

Visual Behaviour In Eyewitness Identification 1

Running head: Visual Behaviour in Eyewitness Identification Tests

An Exploration of Visual Behaviour in Eyewitness Identification Tests

Heather D. Flowe
School of Psychology
University of Leicester

Draft

Abstract

The contribution of internal (eyes, nose and mouth) and external (hair-line, cheek and jaw-line) features across eyewitness identification tests was examined using eye tracking. In Experiment 1, participants studied faces and were tested with lineups, either simultaneous (test faces presented in an array) or sequential (test faces presented one at a time). In Experiment 2, the recognition of previously studied faces was tested in a showup (a suspect face alone was presented). Results indicated that foils were analysed for a shorter period of time in the simultaneous compared to the sequential condition, whereas a positively identified face was analysed for a comparable period of time across lineup procedures. In simultaneous lineups and showups, a greater proportion of time was spent analysing internal features of the test faces compared to sequential lineups. Different decision processes across eyewitness identification tests are inferred based on the results.

An Exploration of Visual Processing Strategies in Eyewitness Identification Tests

The police have a number of procedures that they can use to identify criminal suspects. Showups are widely used in the United States (Flowe, Ebbesen, & Mehta, 2010). In a showup, a suspect, who usually has been apprehended by the police at the crime scene, is presented alone to the eyewitness for identification. Alternatively, a simultaneous lineup can be used, wherein the suspect is presented embedded among foils (i.e. distractors) to control for guessing (Stebly, Dysart, Fulero, & Lindsay, 2003). The sequential lineup procedure has been more recently introduced to further increase the reliability of eyewitness identification (Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). The suspect and foil faces are shown to the eyewitness one at a time in a sequential lineup. The eyewitness makes a yes/no decision to a test face before the next one is shown. The sequential lineup typically concludes once a face has been positively identified or once all of the test faces have been presented, though practices may vary (see McQuiston-Surrett, Malpass, & Tredoux, 2006). In the United Kingdom, video clips of lineup members are presented sequentially to eyewitnesses (see Valentine, Memon, & Darling, 2007).

Meta-analyses of the adult identification literature have found that identification responses vary depending on whether a test face can be compared to other faces. Hits (the rate of positively identifying the guilty suspect) and false alarms (the rate of mistakenly identifying an innocent suspect) are lower in a sequential compared to a simultaneous lineup (Stebly, Dysart, Fulero, & Lindsay, 2001). Additionally, correct rejections are higher in a showup compared to a simultaneous lineup, whereas correct identification rates

are comparable (Stebly et al., 2003). Thus, identification responses are affected by whether foil faces are included at test and by whether the test faces can be visually compared.

The current study explored visual behaviour in simultaneous, sequential and showup identification tests by tracking participants' eye movements. A range of forensic topics has been investigated with eye tracking, from the effects of weapon exposure on remembering (e.g. Loftus, Loftus, & Messo, 1987) to the role of expertise in deception detection (Bond, 2008). More recently, researchers have begun to use eye tracking to study simultaneous lineup decision processes (Flowe & Cottrell, 2010; Mansour, Lindsay, Brewer, & Munhall, 2009). These studies have found that a correct positive identification involves fewer comparisons across the test faces. Additionally, the length of time that a test face is examined is longer if the face is a target (i.e. 'suspect') rather than a foil. The current study extends this previous work by addressing several additional basic but unanswered questions, including: Does the length of time that a test face is analysed vary depending on whether there is a single test face or many test faces from which to choose? Will participants apportion their time across the suspect's facial features in a different manner depending on whether the suspect face can be compared against distractor faces?

Theoretical accounts of eyewitness identification processes suggest that visual behaviour may vary across procedures. Relative judgment has been widely accepted in the forensic literature as the psychological process that underlies mistaken eyewitness identification (Wells et al., 1998). In relative making a judgment, the face in the lineup that is relatively the most similar to the culprit in memory is identified. Errors are minimized, on

the other hand, if eyewitnesses are encouraged to make absolute judgments, or to identify a face solely based on their memory for the perpetrator. The incorrect identification rate is lower in sequential procedure compared to simultaneous lineups because the sequential procedure may encourage absolute judgments (Lindsay & Wells, 1985).

Explanations for differences across lineup procedures in identification performance have also been cast within signal detection or dual process memory frameworks. Ebbesen and Flowe (2001) utilized a signal detection framework, proposing that decision criterion placement is affected by lineup procedure. According to their criterion shift account, the decision standard is more liberal in a simultaneous compared to a sequential or showup procedure. Empirical tests have supported the criterion shift hypothesis (Flowe & Bessemer, 2009; Flowe & Ebbesen, 2007; Meissner, Tredoux, Parker, & MacLin, 2005). Meissner et al. (2005) also tested a dual process account of lineup identification, positing that identification in sequential lineups is based on recollection-based processes, whereas in simultaneous lineups it is based on familiarity processes. Using the remember versus know procedure (Tulving, 1985), they found that simultaneous lineup decisions were more often accompanied by 'know' or familiarity-based responses compared to sequential lineups; differences across procedures were not found, however, with respect to 'remember' or recollection-based recognition. In contrast, Gronlund (2005) found evidence that recollection processes may be superior in a sequential lineup, which suggests that perhaps a dual process framework can best account for differences in accuracy across the procedures (though see Mickes, Wais, & Wixted, 2009Q3, who provide evidence that dual process theory is compatible with signal detection theory).

There appears to be agreement across theories that faces that are lower in their similarity to the perpetrator are more likely to be identified from a simultaneous compared to a sequential lineup. Two possible general explanations for why this might be the case were examined in the present study. First, sequential presentation may lead to a lengthier (and thereby perhaps a more thorough) analysis of the test faces. Second, there may be differences across lineup procedures in the features that are used to make an identification decision. In the next sections, the background literature concerning these explanations is briefly reviewed.

Response time and lineup identification

Previous eye tracking research with simultaneous lineups has found that face dwell times (i.e. the total length of time that a face is fixated during the test) vary across lineup members in relation to identification response (Flowe & Cottrell, 2009; Mansour et al., 2009). Dwell times are longer for the positively identified face compared to any other face in the lineup. This research has also found that an incorrectly positively identified face is analysed for a longer period of time compared to a correctly identified face. Sauer, Brewer, and Wells (2008) obtained the same pattern of results in sequential lineups by measuring response times for individual faces (rather than using eye tracking to measure face dwell times). Taken together, the results of these studies are intriguing because they suggest that dwell time (and response time) for a given face indexes not only face familiarity but also face novelty. The pattern of results that has been found in these previous studies was expected for simultaneous and sequential lineups in the present study.

A given test face was also expected to be evaluated for a longer length of time in a sequential compared to a simultaneous lineup. The basis for this prediction derives from the theoretical analyses of lineup decision making that were presented earlier. Speculatively, if sequential eyewitnesses require a higher degree of match between a test face and the perpetrator in memory before making a positive identification, then they might also analyse a test face more extensively. It is important to point out that there is no way in the present study to directly test whether longer dwell times reflect the operation of a specific kind of memory retrieval process. Face dwell time not only reflects retrieval operations but also other processes, such as visual processing, error checking and deliberation. If lineup procedure is associated with face decision times, it could mean that one or more of these underlying operations are driving the effect.

Facial features and face recognition

As face processing ability develops, face analysis becomes less analytical (or feature based) and becomes more holistic instead (Karatekin, 2007). Unfamiliar faces (i.e. previously unencountered faces that have been studied on a single occasion) may be recognized on the basis of external features (hair-line, cheeks, ears and jaw-line) or internal features (eyes, nose and mouth) (Brunas, Young, & Ellis, 1990; Ellis, Shepherd, & Davies, 1979; Young, Hay, McWeeny, Flude, & Ellis, 1985). The external regions of faces may be more important in face perception and recognition than internal features. People have difficulty recognizing that the internal features of two faces are the same if they are embedded within different external contours (Maurer, Le Grand, & Mondloch, 2002;

Young, Hellawell, & Hay, 1987). Additionally, external features of unfamiliar faces when shown alone make a larger contribution to matching and recognizing faces compared to when only internal features are shown (Ellis et al., 1979; Young et al., 1985). With respect to well-known faces—such as when the face is that of a famous person, evidence from behavioural studies suggests that face recognition relies on internal features to a greater extent than external features (Bonner, Burton, & Bruce, 2003).

Eye movement research, however, has not been able to locate differences in visual behaviour in relation to face familiarity. At learning and at test, analyses of eye movements have revealed that fixations are directed toward the internal regions of the face (Althoff & Cohen, 1999; Henderson, Williams, & Faulk, 2005; Janik, Wellens, Goldberg, & Dell'Osso, 1978; Luria & Strauss, 1978; Mertens, Siegmund, & Grüsser, 1993; Stacey, Walker, & Underwood, 2005), especially the eyes and nose (Stacey et al., 2005, see also Gosselin & Schyns, 2001; Henderson et al., 2005; Janik et al., 1978).

There was no literature to guide predictions with respect to the types of facial features that may play a differential role in face recognition decisions depending on lineup procedure. Some evidence suggests that internal features may support familiarity-based processing, whereas external features may support categorization (Olivares & Iglesias, 2008). However, previous work indicates that recognition of gestalt stimuli, such as faces, is supported by familiarity as well as recollection-based processes (Yonelinas, Kroll, Dobbins, & Soltani, 1999). Additionally, in a study that investigated face recognition using a response deadline and the 'remember versus know' paradigm,

the extent to which participants fixated on internal versus external features was not related to remember versus know judgments (Mäntylä, 1997). As such, even if recollection makes a larger contribution to recognizing faces that are presented in a sequential compared to simultaneous lineup, no differences may be found with respect to where participants in each condition look on the face. Therefore, given the limited and somewhat mixed findings, no specific a priori predictions were made regarding the types of features that might be analysed to a greater extent in a sequential compared to a simultaneous lineup. Instead, a more general prediction was tested, which was that face processing in sequential compared to simultaneous lineups would be more thorough, which was indexed by how evenly participants distributed their visual gaze on internal versus external features.

EXPERIMENT 1

Participants

A total of 24 undergraduates (mean age=21.10, 13 female) participated in the experiment for course credit. All had normal or corrected-to-normal visual acuity.

Design

A 2 (lineup procedure) X 2 (phase) X 2 (target condition) X 3 (face type) X 4 (feature region) mixed design was employed. Lineup procedure (simultaneous or sequential) was controlled between subjects. Phase condition (study or test phase), target condition (target present or absent lineup), face type (target, target look-alike, or foil) and feature region (internal regions (3) versus external region (1)) were analysed within subjects. The internal region of the face included the eyes, nose and mouth regions, whereas all remaining parts of

the face comprised the external region.

Stimuli and apparatus

Similar to other investigations of face perception and memory (e.g. Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Yonelinas et al., 1999), composite drawings were utilized as face stimuli. Stimuli were created using FACES 3.0, a composite drawing software program used by law enforcement. Though these are not actual photographs of real people, and therefore, their generalizability to actual faces may be questioned, they are ideal stimuli for the purposes of the present study. The features that varied across the test faces could be precisely determined. In addition, the use of composite drawings rather than real faces allowed for manipulating the features of a face while holding as constant as possible the configural aspects of the face.

A database of 1000 faces was created by having the program generate faces at random. None of the faces had distinguishing features, facial hair, or head hair so as to focus participants' attention on the facial features. Six person lineups were formed by randomly selecting without replacement six faces from the database. From these six faces, the to-be-identified study face was randomly chosen. Additionally, two versions of the to-be-identified study face were created (please see below). A total of 24 study faces and 48 accompanying lineups (one target present and one target absent for each study face) were established. The foils faces and their position in the lineup were held constant across the target present and the target absent lineups that were created for each study face. Each participant studied 12 faces and received 12 lineups, half of which were target present and

the other half were target absent.

To manipulate target condition (present versus absent), a highly similar version of the study face (hereafter referred to as the 'look-alike') was created for every lineup. This new version of the study face was obtained by substituting one feature from the original study face for another feature; the configuration (distance between the eyes and eye height, location of the mouth and nose) was held constant. Previous lineup research has found that these feature changes reduce the rate at which the 'suspect' is chosen in simultaneous as well as in sequential lineups (Flowe & Ebbesen, 2007). Eye tracking studies have indicated that people are particularly sensitive to changes that are made to a pictorial stimulus between learning and test, with greater visual attention directed to the features that had been altered (Ryan & Cohen, 2004). Therefore, another goal of the present study was to determine whether longer dwell times would be obtained for the feature region that had been altered across study and test. For each participant, the target was present in the array on half of the trials and the look-alike was present on the other half. The order in which the target versus the look-alike appeared across the 12 lineups was randomized for every participant.

The test faces were presented together in a 3X2 array with a number placed beneath each picture for purposes of identification. For each participant, half of the targets/lookalikes appeared early in the lineup (positions 1–3), the other half, late (positions 4–6). The onscreen location of the test faces was held constant across lineup procedures. That is, the first lineup member was displayed in the upper left hand corner of the screen, the second lineup member in the middle of the upper row, etc., in each lineup condition, with the only

differences across lineup conditions being whether all of the test faces (simultaneous) or just one of the test faces (sequential) was displayed, and whether participants made a decision for each face (sequential) or for the entire lineup (simultaneous). Examples of the types of lineups employed can be seen in Flowe and Ebbesen (2007).

Mock witness testing of the lineups by another group of participants ($n=80$) indicated that the lineups were not biased against the suspect. Each mock witness evaluated 12 lineups (half target present, half target absent). They were told that the lineups were being used in a recognition study and were asked to indicate who they thought was the target in the lineup. The rate of selecting the suspect did not exceed chance expectation (.17) in any of the lineups ($M=0.16$, $SD=0.14$); target condition did not affect the results. Additionally, Tredoux' E (Tredoux, 1998) ranged from 5.31 to 6.54 across the lineups, suggesting that the foils were adequate.

Eye movements were recorded by an Eyelink II (SR Research) video-based eye tracker with a temporal resolution of 250 Hz and a spatial resolution of 0.28; the eye tracker default settings for cognitive research were utilized. Data from the eye with less error during calibration were used for analysis.

Procedure

Participants were told that they would be given a list of faces to study and to pay careful attention because they would have to recognize them on a later test. Following a calibration phase, the 12 study faces were presented in succession for 10 seconds each, with an interstimulus interval of 3 seconds. The delay between the study period and the test period was 5 minutes; during the delay, the eye tracker was recalibrated.

In the test phase, each participant was given 12 lineup tests, half of which were target absent. Participants were not aware of the number of lineups that would be target present versus target absent. Before the lineup tests commenced, participants were cautioned that only one face, if any, could be identified from a given lineup. They were further told, 'Reject the lineup if none of the faces was a perfect match to one of the study faces. A study face may or may not be present in the lineup, and not identifying anybody could also be a correct response'. Cautionary instructions such as these have been recommended to the criminal justice system (Malpass & Devine, 1981; Wells et al., 1998). Research indicates that participants adopt a stricter judgment standard when cautioned that the lineup may not contain the perpetrator (Clark, 2005).

Each identification trial began with a drift correction wherein a central fixation point was displayed on the screen followed by the identification test. No response deadline was imposed. Participants verbally indicated their response to the experimenter who entered the response using a keyboard. Eye movement data were recorded from the onset of the stimulus until the participant made a verbal identification response, at which point the experimenter pressed a button to terminate the end of the recording session for the trial.

After a test face was positively identified in the sequential condition, if any test faces remained to be seen, they were shown to the participant. Participants were aware that they could not change their response to the previously identified face. Therefore, they had to reject each of the remaining faces. This procedure was undertaken to equate the number of test faces that were shown across the simultaneous and sequential testing conditions.

Dependent measures and data analysis

Identification responses were coded for each lineup and included: All of the faces were rejected, the 'suspect' (i.e. the target or the target look-alike) was positively identified, or a foil was positively identified. The rejection rate and the rate of identifying the target, the target look-alike or a foil were computed across lineups for every participant.

Several measures were derived from the eye movement data. Total face dwell time was calculated for every face by summing across the duration of every fixation that was made to the internal and external regions of the face during a trial. Feature dwell time was also determined for the eyes, nose, mouth and face regions by summing up the total length of time spent fixating within the respective region during a trial. Total face dwell time and feature dwell time were then separately averaged across lineups for each participant, conditioning the data on face type (target, target look-alike or foil), identification outcome ('suspect' positive identification, foil positive identification or lineup rejection) and target condition (target present or target absent).

Total face dwell time and feature dwell time were expected to vary for the study compared to the test faces and also across lineup conditions. These expected differences had to be controlled for in comparing how people distributed their time across the features of the face at learning and at test. Therefore, a feature distribution score was computed for every feature region by dividing feature dwell time for the respective region by total face dwell time. As a result, the proportion of the trial that a participant spent examining each of the internal feature regions (the eyes, nose and mouth) and the external region were obtained for each of the study and test faces. This methodology has been employed to

gauge the relative importance of different facial features in previous eye tracking studies (e.g. Henderson et al., 2005).

Data analysis

The dependent measures were analysed with ANOVAs and t-tests; partial eta-squared (η_p^2) is provided as a measure of effect size. The dwell time data were positively skewed, which is typical for such data; therefore, the data were square root transformed before submitting them to inferential statistical analysis. Because the pattern of results was the same for the transformed and untransformed scores, the descriptive statistics that are provided are based on the untransformed data.

Finally, the dwell time and feature distribution variables were analysed in relation to the particular feature substitutions that were made to create the target look-alike to examine whether the feature changes affected visual processing. Unfortunately, the manner in which the target and the target look-alikes were examined did not vary as a function of the particular feature substitution that was made; therefore, the feature-substituted look-alikes were treated as a single entity in the analysis.

RESULTS

Preliminary analyses

Identification performance was examined as a function of test trial number (1–12) within each lineup condition to examine the effects of repeated testing. First, choosing was examined in relation to trial number to determine whether participants were more likely to choose on later test trials compared to earlier trials, which would suggest that the decision standard adopted at test was becoming increasingly liberal across test trials;

the relationship was not significant for neither the simultaneous ($r=-.07$, $p=.72$) nor the sequential ($r=.18$, $p=.57$) conditions. Second, accuracy across trial number was examined to test for proactive interference effects. If proactive interference was building as testing progressed, accuracy should decrease as the test trials accumulated in number; no evidence of proactive interference was found, however, in the simultaneous ($r=-.15$, $p=.64$) or sequential ($r=.09$, $p=.82$) conditions. Third, a change in decision strategy across trials would be indicated if face dwell time varied in relation to lineup trial number. Average face dwell time was computed for every lineup and then correlated with trial number; the results indicated no significant association in the sequential condition ($r=-.01$, $p=.82$), and a slight but non-significant increase in face dwell times across trials in the simultaneous condition ($r=.12$, $p=.16$). Therefore, lineup responses and dwell time data were collapsed across trials in all subsequent analyses.

Study and target faces compared

The target test faces (the look-alikes were excluded from the analysis) were compared to the study faces with respect to the response time and feature distribution scores. Data were collapsed across lineup condition. The purpose of the comparison was to examine whether the visual behaviour directed toward a face varied across encoding and retrieval. The results are displayed in Figure 1.

A target face was evaluated more quickly when presented at test compared to when the face was first studied ($M=3.11$ seconds versus $M=10.00$ seconds), $t(23)=22.58$, $p<.01$. There were also differences across study and test in where

fixations were directed. Participants spent a larger proportion of time examining the internal features when the face was presented at study compared to at test ($M=0.80$ versus $M=0.65$, respectively), $F(1, 22)=6.22$, $p<.05$, $\eta_p^2=.22$. At study, a significantly larger proportion of time was spent analysing the eyes ($M=0.48$) compared to the nose ($M=0.30$), face ($M=0.14$) or mouth ($M=0.11$) (all $ps<.01$). In contrast, feature distribution scores were more evenly distributed across the eyes ($M=0.31$), face ($M=0.32$) and nose ($M=0.21$) regions (i.e. all pairwise comparisons of the means were not significant) during the test phase. At test, less time was spent on the mouth ($M=0.14$) compared to the eyes and face (all $ps<.01$).

Lineup procedure and identification outcomes

Identification outcomes across lineup procedure and target condition are provided in Table 1. As shown, the rate of rejecting the lineup was significantly higher in the target absent compared to the target present condition ($M=0.39$ versus $M=0.25$, respectively), $F(1, 28)=7.77$, $p<.05$, $\eta_p^2=.26$. The distribution of identification outcomes did not significantly differ across sequential and simultaneous lineups in the target present condition (suspect: $M=0.35$ and 0.43 , foil: $M=0.36$ and 0.36 , lineup rejection: $M=0.29$ and 0.21 , respectively) or in the target absent condition (suspect: $M=0.31$ and 0.31 , foil: $M=0.36$ and 0.25 , lineup rejection: $M=0.33$ and 0.44 , respectively).

The identification response data suggest that the target look-alike was very similar to the target, as the rates of identifying the target versus the target look-alike were not statistically different (see Table 1). Therefore, in the analyses that follow, data were collapsed across the target present and target absent conditions. The target and the target

look-alike face will hereafter be referred to as the 'suspect' face. The rationale for doing this was three-fold: Identification rates for the target and the target look-alike did not vary. Second, none of the forthcoming results were affected by target condition. Third, excluding the data from the target absent trials would have resulted in a considerable loss of statistical power.

Identification outcome and visual behaviour

Total face dwell time data and feature distribution scores were separately entered into a 2 (lineup procedure) × 2 (face type) mixed ANOVA. Separate analyses were conducted for trials in which the suspect was positively identified, trials in which a foil was identified, and trials in which the lineup was rejected. Descriptive results for total face dwell time by lineup procedure, face type and identification response are presented in Figure 2; descriptive results for the feature distribution data are presented in Figure 3. Suspect identifications

The descriptive data for the suspect identification trials (Figure 2, left panel) suggested that faces were examined for a shorter length of time in simultaneous compared to sequential lineups, especially if the face was a foil. These results are consistent with the prediction that face processing in sequential lineups is more thorough. Indeed, significant effects were found for lineup procedure, $F(1, 22)=8.63, p<.01, \eta_p^2=.28$ and face type, $F(1, 22)=74.04, p<.05, \eta_p^2=.77$, and these variables significantly interacted, $F(1, 22)=15.50, p<.01, \eta_p^2=.41$. Simultaneous participants visually analysed a foil face for a significantly shorter period of time compared to sequential participants ($M=1.11$ seconds versus $M=2.91$ seconds, respectively), $p<.001$, whereas suspect

dwell times did not significantly vary depending on lineup procedure (simultaneous $M=3.58$ seconds and sequential $M=4.07$ seconds). In both the simultaneous and sequential conditions, a foil face was analysed for a significantly shorter period of time than the target face ($p < .01$).

As suggested by the descriptive results (Figure 3, top panel), participants apportioned significantly more time on internal facial features in the simultaneous compared to the sequential condition in making a positive identification of the suspect ($M=0.65$ versus $M=0.48$, respectively), $F(1, 21)=7.67$, $p < .05$, $\eta_p^2=.26$. As for the allotment of time across specific internal facial features, lineup procedure did not produce any significant effects.

Foil identifications

A total of 11 sequential participants were included in the analysis of foil identification trials; one participant was excluded because this person picked a foil just once and that was before they saw the suspect. Hence, eye tracking data for the suspect were not obtained for this participant. Additionally, data were conditioned on three types of faces: non-identified suspect, non-identified foil and positively identified foil.

As was found in the suspect identification trials, simultaneous participants analysed nonidentified faces more quickly than sequential participants (Figure 2, middle panel). The interaction between lineup procedure and face type was significant, $F(2, 42)=4.08$, $p < .05$, $\eta_p^2=.16$. The positively identified foil face was analysed for the same length of time in a sequential compared to a simultaneous lineup ($M=4.04$ versus $M=3.90$, respectively). However, non-identified faces were analysed for a shorter period of time in a

simultaneous lineup (suspect $M=1.70$ and foil $M=1.37$, respectively) compared to a sequential lineup (suspect $M=3.31$ and foil $M=2.82$, respectively), $p < .01$.

The length of time that the positively identified face was examined was further analysed across the suspect and the foil identification trials. Results indicated that there was no difference in dwell time depending on the type of face that was identified in neither simultaneous nor sequential lineups.

Average feature distribution scores as a function of face type and lineup procedure are shown in Figure 3 (middle panel). On the trials in which a foil was identified, internal features were examined to a greater extent in simultaneous compared to sequential lineups ($M=0.67$ versus $M=0.46$, respectively), but the effect was only marginally significant $F(1, 21)=3.67$, $p=.07$, $\eta_p^2=.15$. No other effects on feature distribution scores in relation to lineup procedure were found.

Lineup rejections

Figure 2 (right panel) displays dwell time data across conditions for the rejection trials. Lineup procedure had a significant main effect on total dwell time, $F(1, 22)=20.73$, $p < .04$, $\eta_p^2=.48$. Faces were evaluated more quickly in the simultaneous ($M=1.85$) compared to the sequential ($M=3.27$) condition. A significant main effect was also obtained for face type, $F(1, 22)=9.85$, $p < .01$, $\eta_p^2=.38$. Foil faces were evaluated for a longer period compared to suspect faces ($M=2.82$ versus $M=2.30$), which may indicate that participants may have been deliberating over picking a foil in some cases. Lineup procedure and face type did not interact.

Lineup procedure did not significantly influence the proportion of time spent examining

the internal region of the face (simultaneous $M=0.63$ versus sequential $M=0.53$). Feature distribution results by lineup procedure for rejection trials are summarized in Figure 3 (bottom panel).

DISCUSSION

Participants in both lineup conditions spent a comparable length of time analysing the face that was ultimately positively identified. Simultaneous participants, however, tended to reject non-identified faces faster than sequential participants. Participants mostly allotted their time to processing internal features when learning a face. In comparison, external facial features enjoyed more visual processing at test compared to study, especially in the sequential condition. Taken together these results are consistent with the hypothesis that participants analyse faces more thoroughly if a lineup is presented sequentially.

The results indicated that the visual processing of foil faces was affected by whether the test faces were viewed in an array or one at a time. Experiment 2 addressed whether the visual processing strategy adopted in a sequential lineup is comparable to the analysis strategy undertaken during a showup test. In both showups and sequential lineups, participants view the suspect face without being able to compare him or her to any foils. The primary difference between the procedures, of course, is that no distractor faces are presented in a showup. The next experiment examined visual behaviour that was directed toward the 'suspect' in relation to whether foil faces were presented at the test.

Research suggests that even though sequential participants cannot visually compare the suspect to any of the foil faces, identification responses are nevertheless affected by foils. 'Suspects' are more likely to be identified in a low compared to high similarity lineup

(Clark & Davey, 2005; Flowe & Ebbesen, 2007), regardless of lineup procedure. However, in unfair lineups, the sequential lineup affords greater protection against the false identification of 'innocent suspects' compared to sequential lineups (Carlson, Gronlund, & Clark, 2008). Experiment 2 compared visual behaviour toward the 'suspect' face in a showup to a sequential lineup procedure. The goal was to determine whether foil faces affect visual analysis of the suspect. Speculatively, knowing that there are foil faces may lead participants to examine the suspect's face less thoroughly. Evidence of this would be demonstrated by the suspect being analysed more quickly and fixations being distributed less evenly across facial features in the sequential compared to the showup condition.

EXPERIMENT 2

Participants

A total of 12 undergraduates participated in the experiment for course credit. All had normal or corrected-to-normal visual acuity.

Design

A 2 (study versus test phase) X 2 (target condition) X 4 (feature region) repeated measures design was employed. The independent variables were defined in the same way as Experiment 1.

Stimuli and apparatus

The faces that were presented during the study phase (n=12) were randomly chosen from the database described in Experiment 1. The faces presented at test (n=12) were either targets or look-alikes. Target faces were identical to faces that had been studied. Look-alike faces were a highly similar version of the study face.

Eye movements were recorded using the same procedures described in Experiment 1.

Procedure

The study phase procedures were the same as those employed in Experiment 1: Participants were forewarned that they would be tested on the study faces, which were presented for 10 seconds each, with an interstimulus interval of 3 seconds. The delay between the study period and the test period was 5 minutes; during the delay, the eye tracker was recalibrated.

In exactly the same manner as the lineup participants in Experiment 1, showup participants were cautioned that they should not identify a test face unless it was a perfect match to one of the study faces. Each identification trial began with a drift correction wherein a central fixation point was displayed on the screen followed by the identification test. No response deadline was imposed. Participants verbally indicated their response to the experimenter, who then entered their answer using a keyboard. A total of 12 test faces were shown, half of which were targets, and the other half were the target look-alikes.

Dependent measures and data analysis

Identification responses (positive identification or rejection) were computed across trials for every participant. Accuracy was determined by conditioning the response data on whether the test face was a target or a look-alike. Correct responses included positively identifying a target face or correctly rejecting the look-alike. Inaccurate responses included rejecting the target face or incorrectly positively identifying the look-alike.

For the study and test faces, total face dwell time, feature dwell time and feature distribution were calculated following the procedures described in Experiment 1. The

measures were conditioned on phase (study or test), identification outcome (whether the test face was identified or rejected) and target condition (target or the target look-alike was presented).

The dependent measures were analysed with ANOVAs and t-tests; partial eta-squared (η_p^2) is provided as a measure of effect size. Overall face dwell time and feature dwell times were square root transformed before submitting them to inferential statistical analysis; the provided descriptive statistics are based on the untransformed data.

RESULTS

Study and test faces visual behaviour compared

As found in Experiment 1, the target test faces were analysed for a significantly shorter period of time ($M=5.13$) compared to the study faces (10 seconds), $t(11)=-5.05$, $p<.05$. Additionally, the internal features of the study faces were analysed for a greater proportion of time compared to the target test faces ($M=0.90$ versus $M=0.69$, respectively), $t(11)=3.54$, $p<.01$.

Figure 4 displays the feature distribution scores for the study and target faces. At study, a significantly larger proportion of time was spent analysing the eyes ($M=0.51$) compared to the nose ($M=0.21$), face ($M=0.09$) or mouth ($M=0.11$) (all $ps<.01$). In contrast, at test, a larger proportion of time was spent examining the eyes ($M=0.32$) and face ($M=0.31$) when compared to either the nose ($M=0.21$) or mouth ($M=0.15$) (all $ps<.01$).

Testing procedure and identification outcomes

Identification responses for showups and the sequential lineup condition in Experiment

1 were compared. Accuracy overall was significantly higher in a showup ($M=0.60$)

compared to a sequential lineup ($M=0.34$), $F(1, 22)=18.85$, $p<.05$, $\eta_p^2=.46$.

However, the accuracy rate for the target absent sequential condition also includes foil identifications. Therefore, suspect identification rates in showups and sequential lineups were compared next. The results indicated that suspect identification rates did not significantly vary in relation to identification procedure in neither the target present nor in the target absent condition. In showups, the average rate of identifying the target was 60%, whereas the average rate of identifying the look-alike was 40%. In sequential lineups, the averages were 53 and 34%, respectively.

Is visual behaviour in showups and sequential lineups similar?

Total face dwell time length for the suspect face was lower in the sequential condition, but the difference was not statistically significant, (sequential $M=3.58$ versus showup $M=5.14$), $t(22)=-1.38$, $p=.18$, two-tailed.

Feature distribution scores were analysed with a 2 (identification outcome: identify or reject the suspect) \times 2 (identification procedure) ANOVA. For the sequential lineup condition, only the trials in which the participant saw the suspect and only the trials in which either the suspect was positively identified or the lineup was rejected were included in the analysis. Results indicated a significant main effect for identification procedure, with a greater proportion of time directed to internal facial features in the showup compared to the sequential condition ($M=0.68$ and 0.51 , respectively), $F(1, 22)=9.96$, $p<.01$, $\eta_p^2=.31$.

GENERAL DISCUSSION

Eye tracking was employed to explore whether visual behaviour varies in relation to the type of identification test that is used. Experiment 1 examined whether faces in a sequential compared to a simultaneous lineup are analysed more thoroughly, which was indexed by longer dwell times and fixations that were more evenly distributed across facial features. In Experiment 2, visual behaviour in a sequential lineup was compared to a showup test. The experiment tested whether the 'suspect' is analysed more thoroughly when there are no foil faces presented a test.

The results indicated that dwell time length in both simultaneous and sequential lineups was independent of whether the positively identified face was a suspect or a foil. Faces that were positively identified were analysed for a longer length of time than faces that were not identified, a result that is in keeping with previous research (Flowe & Cottrell, 2010; Mansour et al., 2009; Sauer et al., 2008). Results further indicated that the length of time that a face was visually analysed varied in relation to lineup condition. Faces that were not positively identified were evaluated for a longer period of time in the sequential compared to the simultaneous condition, whereas positively identified faces were analysed for a comparable length of time across conditions. These findings suggest that sequential participants are possibly more thorough in their analysis of faces that they perceive to be novel (i.e. unfamiliar) than are simultaneous participants. There was no difference across procedures in how long participants examined faces that they perceived as familiar enough to positively identify.

Sequential participants may have undertaken a more extensive examination of

unfamiliar test faces than did simultaneous participants because they had only one opportunity to view a given face. Once a decision was made with respect to that face, the next one was shown; they could not review previously seen faces or change any of their previous responses. In a simultaneous lineup, participants may engage in a process of elimination whereby they exclude faces that are relatively less familiar and then analyse the best matches in the lineup more extensively (see Dunning & Stern, 1994Q6). As a consequence, some faces get less attention than others in a simultaneous lineup, whereas faces are analysed more extensively, regardless of their familiarity, in a sequential lineup. Clearly further controlled experimentation is needed to examine these possibilities. As mentioned previously, dwell time not only indexes recognition processes but also includes other processes, such as vision, deliberation, error checking, as well as others. Therefore, caution is warranted in drawing conclusions based on dwell time data about recognition processes.

Experiment 2 tested whether foil faces influence on how long people analyse the 'suspect' face. Including foils at test may literally be distracting, leading people to spend less time analysing the suspect. To evaluate this possibility, 'suspect' dwell times were compared across a showup and a sequential lineup procedure. Though face dwell times were shorter in the sequential compared to the showup condition, the difference was not significant. Therefore, foils in a sequential lineup do not appear to decrease the length of time that a suspect's face is analysed. On the other hand, foils affect how information is extracted from the suspect's face, but dwell times may not be sensitive enough to capture this.

Identification procedure did impact how people distributed their fixations across the

features of the test faces. Sequential participants apportioned more of their time to the external regions of the face compared to simultaneous and showup participants. External regions of faces are important in face perception and recognition. The internal features of two faces will be perceived as being dissimilar if they are embedded within different external contours (Maurer et al., 2002; Young et al., 1987). Additionally, faces are better recognized when solely external features versus solely internal features are presented (Ellis et al., 1979; Young et al., 1985). If external features of faces convey important information concerning identity, perhaps sequential lineups reduce false alarms because the procedure encourages participants to focus on external features of faces to a larger extent compared to other identification procedures.

Whether external features support recollection based processes is a question that should be addressed by future research. Gronlund (2005) proposed that recollection processes may be relied on to a greater extent in sequential lineups. The type of process (familiarity or recollection) encouraged by the test may influence where people look on the face. The difference in how visual behaviour is directed toward internal versus external features could also tie in with the theoretical notion that a different decision standard is adopted depending on procedure as suggested by Ebbesen and Flowe (2001). Previous research has shown that dissimilar external features across face pairs lead people to adopt more conservative decision criteria in face matching tests without affecting discrimination (Fletcher, Butavicius, & Lee, 2008; Patterson & Baddeley, 1977). Applied to criminal identification tests, if the external features of a test face do not strongly match the perpetrator in memory, then the face will more likely be rejected in a sequential lineup

compared to other identification procedures. However, the present study is not a strong test of this notion, and further research is needed to follow-up on the results that were obtained. Such work could examine which aspects of faces are strongly represented in memory and how this ties in with recognition processes across criminal identification tests. Such research would also go a long way toward providing information that could help investigators construct fairer lineups.

The question of whether these findings would hold in applied settings also needs to be addressed. In the present study, the number faces presented at test was equated across the sequential and simultaneous conditions. This procedure may have lowered the decision standard that is typically adopted in sequential lineups. On the one hand, on average across trials, the false alarm rate in the sequential condition did not vary from that obtained in the simultaneous condition, suggesting that decision criterion placement was similar across lineup procedures. Despite this being the case, sequential participants were still more thorough in their analysis of the test faces than were simultaneous participants, suggesting that they had adopted a different decision standard. Perhaps if sequential participants had not seen the remaining faces in the lineup, their tendency to more fully analyse the test faces compared to showup and simultaneous participants may have even more pronounced and false alarm rates would have been significantly lower in the sequential condition.

Another limitation of the current study is that the stimuli were tightly controlled images of faces produced by composite drawing software. The configural relations across the test faces were held constant as a starting point for examining differences across testing procedures in the analysis of facial feature regions. The reasoning behind these

methodological choices was to exert laboratory control so as to maximize internal validity. The visual processing strategy adopted at test is likely to be affected the dimensions of faces that were tightly controlled in the present study. Facial hair (e.g. Read, Vokey, & Hammersley, 1990) and hairstyle (e.g. Memon & Gabbert, 2003) are features that affect recognition performance, though very little lineup research has been done on these factors. Additionally, in eyewitness identification research, facial hair is not usually included on the test faces, and hairstyle and feature configuration is not systematically defined or measured to determine how variable this feature is across the members. Since it is not known how these factors affect recognition performance, the current study eliminated them to examine how facial feature visual analysis contributed to lineup recognition. The dwell time results that were found in the present study map onto those found by Mansour et al. (2009), who used photographs of actual people as stimuli in their analysis of face dwell times in simultaneous lineups. Therefore, it is likely that the effects found in the present study extend to photographic stimuli.

To conclude, the present study found that different visual strategies are adopted across identification tests. Non-identified test faces were analysed more thoroughly in a sequential compared to a simultaneous procedure when a positive identification was made. Additionally, visual processing in sequential lineups could be differentiated from showups, which suggests that face recognition processes may be different in sequential lineups.

References

- Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 997-1010.
- Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition*, 10, 527–536.
- Brunas, J., Young, A. W., & Ellis, A. W. (1990). Repetition priming from incomplete faces: Evidence for part to whole completion. *British Journal of Psychology*, 81, 43-56.
- Carlson, C., Gronlund, S., & Clark, S. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, 14(2), 118-128.
- Clark, S.E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behaviour*, 29, 395–424.
- Clark, S., & Davey, S. (2005). The Target-to-Foils Shift in Simultaneous and Sequential Lineups. *Law and Human Behavior*, 29(2), 151-172.
- Dunning, D., & Stern, L. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, 67(5), 818-835.
- Ebbesen, E. B., & Flowe, H. D. (2001). Simultaneous v. Sequential Lineups: What Do We Really Know? Retrieved February 18, 2009, from <http://wwwpsy.ucsd.edu/%7eeebbesen/SimSeq.htm>.

- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*, 431-439.
- Fletcher, K., Butavicius, M., & Lee, M. (2008). Attention to internal face features in unfamiliar face matching. *British Journal of Psychology, 99*(3), 379-394.
- Flowe, H. D., & Bessemer, A. (2010). The Effects of Target Discriminability and Criterion Placement on Accuracy Rates in Sequential and Simultaneous Target Present Lineups, *Psychology Crime, & Law*, in press.
- Flowe, H. D., & Cottrell, G. W. (2010). An Examination of Simultaneous Lineup Identification Decision Processes Using Eye Tracking. *Applied Cognitive Psychology*, in press.
- Flowe, H. D., & Ebbesen, E. B. (2007). The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law and Human Behavior, 31*, 33-52.
- Flowe, H. D., Ebbesen, E. B., & Mehta, A. (2010). The role of eyewitness identification evidence in felony case dispositions. Manuscript under review.
- Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology, 19*(1), 23-37.
- Gosselin, F., & Schyns, P. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261-2271.
- Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*(1), 98-106.
- Janik, S. W., Wellens, A. R., Goldberg, M. I., & Dell'Osso, J. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Perceptual & Motor Skills, 47*, 857-858.

- Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review, 27*(3), 283-348.
- Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*(3), 556-564.
- Lindsay, R., Pozzulo, J., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21*(4), 391-404.
- Loftus, E., Loftus, G., & Messo, J. (1987). Some facts about 'weapon focus.'. *Law and Human Behavior, 11*(1), 55-62.
- Luria, S. M., & Strauss, M. S. (1978). Comparison of eye movements over faces in photographic positives and negatives. *Perception, 7*, 349-358.
- Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66*(4), 482-489.
- Mansour, J., Lindsay, R. C. L., Brewer, N., Munhall, K. G. (2009). Characterizing visual behavior in a lineup task. *Applied Cognitive Psychology, 23*(7), 1012-1026.
- Mäntylä, T. (1997). Recollections of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1203-1216.
- Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. *Trends Cogn. Sci. (Regul. Ed.) 6*, 255-260.
- McQuiston-Surrett, D.E., 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),

137-169.

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.

Memon, A & Gabbert, F. (2003). Unravelling the effects of a sequential lineup. *Applied Cognitive Psychology, 6*, 703-714

Mertens, I., Siegmund, H., & Grüsser, O. (1993). Gaze motor asymmetries in the perception of faces during a memory task. *Neuropsychologia, 31*, 989-998.

Olivares, E. I., & Iglesias, J. (2008). Brain potentials and integration of external and internal features into face representations. *International Journal of Psychophysiology, 68*, 59-69.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 406-417.

Read, J., Vokey, J., & Hammersley, R. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy-confidence relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 870-882.

Ryan, J. D., & Cohen, N. J. (2004). The nature of change detection and on-line representations of scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30*(5), 988-1015.

Sauer, J., Brewer, N., & Wells, G. (2008). Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Legal and Criminological Psychology, 13*(1), 123-135.

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity:

- Evidence from eye-movement data. *British Journal of Psychology*, 96, 407-422.
- Stebly, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, 25(5), 459-473.
- Stebly, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, 27(5), 523-540.
- Tanaka, J. W., & Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, 46, 225-245.
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, 25, 583-592.
- Tredoux, C.G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior*, 22, 217-237.
- Tredoux, C.G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology*, 13, 9-26.
- Tulving, E. (1985). Ebbinghaus's memory: What did he learn and remember? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(3), 485-490.
- Valentine, T., Darling, S., & Memon, A. (2007). Do strict rules and moving images increase the reliability of sequential identification procedures? *Applied Cognitive Psychology*, 21(7), 933-949.
- Wagenaar, W. A., & Veefkind, N. (1992). Comparison of one-person and many-person lineups: A warning against unsafe practices. In F. Lose, D. Bender, & T. Bliesener (Eds.), *Psychology*

and law: International perspectives (pp. 275-285). Berlin: de Gruyter.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998).

Eyewitness identification procedures: Recommendations for lineups and photospreads.

Law and Human Behavior, 22(6) 603-647.

Yonelinas, A., Kroll, N., Dobbins, I., & Soltani, M. (1999). Recognition memory of

faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review*, 6(4), 654-661.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14, 737-746.

Table 1. Identification outcomes across simultaneous and sequential lineups in Experiment 1.

	Target Present			Target Absent		
	<i>Suspect</i>	<i>Foil</i>	<i>Reject</i>	<i>Suspect</i>	<i>Foil</i>	<i>Reject</i>
Sequential	0.35	0.36	0.29	0.31	0.36	0.33
Simultaneous	0.43	0.36	0.21	0.31	0.25	0.44
	0.39	0.36	0.25	0.31	0.31	0.39

Figure Captions

Figure 1. Mean feature distribution scores (+1 SEM) for the target face across feature regions in the study phase compared to the test phase in Experiment 1.

Figure 2. Mean dwell time across feature regions by face type (foil or suspect) and lineup procedure (simultaneous or sequential) for suspect identification trials (left), foil identification trials (middle) and rejection trials (right) in Experiment 1.

Figure 3. Mean feature distribution scores (+1 SEM) across feature regions by face type (foil or suspect) and lineup procedure (simultaneous or sequential) for suspect identification trials (left), foil identification trials (middle) and rejection trials (right) in Experiment 1.

Figure 4. Mean feature distribution scores (+1 SEM) for the target face across feature regions in the study phase compared to the test phase in Experiment 2.

Figure 1

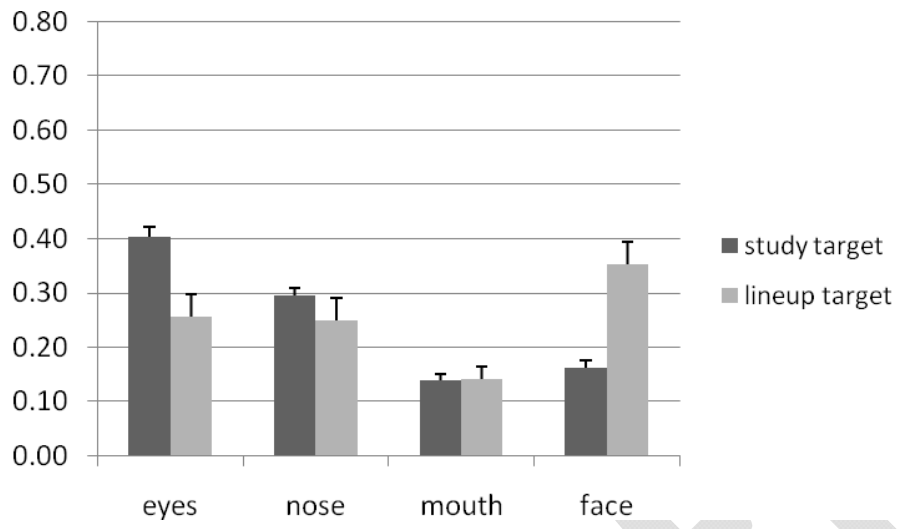
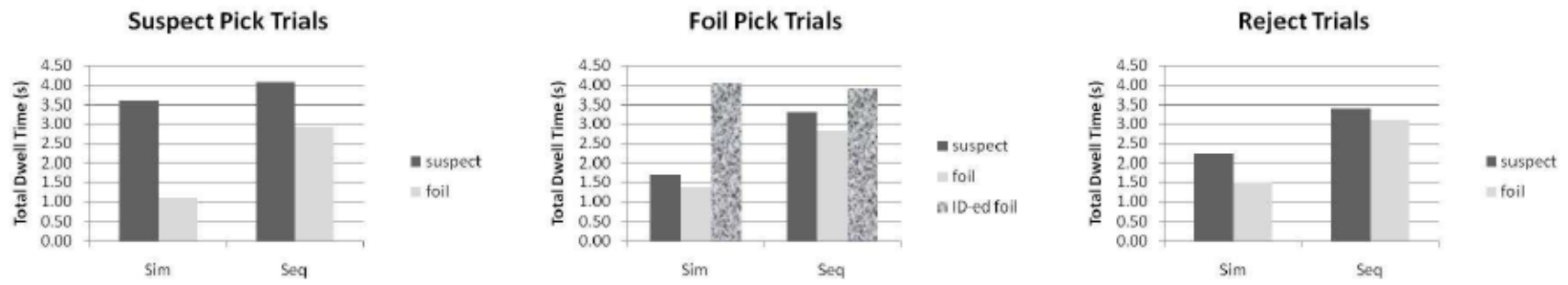


Figure 2



Draft

Figure 4

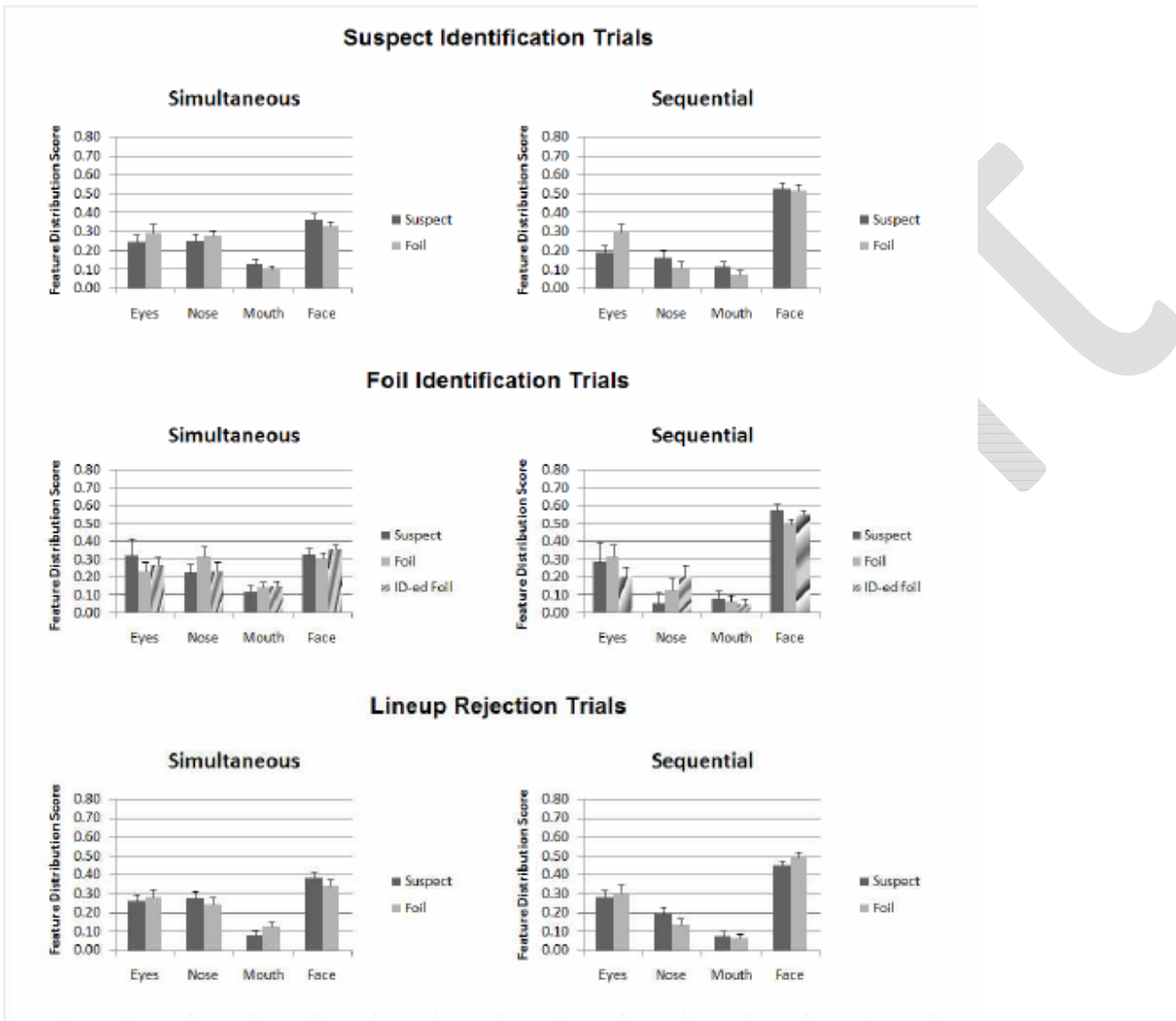


Figure 4

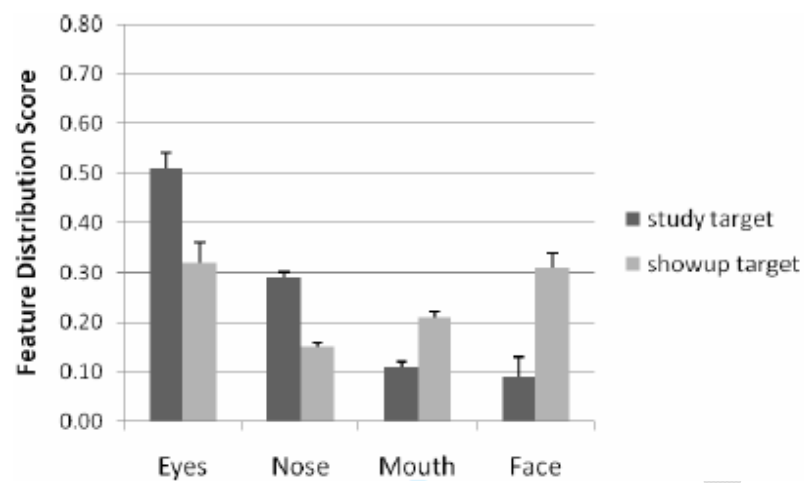


Figure 3

