UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EYEWITNESS IDENTIFICATION: AN INVESTIGATION OF THE FEATURE-DETECTION HYPOTHESIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

SHANNON M. TOFTELY
Norman, Oklahoma
2015

EYEWITNESS IDENTIFICATION: AN INVESTIGATION OF THE FEATURE-DETECTION HYPOTHESIS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY

Dr. Scott Gronlund, Chair

Dr. Michael Wenger

Dr. Lynn Devenport

Dr. Robert Terry

Dr. Susan Sharp

TO MY HUSBAND, JESSE TOFTELY,

whose love emboldens and sustains me.

&

MY PARENTS, JERRY AND TERRI ANDERSON,

who unceasingly love and support me.

Acknowledgements

I would like to recognize the members of my committee: Dr. Scott Gronlund, Dr. Michael Wenger, Dr. Robert Terry, Dr. Lynn Devenport, and Dr. Susan Sharp for critically evaluating my work and supporting my academic undertakings. Each of you has taught and aided me in your own way and I am forever grateful.

I would especially like to thank my advisor and mentor, Dr. Scott Gronlund, for his ceaseless support, kind words of encouragement, and, of course, his meticulous editing. Scott, you shaped me into a successful academic and encouraged me to become a better writer and communicator. Even though I sometimes forgot to incorporate your comments (and you sometimes forgot our 2:30 Friday meetings), I will never forget your gentle words of encouragement. When my anxiety pre-presentation took over you would tell me "You're only nervous because you truly care". And when I was so exhausted and struggling with my dissertation data analysis you reminded me "If you're tired, it means you're close finishing!" and told me that I could do it. You believed in me and, in turn, helped me believe in myself. Most of the time that was all I needed to push through. You have always been there for me. Cheering me on when I needed it most. Thank you so much, Scott.

I also wanted to give special acknowledgments to an additional academic mentor, Dr. Rick Thomas. Rick contributed greatly to this dissertation, despite no longer serving on my committee. His suggestions at my prospectus meeting completely changed the one of my proposed studies into a much-improved, innovative test of a novel theory. He also afforded me applied research opportunities that undoubtedly facilitated my job placement at Pacific Science and Engineering. Rick has been a kind and willing mentor, who often discussed my research for hours at a time, even after leaving OU. I am very grateful to you, Rick.

Without the love and encouragement of my husband, Jesse Toftely, I would not be where I am today. Jesse you have been persistent research assistant, head chef, litter box cleaner, dog feeder, hand holder, Netflix watcher, laundry doer, foot rubber, shoulder to cry on, soft place to land, couch cuddler, compliment giver, loudest cheerleader, car fixer, supreme listener, biggest hugger, housekeeper, best friend, husband, and, most importantly, the love of my life. You did all of this for me while you worked two jobs to support us, pursued a second degree in chemical engineering, and were applying to dental school. Words are not enough to express my eternal appreciation for all you have done and continue to do. Thank you, baby. I love you.

I would also like to acknowledge my parents, Jerry and Terri Anderson, who encouraged me to never give up in my academic pursuits. They shaped me into the hardworking, self-confident woman that I have become. I am also grateful to my siblings, Jessica Anderson (impending rockstar), Jenny Anderson (future laywer), and Jack Cavanaugh (present-day engineer), for always cheering me on. My grandparents, Dan and LaVonne Cavanaugh, also deserve special recognition for their lifelong love and support. I have been blessed to have them in my life and they have taught me some important life lessons; such as, my liver is green, cookies are an acceptable breakfast food, and the Irish <u>never</u> give up. I love you all and thank you.

I would also like to recognize my dear friends Kelsey Suther, Cassandra Cross, Dr. Sara Bozeman, Dr. Stacy Wetmore, and Dr. April Martin. Your late night talks and wine-fueled girls nights help me survive graduate school. I could not have done it without you.

I would also like to thank my friends, Pamela and Bruce Byers, at Gallery Farm for giving me the opportunity to ride horses and work with my hands. Being at your stable were some of the only occasions where I could truly clear my head. I would have gone insane long ago were it not for the long rides on a demanding horse, red dirt up to my knees, sweat dripping down my face, and leaving your farm smelling of horses. You are two of the kindest, most genuine, and hardworking people I know. Thank you both.

Lastly, I would like to thank my pets. I would like to recognize my pug, Lucy in the Sky with Diamonds Kisses Toftely. I would have never made it through the six months away from your Dad, needed to finish this work, without the love and joy you brought to my life. You are a clown, a cuddle bug, a loaf of bread, and the smartest pug in the world (as proven by your undefeated, consecutive best trick wins at Puggerfest OKC). I love you. Last, but not least, I have to acknowledge my cat, Okie Bojangles Kisses Toftely, for tolerating my presence and providing companionship to your Dad and best friend, especially the six months we were apart. I love you too.

Table of Contents

List of Tables

List of Figures

Abstract

In eyewitness research, presenting lineup photos sequentially (one at a time) is thought to be superior to presenting lineup photos simultaneously (all at once); however, recent research has identified a robust simultaneous lineup advantage. This simultaneous advantage has been explained by the feature-detection hypothesis, which posits that simultaneous presentation allows witnesses to discover that particular facial features are non-diagnostic of the perpetrator (i.e., features shared by all lineup members) because lineup members can be directly compared (Wixted & Mickes, 2014). In two experiments, the feature-detection hypothesis was examined with a load manipulation (Experiment 1) and using eye tracking measures (Experiment 2). Although there was a significant simultaneous advantage in both experiments, support for the feature-detection hypothesis was mixed. In Experiment 1, a load manipulation designed to interfere with the processes described by the feature-detection hypothesis, was ineffective, possibly due to poor performing lineups or to the unique processing inherent to faces. In Experiment 2, eye movements, as predicted by the feature-detection hypothesis, was expected to change as a participant focused in on useful features. Eye movements changed as predicted in the simultaneous lineup, but not in the sequential lineup. More work is needed to examine the plausibility of the feature-detection hypothesis. Several alternative experiments are discussed.

Keywords: Eyewitness Identification, Lineups, Feature-Detection Hypothesis, Eye Tracking, Visual Working Memory

Introduction

Faulty eyewitness identification is the primary cause of wrongful convictions in the criminal justice system (see http://www.innocenceproject.org). The goal of eyewitness identification research thus far has been to identify conditions under which the accuracy of identifications is improved, with a major focus of research being on lineup presentation methods. Despite over years 30 of research, only two methods of lineup presentation and one dominant theoretical account have been emphasized in the literature. Wells, Memon, and Penrod (2006) suggest that research in eyewitness identification has been too narrow in its focus on traditional lineup procedures (i.e., placing a suspect among fillers) and pose the question: What if the traditional lineup never existed and instead the criminal justice system asked psychologists to develop methods to extract information from witnesses? They propose that, deprived of an emphasis on the traditional lineup presentation procedures, modern psychology may have developed different methods of identification (e.g., brain-activity measures, eye movements, rapid displays of faces, reaction times, etc.) than are employed today.

In the spirit of departing from the traditional behavioral methods, the goal of my dissertation is to assess eye movements via eyetracking and a load manipulation in order to better understand underlying memory and decision processes involved in eyewitness identification. Specifically, I propose a set of studies to assess a new theory of lineup identification as well as examine the relationship between eye movements and accuracy. A better understanding of the memory and decision processes that underlie lineup identifications should lead to better theoretical accounts and, eventually, more effective methods of identification.

As mentioned, two methods of lineup presentation have been the major focus of experiment. In the sequential lineup, witnesses view suspects one-at-a-time and make a yes/no decision about the current lineup member before seeing the next one. Alternatively, in the simultaneous presentation method, all the lineup members are presented together and the eyewitness is asked if the perpetrator is present in the lineup (for a review see Gronlund, Andersen, & Perry, 2013). Lindsay and Wells (1985) found that presenting lineups sequentially rather than simultaneously reduced the false identification rate for innocent suspects without significantly decreasing the ability of the witness to correctly identify the guilty suspect.

Simultaneous vs. Sequential Lineups

The dominant theory for the existence of a sequential lineup advantage has been the decision strategies account. This account suggests that in a simultaneous lineup, witnesses tend to use a relative decision strategy in which they are inclined to choose the person that looks most like the perpetrator (i.e., adopt a more liberal response criterion). Of course, this is problematic if the police have an innocent suspect who resembles the actual perpetrator. Conversely, participants tend to adopt an absolute decision strategy in a sequential lineup, whereby witnesses compare each lineup member to their memory for the perpetrator. However, Wixted and Mickes (2014) recently proposed that the decisions strategies account is a theory that makes prediction about response bias or willingness to choose, and not about discriminability.

Indicators of accuracy in lineup research have traditionally focused on correct versus false identification rate. The problem with assessing lineup performance by analyzing the correct identification rate and false identification rate (or calculating some combined measure) is that these identification rates are influenced by a witnesses' willingness to choose. Thus, accuracy is confounded with response bias. In order to separate accuracy from willingness to choose, researchers have argued that these traditional measures be replaced by signal detection theory assessments such as Receiver Operating Characteristic (ROC) analysis (Gronlund, Carlson, Neuschatz, Goodsell, Wetmore, Wooten, & Graham, 2012, Gronlund, Wixted & Mickes, 2014; Wixted & Mickes, 2012). An ROC is a plot of correct and false identification rates across varying levels of bias or confidence.

According to Wixted and Mickes (2014), the decision strategies theory predicts that the simultaneous and sequential lineups would fall on the same ROC curve (i.e., have the same discriminability). However, Wixted and Mickes (2012) and Gronlund et al. (2012) have found that the simultaneous lineup ROC curve falls higher than the sequential lineup in ROC space, suggesting that the simultaneous lineup results in better discriminability (i.e. more correct identifications and fewer false identifications for a given level of response bias) than sequential lineups. Consequently, in light of current findings, the decision strategies theory is no longer a satisfactory explanation of lineup identification performance.

Instead of a theory of response bias, researchers need a theory of discriminability to explain the differences in performance across lineup presentation methods. Not only does the data demand an alternative explanation, but a theory of response bias, such as the decision strategies account, is less useful. Response bias can easily be understood and manipulated, whereas accounting for why a particular procedure may lead to differences in discriminability is a much more interesting scientific question. As a result, Wixted and Mickes (2014) recently proposed the feature-detection hypothesis to account for differences in discriminability across lineup procedures.

The Feature-Detection Hypothesis

The feature-detection hypothesis posits that, because lineup members can be directly compared in simultaneous lineup, it allows witnesses to discover that particular facial features are non-diagnostic of the perpetrator (i.e., features shared by all lineup members). Discrimination of the guilty from innocent suspect is improved when non-useful features are disregarded and useful features are emphasized. For instance, consider a description of a perpetrator that includes age, gender, and race. The police are likely to create a lineup containing the suspect and foils that match this description so that all lineup members are the same age, gender, and race as the suspect. Consequently, age, race, and gender are non-useful features because all lineup members share those features. Features that are perpetrator-specific, such as perhaps eye shape and nose size, likely were not mentioned by the witness and therefore do not serve as the basis for the selection of foils. Perpetrator-specific features are less likely to be shared by foils and therefore are more useful for discrimination.

There is prior empirical support for the claim that discriminability is a function of the extent to which participants can identify and utilize the most useful features. Directing participants' visual attention to features useful for discrimination can reduce and even eliminate the own-race face bias. The own-race face bias is the finding that individuals are more accurate and faster at recognizing faces of their own-race compared to other-race faces. Hills, Cooper, and Pake (2013) guided White and Black participants to either the eyes (useful feature for discriminating White faces) or to the tip of the nose (useful feature for discriminating Black faces) via a fixation cross that preceded studied and subsequently tested faces. When the first fixation was directed to the nose, Black faces were better recognized, while a first fixation directed to the eyes predicted improved White face discrimination. The improvement of race discrimination, using the feature fixation method, was found regardless of participant race. When no fixation crosses were present, White participants oriented their first fixation to the eye region and Black participants oriented their first fixation to the nose region. The finding that participants change their eye movements as a result of their own race suggests that eyes versus noses offer differential diagnostic utility depending on face race. Generally, the results of this experiment suggests that diagnostic features differ as a function of race, and that discriminability is a function of the extent to which participants can identify and utilize the most useful features, a basic assumption of the feature detection hypothesis.

In the simultaneous lineup, because all lineup members are viewed at once, useful and non-useful features can be more readily identified (e.g., witness can immediately notice that all lineup members are the same race and age and disregard that information). The ability to discriminate the guilty from innocent suspect (and foils) is a function of the extent to which useful features are identified and non-useful features are discounted. In the sequential lineup, non-useful features are not readily apparent because witnesses view faces one-at-a-time. However, the diagnostic feature-detection hypothesis would predict that when the suspect comes late in a sequential lineup, discrimination could be enhanced, as witnesses will have had time to identify useful and non-useful features.

Gronlund et al. (2012) found support for this prediction using ROC analysis to examine the impact of suspect position on performance. When the innocent or guilty suspect was placed early (position 2) into a sequential lineup, performance was significantly worse than the simultaneous lineup and equivalent to a single-suspect lineup or showup. However, when the suspect was placed late in the lineup (position 5) in the sequential lineup, performance was equal to that of simultaneous lineup performance. The finding that later suspect positions produce better discriminability is in line with the feature-detection hypothesis. As the sequential lineup unfolds, useful cues can be identified, which improves accuracy.

Present Experiments

Two experiments were conducted in order to empirically test the plausibility of the feature-detection hypothesis. The first experiment examines lineup identification performance while participants are engaging in a secondary visual working memory task (i.e., modified symmetry span task). Because the feature detection hypothesis is describe as an active process requiring witnesses' to maintain facial features in working memory, it follows that interrupting that process with a secondary task will affect lineup identification performance.

The second experiment employed the use of an eye tracker to assess how witnesses are interacting with facial information while making an identification. Examining the eye movement behavior of participants viewing a lineup allows for a test of the feature-detection hypothesis by examining how participants distribute their fixations. For example, early in the sequential lineup participants may distribute their fixations across a wider set of features on a face, because participants will not yet have identified useful features, resulting in more overall fixations. However, as the sequential lineup unfolds, participants should begin to identify useful features. As a result, participants will focus on a limited set of useful features, thereby making fewer fixations. In the simultaneous lineup, on the other hand, fixations should be greater in number on initial viewing of lineup members, but later visits to faces should focus increasingly on useful features that aid identification and thereby result in fewer overall fixations.

In addition to identifying potential features useful for discrimination, eyetracking can provide insights into how eye movement behavior can change as a result of experience. Heisz and Shore (2008) examined how eye movement patterns changed as a function of the familiarity of faces. Over three days, participants were asked to study novel faces paired with a name and then recall the name of studied faces in a subsequent test. Each day participants studied and were tested on 10 new faces along with faces learned on the previous day, such that the 10 faces studied on the first day were studied in two additional sessions. On the fourth day, participants performed an old/new recognition test and a recall test for the faces learned on the previous days, as well as for 10 novel faces. Heisz and Shore predicted that as familiarity increased (i.e., faces studied on all three days), there would be a shift in face scanning behavior. They found that in both recall and recognition tasks, performance accuracy improved with experience. Interestingly, this marked increase in performance with experience was supported by a decrease in the average number of fixations to faces that were familiar. Additionally, they found that with increased familiarity, participants made fewer fixations, sampled more from the eyes, and sampled less from the nose, mouth, forehead, chin, and cheek regions. But this change in scanning behavior was more apparent for tasks that required explicit recollection (i.e., recall) than recognition. Similar to predictions made by the feature-detection hypothesis, Heisz and Shore's findings suggest that eye movement patterns change as a function of experience. Specifically, that participants may make fewer fixations as they become familiar with lineup members, which may result from participants focusing in on useful features.

In addition to using eye movements to examine the feature-detection hypothesis, another goal of the second experiment, is to improve on and replicate previous eyetracking work in eyewitness identification. Only three studies have examined lineup presentation and eye movements (Flowe, 2011; Flowe & Cottrell, 2010; Mansour, Lindsay, Brewer, & Munhall, 2009), and only one of those examined both simultaneous and sequential lineups (Flowe, 2011). As this experiment is focused on eye movements in both simultaneous and sequential lineups, the results of this experiment will be compared to the findings in Flowe (2011) in order to determine the robustness of eye movements findings in eyewitness identification.

In Flowe (2011), participants studied 12 computer-generated faces that served as perpetrators in the subsequent lineup task. After a short delay, participants viewed either 12 simultaneous or sequential lineups. Half were target-present (contained the studied perpetrator), and half were target-absent (contained a look-alike). Both target-present and absent lineups shared the same foils, but were presented in separate randomized blocks. Eye movements were assessed in both the encoding and test phases of the experiment. Flowe compared dwell time[1] for foils and suspects in simultaneous and sequential lineups for suspect identification, foil identification, and lineup rejection trials. In general, she found that faces in the simultaneous lineup were examined for a shorter length of time than faces in the sequential lineup, and that foils were dwelled on for a shorter length of time than suspects. Flowe interpreted her results to suggest that face processing was more thorough in the sequential lineup, despite a lower *d*' than simultaneous lineups. From a feature-detection perspective, longer dwell times in the sequential lineup instead suggests an inability to identify useful features, which would explain the poorer identification performance. However, we

[1] Dwell time is calculated by summing up the total length of time spent fixating within a restricted area region.

should be cautious regarding these data because lineup performance was nearly at chance and the stimuli were not real faces.

Experiment 1 and 2 Predictions

I conducted two experiments designed to test predictions made by the feature detection hypothesis. In the first experiment, I manipulated lineup presentation method (simultaneous and sequential), suspect position (early or late), and presence of a visual working memory task (no load or load). The presence of a load manipulation tests the feature-detection hypothesis by requiring participants to hold information in visual working memory while engaged in a lineup identification task. If the feature detection process demands visual working memory in order to identify and maintain featural comparisons across faces, it follows that lineup identification performance should suffer. If the feature-detection hypothesis holds, there will be poorer accuracy across simultaneous and sequential lineup procedures under load, plus a lack of suspect position effects in the sequential lineup because participants will not be able to identify useful features as the lineup unfolds.

In the second experiment, I again manipulated lineup presentation method (simultaneous and sequential) and suspect position (early or late), and added eyetracking to evaluate the feature-detection hypothesis. In order to ensure better performance than Flowe (2011), I used real faces, and provided a better encoding opportunity via a longer presentation duration. In the sequential lineup, I predicted that early in the sequential lineup, participants would distribute their fixations across several features, resulting in more fixations. However, as the lineup unfolds, participants should begin to identify diagnostic features. As a result, they should spend time on fewer features, and as a consequence make fewer fixations, in both target-present and target-absent lineups. In the simultaneous lineup, I predicted that there will be more fixations for the initially examined faces. However, lineup members examined subsequently will engender fewer fixations in both target-present and target-absent lineups.

Method: Load Manipulation Experiment

Design

A 2 Lineup Type (simultaneous or sequential) x 2 Presence of Target (target-present or target-absent) x 2 Target Position (guilty/innocent positions 2,3 or guilty/innocent positions 4,5) x 3 Face Type (guilty suspect, innocent suspect, or foil) design was manipulated within subjects.

Participants

A total of 126 undergraduates at the University of Oklahoma participated in the experiment. All procedures were approved by the IRB of the University of Oklahoma. Participants gave their informed consent, as needed, according to the university IRB policy, and received course credit for completing the experiment.

Materials

The stimuli used in this experiment all served as stimuli in previous eyewitness identification studies. A total of 10 guilty suspects were selected from the pool of materials, which included a mock crime video and fair lineups.

Choosing materials that have been previously tested ensured that assembled lineups have been validated as fair and resulted in adequate levels of performance. Criteria for selection of the stimuli included materials that contain a perpetrator, five description-matched foil faces, and an innocent look-alike that matched the description of the guilty suspects.

The experiment consisted of a study task and a lineup test task. In the encoding task, participants studied photographs from a mock crime. In order to create a distinct memory of the guilty suspects, three screen shots from the video (i.e., one of the scene of the crime and two of the perpetrator's face) were used as to-be-studied stimuli. In the lineup test task, participants were asked to identify studied guilty suspects from a lineup. Target-present lineups were created using the guilty suspects and five foils faces; the target-absent lineups were created by replacing the guilty suspects with the innocent look-alike.

The experiment was broken down into two blocks to reduce the number of guilty suspects participants were required to remember. Half the guilty suspects were studied and tested in block one and the other half were studied and tested in block two. Each study-test block consisted of five target faces (each consisting of a set of three photos), a 3-minute distractor task, and then 10 lineup tests of the previously studied faces. Five of the lineups were target-present (contained a studied target) and the other half were target-absent (contained the innocent suspect). All targets tested as target-present among the first five tests were tested as target-absent lineup amidst the second five tests (and vice versa). Half of the guilty suspects and corresponding innocent look-alike appeared early in the lineup (positions 2 or 3), the other half appeared late (positions 4 or 5) (See Figure 1).

There were two methods of lineup presentation: simultaneous and sequential. In the simultaneous lineup, all of the members were presented at once in a 3x2 array. For example, the first lineup member was displayed in the upper left hand corner of the screen, the second lineup member was placed in the middle-upper row, and so on. A number appeared under each photo of a lineup member, and participants made a decision by pressing a labeled key (i.e., 1-6) on the computer keyboard that corresponded to the face they were choosing, or pressed the not present option (0). In the sequential lineup, the faces also were presented in a 3x2 array, but they appeared one-at-a-time and a decision was made for each face (Yes or No) before the next face appeared.

The load on visual working memory was manipulated via a modified version of Automated Symmetry SPAN (ASSPAN) (Shah & Miyake, 1996). The ASSPAN is a visual working memory task where participants make a symmetry decision, while maintaining spatial information in working memory. The ASSPAN task proceeds as follows: A series of red squares is presented at various positions within a 4 by 4 grid, after which participants must decide whether positioning of the red squares is symmetrical across its horizontal axis. After the symmetry decision, participants signify where and when the red squares appeared and are provided feedback about their recall accuracy (e.g., 3 out of 4 squares recalled correctly). As a consequence, participants' ability to maintain the spatial locations is interfered with by performing the symmetry task.

The ASSPAN task was modified in order to differentially interfere with participants' ability to maintain information about lineup members in visual working memory. I modified this task by asking participants to remember the spatial location of either 1 (low load) or 4 (high load) squares within a 4 x 4 grid. Each square appeared for 1000 ms, disappeared, and a new square then appeared in a novel location. Instead of a symmetry task, like in the ASSPAN, participants were asked to make an identification from a lineup. Following their identification, participants were asked to signify where and when the red squares appeared, and provided accuracy feedback before studying a novel set of square locations.

Procedure

Participants began by completing a short practice trial. In the experiment, participants viewed a set of three photos for each of five perpetrators. Each photo in the set was presented for 5 seconds. After each set of photos, participants were asked to press the space bar when they were ready to view the next set of photos. They were told to try to remember the crime and the corresponding perpetrator because they would be tested on it later. The study phase was followed by a three-minute distractor task where participants were asked to engage in a visual word search on the computer screen.

Prior to the test phase, participants were instructed to imagine that the faces of people they studied went on to commit a crime, and as an eyewitness it is their job to pick them out a lineup. They were explicitly instructed that the target may or not be present in the lineup. Because participants were presented with both a target-present and target-absent lineups containing the same foils, they were instructed that the perpetrator might appear once, twice, or not at all, across the different lineups. In the sequential lineup condition, if participants choose a lineup member, they still got to view the rest of the faces in the lineup. This was done to equate the number of faces viewed in both lineup conditions. Participants were told that once they choose a face, they would not be allowed to change their decision or choose another face.

There were three parts to the test phase: load study, lineup test, and load recall (see Figure 2 for an example of the procedure). In the load study, participants were asked to study the spatial location of 1 or 4 square(s) in the modified ASSPAN task. Next, in the lineup test, participants were asked to make an identification from a single lineup. Before viewing each lineup, a slide appeared containing the crime label from the experiment phase (e.g., car jacking), indicating to participants which target they were to look for in the lineup. When they were ready to view the lineup, participants pressed the space bar.

Participants had as much time as they wanted to make an identification decision. Participants made a decision about faces in the simultaneous lineup by pressing a labeled key (i.e., 1-6) on the computer keyboard that corresponded to the face they were choosing, or they pressed "0" to indicate the suspect was not present. In the sequential lineup, participants were asked to make a yes/no response for each face in the lineup using the keyboard. After making a decision, participants reported their confidence on a Likert scale ranging from 1 (not confident at all) to 6 (extremely confident).

Following their lineup identification decision and confidence judgment, participants were asked to recall the spatial location(s) of the square(s) they had studied prior to the lineup identification. After selecting the location and order they believed the squares appeared in, participants were given feedback about the correctness of their recall and then repeated the test phase for a novel set of modified ASSPAN squares and an unviewed lineup. Participants repeated the test phase procedure until they had made identification decisions about each of the five studied perpetrators' target-absent and target-present lineups (i.e., 10 lineups). The five-study-10-lineup block was then repeated for the second set of perpetrators.

Results: Load Manipulation Experiment

Table 1 presents proportions of correct, false, foil identifications, rejection rates and *d'*. However, taken in isolation**,** correct and false identifications rates are not good indicators of performance because they are influenced by the willingness to choose from a lineup. Therefore, I will focus my discussion of lineup performance on analysis of the ROC curves. An ROC is a plot of correct (guilty suspect) to false (innocent suspect) identifications across varying levels of confidence (see Figure 3). The lower-left point reflects the suspect identifications made with the highest confidence. Each additional point extends the curve across the ROC space with a cumulative record of suspect identification rates across levels of confidence (i.e., the second point reflects the correct and false identifications made with the highest and second-highest confidence, the third point adds identifications from the third-highest confidence, and so forth). The farther away a curve is from the chance diagonal ($d' = 0$), the more accurate the identification procedure. By comparing the area under each curve (AUC), I can determine if one procedure results in significantly better discriminability than another. For lineup ROCs, I need to compute the partial area under the curve (pAUC). A pAUC must be computed because lineup ROCs do not extend over the entire probability space (0-1). In a lineup ROC, participants can choose foils and are not limited to selecting or rejecting a target or innocent suspect. As a result, even if the witness is choosing 100% of the time, the suspect may not be chosen 100% of the time, resulting in a truncated ROC curve.

When there are not enough data to construct a stable ROC curve, researchers compute *d'* values (*d'*= *z*(correct ID rate) − *z*(false ID rate). Relying on *d'* is better than assessing performance based on the ratio of hits to false alarms (e.g., probative value) as these ratios are influenced by a witnesses' willingness to choose (Mickes, Moreland, Clark, & Wixted, 2014)

Lineup Performance Analyses

The simultaneous and sequential lineup ROCs are presented in Figure 3. Using the statistical package *pROC* (Robin et al., 2011), I analyzed the pAUC for a false identification rate from 0 to .18, which subsumes the performance of both curves. The simultaneous pAUC (.16) was significantly greater than the sequential pAUC (.13) D = -2.79, p < .05. D is defined as (AUC1 − AUC2)/s, where s is the standard error of the difference between the two AUCs estimated by a bootstrap.

Load Analyses

Correct recall on the load task meant that participants correctly recalled the location and order of the studied squares following the lineup task. Incorrect recall was coded as any mistake either in location or order. The two levels of load induced different levels of recall performance such that participants were made more errors in the high load condition. Participants were correct in the high load condition 49.7% of the time and were correct in the low load condition 68.7% of the time. Load recall performance was similar across lineup presentation methods. In the sequential lineup high load condition, participants were correct 48.3% of the time; in the simultaneous lineup they were correct 51.2%. In the low load condition, participants were correct 67.3% in the sequential lineup and 68.1% in the simultaneous lineup.

Figure 4 shows the simultaneous and sequential ROC curves for each identification procedure broken down by high and low load conditions. There were no significant effects of load on lineup performance. There was no significant difference between sequential low (.127) and sequential high (.129) pAUCs , $D = .0100$, $p > .05$., and no difference between simultaneous high (.155) and simultaneous low (.156) pAUCs, $D = -.099$, $p > .05$. Overall, the simultaneous pAUCs were significantly greater than the sequential lineup pAUCs, irrespective of the load manipulation. The high load simultaneous pAUC (.155) was significantly greater than the high load sequential lineup pAUC (.129), $D = -1.86$, $p < .05$, the low load simultaneous pAUC (.156) was significantly greater than the high load sequential lineup pAUC (.129), $D = -2.00$, $p < .05$, the high load simultaneous pAUC (.155) was significantly greater than the low load sequential lineup pAUC (.127), $D = -1.91$, $p < .05$, and the low load simultaneous pAUC (.156) was significantly greater than the low load sequential lineup pAUC (.127), $D = -2.07$, $p < .05$. See Table 2 for the choosing rates and *d'* values as a function of the load manipulations.

The lack of an effect of load on lineup performance might indicate that participants abandoned the load task to focus on the lineup identification task. A stronger test of this hypothesis focuses on just those trials where participants correctly recalled the load items. Because there was insufficient data to construct stable ROC curves, lineup performance for correct load recall was assessed via *d*' analyses. When participants were correct on the load task, the simultaneous low load had a higher *d'* (.92) than the sequential lineup low load *d'* (.79); however, this difference was not significant, G= -.16, $p > .05$. Interestingly, under high load, the sequential lineup *d'* (.91) was slightly higher than the simultaneous lineup *d'* (.86), but again, this difference was not significant, G=.04, $p > .05$.

Suspect Position Analyses

Figure 5 shows the simultaneous and sequential ROC curves for each identification procedure, with the lineup data split by suspect position early (positions 2 and 3) and late (positions 4 and 5). Not surprisingly, there was no significant difference between the simultaneous early pAUC (.157) and the simultaneous late pAUC (.155), $D = .641$, $p > .05$. However, contrary to my predictions, sequential late lineup pAUC (.122) was not significantly different from the sequential early lineup pAUC (.134), $D = .99$, $p > .05$. In general, the simultaneous lineup performed significantly better than the sequential lineup regardless of suspect position.

Because there were no suspect position effects when collapsed over early versus late positions, the suspect position was further subdivided by each position (i.e., suspect in the 2nd ,3rd ,4th , or 5th position) . See Table 1 for choosing rates and *d'* for the suspect positions. Gronlund et al. (2009; Carlson et al., 2008) found a suspect position effect when the suspect was in the 2nd position versus the 5th position in the sequential lineup. The *d'* values were significantly different for the sequential position 2 (*d'*=1.12) and sequential position 5(*d'*= 0.32), G= -.38, *p*>.05, with suspect position 2 performing better than position 5. But this is contrary to what Gronlund et al. (2009) found.

Discussion: Load Manipulation Experiment

The goal of the first experiment was to test the feature-detection hypothesis via the presence of a load manipulation in which participants had to maintain visual information in working memory while engaged in a lineup identification task. The results indicated that there was not a negative effect on lineup performance (simultaneous or sequential) under high load, despite high load clearly being more difficult than low load. Why is it that the load manipulation did not affect lineup performance, and what does this result mean for the plausibility of the feature-detection hypothesis?

One reason there was no effect of load on lineup performance could be due to the low overall performance on lineups (a floor effect). Perhaps using fewer lineups in a multiple lineup paradigm, increasing the study duration, or presenting only a single lineup, would have improved performance and resulted in a detectible effect of load or suspect position. A second possibility is, that the use of a multiple lineup paradigm might have led participants to rely on familiarity, instead of recollection. Familiarity is thought to require few cognitive resources (Yonelinas, 2002); therefore, a load manipulation would have little effect on performance if familiarity dominates lineup decisions in this experiment. It is possible that the improved performance via the identification of useful features (as described by the feature-detection hypothesis) relies on contributions from a recollective process. More work is needed to understand the potential relationship between recollective processes and the feature-detection hypothesis.

Of course, another possibility is that the predictions made by the feature-detection hypothesis are incorrect. Perhaps visual working memory is not required, or is not overloaded, by the identification and maintenance of comparisons across faces. This might be the case due to people being experts at facial recognition (McKone, Kanwisher, & Duchaine, 2007). Experts in a particular domain are able to automate much of their processing. Utilizing less working memory capacity to execute a task allows participants to maintain more information (Lewandowsky, Little, & Kalish, 2007). For example, Chase and Simon (1973) demonstrated that chess experts could encode, store, and maintain complex chunks of pieces on a chessboard, and use that information to retrieve more information than novices. Curby and Gauthier (2007) demonstrated that visual short-term working memory capacity was greater for upright faces (more faces could be maintained) than for inverted faces due to the reliance on holistic processing of upright faces. Perhaps it is because humans are face experts that facial recognition ability is little affected by a load on visual working memory. Conceivably, if the lack of a load effect is due to expertise with faces, a face inversion manipulation may eliminate the expert face advantage, resulting in an increased sensitivity to a load manipulation.

Despite not observing an effect of load on lineup performance, an examination of eye movements provides an alternative method of assessing the predictions. Experiment 2 further evaluates the feature-detection hypotheses via the examination of eye movements.

Method: Experiment 2: Eyetracking Experiment

Design

A 2 Lineup Type (simultaneous or sequential) x 2 Presence of Target (target-present or target-absent) x 2 Target Position (guilty/innocent positions 2,3 or guilty/innocent positions 4,5) x 3 Face Type (guilty suspect, innocent suspect, or foil) design was manipulated within subjects.

Participants

A total of 40 undergraduates at the University of Oklahoma participated in the experiment. Participants had normal or corrected-to-normal vision. All procedures were approved by the IRB of the University of Oklahoma. Participants gave their informed consent according to the university IRB protocol, and received course credit for completing the experiment.

Stimuli and Apparatus

The same stimuli were used in this experiment as in Experiment 1. Eye movements were recorded via an EyeLink 1000 eyetracker (SR Research). Stimulus presentation and data recording were controlled via the Experiment Builder software. A keyboard was used to collect manual responses during the experiment.

Procedure

The procedure in Experiment 2 was the same as Experiment 1, except that the eye tracker was calibrated before beginning the experiment. In addition, a central fixation appeared before the beginning of each lineup test to recalibrate participants' eyes.

Results: Eyetracking Experiment

Due to the smaller sample size resulting from having to run participants one at a time**,** d' values ($d'= z$(correct ID rate) − z(false ID rate), instead of pAUCs, were compared to assess lineup performance and the suspect position effect.

Lineup Performance and Suspect Position Analyses

Discriminability from simultaneous lineup (d'=1.03) was significantly better than from the sequential lineup (d'=0.48), G=11.74, $p < .05$. Suspect position effects were assessed comparing sequential lineup position 2 to sequential lineup position 5. Suspect position 2 (d'=.65) discriminability was significantly higher than the sequential lineup position 5 (d'=.22), G= - 3.77, p < .05, opposite of the predicted effect.

Eye Movement Behavior: Feature-Detection Hypothesis.

Each face in a lineup served as its own interest area, and several measures were derived from the eye movement data. The main dependent variable was the number of fixations made to an interest area. With this variable, I could assess how fixation behavior changed over the course of making a lineup identification. The number of fixations was calculated by summing the total number of fixations made within an interest area.

In order to assess how fixation behavior changed over time in the simultaneous lineup, the number of fixations for the first and last run was compared. A run is defined as a group of consecutive fixations that are directed towards the same interest area. In the simultaneous lineup there are six interest areas (i.e., faces). Imagine, for example, a participant made the following 10 fixations: Face1, Face1, Face2, Face1, Face1, Face1, Face3, Face1, Face4, Face5. There are seven runs is this example (because fixations to the same face belong to the same run). The average number of fixations for the first run can be compared to the average number of fixations in the last run to assess how fixation behavior to an interest area changed over time. Thus, in the above example, Face1 had two fixations in the first run, but one fixation in the last run.

In the simultaneous lineup, I predicted that there would be more fixations for the initial runs to faces, but lineup members examined on subsequent runs would engender fewer fixations, irrespective of the eventual identification decisions. The total number of fixations made to faces was then separately averaged across lineups for each participant, conditioning the data on run (first or last). For each participant, the average number of first run fixations was compared to average number of last run fixations in simultaneous lineups. In the simultaneous lineup, average lineup fixations for the first run (M=11.15, SD=1.42) were significantly higher than the last run (M=8.40, SD=2.45), $t(39)=7.38$ $p < .05$, supporting the prediction that faces examined initially received more fixations.

In the sequential lineup, I predicted that guilty suspects viewed early in the sequential lineup (positions 2 or 3 would result in more fixations because participants have not identified what features are useful. However, as the lineup unfolds, participants can begin to identify useful features. As a result, participants should narrow in on useful features and make fewer fixations later in the lineup. In order to assess how eye movements changed as the sequential lineup unfolded, the average number of fixations to a guilty suspect in positions 2, 3, 4, and 5 were compared using a one-way repeated measures ANOVA. The number of fixations to the suspect was not significantly different across suspect positions, Wilks' Lambda = .874, $F(3,37) = 1.78$, $p > .05$. Despite not reaching significance, there was a trend toward more fixations for suspects in position 2. See Figure 6 for means and standard errors.

Eye Movement Behavior: Lineup Performance

A secondary goal of Experiment 2 was to attempt to build on and replicate the limited previous work examining lineup performance using an eyetracker. Only one study by Flowe (2011) examined the relationship between eye movements and lineup performance. Flowe's study focused on face dwell time instead of fixations. However, average dwell time was strongly correlated with the average number of fixations to interest areas $r(40)=.53$, $p < .01$. As a result, the analyses presented will focus on the number of fixations to an interest area.

In an attempt to replicate Flowe's (2011) findings, interest area fixations were averaged across lineups for each participant, conditioning the data on face type (suspect or foil) and identification outcome (suspect/target identification, foil identification or lineup rejection). Interest area fixation data were then submitted to a 2 (lineup procedure) x 2 (face type) within-subjects repeated measures ANOVA, and separate analyses were conducted for trials in which the suspect was identified, trials in which a foil was identified, and trials in which the lineup was rejected.

For suspect identification trials, there was a significant main effect of lineup, Wilks' Lambda = .741 $F(1,35) = 12.23$, $p<.05$, and face type, Wilks' Lambda = .484, $F(1,35) = 37.36$, $p <.05$., and no significant interaction between these variables. Follow-up comparisons indicated that suspects in the sequential lineup received significantly more fixations than foils, $t(36)=3.75$ $p < .05$, as did suspects in the simultaneous lineup, $t(38)=5.17$ $p < .05$ (see Figure 7). Lineup rejection trials resulted in a significant main effect of face type only, Wilks' Lambda = .870 $F(1,35) = 5.25$, $p <.05$, with suspects receiving more fixations than foils (see Figure 8). In foil identification trials, there was no significant main effect found for lineup, Wilks' Lambda = .996 $F(1,34) = .14$, $p >.05$ or face type Wilks' Lambda = .99, $F(1,34) = .039$, $p >.05$ (see Figure 9).

Discussion: Eyetracking Experiment

The goal of Experiment 2 was to evaluate the feature-detection hypothesis by examining the eye movement behavior of participants viewing a lineup. Support for the feature-detection hypothesis was mixed. In the simultaneous lineup, average fixations to faces in the first run were reliably greater than in the last run, suggesting that participants were potentially narrowing in on useful features. In the sequential lineup, there was a trend toward more fixations to the suspect if he or she was in the second position compared to later positions; however, this effect was not reliable. The finding trended in the predicted direction, but the multiple lineup paradigm used in this experiment did not induce suspect position effects, which might be why there was no effect on eye movements. More work is needed to examine the conditions under which suspect position effects appear and whether eye movements changes when reliable suspect position effects are found. One really cannot evaluate if participants are narrowing in on a set of useful features if there is little evidence of learning as the sequential lineup unfolds.

As in Experiment 1, participants may have been relying on familiarity to make their lineup IDs. Evidence supporting familiarity-based processing comes from the quick examination of lineup members, an average of around 1000ms each, suggesting that participants may not have examined lineup members thoroughly enough to recollect information. If identification performance was improved through better encoding at study, the eye movements findings may become more pronounced. Heisz and Shore (2008) demonstrated that eye behavior, specifically the narrowing in on facial features, was more pronounced for facial memory tests that required more recollection. Improved encoding, or presenting a single lineup, would allow participants to potentially access useful features, resulting in a stronger test of the feature-detection hypothesis.

A secondary goal of this project was to attempt to replicate previous work examining lineup performance using an eyetracker (e.g., Flowe, 2011). Although Flowe (2011) examined dwell time, recall that it was strongly correlated with the number of fixations, which were assessed in this experiment. For suspect identification trials, Flowe found that faces in the simultaneous lineup were examined for a shorter length of time compared to the sequential lineup, and that suspects were dwelled on longer than foils. I found a similar pattern of results: More fixations were made in the sequential lineup compared to the simultaneous lineup, and suspects received more fixations than foils. Flowe interpreted her results to suggest that face processing in the sequential lineup was more thorough. However, given the results of my experiment, and the poor discriminability from the sequential lineup, more fixations when making a suspect identification might actually have been the result of an inability to detect useful features. Consequently, these results may actually be consistent with predictions made by the feature-detection hypothesis; more fixations suggest participants have not identified useful features, and that is why performance was poor from the sequential lineup.

General Discussion

The goal of these experiments was to test the validity of the feature-detection hypothesis (Wixted & Mickes, 2014), by way of a load manipulation (Experiment 1), and by assessing eye-tracking measures (Experiment 2). Support for the feature-detection hypothesis was mixed. In Experiment 1, the prediction that load would have an effect on lineup identification performance was not supported, potentially due to the unique nature of facial recognition or poor performance emanating from the multiple lineup paradigm. In Experiment 2, in the suspects viewed early in the sequential lineup did not receive reliably more fixations than suspects that appeared late in the sequential lineup. However, the simultaneous lineup participants made more fixations to lineup members on the first run than on the last run, as predicted.

In both Experiments 1 and 2, a multiple lineup paradigm was used, potentially contributing to the poor performance and thereby providing a weak test of the feature-detection hypothesis. Studying multiple lineup members before making any identifications may weaken memory, thereby making it difficult for participants to identify a perpetrator. A stronger memory for perpetrators may be necessary before participants are able to segregate useful from non-useful features to aid identification. Future work should consider examining the feature-detection hypothesis using a traditional single lineup paradigm or a modified multiple lineup paradigm. Researchers also might explore how memory quality influences the processes described in the feature-detection hypothesis by varying presentation time or the number of repetitions of the perpetrator.

Additionally, contrary to past research (Gronlund et al., 2012; Carlson et al., 2008), no suspect position effects were obtained, as was predicted by the feature-detection hypothesis. The lack of suspect position effects also could be due to use of the multiple lineup paradigm. In the multiple lineup paradigm, the large number of faces a participant views could interfere with the learning processes described by Goodsell, Gronlund, and Carlson (2010) and predicted by the feature-detection hypothesis. More work is needed to examine the robustness of the suspect position effect.

Of note, in both experiments, a significant simultaneous discriminability advantage was found. The discovery of two simultaneous advantages, using a non-traditional paradigm, adds important support to the growing number of studies that find a significant simultaneous lineup advantage. A recent National Academy of Science report (2014) concluded, "there should be no debate about the value of greater discriminability – to promote a lineup procedure that yields less discriminability would be akin to advocating that the lineup be performed in dim instead of bright light"(p. 80). The simultaneous lineup appears to provide greater discriminability. But it is important that future eyewitness reform recommendations be supported by theory (McQuiston-Surrett, Malpass, & Tredoux, 2006), which is why it is important to test the feature-detection hypothesis as an explanation for the SIM advantage.

More work is needed to explore explanations for the underlying cognitive mechanisms involved in eyewitness identification. The decision strategies theory was the leading explanation for the sequential lineup advantage for nearly 30 years, and little literature challenged the dominant account. The recent discovery of a robust simultaneous advantage, using ROC analysis, demanded a new theoretical explanation and as a result produced the feature-detection hypothesis. These experiments are the first to offer a test of this hypothesis, and provides a starting point for researchers to begin testing, not only the plausibility of the feature-detection hypothesis, but other possible explanations for the simultaneous advantage.

Alternative behavior methods, such as eye tracking, are powerful tools to test theoretical questions. The predictions that can be made using eye movements as a dependent variable are vast, and have the potential to lead to a greater understanding of identification processes than choosing rates alone. If researchers can identify the underlying mechanisms that influence lineup identification accuracy, perhaps improved methods can be designed. For example, Hannula et al. (2012) demonstrated that eye movements could distinguish between correct and incorrect decisions when identifying studied faces from highly similar morphs, even when participants explicit identification responses were incorrect. Perhaps future identification procedures could employ eye movements measures to identify who and if a witness should choose, instead of relying on the often error-prone human memory system.

References

Andersen, S. M., Carlson, C. A., Carlson, M. A., & Gronlund, S. D. (2014). Individual differences predict eyewitness identification performance. *Personality and Individual Differences*, *60*, 36-40.

Carlson, C. A., & Gronlund, S. D. (2011). Searching for the sequential line-up advantage: A distinctiveness explanation. *Memory*, *19*(8), 916-929.

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 14(2),* 118-128. doi:10.1037/1076-898X.14.2.118

Carlson, C. A., & Gronlund, S. D. (2011). Searching for the sequential lineup advantage: A distinctiveness explanation. *Memory*, *19*, 916-929. doi:10.1080/09658211.2011.613846

Chase, W. G., & Simon. HA (1973). The mind's eye in chess. *Visual information processing*, 215-281.

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic bulletin & review*, *14*(4), 620-628.

Flowe, H. (2011). An exploration of eye movements in eyewitness identification tests. *Applied Cognitive Psychology*, *25*(2), 244-254.

Flowe, H., & Cottrell, G. W. (2011). An examination of simultaneous lineup identification decision processes using eye tracking. *Applied Cognitive Psychology*, *25*(3), 443-451.

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and human behavior*, *34*(6), 445.

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology*, *19*(1), 23-37. doi:10.1002/acp.1047

Gronlund, S. D., Andersen, S. M., & Perry, C. (2013). Presentation methods. B. Cutler (Ed.), *Reform of eyewitness identification procedures*, (pp. 113-138). Washington, DC: APA Publications. doi: 10.1037/14094-000

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

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating Eyewitness Identification Procedures Using Receiver Operating Characteristic Analysis. *Current Directions in Psychological Science*, *23*(1), 3-10.

Hannula, D. E., Baym, C. L., Warren, D. E., & Cohen, N. J. (2012). The Eyes Know Eye Movements as a Veridical Index of Memory. *Psychological science*,*23*(3), 278-287.

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision*, *8*(1), 1-10. doi: 10.1167/8.1.9.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: An eye-tracking experiment. *Visual Cognition*, *21*(7), 876-898.

Lewandowsky, S., Little, D. R. & Kalish, M. L. (2007). Knowledge and expertise. In F. T. Durso, R. Nickerson, S. Dumais, S. Lewandowsky, & T. Perfect (Eds.). Handbook of applied cognition, 2nd Ed. (pp. 83 - 110). Chicester: Wiley.

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*, 556–564. doi:10.1037/0021-9010.70.3.556

Mansour, J. K., Lindsay, R. C. L., Brewer, N., & Munhall, K. G. (2009). Characterizing visual behaviour in a lineup task. *Applied Cognitive Psychology*,*23*(7), 1012-1026.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces?. *Trends in cognitive sciences*, *11*(1), 8-15.

McQuiston-Surrett, D., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. Simultaneous Lineups: A Review of Methods, Data, and Theory. *Psychology, Public Policy, and Law*, *12*(2), 137.

Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition*, *33*(5), 783-792.

Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *The Journal of neuroscience*, *26*(43), 11187-11196.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12:77*. doi: 10.1186/1471-2105-12-77

Sauer, J. D., Brewer, N., & Weber, N. (2012). Using confidence ratings to identify a target among foils. *Journal of Applied Research in Memory and Cognition*, *1*(2), 80-88.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. Journal of Experimental Psychology: General, 125, 4–27.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence improving its probative value. *Psychological Science in the Public Interest*,*7*(2), 45-75.

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science*, *7*(3), 275-278.

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological review*, *121*(2), 262.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, *46*, 441-517. doi: 10.1006/jmla.2002.2864

Appendices

1 Study 1	2 Study 1	3 Study 1	4 Study 1
Block 1	**Block 1**	**Block 1**	**Block 1**
1st Tests	**1st Tests**	**1st Tests**	**1st Tests**
SIM TP 4	SEQ TP 4	SEQ TP 2	SIM TP 2
SIM TA 2	SEQ TA 2	SEQ TA 4	SIM TA 4
SIM TP 2	SEQ TP 2	SEQ TP 4	SIM TP 4
SIM TA 5	SEQ TA 5	SEQ TA 3	SIM TA 3
SIM TP 3	SEQ TP 3	SEQ TP 5	SIM TP 5
2nd Tests	**2nd Tests**	**2nd Tests**	**2nd Tests**
SIM TA 4	SEQ TA 4	SEQ TA 2	SIM TA 2
SIM TP 2	SEQ TP 2	SEQ TP 4	SIM TP 4
SIM TA 2	SEQ TA 2	SEQ TA 4	SIM TA 4
SIM TP 5	SEQ TP 5	SEQ TP 3	SIM TP 3
SIM TA 3	SEQ TA 3	SEQ TA 5	SIM TA 5
Study 2	Study 2	Study 2	Study 2
Block 2	**Block 2**	**Block 2**	**Block 2**
1st Tests	**1st Tests**	**1st Tests**	**1st Tests**
SEQ TP 4	SIM TP 4	SIM TP 2	SEQ TP 2
SEQ TA 3	SIM TA 3	SIM TA 5	SEQ TA 5
SEQ TP 2	SIM TP 2	SIM TP 4	SEQ TP 4
SEQ TA 5	SIM TA 5	SIM TA 3	SEQ TA 3
SEQ TP 5	SIM TP 5	SIM TP 3	SEQ TP 3
2nd Tests	**2nd Tests**	**2nd Tests**	**2nd Tests**
SEQ TA 4	SIM TA 4	SIM TA 2	SEQ TA 2
SEQ TP 3	SIM TP 3	SIM TP 3	SEQ TP 3
SEQ TA 2	SIM TA 2	SIM TA 4	SEQ TA 4
SEQ TP 5	SIM TP 5	SIM TP 5	SEQ TP 5
SEQ TA 5	SIM TA 5	SIM TA 5	SEQ TA 5

Figure 1. Counterbalancing of blocks and suspect position.

Note: Order of block tests were counterbalanced across participants and lineups were randomized within a block half.

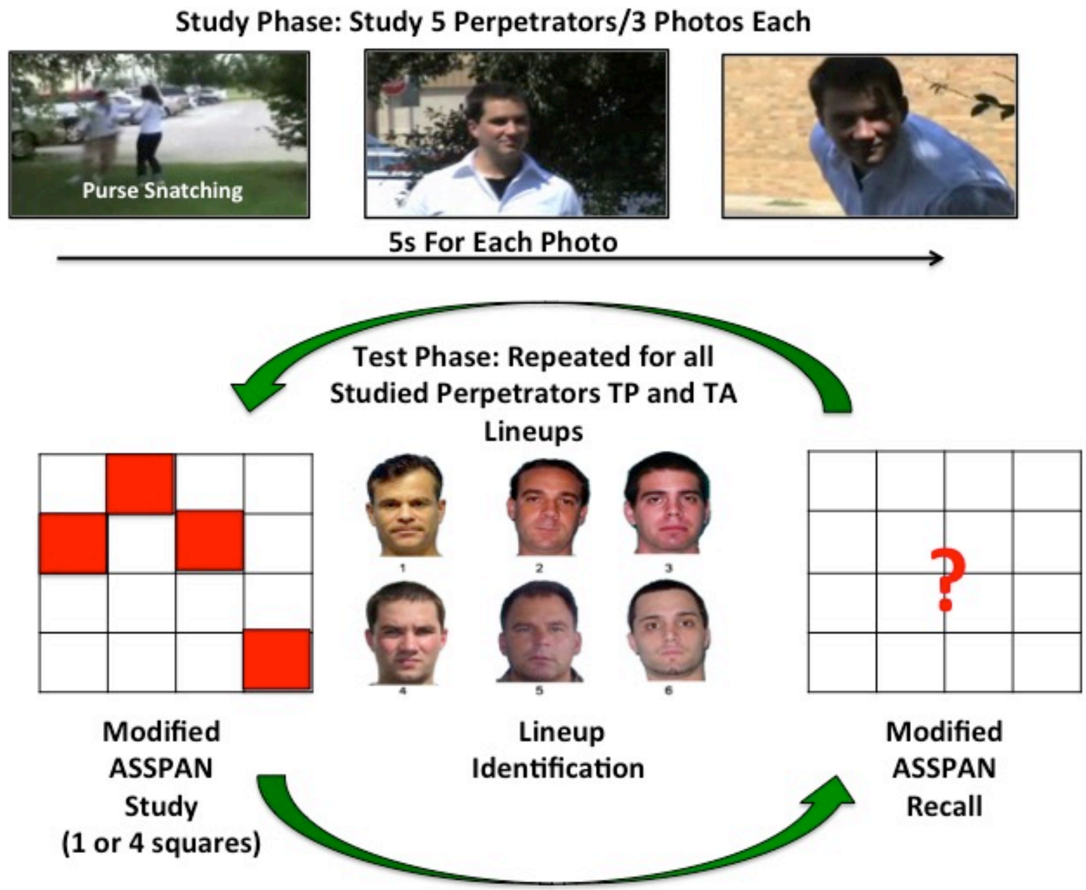

Figure 2. Load experiment study-test procedure.

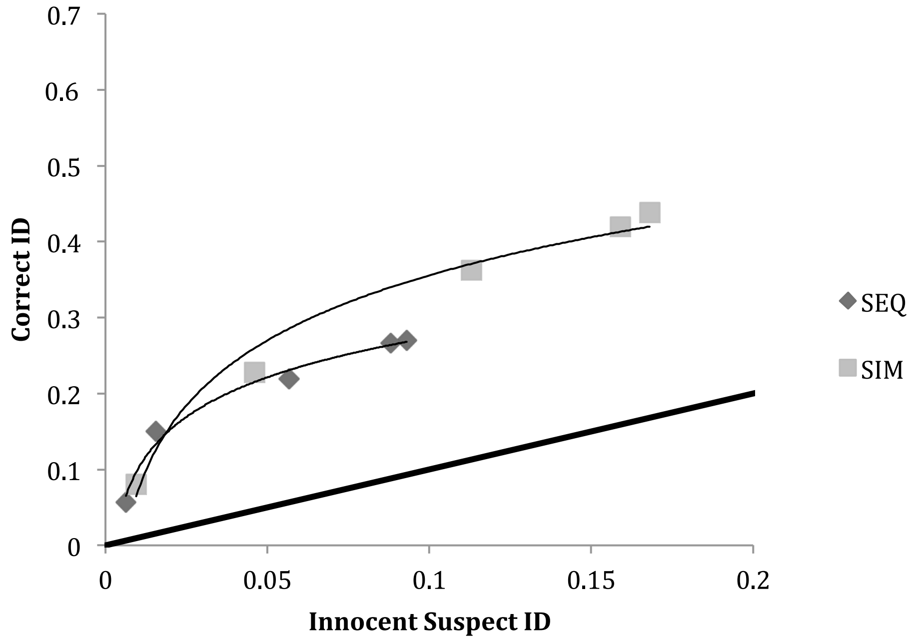

Figure 3. ROC curves comparing simultaneous (SIM) and sequential (SEQ) lineups. The solid line reflects the chance diagonal.

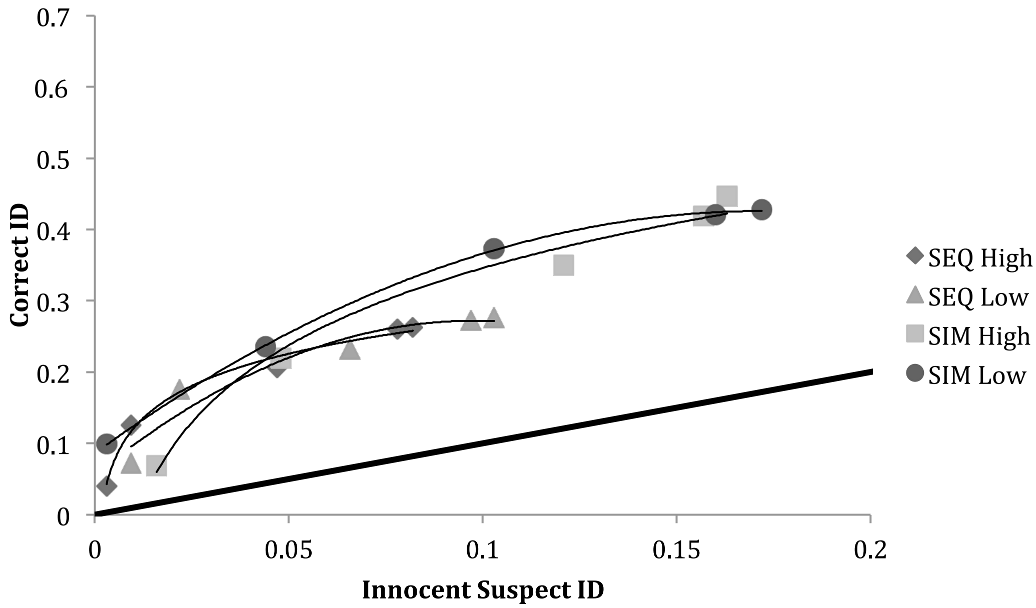

Figure 4. ROC curves depicting simultaneous (SIM) and sequential (SEQ) lineup performance under high and low load conditions.

Note: The solid line reflects the chance diagonal.

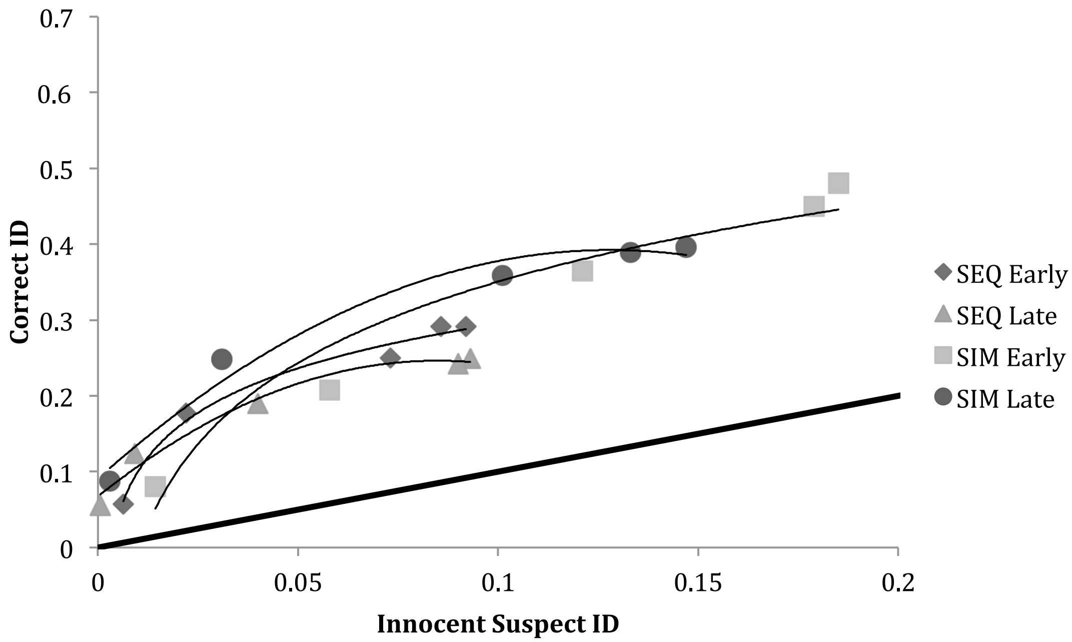

Figure 5. ROC curves comparing simultaneous (SIM) and sequential (SEQ) lineups for early and late suspect positions. The solid line reflects the chance diagonal.

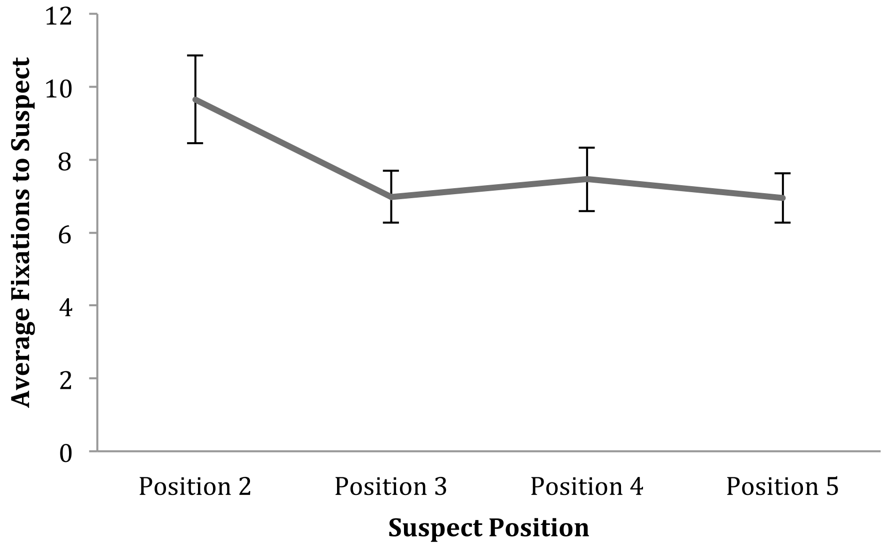

Figure 6. Means and standard errors for fixations to suspects across positions in the sequential lineup.

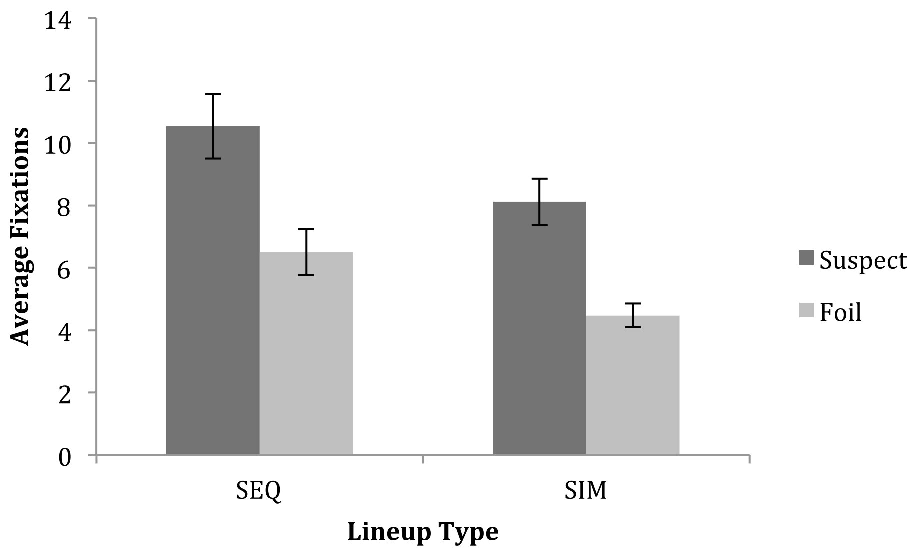

Figure 7. Mean fixations to interest areas by face type (foil or suspect) and lineup procedure (simultaneous or sequential) for suspect identification trials.

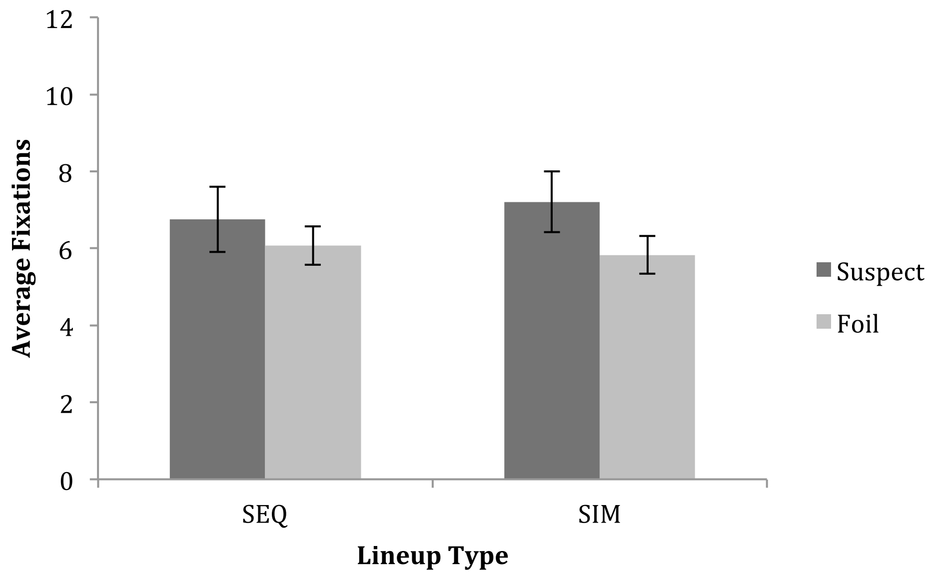

Figure 8. Mean fixations to interest areas by face type (foil or suspect) and lineup procedure (simultaneous or sequential) for lineup rejections.

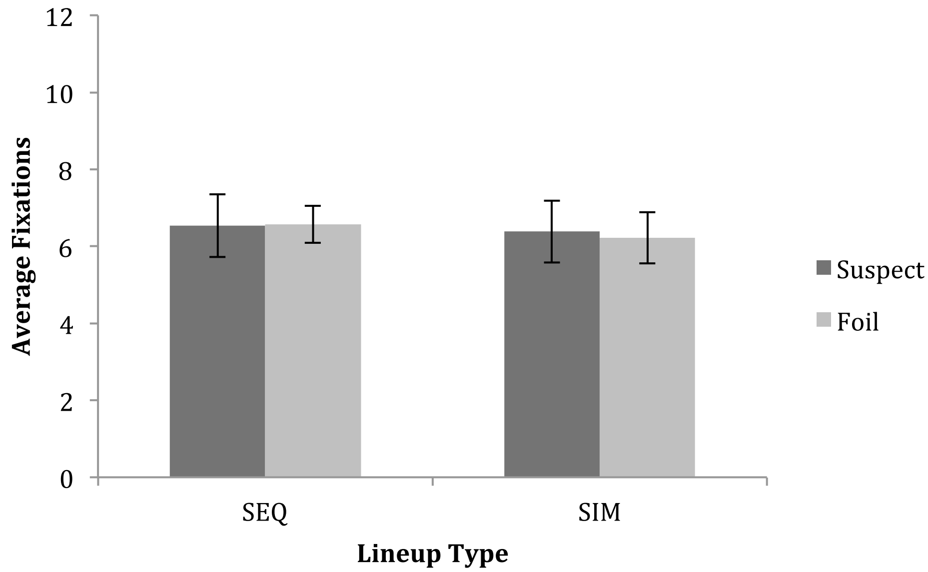

Figure 9. Mean fixations to interest areas by face type (foil or suspect) and lineup procedure (simultaneous or sequential) for foil identifications.

Table 1. *Overall and Suspect Position Performance Measures for Load Experiment*

	Presence of target and identification decisions					
	Target-present			Target-absent		
	Correct ID	Foil ID	Incorrect Rejection	False ID	Foil ID	Correct Rejection
Overall						
SIM	.44	.36	.20	.17	.42	.41
SEQ	.27	.41	.32	.09	.44	.47
Suspect Position						
SIM 2	.56	.24	.20	.19	.34.	.47
SIM 3	.35	.50	.15	.17	.50	.33
SIM 4	.51	.24	.25	.15	.39	.46
SIM 5	.23	.50	.17	.15	.53	.32
SEQ 2	.36	.27	.37	.07	.40	.53
SEQ 3	.19	.50	.31	.13	.44	.44
SEQ 4	.32	.37	.31	.10	.44	.46
SEQ 5	.14	.60	.26	.08	.51	.41
SEQ Early	.29	.36	.35	.09	.42	.49
SEQ Late	.25	.46	.29	.09	.47	.44
SIM Early	.48	.34	.18	.18	.40	.42
SIM Late	.40	.38	.22	.15	.45	.40

Note. SIM = simultaneous lineup, SEQ = sequential lineup

Table 2. *Lineup Performance as a Function of Load*

	Presence of target and identification decisions					
	Target-present			Target-absent		
	Correct ID	Foil ID	Incorrect Rejection	False ID	Foil ID	Correct Rejection
Load						
SIM High	.45	.35	.20	.16	.41	.43
SIM Low	.43	.38	.19	.17	.44	.39
SEQ High	.26	.43	.31	.08	.45	.47
SEQ Low	.28	.39	.33	.10	.44	.46
Correct/Incorrect Load						
SIM High Incorrect	.42	.40	.18	.15	.42	.43
SIM Low Incorrect	.33	.46	.21	.19	.50	.31
SEQ High Incorrect	.27	.46	.27	.10	.43	.47
SEQ Low Incorrect	.24	.38	.38	.12	.40	.48
SIM Low Correct	.47	.34	.18	.16	.42	.42
SIM High Correct	.48	.29	.23	.18	.39	.43
SEQ Low Correct	.29	.40	.31	.09	.46	.45
SEQ High Correct	.26	.39	.35	.06	.46.	.48

Note. SIM = simultaneous lineup, SEQ = sequential lineup.

Table 3. *Overall and Suspect Position Performance Measures for Eyetracking Experiment*

	Presence of target and identification decisions					
	Target-present			Target-absent		
	Correct ID	Foil ID	Incorrect Rejection	False ID	Foil ID	Correct Rejection
Overall						
SIM	.42	.30	.28	.11	.34	.55
SEQ	.33	.37	.30	.18	.36	.46
Suspect Position						
SIM 2	.46	.20	.34	.06	.24	.70
SIM 3	.30	.46	.24	.18	.33	.49
SIM 4	.48	.25	.27	.10	.37	.53
SIM 5	.37	.35	.28	.16	.46	.37
SEQ 2	.35	.31	.34	.15	.33	.52
SEQ 3	.19	.51	.30	.08	.43	.49
SEQ 4	.43	.28	.29	.23	.36	.41
SEQ 5	.29	.44	.27	.22	.37	.41

Note. SIM = simultaneous lineup, SEQ = sequential lineup