence, identification latency, description accuracy, decision process and search style, witness response to initial presentation of a blank lineup or mugshot, event-related potentials (ERP), eye movements, and face recognition ability (see Smith, Lindsay, & Pryke, 2000; Valentine, 2013, for reviews). Witnesses' ability to correctly ID a criminal was also affected by some uncontrollable variables in the real world including personality characteristics (Granhag, Ask, & Giolla, 2013; Valentine, 2013) and individual differences in face identification ability (Olsson & Juslin, 1999; Russel, Duchaine, & Nakayama, 2009). In recent work, Kantner and Lindsay (2012) obtained evidence that recognition memory response bias was a stable individual difference; some people tended to have a liberal bias (i.e., more often classify both old and new items as old than do their conservative counterparts), whereas others tended to be conservative and yet others tended to be neutral. This measure of response bias additionally correlated weakly with witness quality, especially when recognition memory stimuli were faces.

The current study aimed to contribute to the development of measures of individuals' lineup-identification skills. More specifically, we assessed the extent to which a test indexing both face recognition accuracy and proclivity to choose predicted choosing and confidence on target-absent lineups. To put this work in context, the following pages provide a brief review of prior research on individual differences as predictors of eyewitness accuracy. For reviews of research on the predictive value of

witnessing conditions (e.g., lighting, duration) and testing conditions (e.g., delay, lineup composition, instructions) see Granhag, Ask, and Giolla (2013) and Valentine (2013).

The Extant Literature

Confidence and ID Latency

The understanding of eyewitness confidence in the literature has developed over time (see Valentine, 2013 for a comprehensive review). Whereas the relationship appeared non-significant at first, it became clear that there was a weak but robust positive relationship between witness confidence and accuracy. This relationship was especially clear for lineup choices (rather than rejections) and when the confidence rating was taken after the lineup had been presented. The lack of a stronger correlation seemed to be partially due to a fault in most eyewitnesses' calibration of their confidence. In some studies, anyone making a positive ID (especially children) tended to be overconfident in the choice. Juslin, Olsson, and Winman (1996) found evidence that asking witnesses to ruminate on the crime and the criminal before the lineup helped them more accurately calibrate their confidence levels. Lindsay, Nilsen, and Read (2000) argued that the weak correlation may also be due to a truncation in the range of ability of eyewitnesses to make accurate identifications.

Measurements of the speed of an eyewitnesses' identification (or ID latency) revealed that, as most recognition memory researchers would suppose, faster witnesses were more often correct (Valentine, 2013). Dunning and Peretta (2002) proposed a hard cut-off for the difference between correct and incorrect identifications at 12 seconds after the presentation of the lineup, but their cut-off was found imperfect by other researchers (see Weber et al., 2004, for a thorough discussion of possible moderators of the cut-off and Valentine, 2013, for a mention of other reanalysis that refutes the original rule). Though a measure with a cut-off would be the most useful in real-world application, the

large amount of evidence detailing the negative correlation between latency and accuracy in eyewitness identification could certainly be useful within a set of other additionally informative measures. It also lends to the possibility of future studies perhaps gaining an individual measure of a witness' latency when accurate and inaccurate for a different set of stimuli and using that to inform judgments of ID latency on their real lineup choice.

Eyewitness confidence and identification latency were investigated and discussed together by many researchers (e.g. Dunning & Peretta, 2002; Pozzulo, Crescini, & Lemieux, 2008; Sauer, Brewer, & Wells, 2008; Sauerland, Sagana, & Sporer, 2012), but were discussed perhaps most simply in relation to predicting witness identification accuracy by Sauerland and Sporer (2007, 2009). In the 2009 paper, confederate targets stopped participants at a mall and asked for directions. The researchers then approached the participant and asked whether they would like to participate in a psychology study and were shown a TP or TA lineup for the confederate target (their identification latency was measured and a confidence rating was gathered before the lineup was shown). Participants then made a confidence judgment between 1 and 100, a remember/know/familiar judgment, and a judgment of how fast they were in making the ID. Chi-square analyses indicated that choosers presenting confidence levels above 90 and identifying within the first 6 seconds represented the group most often producing accurate identifications. Despite the elimination of most incorrect choosers, this combination of cut-off points left out a fair number of correct choosers. Pre-decision confidence correlated with accuracy in Sauerland and Sporer's data, but the slope was so low that the differences were not vast between the most and least confident in terms of accuracy. The same was true for post-decision confidence for a non-chooser (that is,

someone who rejected the lineup). The challenge remaining for researchers is to take what seems to be a stable correlation (for choosers) and find some usefulness for practitioners, as a hard cut-off point for accuracy had eluded us thus far. Confidence and latency, then, seem to be moderately but dependably informative measures of eyewitness accuracy.

Witness Demographics

Researchers have also investigated pre-existing characteristics of eyewitnesses as a predictor of accuracy. Though many studies found no gender differences for eyewitnesses (see Valentine, 2013, for a review), some have shown general age differences and many have found the Other Race Effect (ORE) in eyewitness ID situations. As Valentine discussed, identification accuracy in general has been shown to start to decrease around the age of 50 along with other general cognitive decline. Separately, children have typically performed more poorly on lineup tasks and were more prone to misinformation than adults, but they were aided by the inclusion of a blank "wild card" face in the lineup (Valentine, 2013). Others have recently presented data suggesting that face processing ability matures linearly with age, but more information is needed to make any certain conclusions (Susilo, Germaine, & Duchaine, 2013). Granhag, Ask, and Giolla (2013) referred to an Own Age Effect similar to the ORE in that there was some evidence of increased accuracy when the criminal was nearer to the age of the eyewitness. The Other Race Effect is a well-established phenomenon of face recognition ability (see Meissner & Brigham, 2001); the effect has presented as well in the eyewitness literature in the identification context, though it did not present for event recall (Granhag et al., 2013). In an interesting twist, descriptions generated by

participants for own-race faces more often led to successful later lineup creation and identification by another participant than descriptions generated in cross-race conditions. Like confidence and latency, witness characteristics did not present an individually strong predictor for eyewitness accuracy.

Personality and Processing Style

Macrae and Lewis (2002) attempted to alter eyewitness memory quality rather than predict accuracy via a method inducing either a global or a local visual processing preference by presenting one or another set of stimuli created by Navon (1977). The authors theorized that a global orientation would assist performance while a local orientation would impair performance on a succeeding lineup task. Navon's stimuli are sets of large letters constructed of smaller letters that may or may not be the same as the large letter. By presenting inconsistent Navon letters (i.e. a letter made of many of one smaller letter) and instructing the participant to identify either the large or the small letter, one can orient the participant toward global or local processing respectively. Macrae and Lewis found that a set of orienting Navon letters presented between a crime video and an 8-person photospread lineup significantly changed performance in that globally oriented participants performed significantly better than controls and locally oriented participants performed significantly worse than both controls and globally oriented participants. This method represents one possible tool that may alter eyewitness accuracy but is not exhaustive. Other labs have found results similar to Macrae and Lewis, though not in every experiment they conducted (Perfect, 2003; Perfect, Dennis, & Snell, 2007; Perfect, Weston, Dennis, & Snell, 2008).

Certain dimensions of personality were tested as an eyewitness performance predictor by Geiselman, Haghighi, and Stown (1996). Self-rated high self-monitors made more positive ID's than low self-monitors when viewing target-present (TP) lineups, but no published replications appear since. Conversations with colleagues have revealed that other attempts to link personality measures such as impulsivity to eyewitness accuracy performance have failed to find any relationship, as can be seen in the lack of published studies with significant results (see Hosch, 1994 for a discussion). Megreya and Bindemann (2013) found that face matching ability correlated negatively with anxiety and other measures that make up neuroticism as a construct, but only for their female participants. Though these two results seem consistent, more replication within eyewitness identification paradigms is necessary before they would be considered reliable.

Individual Differences in Memory for Faces

Perhaps the most logical individual trait to relate to eyewitness performance was face recognition performance, as the two skills seem quite similar. This measure could only be useful if individual differences in face memory actually exist. Olsson and Juslin (1999) reported that participants in an eyewitness paradigm who self-reported using more holistic (rather than analytic) encoding styles showed a stronger relationship between confidence and accuracy in their lineup choices as well as generally more accurate positive identifications. This finding has been supported in the traditional face recognition literature (e.g. Tanaka & Farah, 1993), in which memory for faces holistically encoded has consistently outperformed that for faces encoded in an analytic or part-by-part method. The same result was found for participants who rated themselves as better-

than-average at all tasks of face recognition in daily life. While folks who fancy themselves good face recognizers might also have been the more confident eyewitnesses, these results are interesting in that non-informed participants rated themselves differently from each other in terms of processing method in a way that enabled Olsson and Juslin to establish definitive groups. Most introductory psychology students would not know that individual differences exist in face recognition, or that processing method can affect accuracy as well. Additionally, that some participants did not use the theoretically natural and easy method when presented with a set of faces to study suggests that the group using any non-holistic encoding method must have performed worse than their holistic-encoding colleagues. Another example of individual differences in face recognition was Russell, Duchaine, and Nakayama's (2009) display of the existence of super-recognizers; the discovery of such people who appeared to have a unique ability to recognize faces very easily provided additional existence of differences among people in face recognition ability. More recently, Bindemann, Avetisyan, and Rakow (2012) presented data from participants who varied enormously from one face matching test to another, both in general and for specific faces when retested. Some participants, however, were far more accurate overall than others (also highly accurate participants tended to be more consistent). Taken together, these results are convincing evidence of the presence of individual differences in face recognition.

Individual face recognition ability has been used as a predictor of eyewitness accuracy with some success. Geiselman et al. (2001) used the short form of a face-matching task known as the Benton Facial Recognition Task and found a moderate correlation between BFRT performance and lineup accuracy when witnessing conditions

were poor. Morgan et al. (2007) observed a positive relationship between face recognition ability and eyewitness accuracy for a group of Army trainees when using the Weschler Face Test to predict the trainee's ability to identify the officer from a stressful interrogation each trainee experienced earlier in the day. Bindemann, Brown, Koyas, and Russ (2012) recruited eighty participants to complete an eyewitness identification task and a face recognition test of their own design. Participants who saw a TP lineup and made a correct identification generated more face recognition hits than inaccurate participants in the same condition, but only nine participants generated a positive identification. In a second experiment with more subjects (185) and both TA and TP conditions, participants produced more correct rejections in subsequent face recognition when the initial lineup response was correct, and choosers on TP lineups again provided a better index of their later face recognition accuracy than non-choosers who saw TA lineups. No other relationships were clear, as the authors struggled with generally low accuracy rates and a dearth of respondents in certain cells of the dataset. If some face recognition test could predict or postdict eyewitness accuracy reliably, the field may eventually have a solid recommendation to make to police in this line of research.

Individual Differences in Response Bias

The research on individual differences in face recognition is backed up by similar findings for individual differences in response bias. Individual differences in response bias have been found by several researchers (e.g. Gillespie & Eysenck, 1980; Huh et al., 2006), and much research has been done in this vein via the Deese/Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). Qin, Ogle, and Goodman (2008) found that implantation of false childhood memories was more successful in participants

who had displayed a more liberal response bias on noncritical DRM trials, displaying a liberal bias' carryover to other tasks. Kantner and Lindsay (2012) discussed evidence of recognition memory response bias as a stable cognitive trait within a participant. They found support of the trait's stability in memory for a variety of different stimuli, including faces and words. When an individual's response bias was compared to his/her proclivity to choose from four target absent (TA) lineups related to earlier-witnessed crimes, a negative correlation was found between response bias and proclivity to choose, that is the more conservative recognizers (those with a higher threshold of evidence in memory for positive recognition) also tended to correctly reject more of the lineups than their liberal counterparts. Taken together, these results suggest that response bias may be a useful measure of ability on tasks outside recognition memory.

Still other attempts to predict eyewitness performance have come from participants' performance on tasks other than personality inventories or self-reports of the memory experience. After a study showing that many participants responded to a lineup based on an earlier live staged crime by choosing an innocent suspect shown in an intervening mugshot, Brown, Deffenbacher, and Sturgill (1977) proposed the use of a blank lineup (i.e. a lineup based on the target but not containing the target) before the actual lineup to attempt to identify "choosers," participants who would choose on a TA lineup because of a lower personal decision criterion. The nature of this individual difference was theorized to be that certain individuals had a more lax criterion than others for recognition and displayed this proclivity to choose on the initial blank lineup, displaying their lack of credibility for the second lineup compared to those resisted choosing from the blank lineup. Wells (1984) used an initial blank lineup as a screening

tool, and found that initial choosers were more likely to identify a foil from a subsequent TA lineup than initial non-choosers and less likely to identify the suspect than non-choosers or witnesses who had no blank lineup. Wells concluded that initial choosers adopt a relative judgment strategy and have poorer memory of the culprit. Palmer, Brewer, and Weber (2012) replicated Wells' findings (though they found a weaker effect) and found a similar pattern of results when forcing participants to respond to the blank lineup via instructions to choose the person who looks most like the perpetrator if he is not present. The lack of any strong conclusions from these studies on blank lineups suggests that the method may be best used in concert with others to more fully predict witness accuracy.

The Current Project

It is from the possibility of individual proclivity toward choosing as a cognitive trait and of face recognition and eyewitness memory correlations that the current project sprang. With evidence from our own lab (Kantner & Lindsay, 2012) and others that recognition bias is a stable individual trait, we looked to relate eyewitness memory performance on five crimes and lineups presented one after another to performance on a 2-alternative forced choice (2AFC) memory task for a set of face stimuli with a set of "trick" items in which neither face shown was originally studied, akin to a TA lineup. The 2AFC method helpfully paired a bias-free measure of facial recognition skill (accuracy on pairs containing one studied face and one non-studied faces) with a measure of proclivity to choose (pairs containing two non-studied faces). Confidence and decision latency for any positive responses during the face test should also be informative.

Kantner and Lindsay's (2012) participants first completed a face recognition study/test cycle on a computer, judging each face as "old" or "new" during the test as well as rating their confidence on a scale from 1 to 3. Next, they viewed 5 crime videos obtained from other researchers. The videos were between 45 and 75 seconds in length. Participants were told before the videos that they would later be expected to identify the culprits in six-person lineups. After watching the videos, participants completed a word recognition study/test cycle with the same format as the face recognition cycle. They then completed the lineup task by writing the number of their chosen suspect in a box or checking another box that indicated the perpetrator was not in the lineup. One lineup contained the criminal and was designated target present (TP); the rest did not contain their respective criminals and were designated target absent (TA). Participants also rated

confidence and decision ease on separate 1-10 scales. There was a significant negative correlation between criterion on the face recognition test and number of suspects identified, $r(63) = -0.30$, $p < .05$, indicating that more liberal recognizers tended to make more lineup identifications. This relationship was slightly weaker for the word recognition test but remained significant, $r(63) = -0.24$, $p < .05$. Results from Kantner and Lindsay's study indicated that response bias on either recognition test may correlate reliably with the likelihood of choosing in a lineup identification task.

Experiment 1

We elected to move forward in studying the relationship between likelihood of choosing in a lineup identification task and response bias for faces, as this relationship was stronger than that with response bias for words and it seemed more intuitively plausible that face recognition ability would be related to lineup identification ability. Our central aim in this project was to develop a procedure that yields measures of proclivity toward choosing foils and of ability to recognize faces. Our basic approach was influenced by personal communications with Larry L. Jacoby, who was, in a completely different context, testing people on 2AFC tests that included "trick" trials in which both of the words were new. In our procedure, subjects study a series of faces and then are shown pairs of faces including some "trick" trials. We use responses on trick trials (New/New pairs) as an index of proclivity to chose, and we used accuracy on non-trick trials (Old/New pairs) as an index of face-recognition ability. Participants in Experiments 1-3 were all UVic undergraduates who participated in exchange for optional bonus points in psychology courses. Approximately 75% of this pool was female, and 95% were 17 to 25 years of age.

Method. In Experiment 1 ($N = 60$), participants entered the lab and viewed the same five crime videos as in Kantner and Linday (2012) on a 19'' LG Flatron monitor (1,280 x 1,024 max resolution) running a Windows system through E-Prime software. The videos depicted a store theft committed by a young woman, a house theft by a young man (both created by Neil Brewer), a man stalking a woman at an ATM, a man stalking a woman at a park (both stalker videos by Fiona Gabbert), and a young man planting a bomb on a roof (created by Gary Wells), The videos were followed by a study list of 40 photos of Caucasian male faces (with no obviously distinctive features such as tattoos or scars) shown one at a time from the shoulders up with neutral facial expressions against a white backgrond in an order randomized anew for each participant. The photos were of undergraduate students from a previous introductory psychology course who gave consent for their photos to be used in future studies. Memory for these faces was tested after a short distractor task via a two-alternative forced choice (2AFC) recognition test in which the same or different Caucasian males were pictured smiling, so as to guarantee face recognition was being tested rather than photo recognition, in side-by-side pairs (Bindemann, Avetisyan, & Rakow, 2012). Test faces were also randomized for each participant. The first and last two faces in the study list were not used in the test list to avoid primacy and recency effects. Thirty-six test pairs consisted of one studied (or old) and one non-studied face (or new, creating an Old/New pair), and 4 test pairs consisted of two non-studied faces (a New/New pair, akin to Jacoby's "trick" pairs). Test instructions did not mention the possibility of a New/New pair, and rejecting the pair was not a response option. Participants rated confidence in each response on the test list on a 5-point scale and finished by completing the lineup task with one TP lineup and four TA

lineups (obtained from previously mentioned researchers) with the same photos and post-decision ratings used by Kantner and Lindsay (2012). Responses to Old/New pairs were excluded on a trial-by-trial basis when output indicated reaction times faster than 500ms (agreed upon a priori) or slower than 15000ms (selected after all studies were completed based on typical RT ranges).

Results. Figure A1 is a jittered scatterplot of confidence for New/New pairs and proportion correct in the lineup task, $r(60) = -0.18$, $p < .05$. Table A1 presents reliability analyses indicating a possible maximum correlation of $r = -0.38$ based on Cronbach's Alpha values of 0.46 and 0.47 for the predictor and outcome variables, respectively. Reanalysis without the one TP lineup resulted in a maximum possible correlation of r=-0.33. However, such a correlation would not make logical sense as more confident choosers on New/New face test pairs should be displaying a higher proclivity to choose on lineups.

Figure A2 shows a weak albeit significant relationship between confidence for New/New pairs and confidence on lineups from which the participant chose (16 participants' confidence ratings were not recorded due to a programming error), $r(44) = 0.30$, $p < .05$. Reliability analyses indicated a possible maximum correlation of $r = .47$ based on Cronbach's Alpha values of 0.50 and 0.82 for the predictor and outcome variables, respectively. This finding suggests that there may be individual differences in confidence that hold across tasks (e.g., Kasperski & Katzir, 2013).

Statistics in Table A2 show that confidence ratings in response to Old/New pairs on the face recognition test were higher when choosing correctly than incorrectly when examined with a paired-samples T Test, means 3.20 and 2.52 respectively, $t = 9.17$, $p <$

.001. Also, participants with a higher correlation between their confidence and accuracy for Old/New pairs during the face test were no more or less likely to produce false alarms on the lineups, $r(60) = -0.06$, $p > .05$.

Experiment 2

As the expected relationship was not observed between New/New trial confidence and lineup rejection rate, several changes were made for Experiment 2. We speculated that giving participants an answer choice analogous to lineup rejection would aid them in making the analogy between the two tasks. Thus, we hypothesized that there might be a relationship between rejection rates for New/New pairs and rejection rates for lineups.

Method. The procedure for Experiment 2 ($N = 53$) was the same as that of Experiment 1 with the notable exceptions that participants were warned of the possibility of New/New pairs during the face recognition test and given the option to reject any of the pairs via a "Neither" option presented between the two faces. Data were also collected by a different research assistant and sampled from a new pool of UVic psychology undergraduates in a different semester.

Results. Figure A3 is a jittered scatterplot proportion correct on N/N pairs and proportion correct on lineups, $r(53) = 0.01$, $p > .05$. Table A1 presents that reliability analyses indicated a possible maximum correlation of r=-.04 based on Cronbach's Alpha values of 0.40 and 0.18 for the predictor and outcome variables, respectively. Reanalysis without the one TP lineup resulted in a maximum possible correlation of r=0.26.

Figure A4 shows a weak albeit significant relationship between confidence ratings when identifications were made in lineups and New/New face recognition trials, $r(51) = .36$, $p < .01$ ($N = 51$ because some participants made no false positive responses on one

task or the other). Reliability analyses are not prudent here because subjects sometimes correctly rejected New/New pairs resulting in automatic removal of too many cases to conduct the analyses. The correlation, however, adds evidence to the suggestion that confidence may be a stable individual difference across tasks.

Statistics in Table A2 show that confidence ratings in response to Old/New pairs on the face recognition test were higher when choosing correctly than incorrectly when examined with a paired-samples T Test, means 3.42 and 2.60 respectively, $t = 13.48$, $p < .001$. Also, participants with a higher correlation between their confidence and accuracy for Old/New pairs during the face test were no more or less likely to produce a false alarm on the lineups, $r(53) = -0.05$, $p > .05$.

Experiment 3

Method. To encourage participants to approach the 2AFC face memory task in the same way they approached the lineup task, In Experiment 3 (N = 50) participants were told before beginning the 2AFC test that the aim of the study was to develop a measure of eyewitness reliability. The 2AFC test was referred to as a series of "mini-lineups," with instruction that some of those 2-person lineups had no culprit and hence should be rejected; the instructions also made clear the fact that we would use their performance on the mini-lineups to predict their ability on the full lineups. It was theorized that highlighting the parallels between the two tests would increase the extent to which performance on the 2AFC test predicted performance on the 6-person TA lineups. Another advantage of straightforward, above-board instructions in such a study would be relatively easy implementation in the real world. Crime and lineup order were counterbalanced for this experiment, and the one previously TP lineup was changed to a

TA lineup for ease of analysis and in an attempt to gain the most data possible from each participant. Data were collected from a new sample of Uvic undergraduates in a different semester by the same research assistant as Experiment 2. Crime and lineup order were counterbalanced.

Results. Figure A5 is a jittered scatterplot proportion correct on N/N pairs and proportion correct on lineups, $r(50) = -.006$, $p > .05$. Table A1 presents that reliability analyses indicated a possible maximum correlation of $r = -.02$ based on Cronbach's Alpha values of 0.31 and 0.20 for the predictor and outcome variables, respectively.

Figure A6 shows a relationship between confidence ratings when identifications were made in both lineups and New/New face recognition trials, $r(49) = .43$, $p < .01$. This finding adds more evidence to the suggestion that confidence may be a stable individual difference across tasks.

Statistics in Table A2 show that confidence ratings in response to Old/New pairs on the face recognition test were higher when choosing correctly than incorrectly when examined with a paired-samples T Test, means 3.29 and 2.66 respectively, $t = 11.22$, $p < .001$. Also, participants with a higher correlation between their confidence and accuracy for Old/New pairs during the face test were more likely correctly reject the lineups, $r(50) = 0.34$, $p < .02$.

Experiment 4

Method. Perhaps the reason for a lack of the theorized relationship between face recognition ability and eyewitness accuracy may be due to a lack of power stemming from the small number of observations gained from each participant (with four New/New pairs and five lineups). Experiment 4 (N=65), therefore, marked several major changes in

method. A new set of five video clips was created from British television crime dramas, all of which depicted middle-aged Caucasian male culprits committing crimes. A clip of a man breaking into a home was obtained from "Vincent," a clip of a man and woman arguing and a clip of a woman's car exploding as she leaves her home were obtained from "MI-5," a clip of a man destroying a set of china with a shotgun was obtained from "Dalziel and Pascoe," and a clip of a man shooting another man was obtained from "Murder City." Lineups for the new crimes were created by finding the first results when scrolling through the State of Florida's online database of criminal mugshots that matched the description of the culprit from the video. Many faces were collected, then face sets were pared down to include only the most similar faces in the lineups as judged by the experiment and a Research Assistant. The most similar face was designated the "innocent suspect," which replaced the culprit's face for the TA lineup. It was theorized that this significant change in perpetrator age would help prevent any confusion between the faces for the two tasks, as the faces studied for the face recognition test were all University undergraduates. The list of faces was also expanded to include Caucasian females from the same set of previous introductory psychology students. Some students in the new set appeared wearing a smock to conceal their clothing. The study list was shortened to 30 faces, and the test list was expanded to 60 pairs, 30 of which were Old/New pairs and 30 of which were New/New pairs. This was perhaps the most important change for Experiment 4, as the large increase in New/New pairs was meant to offer the necessary amount of data from each participant to investigate whether the originally theorized relationship between N/N pair rejections and lineup rejections can be found in the laboratory setting. The study instructions referred to the face recognition test

as a Lineup Skills Test and informed participants before the test phase that it was meant to measure their ability on the forthcoming lineup task. Finally, crime and lineup order were counterbalanced and the face recognition study and test phases were presented in a fixed random order that was different for each version of the counterbalance.

Participants. This study also marked a change in recruitment strategy. Whereas the previous three all were conducted via PST's E-Prime software and recruited from the local population of UVic psychology undergraduates, this study was conducted online via the survey-hosting site Qualtrics and recruited participants through Amazon's Mechanical Turk work exchange website (see Mason & Suri, 2012, for discussion of M. Turk recruitment). Workers completed the task in exchange for $0.60 USD, effectively about a $1.20/hour wage. Participants who did not complete the task were eliminated from analysis, as were participants who confessed to major distractions, responded incorrectly to immediate questions regarding content of the videos or did not stay on the video pages long enough to watch them.

Participants self-reported demographics. The question regarding age appeared to have malfunctioned in its drop-down menu, as many participants reported age levels below 18 and even 10 years of age. Of those for whom the age menu appears to have worked, average age was 31.7 years. Forty of the 65 participants were female, 58 reported English as their native language, 33 reported having taken at least some university courses, and 61 reported no university courses in psychology.

Results. Figure A7 is a jittered scatterplot displaying proportion correct on N/N pairs and proportion correct on lineups, $r(65)= 0.45$, $p < .001$. Table A1 presents that reliability analyses indicated a possible maximum correlation of $r = 0.88$ based on

Cronbach's Alpha values of 0.91 and 0.30 for the predictor and outcome variables, respectively. The score of .91 for the N/N face pairs is the first such score high enough to consider individual difference analyses appropriate (Susilo, Germaine, & Duchaine, 2013).

Figure A8 shows a lack of relationship between confidence ratings when identifications were made in both lineups and New/New face recognition trials, $r(65) = .16$, $p > .05$. This finding is evidence against the suggestion that there may be individual difference in confidence that hold across tasks, but in light of the presentation of the correlation on three of four experiments, one might begin to find this relationship worthy of further investigation. Another weak yet significant relationship was found between Old/New pair accuracy when choosing and lineup rejection rates, $r(65) = 0.33$, $p < .05$, suggesting that face recognition sensitivity might be a predictor of lineup rejection ability. A similar relationship was discovered between Old/New pair rejection rate and lineup rejection rates, $r(65) = 0.38$, $p < .01$. These correlations may suggest that other information gained from the face recognition task might be informative when predicting lineup rejections. However, a hierarchical regression of lineup rejection rate treating New/New rejection rate as the predictor in model 1 did not benefit significantly from the addition of Old/New accuracy and rejection rate as predictors in model 2, $F(61) = 1.40$, $p > .05$.

Statistics in Table A2 show that confidence ratings in response to Old/New pairs on the face recognition test were higher when choosing correctly than incorrectly when examined with a paired-samples T Test, means 3.69 and 2.74 respectively, $t = 6.60$, $p < .001$. Also, participants with a higher correlation between their confidence and accuracy

for Old/New pairs during the face test were no more or less likely to produce a false alarm on the lineups, $r(65) = 0.14$, $p > .05$. This final relationship taken across studies seems to indicate no strong relationship between these two variables.

General results. If our "mini-lineups" were processed like regular lineups, the faster and more confident choosers should be accurate more often for Old/New pairs than those choosing more slowly and with less confidence (Sauerland & Sporer, 2007 & 2009). Figure A9 (created with Cumming's ESCI module) displays the results of a random effects meta-analysis showing that in Experiments 1, 3, and 4 accurate choices were made a bit faster than inaccurate choices, leading to a grand mean of 176.98 milliseconds faster (CI = -67.71, 421.67) collapsed across studies (Cumming, 2012). Figure A10 (also via ESCI) displays another random effects meta-analysis showing that in all four Experiments accurate choices were made with more confidence than inaccurate choices, leading to a grand mean difference of 0.73 points more confident on a scale 1-5 (CI = 0.61, 0.85).

Discussion. The current findings add to the growing literature (Geiselman et al., 2001; Morgan et al., 2007; Bindemann, Brown, Koyas, and Russ, 2012) supporting the idea that face recognition may be a reliable predictor of eyewitness identification performance ability. We may begin to discuss the possibility that proclivity to choose is a stable individual trait and may eventually suggest that police departments run their eyewitnesses through a face recognition protocol in order to predetermine their likelihood to produce a false positive. This type of test would prove useful in identifying witnesses who, regardless of the quality of their view of the culprit, have a lower criterion for face recognition in a lineup-type situation and would be more likely to accidentally send an

innocent suspect to jail. This method would make the eyewitness identifications of pre-established choosers less informative, but would make any lineup rejections produced by a chooser indicative of a more definite lack of the culprit in that lineup. The method would also enable identification of witnesses who show a strong ability to resist production of false positive identifications, turning whatever ID they do make into strong evidentiary support and enabling any rejection made by such a non-chooser to be taken in the proper context.

In addition to the hypothesized correlation, we found evidence that our "mini-lineups" elicited behavior similar to that elicited by regular lineups in that our participants responded faster in three of four experiments and were more confident in all experiments when choosing correctly than when choosing incorrectly on the face recognition test pairs (with rejections and New/New trials removed). Despite that the encoding conditions were quite different from that of the videoed crimes; participants seem to have reacted in a similar fashion. These findings fit within the existing research on confidence and RT for lineup tasks and general recognition memory tasks. Such data is not informative from the actual lineup identifications because no accurate positive ID's were made (other than those on the accidental TP lineup). We also found that confidence when choosing on New/New pairs correlated with confidence when choosing on lineups, most likely indicating that individual participants use the confidence scales differently in that some use the higher side and some use the lower side of the scale. As this tendency is an individual difference as is proclivity to choose, it stands to reason that this correlation would be present.

Most weaknesses in the first three experiments were eliminated for Experiment 4. These included possible confusion between the students in the crime videos and the students in the slideshow, which was corrected by the creation of different crime videos that all depicted middle-aged Caucasian male criminals rather than the university-aged criminals in the original videos and in the face recognition test. Another issue in Experiments 1-2 was the accidental inclusion of a TP lineup, which has been accounted for in previous sections. The third major issue with Experiments 1-3 was the small number of New/New trials and the lack of observations with which we attempted to predict lineup rejections. An additional issue may be the real-world implementation of this exact method, as police officers would not necessarily be willing to expose an eyewitness to 120 new faces before they attempt to identify their criminal (especially given the research on blank lineups' influence of future choices). The current procedure was chosen to place time and memory load between the crime videos and lineups, but modifications such as presenting lineups before the entire face recognition task might be needed if this were to ever be implemented in an actual police station. Perhaps a more ecologically valid method would be to first expose participants to the crime videos then have them complete lineups up to a week later followed immediately by the face recognition test. One final issue may be that the one study in which the predicted relationship was found was sampled from a different population from the other three. Some studies on the representative capabilities of Mechanical Turk samples have shown similar behavior to those of undergraduate populations (Behrend, Sharek, Meade, & Wiebe, 2011), and in this case a more representative sample should be more desirable. Still, these findings may benefit from replication in the lab or with different materials. As

has been stated, the next step is replication of Experiment 4. Beyond that step, we will look to parametrically vary the presence and absence of culprits in their lineups to attempt to find a sweet spot from which we can not only identify participants as choosers but as often-accurate or often-inaccurate choosers. We will also increase the ecological validity of the study by presenting lineups before the face recognition test, adding a delay between viewing the crime videos and responding to the lineups, and eventually using a live mock crime procedure for maximum realism. Additionally, we plan to conduct a test phase using stimuli other than faces to minimize interference (e.g., words or greebils), and perhaps lengthen the test further by using two study/test cycles.

This study supports previous findings of the relationship between face recognition performance and lineup accuracy, specifically that proclivity to choose on TA face recognition trials predicts proclivity to choose on later TA lineups. The current data also support previous findings of the relationship between confidence and accuracy and between reaction time and accuracy on face recognition trials. With most procedural issues eliminated for Experiment 4, future studies will focus on increasing ecological validity and realism for participants as well as looking for links between face recognition performance and proclivity to choose from TP lineups.

References

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800-813.

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*, 96-103.

Bindemann, M., Avetisyan, M., & Rakow, T. (2012). Who can recognize unfamiliar faces? Individual differences and observer consistency in person identification. *Journal of Experimental Psychology: Applied, 18*(3), 277-291.

Brewer, N., Weber, N., Clark, A., & Wells, G. L. (2008). Distinguishing accurate from inaccurate eyewitness identifications with an optional deadline procedure. *Psychology, Crime & Law*, *14*(5), 397-414.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, *62*(3), 311-318.

Chance, J. E., & Goldstein, A. G. (1979). Reliability of face recognition performance. *Bulletin Of The Psychonomic Society*, *14*(2), 115-117.

Charman, S. D. & Cahill, B. S. (2012). Witnesses' memories for lineup fillers postdicts their identification accuracy. *Journal of Applied Research in Memory and Cognition*, *1*(1), 11-17.

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, *35*, 364-380.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals,*

and meta-analysis. New York: Routledge.

Darling, S., Martin, D., Hellmann, J. H., & Memon, A. (2009). Some witnesses are better than others. *Personality and Individual Differences*, *47*, 369-373.

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law And Human Behavior*, *3*(3), 287-307.

Dunning, D., & Stern, L. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal Of Personality And Social Psychology*, *67*(5), 818-835.

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal Of Applied Psychology*, *87*(5), 951-962.

Emmett, D., Clifford, B. R., & Gwyer, P. (2003). An investigation of the interaction between cognitive style and context reinstatement on the memory performance of eyewitnesses. *Personality and Individual Differences*, *34*, 1495-1508.

Geiselman, R. E., Tubridy, A., Bkynjun, R., Schroppel, T., Turner, L., Yoakum, K., & Young, N. (2001). Benton Facial Recognition Test scores: Index of eyewitness accuracy. *American Journal of Forensic Psychology*, *19*(1), 77-88.

Geiselman, R. E., Haghighi, D., & Stown, R. (1996). Unconscious transference and characteristics of accurate and inaccurate eyewitnesses. *Psychology, Crime & Law*, *2*(3), 197-209.

Gorenstein, G. W. & Ellsworth, P. C. (1980). Effect of choosing an incorrect photograph on a later identification by an eyewitness. *Journal of Applied Psychology*, *65*(5),

616-622.

Granhag, P. A., Ask, K., & Giolla, E. M. (2013). Eyewitness recall: An overview of estimator-based research. In D. S. Lindasy & T. J. Perfect (Eds.), *The SAGE Handbook of Applied Memory* (pp. N/A). New York, NY: SAGE Publications.

Gronlund, S. D & Carlson, C. A. (2013). System-based research on eyewitness identification. In D. S. Lindasy & T. J. Perfect (Eds.), *The SAGE Handbook of Applied Memory* (pp. N/A). New York, NY: SAGE Publications.

Hosch, H. (1994). Individual differences in personality and eyewitness identification. In D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 328-347). New York, NY: Cambridge University Press.

Huh, T. J., Kramer, J. H., Gazzaley, A., & Delis, D. C. (2006). Response bias and aging on a recognition memory task. Journal of the International Neuropsychological Society, 12, 1–7.

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

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, *40*(8), 1163-1177.

Kasperski, R., & Katzir, T. (2013). Are confidence ratings test- or trait-driven? Individual differences among high, average, and low comprehenders in fourth grade. *Reading Psychology*, *34*(1), 59-84.

Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witness' versus investigators' confidence in the accuracy of witness' identification decisions. *Law and Human Behavior, 24*, 685-697

Mansour, J. K., Lindsay, R. C. L., & Beaudry, J. L. (29 June, 2013). Comparing choosing accuracy and confidence in single- versus multiple-trial lineup experiments. Paper presented at the 10th biannual meeting of the Society for Applied Research in Memory and Cognition.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1-23.

Megreya, A. M., & Bindemann, M. (2013). Individual differences in personality and face identification. *Journal Of Cognitive Psychology*, *25*(1), 30-37.

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

Morgan, C. A., Hazlett, G., Baranoski, M., Doran, A., Southwick, S., & Loftus, E. (2007). Accuracy of eyewitness identification is significantly associated with performance on a standardized test of face recognition. *International Journal of Law and Psychiatry*, *30*, 213-223.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353-383.

Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identifications? *Journal of Applied Psychology*, *84*(1), 42-49.

Palmer, M. A., Brewer, N., & Weber, N. (2012). The information gained from witnesses' responses to an initial "blank" lineup. *Law and Human Behavior*, *36*(5), 439-447.

Perfect, T. J. (2003). Local processing bias impairs lineup performance. *Psychological Reports*, *93*, 393-394.

Perfect, T. J., Dennis, I., & Snell, A. (2007). The effects of local and global processing orientation on eyewitness identification performance. *Memory*, *15*(7), 784-798.

Perfect, T. J., Weston, N. J., Dennis, I., & Snell, A. (2008). The effects of precedence on Navon-induced processing bias in face recognition. *The Quarterly Journal of Experimental Psychology*, *61*(10), 1479-1486.

Pozzulo, J. D., Crescini, C., & Lemieux, J. M. T. (2008). Are accurate witnesses more likely to make absolute judgments? *International Journal of Law and Psychiatry*, *31*, 495-501.

Qin, J., Ogle, C. M., & Goodman, G. S. (2008). Adults' memories of childhood: True and false reports. Journal of Experimental Psychology: Applied, 14, 373–391.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 803–814.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252-257.

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, *137*, 528-547.

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law*, *13*(6), 611-625.

Sauerland, M., Sagana, A., & Sporer, S. L. (2012). Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law & Human Behavior*, *36*(5), 394-403.

Sauerland, M. & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, *15*(1), 46-62.

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminological Psychology*, *5*, 219-235.

Smith, S. M., Lindsay, R. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology*, *85*(4), 542-550.

Susilo, T., Germine, L., & Duchaine, B. (2013). Face recognition ability matures late: Evidence from individual differences in young adults. *Journal Of Experimental Psychology: Human Perception And Performance*, *39*(5), 1212-1217

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal Of Experimental Psychology A: Human Experimental Psychology*, *46A*(2), 225-245.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive*

Psychology, *17*, 969-993.

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal Of Experimental Psychology: Applied*, *10*(3), 139-147.

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, *14*, 89-103.

Wells, G. L., (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology*, *70*(4), 619-626.

Weston, N. J., Perfect, T. J., Schooler, J. W., & Dennis, I. (2008). Navon processing and verbalisation, A holistic/featural distinction. *European Journal of Cognitive Psychology*, *20*(3), 587-611.

Macrae, C., & Lewis, H. L. (2002). Do I know you?: Processing orientation and face recognition. *Psychological Science*, *13*(2), 194-196.

Valentine, T. (2013). Estimating the reliability of eyewitness identification. In D. S. Lindasy & T. J. Perfect (Eds.), *The SAGE Handbook of Applied Memory* (pp. N/A). New York, NY: SAGE Publications.

Appendix A

Figure A1

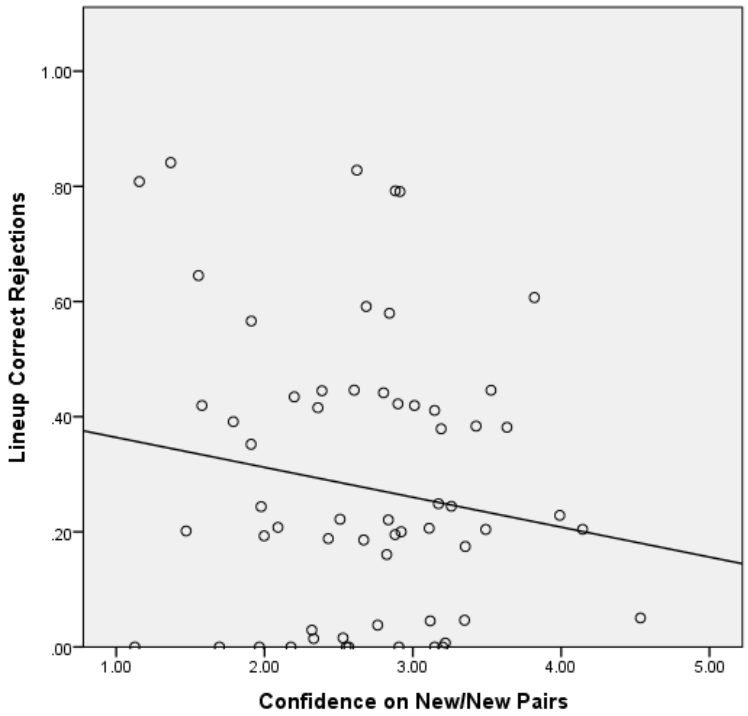

Figure A2

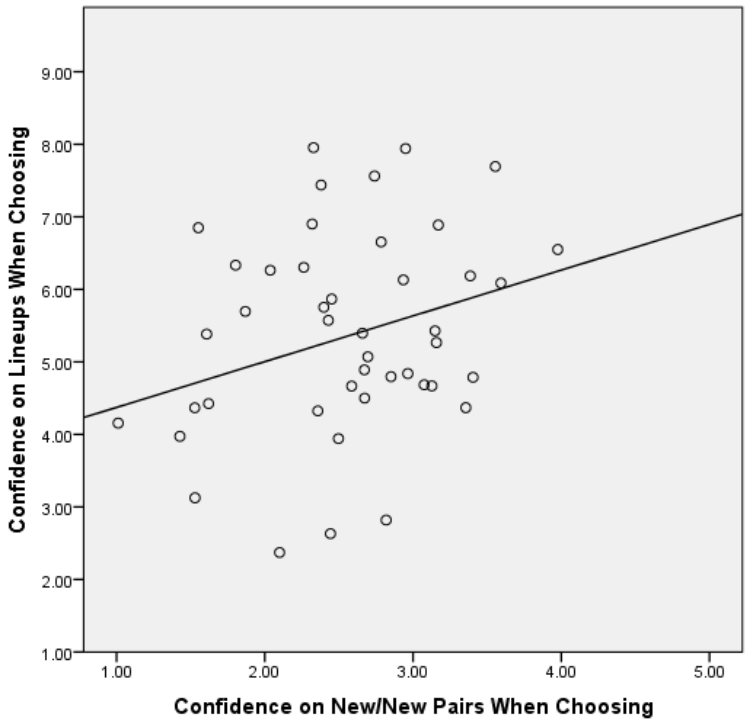

Figure A3

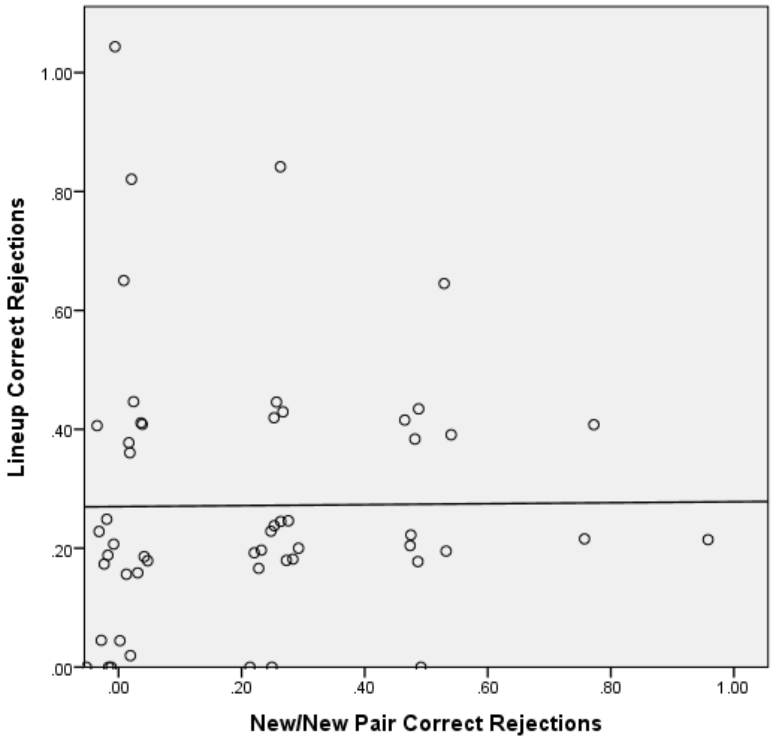

Figure A4

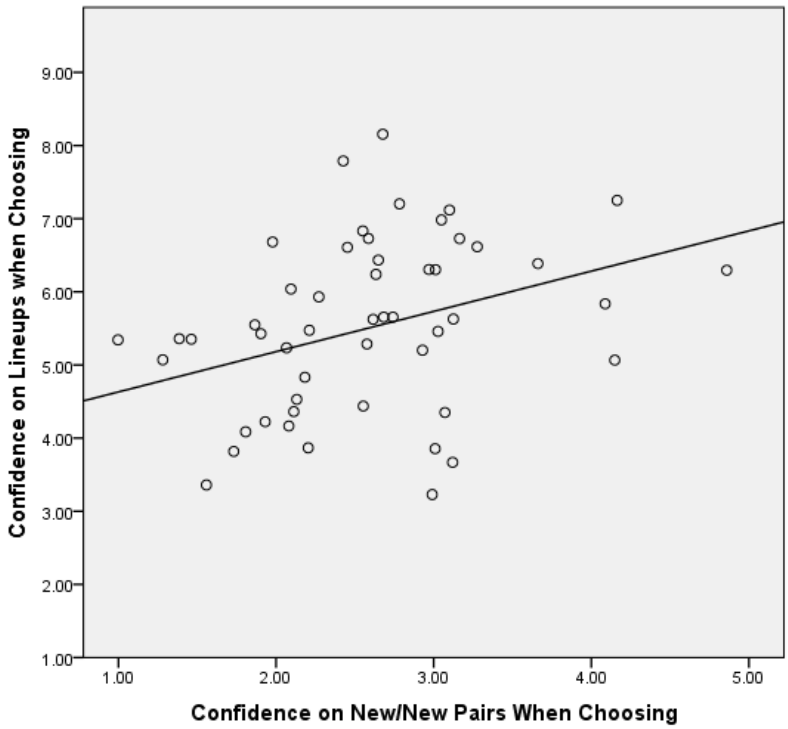

Figure A5

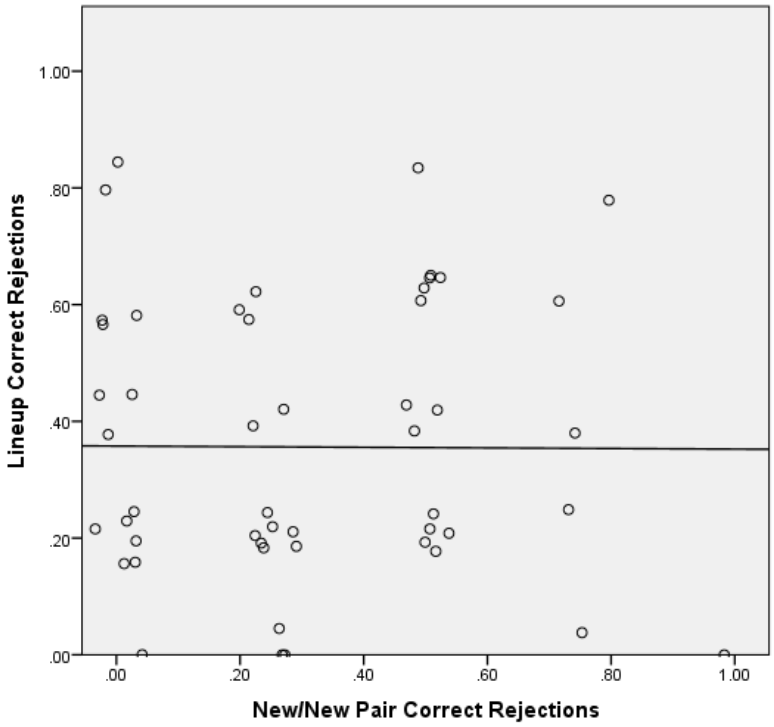

Figure A6

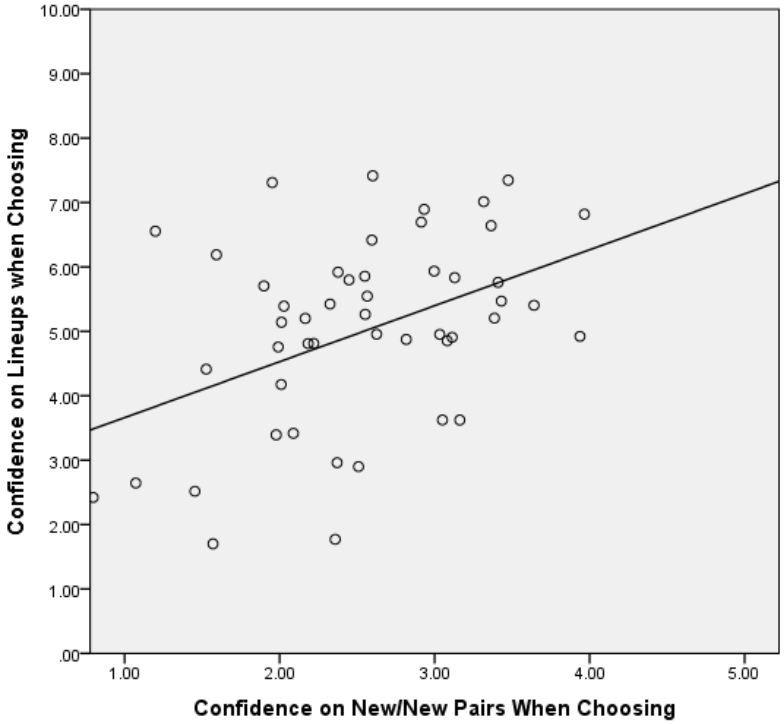

Figure A7

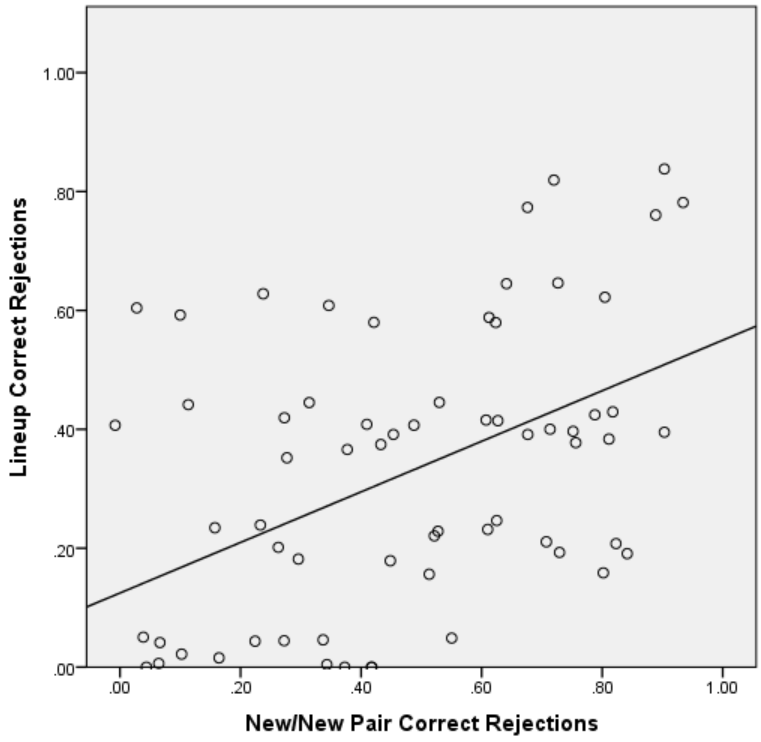

Figure A8

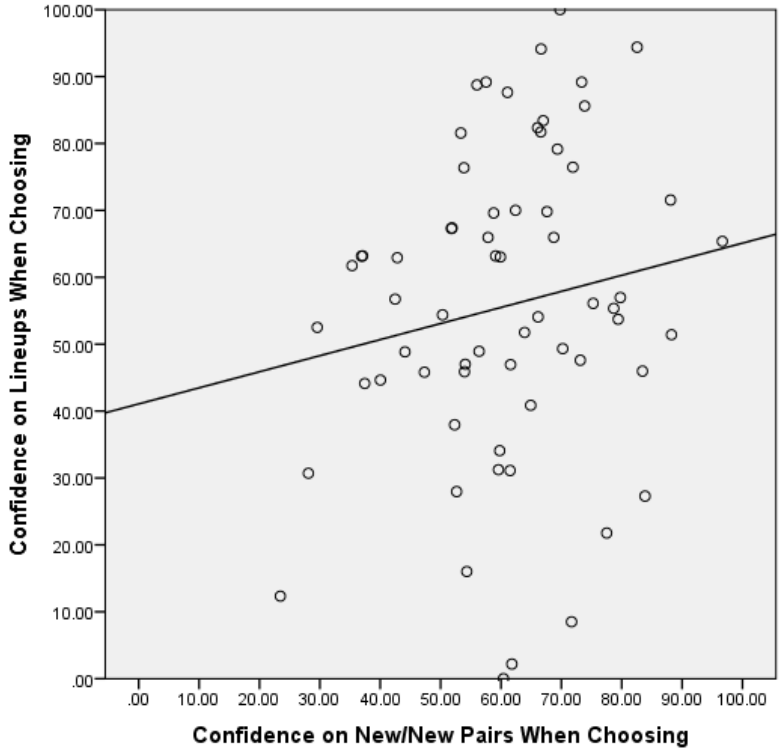

Figure A9

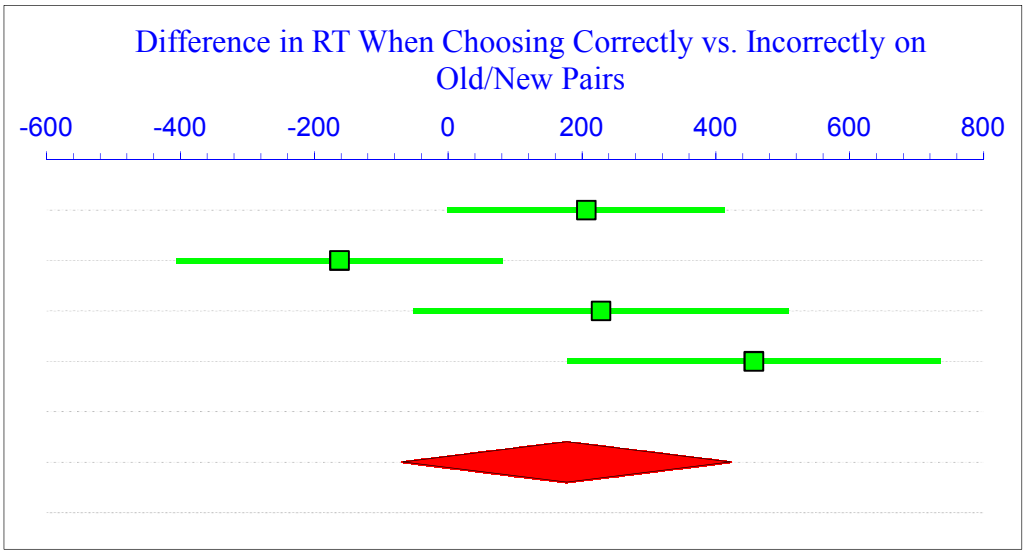

Figure A10

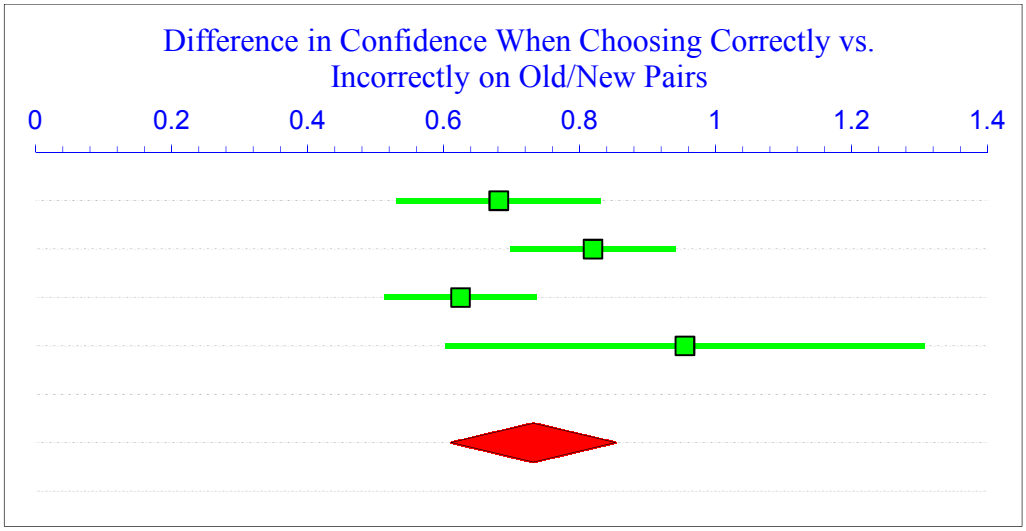

Table A1: Correlation given individual Cronbach's alpha

PID 1	With lineup 4 rejection	Without lineup 4
Original R	-0.18	-0.16
Face Alpha	0.46	0.46
Lineup Alpha	0.47	0.53
Final R	-0.38	-0.33
PID 2	With lineup 4 rejection	Without lineup 4
Original R	0.01	0.07
Face Alpha	0.40	0.40
Lineup Alpha	0.18	0.19
Final R	0.04	0.26
PID 3		
Original R	-0.01	
Face Alpha	0.31	
Lineup Alpha	0.20	
Final R	-0.02	
PID 4		
Original R	0.45	
Face Alpha	0.91	
Lineup Alpha	0.30	
Final R	0.88	

Table A2: Confidence ratings

Experiment	Status	Mean	SD	SEM	CI Upper Bound	CI Lower Bound	t	p<
1	Correct	3.198	0.463	0.060	4.123	2.273	9.17	0.001
	Incorrect	2.517	0.537	0.069	3.591	1.442		
2	Correct	3.417	0.629	0.086	4.675	2.159	13.48	0.001
	Incorrect	2.598	0.614	0.084	3.826	1.369		
3	Correct	3.285	0.584	0.083	4.454	2.116	11.22	0.001
	Incorrect	2.660	0.658	0.093	3.975	1.345		
4	Correct	3.692	0.771	0.096	5.233	2.150	6.60	0.001
	Incorrect	2.737	1.476	0.183	5.689	-0.216		

Table A3: Distribution of Lineup Choices in Exp 1-3 (Suspect in bold)

Lineup Member	Store Theft	ATM Stalker	Roof Bomb	Park Stalker	House Theft
1	4	7	12	5	7
2	10	17	6	21	22
3	5	7	26	1	15
4	**44**	10	14	11	**32**
5	20	**39**	14	19	21
6	22	33	**45**	**81***	6
Rej	59	51	47	26	61

*In Exp 1&2 this was the perp

Table A4: Distribution of Lineup Choices in Exp 4 (Suspect in bold)

Lineup Member	Car Bomb	Stabbing	Damaged China	Gunshot Murder	Break & Enter
1	3	11	2	10	6
2	20	**7**	5	12	15
3	**8**	5	11	5	7
4	12	12	4	**6**	**12**
5	4	2	**12**	10	5
6	1	5	4	1	1
Rejections	17	23	27	21	19

Table A5: Raw Number of Face Recognition Responses by Participant (Exp 4)

Face Recognition				Lineups		
Hits	Foil Selections	False Rejections	Correct Rejections	Suspect Selections	Foil Selections	Correct Rejections
23	13	24	13	0	2	3
22	17	21	12	2	1	2
15	12	33	18	0	4	1
3	7	50	27	0	1	4
10	20	30	22	1	3	1
24	19	17	9	2	1	2
27	14	19	8	0	2	3
13	13	34	22	2	1	2
11	21	28	21	0	1	4
19	14	27	16	1	3	1
10	22	28	20	0	4	1

10	13	37	20	0	3	2
11	18	31	24	0	3	2
17	25	18	15	1	3	1
19	18	23	18	0	2	3
37	8	15	4	0	2	3
8	24	28	25	2	0	3
33	19	8	5	0	4	1
34	12	14	7	1	4	0
27	10	23	12	0	5	0
19	11	30	15	1	2	2
27	20	13	10	1	4	0
20	17	23	15	1	3	1
34	18	8	6	2	3	0
22	21	17	12	0	4	1
27	17	16	8	1	2	2
36	19	5	3	0	3	2
26	16	18	10	1	4	0
12	17	31	21	0	2	3
11	23	26	19	0	2	3
17	11	32	17	0	5	0
21	17	22	14	0	3	2
21	18	21	14	0	3	2
5	10	45	27	0	1	4
42	12	6	1	1	4	0
18	26	16	13	0	5	0
12	19	29	21	0	1	4
10	9	41	23	1	3	1
12	18	30	23	1	2	2
9	17	34	23	0	3	2
16	11	33	18	1	2	2
23	21	16	11	2	1	2
4	10	46	26	0	1	4
41	13	6	3	3	2	0
27	18	15	11	1	4	0
15	11	34	20	0	2	3
34	10	16	8	2	3	0
37	22	1	0	0	3	2
4	19	37	26	1	2	2
5	18	37	25	0	3	2
7	15	38	24	2	2	1
6	19	35	24	0	4	1
10	21	29	20	1	2	2
18	21	21	14	1	2	2
36	18	6	3	1	4	0
28	21	11	8	1	3	1
39	17	4	3	2	3	0

39	19	2	2	2	3	0
39	19	2	1	0	2	3
29	17	14	9	1	2	2
28	21	11	6	1	3	1
22	5	33	17	1	3	1
30	18	12	10	1	3	1
26	13	21	12	0	2	3
7	23	30	23	0	3	2

Appendix B: Lineups

Lineups from Experiments 1-3

Store Theft

1	2	3	4	5	6

ATM Stalker

1	2	3	4	5	6

House Theft

1	2	3	4	5	6

Bomb Scenario

1	2	3	4	5	6

Lineups from Experiment 4

1 2 3

4 5 6

1 2 3

4 5 6


1
2
3
4
5
6


1
2
3
4
5
6