

Running head: VISUAL BEHAVIOR ON LINEUPS

**This is the peer reviewed version of the following article:

Mansour, J. K., Lindsay, R. C. L., Brewer, N., & Munhall, K. G. (2009). Characterizing visual behavior on a lineup task. *Applied Cognitive Psychology*, 23, 1012 - 1026. doi: 10.1002/acp.1570

which has been published in final form at <http://onlinelibrary.wiley.com/doi/10.1002/acp.1570/abstract>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

Characterizing Visual Behavior in a Lineup Task

Jamal K. Mansour^{1,2} and R. C. L. Lindsay¹

Queen's University

Neil Brewer³

Flinders University

K. G. Munhall^{1,2,4}

Queen's University

¹Department of Psychology, Queen's University; ²Queen's Biological Communication Centre, Queen's University; ³School of Psychology, Flinders University; ⁴Department of Otolaryngology, Queen's University.

Portions of this research were presented at the 2006 meeting of the American Psychology-Law Society and at the 2007 meeting Off the Witness Stand: Using Psychology in the Practice of Justice. This research was supported in part by grants from the Social Sciences and Humanities Research Council of Canada to Jamal Mansour (766-2005-0244) and Rod Lindsay (410-04-1298) and a grant from the National Institute on Deafness and other Communication Disorders (DC-00594) to K. G. Munhall.

Correspondence concerning this article should be addressed to Jamal K. Mansour, Department of Psychology, Queen's University, Kingston, Ontario, Canada, K7L 3N6; E-mail: 4jkm@queensu.ca.

Summary

Eye tracking was used to monitor participants' visual behavior while viewing lineups in order to determine whether gaze behavior predicted decision accuracy. Participants viewed taped crimes followed by simultaneous lineups. Participants ($N = 34$) viewed 4 target-present and 4 target-absent lineups. Decision time, number of fixations, and duration of fixations differed for selections versus non-selections. Correct and incorrect selections differed only in terms of comparison-type behavior involving the selected face. Correct and incorrect non-selections could be distinguished by decision time, number of fixations and duration of fixations on the target or most-attended face, and comparisons. Implications of visual behavior for judgment strategy (relative versus absolute) are discussed.

Characterizing Visual Behavior in a Lineup Task

In recent years methods for collecting eyewitness evidence have improved. Various efforts have been made to address the important issue of mistaken identification, including measuring confidence in one's decision and time taken to make a decision (Leippe & Eisenstadt, 2006; Sporer, 1993). Another effort toward this end is the study of strategies employed by eyewitnesses who view lineups (Dunning & Stern, 1994; Kneller, Memon, & Stevenage, 2001; Lindsay & Bellinger, 1999). Wells (1984) proposed that use of a strategy where witnesses compare lineup members and pick the member that most closely matches their memory of the criminal (a relative strategy) leads to a lower level of accuracy than one that leads witnesses to compare lineup members only to their memory of the criminal (an absolute strategy). The present manuscript adds to this literature on strategy by using a sensitive measure of real-time visual behavior – eye tracking.

Techniques to assess strategy use (such as characterizing decisions as automatic or deliberative, cf. Dunning & Stern, 1994) often rely on retrospective, self-report measures, which have many inherent problems (e.g., Nisbett & Wilson, 1977). Eye tracking may be an alternative that is well suited for addressing the eyewitness decision process. First, eye movements are an essential component of facial recognition. Henderson, Williams, and Falk (2005) showed empirically that participants recognized faces 1.28 times better when the faces were viewed freely rather than fixating a center point during the learning phase. Second, eye tracking provides real time information about the decision process because it involves monitoring participants' gaze while they view some event or complete some task. While cognitive processing need not be focused on the location of an individual's fixation, research shows that fixation on a target (overt attention) generally indicates processing of

that target (Findlay & Gilchrist, 2003). Finally, current eye tracking systems obtain a large amount of information in each trial and are much less cumbersome and more accurate than in the past (Collewyn, 1999). Facial recognition research commonly uses eye tracking to assess decision strategies and the information used for face recognition (Everdell, Marsh, Yurick, Munhall, & Pare, 2007; Henderson, Falk, Minut, Dyer, & Mahadevan, 2001; Mertens, Siegmund, & Grusser, 1993). For example, Bloom and Mudd (1991) present evidence that the number of features encoded, neither the amount of time for viewing nor the level of processing, dictate later recognition of faces.

Eye tracking has not been widely used in the eyewitness field. Some research using eye tracking has focused on behavior in response to emotion-invoking stimuli (Calvo & Lang, 2004; Christianson, Loftus, Hoffman, & Loftus, 1991) and scene perception (Loftus & Mackworth, 1978). Loftus, Loftus, and Messo (1987) tracked participants' gaze while they viewed a video of an interaction where a customer waved either money or a gun at a store clerk. The gaze data showed that more visual attention was allocated to the weapon than the money, suggesting recall deficits in weapon focus research may occur because of a lack of attentiveness to aspects of the scene other than the weapon. Hope and Wright (2008) recently showed that unusual and threatening objects in a scene elicit different patterns of eye movements, suggesting that the weapon focus effect is not simply a product of the unusualness of the presence of a weapon in a scene.

Thus, eye tracking has the potential for informing about the types of strategies used by mock-witnesses by providing explicit information about where overt attention is directed while they make decisions. When asked to view a lineup and select the perpetrator (if present), a witness must go through a decision process. This process can have a positive

outcome (select a guilty suspect or select no one from a perpetrator-absent lineup) or a negative one (select an innocent person or fail to select a guilty suspect). As such, a selection response may be correct or incorrect and a non-selection may be correct or incorrect. Knowing something about the process engaged in by witnesses doing this task can inform our understanding of the strategies used by eyewitnesses under different circumstances and may lead to procedures that minimize the likelihood that inadequate strategies are used.

Wells (1984) has proposed a dichotomy in witness strategies that should be reflected in discernible behavioral differences during the process of decision making. An absolute strategy implies review of lineup faces without comparisons among the images (comparison to memory only), while the relative strategy is defined as comparisons among lineup members. Because an absolute strategy involves comparing to memory only, witnesses need not view all faces in a target-present lineup. Thus, an exhaustive search of a lineup could be considered indicative of relative strategy use. Using similar logic, visual behavior that is suggestive of comparisons, such as looking at a face, then looking at another, and then returning to the first face could be useful for assessing the degree of relative strategy use. However, it is worth noting that while overt attention is the normal means through which we attend, covert attention involves attending to locations without fixations and thus these measures cannot unambiguously represent attention (Findlay & Gilchrist, 2003).

We had four general hypotheses about how gaze and accuracy would interact. First, selecting from a lineup is associated with recognition (accurately or inaccurately) while not selecting from a lineup is indicative of a failure to recognize (accurately or inaccurately).

Thus, we expected that when participants did not select a face, versus did select a face, they would devote more visual attention to the lineup reflecting a search for a familiar face but a failure to find one.

Second, when a selection is accurate, it should reflect a stronger recognition response than when a selection is inaccurate (assuming no foil or innocent suspect is of clone-like similarity to the target). A true match should elicit a stronger recognition response than when a face is similar but not a true match. Moreover, recognition should occur more quickly when a selection is accurate (Sporer, 1993; Weber, Brewer, Wells, Semmler, & Keast, 2004). We, therefore, expected less visual attention to be devoted to lineups when correct selections were made than when incorrect selections were made.

Third, because a potentially perfect match to a participant's memory is provided in a target-present lineup, participants should experience some familiarity with the target, even if their memory trace for the face is not particularly strong. In contrast, participants viewing target-absent lineups should experience little familiarity with the faces because they have never viewed any of the faces before. A lack of familiarity may lead participants viewing a target-absent lineup to distribute their gaze relatively equally across images when they make a non-selection, while gazing more at the weakly familiar image (target) when viewing a target-present lineup.

Fourth, because simultaneous lineups are associated with relative judgments, we expected gaze behavior suggestive of this would be detected (Wells, 1984), with: a). More evidence of exhaustive searches than non-exhaustive searches and; b). More decisions involving comparisons between faces than decision not involving comparisons. Because absolute and relative judgments are associated with similar correct identification rates, we

expected c). All searches to produce similar proportions of correct identifications. Conversely, because absolute judgments are associated with a lower false identification rate, we expected d). Non-exhaustive searches to produce a smaller proportion of incorrect selections than exhaustive searches and e). Non-comparative searches to produce a smaller proportion of false selections than comparative searches (Kneller et al., 2001; Lindsay & Bellinger, 1999; Smith, Lindsay, & Pryke, 2000; Sporer, 1993).

Method

Participants

Participants were 34 introductory psychology students recruited from the Psychology Subject Pool to participate for course credit or pay (14 men and 19 women (1 not recorded), M age = 19.00 years, $Range$ = 18 - 29 years). The ethnicity of participants was mostly European ($N = 30$) with the remaining participants listing their ethnicity as Asian or Native American, and one providing no information.

Design

The only manipulated variable was the presence versus absence of the criminal in the lineup (within-participants).

Stimuli

Vignettes. Twelve short videos used in previous studies were utilized for the current studies (Bertrand, Beaudry, Mansour, & Lindsay, 2006). Each video took approximately 30 seconds and depicted the same scene with a different actor. The video showed a male adult (approximately 19 years old, of European descent) enter a room containing a desk with a purse on it. The male indicated to a female off screen that he was supposed to pick up a VCR for repair. The female left to find the VCR during which time the male rummaged

through the purse and removed money. The female returned and indicated the VCR could not be found. The male responded that he would return later and left. The actor was visible throughout the scene.

Lineups. Participants attempted to identify the actors (criminals) in the videos from lineups. Foils were selected using a modified match-to-description strategy: A median description of each of the eight targets was constructed from descriptions obtained from not fewer than 11 participants (who did not participate in any other part of the study) immediately after viewing each crime video. Lineups were constructed by choosing, from a pool of photos, the six faces that best matched descriptions of the targets obtained previously and that were not being used in any other lineup (McQuiston-Surrett, Malpass, & Tredoux, 2006).

As per Figure 1, each lineup consisted of six head and neck images (10 cm x 12 cm) presented simultaneously in two rows of three. The head and neck portion of each image was approximately 4.5 cm and x 6.5 cm. Because, each participant was seated approximately 57cm from the screen, the rectangles containing the faces subtended 10×12 degrees of visual angle and each face subtended approximately 4.5×6.5 degrees.

Rectangles were separated by approximately 3 horizontal visual degrees and 1 vertical visual degree. Faces were separated from their nearest neighbor by approximately 6.5 horizontal visual degrees and 5 vertical visual degrees. Target position was counterbalanced across the studies such that each target appeared approximately equally often in all lineup positions. One lineup was shown for each video. Participants responded to lineups with a key press (1-6, not there, don't know) and gave a verbal rating of confidence in their

decision¹.

Care was taken to ensure lineups were not biased such that participants would be cued to select the target or any other face regardless of previous experience, using the modified match-to-description procedure described above. Fairness of lineups was assessed by another 60 students (mock witnesses) who selected the person they thought was “guilty” from the lineups based on reading the descriptions of the targets. Each mock witness saw half of the targets (four) in target-present lineups such that 30 of the mock witnesses saw the first four targets and the other 30 saw the remaining four targets. The mean functional lineup size was 4.87 ($SD = 4.18$; $Range = 2.50 - 15.00$). The mean effective size for the lineups was 3.87 ($SD = .74$; $Range = 2.93 - 5.00$).

Questionnaire. Following completion of the trials, participants read conceptual definitions of relative and absolute judgments and indicated which strategy they used on the final trial of the experiment (Dunning & Stern, 1994; Kneller et al., 2001; Lindsay & Bellinger, 1999). The definitions read as follows: “**Absolute** judgments are those where you compare each picture in a lineup to your memory of the culprit. **Relative** judgments are those where you compare the images in a lineup to each other to find the one that most closely matches your memory for the culprit. Which type of judgment best characterizes your decision?” The response options were “absolute” and “relative.”

Apparatus

Eye tracker. Gaze was measured using a head-mounted eye tracking device, the EyeLink II, Version 1.0.5 (SR Research Ltd., Osgoode, Ontario). Eye and head position were recorded by three cameras located on a headband. Participants were seated approximately 57cm from a 27cm by 37cm (20 inch) video monitor. Head movements were

minimized with the use of a chin rest. This also ensured participants viewed videos and lineups from a standard location. The eye tracker recorded the time and position of the participant's gaze on the screen while the video ran at the standard NTSC frame rate (29.97 frames per second). The Eyelink II has a very high resolution for fixation location detection (less than .01 degrees of visual angle error) and a high sampling rate (500 Hz), providing us with excellent accuracy for detecting gaze.

Prior to recording eye movements, the eye tracker was calibrated. Calibration involved participants fixating a dot on the screen which changed location when a fixation was detected.² The dot appeared in nine locations and the procedure was conducted twice: once to calibrate and once to validate the calibration. If the computer's error estimating any of the fixation points was greater than 1.00 degree, the procedure was repeated. Prior to each video and each lineup, participants fixated a static dot and the computer adjusted its calculations to account for how much the difference between a participant's gaze and a central point had "drifted" over a short time period (drift correction), with a maximum acceptable error of 1.5 visual degrees. Calibration and drift correction ensure the accuracy of gaze measurements.

INSERT FIGURE 1 ABOUT HERE

Procedure

Participants first completed a consent form and then heard an explanation of how eye tracking works. The eye tracker was then fitted to their head and the system calibrated. Next an explanation of the task was given. Participants were instructed that "the person you

saw in the video may or may not be present in the array of photos. If he is present, push the button associated with his face, otherwise choose ‘not there’ or ‘don’t know.’” In order to ensure they understood the relation between the keys and the face locations, participants were then shown a lineup and asked which button they would push if the culprit appeared in different positions. Two practice trials were completed followed by eight experimental trials. Each trial began with a drift correction. Participants viewed the crime video and then solved anagrams for 30 seconds. Following this distractor task participants did another drift correction and then viewed a lineup and made a key press response in the eight-alternative fixed-choice task (which picture was of the actor from the crime video, not there, or don’t know). Next, the lineup disappeared and participants gave the verbal rating of confidence. Participants completed a questionnaire following the eye tracking trials. After completing the questionnaire, participants were debriefed and thanked.

Data Analysis

Eye movements. A number of measures were calculated from the eye movement data. A rectangular region of interest was defined for each lineup face corresponding to the rectangle containing the face (See Figure 1). This area was selected because previous research indicates that when people view an image they are able to obtain information from their periphery (via covert attention) that can be used for identification purposes (Mäkelä, Näsänen, Rovamo, & Melmoth, 2001). Analyses were also conducted with a rectangle surrounding just the face but the pattern of results did not differ. The total time fixating (*duration of fixations*) and the *number of fixations* in each area were recorded by computer. Single or multiple fixations could occur in an area (See Figure 1B. upper right and upper left, respectively). In order to examine the sequential patterns of examination of faces, the

number of comparison were determined. Comparisons were defined as the number of times a participant fixated a face, fixated a new face or faces, and then returned to the original face. Thus, a *first-order comparison* involved fixating a particular face, a face other than the first face, and then the first face again (A-B-A) as shown in Figure 1E. A *second-order comparison* involved fixating a particular face, a second face, then a third face, and then the first face again (A-B-C-A) as shown in Figure 1F. Although higher order comparisons were also counted, all had a mean frequency of less than one per trial, and so were excluded from analysis.

Each trial was categorized as having at least one face not fixated (*non-exhaustive search*) or as having each face fixated at least once (*exhaustive search*). Figures 1A and 1C illustrate non-exhaustive searches while Figures 1B and 1D illustrate exhaustive searches. Each trial was also independently categorized as having no comparisons (*non-comparative search*) or some comparisons (*comparative search*), where both first- and second-order comparisons were counted. Figures 1A and 1B illustrate non-comparative searches while Figure 1C and 1D illustrate comparative searches. For example, a given trial could have been exhaustive with comparisons, or exhaustive with no comparisons. These two dichotomies were used to reflect how relative and absolute search strategies might be manifested in visual behavior. That is, if participants compare each face to their memory, they should stop looking at faces when a match to their memory is found; thus, that lineup search would be non-exhaustive (However, one could use this strategy but recognize the last person viewed in a lineup. This would not be captured as a non-exhaustive search by our operationalization; thus, some absolute judgments would not be captured by our definition, though the majority should be). Similarly, an absolute judgment strategy implies

no comparisons between lineup members. Thus, we categorized trials into ones in which comparative returns to faces were made and ones in which no comparison-like returns to faces were made. (However, a person might look back and forth between faces because they did not get a good look at a face originally, thus our definition will capture more than just comparisons).

Behavioural Measures. For each participant we calculated the number of responses that were (a) correct selections, (b) incorrect selections from target-present lineups, (c) incorrect selections from target-absent lineups³ (d) non-selections from target-present lineups, and (e) non-selections from target-absent lineups. We also calculated the number of each participant's responses that were selections versus non-selections. "Don't know" and "Not there" responses were collapsed into a "Non-selections" category because there were very few "Don't know" responses (.03 of all responses).

We tested for differences in response types using repeated measures multivariate analyses of variance (MANOVAs) with Trial duration, Duration of fixations, Number of fixations, Number of first-order comparisons, and Number of second-order comparisons as the dependent measures. This statistical model was applied to test for differences between visual behavior for choices and non-selections from lineups, visual behavior for correct and incorrect selections from the lineup and on the selected face, and visual behavior for correct and incorrect non-selections from the lineup and on the target or most attended face.

Results are reported only where the omnibus test was significant. The degrees of freedom for these analyses vary because participants occasionally were dropped from an analysis when they did not provide data for all of the within-subjects conditions tested. For example, a participant who made no incorrect selections could not be included in a comparison of the

total fixations per trial for correct versus incorrect selections. Note also that this means that proportions of correct selections, incorrect selections, correct rejections, and incorrect rejections will not add up to 1.0.

Questionnaire. The proportions of participants indicating they made absolute and relative judgments were calculated overall and by whether the trial was exhaustive/non-exhaustive and comparative/non-comparative.

Results

The mean rate of correct selections in the target-present condition was .40 ($SD = .21$) and the mean rate of correct non-selections in the target-absent condition was .37 ($SD = .30$). The mean rate of incorrect selections in the target-present condition was .18 ($SD = .21$) and .57 ($SD = .29$) in the target-absent condition. The mean non-selection rate in the target-present condition was .39 ($SD = .18$).

Using a repeated measures design can cause early trials to differ from late trials because of practice effects or proactive interference (Perfect, Hollins, & Hunt, 2000; Peterson & Peterson, 1959). We conducted a logistic regression with Response (correct, incorrect) as the criterion and Trial Number as the predictor, with Participant partialled out to account for the within-subjects design. There was no significant pattern of change in responding associated with trial number, $\chi^2(1, N = 266) = 1.55, p = .21$. The proportion of correct decisions from trials 1 to 8 were .35, .41, .35, .38, .38, .41, .41, .53, respectively.

Gaze Patterns

The mean time spent on a trial was 12.88 seconds ($SD = 8.69, Range = 2.34 - 56.85$ seconds) where a mean of 10.22 seconds ($SD = 7.87, Range = 1.37 - 50.46$ seconds)

or .81 of that time was spent on the lineup faces with a mean of 35.21 ($SD = 23.80$, $Range = 4 - 144$) fixations. Participants made a mean of 2.53 ($SD = 2.28$, $Range = 0 - 11$) first-order comparisons to faces and 1.94 ($SD = 2.09$, $Range = 0 - 11$) second-order comparisons.

Participants also showed a slight spatial bias in location of fixations whereby they spent a greater amount of time looking at faces in the upper middle of the lineups (.18) than other face locations in the lineup. Each other area received approximately .14 - .16 of the fixations, except for the lower right, which received only .11 of fixations. However, this may have occurred because of the design of the study: Participants began each trial fixated on a dot in the center of the screen. Consistent with this interpretation, .83 of first fixations were to the upper middle face in the lineup and .11 were to the lower middle face. Analysis of the relationship between initial fixation and decision accuracy was conducted, but indicated no relationship between the two.

Selecting versus Non-selecting

The omnibus repeated Measures MANOVA with Type of Decision (selecting, non-selecting) was significant, $F(7, 26) = 6.45$, $p < .001$, $\eta^2 = .63^4$, with participants devoting more visual attention to lineups when they did not select versus did select. Participants took significantly more time to make a decision ($M = 12.73$ seconds, $SD = 5.76$) when they did not make a selection compared to when they did ($M = 10.71$ seconds, $SD = 6.28$), $F(1, 32) = 7.36$, $p = .01$, $\eta^2 = .19$, and there was no difference in first-order ($p = .68$) or second-order comparisons ($p = .80$). Non-selections involved more time looking at lineup faces ($M = 10.22$ seconds, $SD = 5.79$) than selections ($M = 8.46$ seconds, $SD = 5.85$), $F(1, 32) = 7.94$,

$p = .008$, $\eta^2 = 20.06$, and there was no difference in the number of fixations made ($p = .19$).

These data partially support Hypothesis 1, that not selecting a face would involve more visual attention to the lineup than selecting a face. Specifically, they indicate that, when participants did not select, they spent more time considering faces than when they did select, but looked at faces just as often.

Distinguishing Accurate and Inaccurate Selections

We were interested in whether visual behavior distinguishes correct selections and selections from target-absent lineups. We hypothesized that more visual behavior would be devoted to the selected face for incorrect selections than correct selections (Hypothesis 2). Only the results for second-order comparisons were consistent with this hypothesis. The omnibus test for visual attention to the selected lineup face (comparing target-present correct selections to target-absent selections) was significant, $F(4, 27) = 4.46$, $p = .007$, $\eta^2 = .40$ with less visual attention being devoted to selections from target-present than target-absent lineups. However, at the univariate level, only second-order comparisons significantly differentiated correct selections and incorrect selections (target-absent lineups). Correct selections involved fewer second-order comparisons to the selected face ($M = .81$, $SD = 1.09$) than did incorrect selections ($M = 1.92$, $SD = 2.02$), $F(1, 30) = 8.55$, $p = .007$, $\eta^2 = .22$. The result for first-order comparisons was in the same direction but not significant and both number and duration of fixations were non-significantly greater for correct than incorrect selections.

Distinguishing Accurate and Inaccurate Non-selections

Incorrect non-selections were expected to elicit more visual attention to the target-face

than correct non-selections on any one face (Hypothesis 3). The omnibus model for visual attention to a particular face (target face or the most fixated face) was significant, $F(4, 23) = 2.96, p = .04, \eta^2 = .34$. However, participants did not attend to a guilty suspect more than any other foil. For non-selections from a target-present lineup, the mean duration of fixations ($M = 2.16$ seconds, $SD = 2.71$) on the target was marginally significantly less than the mean duration of fixations ($M = 3.04$ seconds, $SD = 1.95$) on the most fixated person in a target-absent lineup, $F(1, 26) = 3.70, p = .07, \eta^2 = .12$. The number of fixations on the target face for non-selections from target-present lineups ($M = 7.2, SD = 6.30$) was less than the number of fixations on the most fixated face for non-selections from a target-absent lineup ($M = 10.26, SD = 4.90$), $F(1, 26) = 6.02, p = .02, \eta^2 = .19$. However, correct non-selections involved significantly more first- and second-order comparisons to the most compared face than incorrect non-selections on the target face. Correct non-selections involved a mean of 1.23 ($SD = .96$) first-order comparisons to the most compared face versus .67 ($SD = 1.26$) for incorrect non-selections, $F(1, 26) = 4.79, p = .04, \eta^2 = .16$. Similarly, correct non-selections had a mean of .91 ($SD = .54$) second-order comparisons to the most compared face versus .48 ($SD = .76$) second-order comparisons to the target face when incorrect non-selections were made, $F(1, 26) = 5.65, p = .03, \eta^2 = .18$. Thus our hypothesis that visual attention to the target face would be greater than attention to any particular face was not supported. In fact, the data suggest the opposite pattern.

Exhaustiveness of Search

The overall proportion of all judgments involving non-exhaustive searches was .09 with .44 of all participants engaging in at least one non-exhaustive search. This is consistent with our expectation of more evidence of relative than absolute judgments (Hypothesis 4a).

We expected that non-exhaustive and exhaustive searches would elicit similar proportions of correct responses (correct selections and non-selections from target-absent lineups) and consistent with this, .36 ($SD = .18$) of non-exhaustive searches and .34 ($SD = .41$) of exhaustive searches resulted in correct decisions, $t(14) = .16$, $p = .88$, $d = .04$ (Hypothesis 4c). We also predicted with Hypothesis 4d, that non-exhaustive searches as compared to exhaustive searches would be associated with a smaller proportion of incorrect selections because absolute judgments elicit more fewer identifications (Lindsay & Wells, 1985). However, the proportion of false identifications (from target-present and target-absent lineups) for non-exhaustive searches ($M = .66$, $SD = .42$) was greater than the proportion for exhaustive searches ($M = .39$, $SD = .21$), $t(14) = 2.41$, $p = .03$, $d = .62$. This may be due to the low frequency of non-exhaustive searches.

Comparing Faces

Fewer trials involved non-comparative than comparative searches: only .11 of trials were non-comparative, consistent with Hypothesis 4b. Only .38 of participants engaged in at least one trial that involved no comparisons. We found similar proportions of correct decisions when non-comparative ($M = .42$, $SD = .49$) and comparative ($M = .32$, $SD = .12$) searches were conducted, $t(12) = .78$, $p = .50$, $d = .22$, also consistent with Hypothesis 4c. Non-comparative searches ($M = .42$, $SD = .49$) and comparative searches ($M = .26$, $SD = .14$) did not differ in their rates of false selections, $t(12) = 1.11$, $p = .27$, $d = .32$, across both target-present and target-absent lineups, inconsistent with Hypothesis 4e.

Self-Report of Judgment Strategy

Participants were questioned about their judgment strategy on the final trial. On this trial, .10 of searches were non-exhaustive and .37 had no comparisons, though .63 of participants indicated they used an absolute judgment strategy on the final trial. Lindsay and Bellinger (1999) found similar results for self-report of judgment strategy (.66). Of those indicating they used an absolute judgment, .84 conducted an exhaustive search and .68 conducted a comparison search. Of the only nine participants who indicated they made a relative judgment, none made a non-exhaustive search of the lineup and .56 made no comparisons.

Discussion

The purpose of this study was to explore the visual behavior of eyewitnesses viewing lineups and to see whether visual behavior could differentiate correct and incorrect decisions and/or indicate the type of judgment process witnesses engaged in. In summary, visual behavior was a poor predictor of decision accuracy and the visual behavior of witnesses viewing simultaneous lineups was suggestive of a relative judgment strategy most of the time. There were some significant differences in visual behavior between decision types (selections versus non-selections, correct versus incorrect selections, correct versus incorrect non-selections) but these differences were temporally and numerically small (on the order of one second, one comparison, or a few fixations) and thus have limited probative value.

There were statistically significant differences in visual behavior for each of the decision types compared. For example, participants took significantly longer to make a decision when they did not select compared to when they did. This suggests differing

amounts of search or consideration between the two decision types and is consistent with other eyewitness literature showing that selections from simultaneous lineups occur more quickly than lineup rejections (Sporer, 1993). Unfortunately, the difference in time looking at faces was only about two seconds. Given the substantial variability between subjects, this may not be an easily detected difference in applied settings. Similarly, the only differences between correct and incorrect selections and between correct and incorrect non-selections were small differences in the number of comparisons. For selection decisions, this was consistent with our expectation that less visual search would occur when a participant correctly recognized a target compared to when they incorrectly selected a foil. Again, however, the difference was practically small: participants made approximately one second-order comparison when they made a correct selection versus two when they made an incorrect selection.

With regard to non-selections, the finding that only comparisons differed between correct and incorrect decisions was contrary to our hypothesis. We expected that when non-selections eventually occurred, a sense of familiarity with a face would lead to more visual attention to it than a non-familiar face. The absence of this expected effect suggests that witnesses who fail to identify a guilty suspect may have such poor quality memory for the target that they do not experience familiarity. This inference could be tested with future research by varying the conditions under which participants view “crimes.”

Still, the small differences we found suggest potential cautions for interpreting the value or accuracy of identification decisions and such cautions are consistent with other eyewitness findings. Specifically, when witnesses make many comparisons between faces or take a long time with lineups, these behaviors may be cues that the lineup does not

contain the perpetrator. Researchers have advocated for some time that taking a long time to make a decision is indicative of an innocent suspect (e.g., Dunning & Stern, 1994).

Relative judgments, or judgments involving the search for a closest match among many options, have been associated with false positive selections (e.g., Lindsay & Bellinger, 1999).

The prevalence of exhaustive searches regardless of whether participants selected or did not select and whether their decision was correct or not, may reflect the use of a common search strategy for all lineups. In general, visual behavior was more often indicative of relative judgments than absolute. That is, participants made more relative judgments as defined by our operationalizations of relative judgments (exhaustive search, comparisons of faces). For example, only .02 of trials were devoid of comparisons and did not involve looking at all the lineup faces. Interestingly, participants believed they were making absolute judgments even though their visual behavior was not consistent with their self reports. Lindsay and Bellinger (1999) noted that while .67 of participants viewing a lineup indicated they used an absolute strategy, many engaged in processes suggestive of a relative strategy, such as flipping back and forth through an album of faces. Consistent with this, .63 of participants in this study indicated they used an absolute judgment process on their final lineup decision but only .10 of these trials were absolute judgments according to our search exhaustiveness definition and .37 according to our lack of comparisons definition. That said, the small proportion of participants who conducted non-exhaustive searches or made no comparisons on the final lineup they viewed never indicated that they had used a relative judgment strategy. Furthermore, when participants indicated they were making a relative judgment, their visual behavior always supported this. Unfortunately, because data

were collected only for the final trial there were limited data available for analysis. There is no way to say with certainty whether participants who claimed to have used an absolute process actually did so, but their visual behavior does not support this position.

Two factors temper this conclusion about the relative/absolute judgment strategy interpretation. First, recent eyewitness research on search and decision strategy suggests that the absolute/relative dichotomy may not accurately reflect the complexity of the cognitive processes engaged by witnesses when examining lineups. Charman and Wells (2007) suggest automatic and deliberative processing are analogous to absolute and relative processing and that the two processes occur in tandem. This is consistent with our results whereby participants who engaged in behaviors suggestive of absolute processing also engaged in behaviors suggestive of relative processing (e.g., non-exhaustive search with comparisons). This is also consistent with Clark's (2003) finding that a mix of relative and absolute judgment strategy best predicted modeled lineup outcomes. Perhaps strategy use changes dynamically over the course of a lineup or multiple lineups. Witnesses may use an absolute strategy to make a tentative decision and follow it up with a confirmatory relative process. Alternatively, they may vacillate between absolute and relative processes giving priority to one or another mode of consideration at different times or shifting the weight given to each contributor depending on some other factor(s). Even if the absolute/relative dichotomy is appropriate, our definitions of strategies may not have captured the nature of absolute and relative judgments.

A second reason for further consideration of relative strategy is a limitation in the data analysis itself. Eye movement data are very rich indices of human real time behavior and our extensive measures are not a complete description of our subjects' gaze behavior.

Specifically, temporal patterns of gaze were not captured in our analysis. When people look at objects or scenes the sequence of fixations produce a ‘scan path’ that tells how information is selected sequentially. Differences in strategy may be reflected in complex visual behavior patterns that were not investigated in this research. For example, Gilchrist and Harvey (2006) found that the display structure (how regular a pattern of items the participants had to search through) modulated how systematic participants’ visual scanning was.

Perhaps witnesses begin with a strategy of scanning the lineup until a face stands out. If this happens, they may revise their strategy based on the similarity of the face to their memory. Pomplun et al. (2001) found evidence for use of a similar strategy when comparing two images. They identified “search and comparison” and “detection and verification” processes occurring during a difference detection task with two images. These two processes were differentiated by the proportion of very long fixations (greater than 500ms) that occurred within each phase.

Witnesses may first engage in a simple scan of a lineup to detect possible matches to their memory and then engage in a relative judgment process consisting of comparisons and consideration only with these potential matches. Alternatively, witnesses might scan lineups until a possible match is found and continue scanning until another possible match is encountered. They might then switch to a comparative process ending in selection of one of the faces and reengagement of the lineup scan. Such a process might loop many times depending on the similarity of foils and innocent suspects to the witness’ memory for a perpetrator. Many such patterns are possible given the fluid nature of visual search and the complex nature of attention. The revised new strategy may be indicative of accuracy. That

is, there may be specific scan paths (sequences of fixations) associated with selections and others associated with whether a correct versus incorrect selection was made. However, our data clearly indicate these are more complicated than simple-comparison style sequences.

From an applied perspective, findings using visual behavior would have to be clear-cut to warrant the expense, time, and effort necessary to implement eye tracking systems as a means of diagnosing identification accuracy. At this time, visual behavior does not sufficiently differentiate accurate and inaccurate eyewitnesses to be useful in applied settings. However, visual behavior provides an interesting picture of witnesses that can inform future studies. Witnesses frequently engaged in what looked like relative judgments with simultaneous lineups. Even judgments categorized as absolute often evidenced what appeared to be comparisons. This strongly suggests that simultaneous lineups can fail with innocent suspects because witnesses often choose a best match to their memory from the lineup, even if they know they should or intend to engage another strategy or that the perpetrator may be absent.

Clearly a behavioral definition of relative and absolute judgments requires much more consideration, if one is to be found at all. Future research might consider comparing different behavioral definitions and different self-report measures to assess ways in which to bring the two together with the ultimate goal of successfully instructing witnesses to use an absolute strategy. Similarly, research should consider whether the absolute/relative dichotomy is appropriate at all. Charman and Wells (2007) suggest the two conceptualizations are better seen as competing processes, rather than mutually exclusive ones. Finally, an important purpose of eyewitness research is to develop procedures for decreasing the likelihood of false identification and increasing the likelihood of correct

identifications. While the visual behaviors we considered did not usefully differentiate accurate and inaccurate identifications or the type of strategy utilized by witnesses, there may be more complex scan paths that can.

References

- Bertrand, M., Beaudry, J., Mansour, J., & Lindsay, R. C. L. (2006, March). *Is increasing lineup size an alternative to sequential presentation? The question revisited and revamped*. Poster presented at the American Psychology-Law Society, St. Petersburg, Florida.
- Bloom, L. C., & Mudd, S. A. (1991). Depth of processing approach to face recognition: A test of two theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 556-565. DOI: 10.1037/0278-7393.17.3.556
- Calvo, M. G., & Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: Motivationally biased attention. *Motivation and Emotion*, *28*, 221-243. DOI: 10.1023/B:MOEM.0000040153.26156
- Charman, S., & Wells, G.W., (2007). Applied Lineup Theory. In R. Lindsay, D. Ross, J. Read, & M. Tolia (Eds.), *Handbook of Eyewitness Psychology: Memory for People* (219-254). New Jersey: Lawrence Erlbaum.
- Christianson, S. A., Loftus, E. F., Hoffman, H., & Loftus, G. R. (1991). Eye fixations and memory for emotional events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 693-701. DOI: 10.1037/0278-7393.17.4.693
- Clark, S. E., (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629 - 654. DOI: 10.1002/acp.891
- Collewijn, H. (1999). Eye movement recording. In RHS Carpenter & JG Robson (Eds.), *Vision research: A practical guide to laboratory methods* (pp. 245-285). Oxford: Oxford University Press.
- Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness

- identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, 67, 818-835. DOI: 10.1037/0022-3514.67.5.818
- Everdell, I.T., Marsh, H., Yurick, M. D., Munhall, K. G., & Paré, M. (2007). Gaze behavior in audiovisual speech: Asymmetrical distribution of face-directed fixations. *Perception*, 36, 1355-1545. DOI: 10.1068/p5852.
- Findlay, J. M., & Gilchrist, I. D. (200?). *Active vision: The psychology of looking and seeing*. New York: Oxford University Press.
- Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, 14, 704 – 715. DOI: 10.1080/13506280500193719
- Henderson, J. M., Falk, R., Minut, S., Dyer, F. C., Mahadevan, S. (2001). Gaze control for face learning and recognition by humans and machines. In T. F. Shipley, & P. J. Kellman (Eds.), *From Fragments to Objects – Segmentation and Grouping in Vision* (pp. 463-481). Amsterdam: Elsevier Science.
- Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, 33, 98-106.
- Hope, L. & Wright, D. (2008). Beyond unusual? Examining the role of attention in the weapon focus effect. *Applied Cognitive Psychology*, 21, 957- 961. DOI: 10.1002/acp.1307
- Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology*, 15, 659-671. DOI: 10.1002/acp.739

- Leippe, M. R. & Eisenstadt, D. (2006). Eyewitness confidence and confidence-accuracy: Should we be concerned? In R. Lindsay, D. Ross, J. Read, & M. Tolia (Eds.), *Handbook of Eyewitness Psychology: Memory for People* (pp. 377-425). New Jersey: Lawrence Erlbaum.
- Lindsay, R. C. L., Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84*, 315-321. DOI: 10.1037/0021-9010.84.3.315
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*, 556-564. DOI: 10.1037/0021-9010.70.3.556
- Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus." *Law & Human Behavior, 11*, 55-62. DOI: 10.1007/BF01044839
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 565-572. DOI: doi:10.1037/0096-1523.4.4.565
- Mäkelä, P., Näsänen, R., Rovamo, J., & Melmoth, D. (2001). Identification of facial images in peripheral vision. *Vision Research, 41*, 599-610. DOI: 10.1016/S0042-6989(00)00259-5
- 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*, 137 – 169. DOI: 10.1037/1067-8971.12.2.137
- Mertens, I., Siegmund, H., & Grusser, O.-J. (1993). Gaze motor asymmetries in the

- perception of faces during a memory task. *Neuropsychologia*, 31, 989-998. DOI: 10.1016/0028-3932(93)90154-R
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259. DOI: 10.1037/0033-295X.84.3.231
- Perfect, T. J., Hollins, T. S., & Hunt, A. L. R. (2000). Practice and feedback effects on the confidence-accuracy relation in eyewitness memory. *Memory*, 8, 235-244. DOI: 10.1080/096582100406793
- Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.
- Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G., & Ritter, H. (2001). Comparative visual search: A difference that makes a difference. *Cognitive Science*, 25, 3 – 36. DOI: 10.1016/S0364-0213(00)00037-9
- Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology*, 85, 542-550. DOI: 10.1037/0021-9010.85.4.542
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, 78, 22-33. DOI: 10.1037/0021-9010.78.1.22
- Stampe, D. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking system. *Behavior Research Methods, Instruments, & Computers*, 25, 137-142. DOI: 10.1016/S0165-0270(01)00527-1

- 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*, 139-147. DOI: 10.1037/1076-898X.10.3.139
- Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, *14*, 89-103. DOI: 10.1111/j.1559-1816.1984.tb02223.x

Figure Captions

Figure 1. Example lineups viewed by participants with gaze patterns superimposed. A. A non-exhaustive search without comparisons. B. An exhaustive search without comparisons. C. A non-exhaustive search with comparisons. D. An exhaustive search with comparisons. E. A first-order comparison. F. A second-order comparison.

Endnotes

¹ Because confidence in a lineup decision is normally collected following a lineup decision but is not central to our analysis, the data were collected but are not reported here.

² For EyeLink recordings, the eye-event detection is based on an internal heuristic saccade detector built in the EyeLink tracker program. Fixations are defined by the EyeLink II as anything that is not a blink or a saccade. A blink is a period of saccade-detector activity with the pupil data missing for three or more samples in a sequence. A saccade is a period of time where the saccade detector is active for 2 or more samples in sequence and continues until the start of a period of saccade detector inactivity for 20 milliseconds (Stampe, 1993).

³ Because we were interested in how visual behavior varied when choosing a previously seen person compared to when choosing a previously not seen person, any selection from a target-absent lineup was considered a “false identification”. It is hard to know how visual behavior could discriminate between selection of two people not seen before (i.e., between a known foil and an innocent suspect). Such information may tell us which image more closely matches a participant’s encoded image of the perpetrator, but this image may or may not actually resemble the perpetrator.

⁴ All reported eta values are partial eta squared values.