5-14-2023

Comparing Response Outcomes for Simultaneous and Sequential Lineup Procedures

Gabriella Larson
Syracuse University

Abstract

After witnessing a crime, eyewitnesses are typically presented with a six-person lineup, either simultaneous or sequential, and asked to pick out the perpetrator from the six faces presented. These eyewitnesses may or may not be asked to provide a confidence rating for their decision. Current research remains split on if simultaneous or sequential lineups provide the best opportunity for correct identifications of the perpetrator (hits) while limiting incorrect identifications of innocent lineup members (false alarms or foil IDs), though most recently there has been a shift towards adopting the sequential procedure in police departments. Furthermore, it is not clear how accuracy shifts in the absence of a sequential stopping rule and if the act of giving confidence ratings impact lineup response outcomes and therefore should or should not be mandated for police eyewitness tasks. The experiment used a 2x2x2 design testing lineup type, sequential stopping rule procedure, and giving confidence ratings. Signal Detection Theory modeling is used to determine which popular measurement models best fit the observed data and the implications for their theoretical underpinnings. Results from the current study do not indicate a sequential advantage for discriminability over simultaneous lineups, nor a significant difference when using a stopping rule or not. Concurrent with expectations, the results indicate that giving confidence ratings do not impact the proportion of hits or false alarms across lineup types and procedures. Additionally, results lend support towards the application of Diagnostic Feature Detection Theory (Wixted & Micked, 2014) towards both simultaneous and sequential lineups.

Comparing Response Outcomes for Simultaneous and Sequential Lineup Procedures

by

Gabriella Larson

B.S., Suffolk University, 2020

Thesis

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Psychology

Syracuse University

May 2023

Table of Contents

Master's Thesis: Comparing Response Outcomes for Simultaneous and Sequential Lineup Procedures

Overview

A goal of eyewitness identification research is to identify which procedures optimize identification results by increasing correct identifications and simultaneously decreasing misidentifications. Historically, simultaneous lineups were more commonly used over sequential lineups by law enforcement until new research proposed the sequential superiority effect, claiming that sequential lineups provided more correct identifications and fewer misidentifications, which led to an increase in the implementation of sequential lineups (Lindsay & Wells, 1985; Wixted et al., 2019). Researchers in the field of eyewitness identification remain split and continue to claim that this change may have been adopted too soon, with research suggesting that simultaneous lineups may still outperform sequential lineups (Clark 2012; Mickes, Flowe, Wixted, 2012). Furthermore, this increased adoption of sequential lineups has necessitated further research on task dependent features, such as using a stopping rule to terminate lineup progression, which has been overlooked in current research. Current research tends to implement a stopping rule, or in the absence of one, to still count the first identification made (Wilson et al., 2019; Dunn et al., 2022). This leaves the question of if multiple identifications were allowed in a single sequential lineup, if subsequent identifications could be counted as corrections to previously mistaken identifications. This debate is usually enhanced through the use of measurement models to quantify eyewitness performance in the lab, as understanding which popular measurement models and their corresponding theories best fit eyewitness identification data can assist in understanding how an eyewitness formulates their decision on either choosing a lineup member or rejecting a lineup. The current study aims to further this debate by identifying which lineup procedure, simultaneous or sequential, is preferred, while additionally comparing lineup results with and without a sequential stopping rule and the use of confidence judgements. This thesis also asks which signal detection theory model provides the best fit to the data.

Background

Eyewitness identifications have a large impact on the justice system. A successful identification increases the possibility of keeping a perpetrator off the streets and a wrongful identification can assist in ruining an innocent person's life by resulting in a wrongful conviction. The National Registry of Exonerations keeps a running list of all exonerations since 1989 and out of the 3063 exonerations listed, 829 of the convictions involved eyewitness misidentifications, of which 311 relied on DNA evidence for the exonerations. The only other contributing factors that resulted in overall higher rates of exonerations than eyewitness misidentifications were cases where people other than the defendant committed perjury or where police, prosecutors, or other government officials abused their power. When only considering DNA exonerated cases, eyewitness misidentification was the single highest contributing factor (Gross, Shaffer, & National Registry of Exonerations 2012).

In 2010 the Police Executive Reform Forum distributed 1377 surveys to police and law enforcement precincts with the hopes of collecting information on how eyewitness identifications are handled at each level. Six hundred and nineteen of the surveys were returned for analyses which overwhelmingly indicated that the responding departments had no formal written policies for how showups, photo lineups, or live (in person) lineups are to be conducted (77.1 percent, 64.4 percent, and 84.2 percent, respectively). Roughly 85 percent of departments that use showups or photo lineups request, but do not require, witnesses to give statements of certainty if a positive identification is made even though the majority of eyewitness research implement these statements of certainty in their research tasks (Police Executive Reform Forum, 2014). This highlights a disconnect between research and public policy, leading to a need for formally accepted and standardized proceedings. Should these statements of clarity be mandated for police to obtain with every identification, or does the act of giving these judgements impact how an eyewitness responds to a lineup? Understanding the basis of how these identifications, and misidentifications, are made is necessary in order to provide acceptable best practices for law enforcement to maximize correct identifications while concurrently minimizing misidentifications.

Simultaneous and Sequential Lineup Procedures

After viewing a crime, there are three main ways that law enforcement may have an eyewitness attempt to make an identification. The first of the three ways is through a showup, where a single person is presented and the eyewitness must decide if that person is recognized or not. The other two procedures are simultaneous and sequential lineups, which present the same number of faces (usually six in the United States) but in varying ways. In a simultaneous lineup the eyewitness must search through all faces presented at the same time for a recognized perpetrator (the target face) among novel faces (foils or lures). Simultaneous lineups require the eyewitness to make up to two distinct judgements: if the perpetrator is present or not, and if so, which face is the perpetrator. Sequential lineups differ in that faces are presented one after the other and a recognition decision must be made for each presented face before the subsequent one can be shown.

Simultaneous and sequential lineups are compared to determine which procedure can yield the greatest number of correct identifications (when an eyewitness properly identifies the perpetrator) while concurrently decreasing misidentifications (when an eyewitness identifies a face other than the perpetrator). Lineups where the perpetrator, or target face, are included are referred to as target present lineups, while lineups that consist only of lures are target absent lineups. Differentiating response options to target present and target absent lineups allow for comparisons on correct and incorrect identifications across lineup types.

The use of simultaneous lineups has been supported by the Diagnostic Feature Detection Theory (DFDT) proposed by Wixted and Mickes (2014). This theory predicts better discriminability for simultaneous lineups over sequential lineups due to eyewitnesses having the ability to directly compare all faces in a simultaneous lineup and distinguish between which features are lineup defining (shared features needed in order to be part of the lineup) or diagnostic (differing between all lineup members). In contrast, past work on comparative judgements predicts that sequential lineups outperform simultaneous lineups (Lindsay & Wells 1985; Wells, Memon, & Penrod 2006). This prediction is based on sequential lineups relying upon absolute judgements and having lower instances of misidentifications compared to simultaneous lineups which rely on relative judgments.

Stopping Rules

Sequential lineups include a stopping rule where the first positive identification terminates the lineup and any remaining faces in the lineup are not presented to the eyewitness. Little research has been done on sequential lineups in the absence of a stopping rule and the question of what multiple identifications in these contexts mean and whether or not a stopping rule is the best practice for law enforcement to use still remains.

The vast majority of research including a sequential lineup includes the use of the stopping rule, as it is one of the main features differentiating sequential and simultaneous lineups and has been included since the proposal of the benefits of sequential lineups by Lindsay and Wells (1980). A sequential stopping rule occurs when a participant makes a positive identification and the lineup is halted and any remaining faces are not presented to the eyewitness. However, studies designed to not implement a stopping rule may have still unintentionally allowed for participants to adopt an internal stopping rule all on their own. The study by Wilson and colleagues (2019) aimed to further understand sequential lineups by removing the stopping rule and allowing participants to make multiple identifications within a single lineup. However, this was done by removing the instructions telling participants that only their first identification would be counted and allowing participants to view all lineup members and make subsequent positive identifications, which would not actually be counted and were effectively canceled out by the previous positive identifications. Furthermore, this study served as the basis for multiple other re-analyses aiming to compare stopping rule and no stopping rule sequential lineups, however, due to the potential for participants implementing their own internal stopping rule, further research is still necessary for comparison to the current re-analyses (Dunn, Kaesler, Semmler 2022; Kellen, MacAdoo 2022). Police lineups include a single suspect, a concept that research participants are well aware of, and may lead to participants only making a single identification unless explicitly informed of the opportunity for multiple identifications within a single lineup. The present study aims to manipulate the presence of a stopping rule in sequential lineups.

Empirical Aims

The current research aims to further unite the fields of eyewitness research and recognition memory research by borrowing techniques standard in recognition memory literature. Present eyewitness identification studies tend to use a single video of a mock crime and have a large participant list respond to one type of lineup per participant. While this procedure provides as realistic of an experience as possible within a lab room, these studies yield one data point per participant, require large numbers of participants, and only allow for between subject comparisons for lineup types, yielded discriminability, and model fitting. Recognition memory tasks create a more artificial setting where participants are provided with a list of a type of stimuli to memorize (usually words) and later asked to identify the previously studied target stimuli with the addition of new lures. Finley, Wixted, and Roediger (2020) bridged typical recognition memory research and eyewitness identification research through the use of DRM (Deese-Roediger-McDermott) word lists as stimuli. This study included multiple trials (sequential and simultaneous) for each participant and was able to adequately evaluate discriminability across lineup type but ultimately, the use of words as stimuli as opposed to faces resulted in the study being closer to typical recognition memory research rather than a compromise between it and eyewitness research.

The current study provides a better compromise between the two fields and provides further insight on how individual participants respond to both simultaneous and sequential lineups. The study allowed for a newer and less commonly used approach by having the same participant respond to multiple trials of both simultaneous and sequential lineups that use faces as stimuli which provided a closer fit of the procedure that eyewitnesses actually experience. Furthermore, this current study used clear instructions to further investigate the stopping rule and reasons as to why participants may make multiple IDs within the same sequential lineup, which has not been adequately done prior.

Research in the eyewitness identification field relies heavily upon the use of confidence ratings. However, the question of whether providing confidence ratings affects how participants respond to lineups when required to provide ratings for their confidence in their responses remains unaddressed in eyewitness research. Additionally, with so many police departments failing to have formal, written policies on how lineups are conducted it is imperative to determine if obtaining confidence ratings from eyewitnesses should be included when these policies are written.

The research questions and corresponding hypothesis that this study aimed to answer are as follows:

1) Which lineup type (simultaneous or sequential) yields the highest discriminability?

- Hypothesis 1: The use of absolute judgements predicts that sequential lineups will provide superior discriminability compared to simultaneous lineups.

2) Does incorporating a stopping rule in sequential lineups improve discriminability (between subjects comparison of stopping rule)?
   - Hypothesis 2: The diagnostic feature detection theory predicts that the no stopping rule will provide greater discriminability compared to the stopping rule sequential lineup procedure due to participants learning to distinguish between diagnostic and nondiagnostic features as the lineup progresses and making corrections to previous misidentifications.
3) Does the act of obtaining confidence ratings impact lineup decisions?
   Hypothesis 3: Providing confidence ratings will not significantly impact the response tendencies of participants.

Theoretical Aim

In addition to the practical goal of identifying the methods that best support people in making eye-witness decisions, this work also addresses the model that best fits the data. Previous research has indicated varying success for several different models, several of which are considered here. Each of the below are varieties of signal detection theory (SDT). SDT is a theoretical framework geared towards assessing and predicting how well individuals can recognize previously seen information when surrounded by brand new items. Within this framework, a criterion exists where if a memory signal for a given item exceeds the criterion the decision maker will distinguish the given item as old, whereas if the given item falls below the criterion it will be considered a new item.

Each SDT model discussed in this paper, the Independent Observations model, Integration model, Ensemble model, and First Above Criterion Model, all vary in the decision variable and decision rule[1]. The Independent Observations model predicts that whichever face in a lineup generates the MAX, or highest memory strength signal, will be identified if the MAX signal exceeds the criterion. For this model, the MAX signal is the decision variable and exceeding the criterion is the decision rule. In the Integration model, the decision variable is the face that produces the MAX signal while the decision rule is that an identification of the MAX face will only be made if the summed memory strength signals for all lineup members exceeds a given criterion. The Ensemble model uses a decision variable of the MAX difference in memory strength scores of a lineup member and the average memory strength signals of all lineup members and a decision variable of that MAX difference exceeding the criterion. The First Above Criterion model is applied to sequential lineups and is most similar to the Independent Observations model. This model uses a decision variable of the produced memory strength signal but a decision rule stating that the first face to produce a memory strength signal above the criterion will be identified, regardless of any remaining lineup members. Although the Integration model is the most commonly applied model in eyewitness literature, recent

[1] A more in-depth comparison of these models can be seen in the results section for hypothesis four. For more information on how the decision variables, decision rules, and how discriminability is calculated across models, reference Table 5 in the results section.

research has shown support for the Ensemble model when applied to simultaneous lineups (Wixted et al., 2018)

Each of the previously mentioned models give different insight to the decision-making process behind responding to a lineup. Identifying the model that best fits the data, or perhaps more importantly ruling out models that do not fit the empirical data well, can lend support to the theories behind the models or lend support towards rejecting poorly fitting models and their corresponding theories, and hopefully highlight differences in how participants tend to respond to each lineup procedure. The theoretical aim of this work is to ask how well do these models fit the data?

- Hypothesis 4: The diagnostic feature detection theory predicts that the ensemble model will provide the best fit to simultaneous lineup data, as this theory is the mathematical representation for the ensemble model. Furthermore, previous research and the emergence of the first above criterion model predicts an adequate fit to sequential lineup data.

Method

Participants

Participants were 285 Syracuse University students recruited through the Sona Systems (**https://www.sona-systems.com/**). Participants received course credit in exchange for their participation in the study.

Materials

One hundred and fifty-six photos from the Adelaide Lineup Database were used to create the 24 lineups and study images. Twelve six-person lineups were created for both the simultaneous lineups and for the sequential lineups, totaling 144 faces. Lineup members were chosen on the basis of similar facial features and general appearance (race, gender, eye color, nose shape, hair style). Twelve target faces were identified and five similar lineup members were chosen for each of the target present lineups. For the target absent lineups six fillers were used and a seventh face that matched the lineup members' characteristics was included in the study phase, but never appeared in the lineups given to participants. These 12 faces that appeared in the study phase but not in any lineups seen by participants represent instances where an eyewitness saw a perpetrator commit a crime, gave a description of the person to law enforcement, but the resulting lineup included an incorrect (innocent) suspect instead of the true perpetrator.

Procedure

The study followed a 2 x 2 x 2 factorial design. The first variable was lineup type, with the two levels being simultaneous or sequential, and was a within subject design as all participants responded to both simultaneous and a type of sequential lineups. The sequential stopping rule implementation followed a between subjects design, participants only responded to sequential lineups with a stopping rule or without one. Lastly, the variable assessing confidence judgements was done between subjects. Participants either used confidence judgements for all positive identifications or for none.

As all participants responded to both lineup types, the 2 x 2 x 2 factorial design was collapsed into four conditions on the basis of the remaining two variables, sequential stopping rule and confidence ratings. Upon beginning the study, participants were randomly assigned to one of the four conditions. Participants were categorized to either respond to: condition a) sequential lineups *without* a stopping rule and *with* confidence ratings (N=70), condition b) sequential lineups *with* a stopping rule and *with* confidence ratings (N=64), condition c) sequential lineups *without* a stopping rule and *without* confidence ratings(N=75), or condition d) sequential lineups *with* a stopping rule and *without* confidence ratings (N=76). Participants completed the research experiment in person on a university lab computer and received standardized instructions based on their condition placement.

The order of the study phase was randomized for all participants, but the same 24 faces were always shown as target faces. The ordering of simultaneous and sequential lineups was randomized and intermixed for all participants, but the constructed lineups were not randomly assigned to being either simultaneous or sequential across participants and instead remained fixed. The position location of the target face in target present trials and the fillers in all trials were randomized. The study phase presented the target faces one at a time for one and a half seconds each. In total, participants responded to a random order of 12 simultaneous target present lineups, 12 simultaneous target absent lineups, 12 sequential target present lineups, and 12 sequential target absent lineups.

All post study phase instructions provided to participants can be viewed in Appendices 1 through 4. After viewing the study phase, participants in all four conditions received the same instructions stating that only one face per lineup may have previously appeared in the study phase or that none of the faces per lineup were previously encountered. Similarly, all participants received the same instructions regarding the simultaneous lineups stating that if participants recognized one of the six presented faces to choose that face, otherwise if no face was recognized the lineup should be rejected. Participants in the two confidence rating conditions were given additional instructions stating if a positive identification was made (by selecting “yes” to a presented face) then a confidence rating would need to be given on a scale from zero to one hundred indicating the confidence in that positive identification. This was done using a sliding scale with zero indicating “very unsure” with the decision and one hundred indicating the participant felt “very sure” in their decision. Participants in the two non-confidence rating conditions viewed instructions stating that positively identifying a face or rejecting a lineup by not identifying a face would result in the presentation of the next trial.

For the sequential lineups, participants in the stopping rule condition were instructed that they would only be able to make one positive identification per lineup. Additionally, making an identification would terminate that lineup trial and any remaining faces would not be shown. Participants in the no stopping rule condition were informed that for each sequential lineup presented, multiple positive identifications could be made, but only the final identification in each trial would be counted as the identification for that trial. Essentially, making an additional positive identification after the first one in a sequential lineup was seen as a correction to a prior mistaken identification. If a participant selected “no” for all six faces in a sequential lineup (regardless of stopping rule conditions) the lineup was considered to be rejected.

Assessing Lineup Performance

In eyewitness research, the target face is referred to as the guilty suspect and the lure that most closely approximates the target face within a target absent lineup is the innocent suspect. When presented with a lineup, making a positive identification is considered a hit if the face identified is the target during a target present lineup, a false alarm if the identification is made during a target absent lineup, or a foil identification if a lure face is identified from a target present trial. If no positive identification is made during a target present lineup it is categorized as a miss and failing to make a misidentification and correctly rejecting a target absent lineup is termed a correct rejection. The proportion of target present lineups resulting in a hit is referred to as the hit rate (HR) while the false alarm rate (FAR) is calculated by dividing the number of false alarms by lineup size, then dividing the result by the number of target absent lineups.

The sequential superiority effect emerged from the seminal paper by Lindsay and Wells in 1985. This term was used to describe the resulting decrease in FAR for sequential lineups as opposed to simultaneous lineups. A point of contention between the historic literature and more recent studies is that this effect relied upon the use of diagnosticity ratios (DR) as a measure of statistical accuracy, as was common in previous literature (Lindsay & Wells, 1980; Lindsay & Wells, 1985; Wilson et al., 2019; Wixted & Mickes, 2015; Rotello & Chen, 2016). However, current research indicates that the DR does not accurately measure diagnostic accuracy and instead confounds accuracy with response bias, or how willing someone is to make a positive identification, and therefore needs to be replaced with SDT based receiver operating characteristic (ROC) analyses (Gronlund et al., 2014; Wixted & Mickes, 2015; Rotello & Chen, 2016). The DR favored a conservative response bias, where participants are less likely to make identifications, and supported the sequential superiority effect.

Receiver operating characteristic curves plot the hit rate as a function of the false alarm rate for either varying confidence ratings or for varying response biases. Confidence ratings tend to be obtained through the use of a numeric scale while response biases are manipulated through the instructions participants receive prior to responding to a lineup. The measure of area under the curve (AUC) or partial area under the curve (pAUC) for ROC curves has been proposed as a superior way to measure accuracy empirically for different lineup procedures without conflating response bias (Macmillan & Creelman, 1991; Clark 2005; Macmillan & Creelman, 2005; Rotello & Chen 2016). Varying AUC scores for different lineup procedures can be compared according to the standard Delong et al. (1988) methodology and can be used to assess empirical discriminability.

When assessing ROC data, the top left corner of the graph represents perfect discriminability with a hit rate of one and a false alarm rate of zero. Whichever curve on the graph tends toward this top left-hand corner and correspondingly contains the highest AUC demonstrates greater diagnostic accuracy. Figure 1 is taken from Gronlund and colleagues (2014) which plots hypothetical ROC data with the addition of the rightmost data point for each curve originating from Lindsay and Well's (1985) influential paper on the sequential superiority effect (seen in Table 1). However, when assessing these data points that were previously considered evidence of the superiority of sequential lineups the ROC curves support simultaneous lineups leading to greater discriminability, despite sequential lineups providing a greater DR. Due to the previously mentioned findings, this paper will use ROCs and AUC to determine discriminability, not diagnosticity ratios.

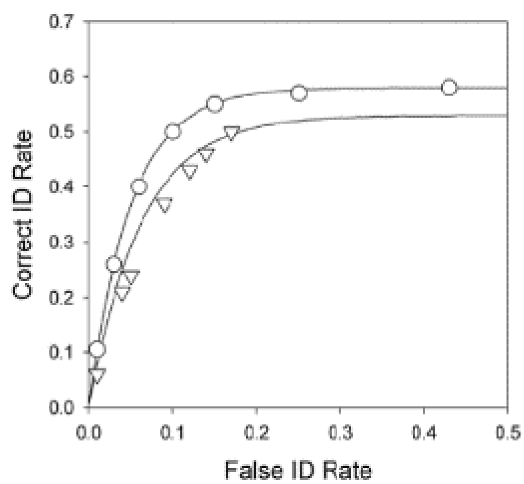


Figure 1: Hypothetical ROC date originally seen in Gronlund et al. (2014) with the right most points on each curve originating from Lindsay and Wells (1985) which proposed a sequential superiority effect where sequential lineups appear to decrease false alarm rates far more than simultaneous lineups while only minimally reducing hit rate. Circles indicate simultaneous lineup data; triangles indicate sequential lineup data.

Table 1

Hit and False Alarm Rates from Lindsay and Wells (1985)

	Simultaneous Lineup	Sequential Lineup
Hit Rate	.58	.50
False Alarm Rate	.43	.17
Diagnosticity Ratio	1.35	2.94

Note. Adapted from Lindsay and Wells (1985) which are the data points for hit rates and false alarm rates used as the right-most points of the curves in Figure 1.

Sequential Lineups

Sequential lineups pose additional challenges over simultaneous lineups due to the addition of the stopping rule and the nature of the sequential presentation of data. Instead of comprising one decision like simultaneous lineups, sequential lineups instead include a separate decision for each face that the eyewitness encounters within the lineup. Each face presented must result in a yes or no decision from the eyewitness and the first instance of a yes response terminates the lineup procedure. Because of this, aggregating sequential lineups to include a single overall ROC curve results in the loss of data surrounding all of the individual yes/no decisions made per lineup.

In order to separate sequential data into six separate serial position curves any positive identifications made prior to the presentation of the target face must be excluded. Essentially, this means that if a target face was in position four in a lineup but the participant made a foil ID for the face at position two and did not get the opportunity to respond to the target face then that lineup would be excluded from the ROC curves. This then allows for the distinction of six different curves which can be used to measure changes in discriminability as serial position shifts.

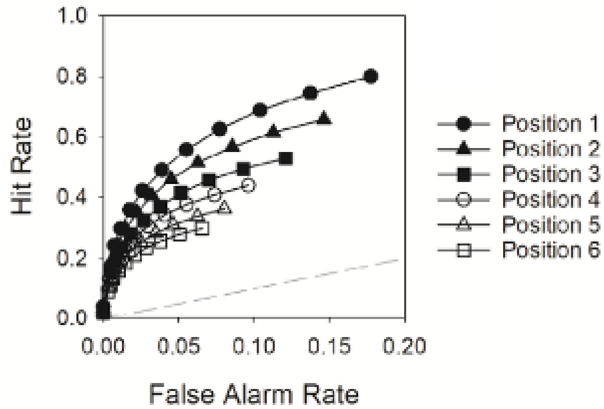


Figure 2: Adapted from Wilson and colleagues (2019) showcasing how discriminability changes with serial position in a sequential lineup. As serial position increases discriminability decreases.

Diagnostic Feature Detection Theory (DFDT), developed by Mickes and Wixted (2014) proposed that eyewitnesses can learn to only focus on features that vary between lineup members while discounting identical, lineup deciding features. In sequential lineups this was hypothesized to be seen through discriminability increasing as serial position increases which would indicate that participants improve their learning of which features are diagnostic the more faces are presented. However, this trend has only been empirically demonstrated once thus far by Wilson and colleagues (2019). Instead, sequential lineups tend to demonstrate a decrease in discriminability, as measured by pAUC, when serial position increases. A potential cause could be output interference, which robustly demonstrates a decline in accuracy in recognition memory tasks as trials increase, likely due to interference from previously presented faces (Criss, Malmberg, Shiffrin, 2011).

When sequential lineup ROC data is separated out into six curves the curve for position one essentially represents a showup and displays the highest level of discriminability as assessed by the largest AUC. When sequential lineup data is aggregated across serial position the resulting curve represents an average of all the six curves and lays somewhere in the middle of where the six curves would be on the graph. However, previous research has shown that sequential lineups (with aggregated data) outperform showups (Steblay, Dysart, Fulero, Lindsay 2001; Steblay, Dysart, Wells 2011; Gronlund et al 2012) which indicates a further need to investigate sequential lineups in order to better understand decision making processes occurring.

Regardless of the varying ideas regarding sequential lineups, much remains unknown about the theories supporting how these decisions are made. This is largely in part due to the difficulty surrounding model development for sequential lineups. This lack of knowledge is further highlighted by a lack of research aimed at further understanding the intricacies of the sequential lineup procedure such as the stopping rule and how they impact diagnostic accuracy.

Results and Discussion

The aim of this study was to investigate four research questions including a comparison of simultaneous and sequential lineup procedures and performance. Participants responded to multiple simultaneous and sequential lineups and their response outcome tendencies were measured to determine which lineup type and procedure produce the best discriminability, if confidence ratings impact response outcome proportions, and which measurement models best fit the observed data.

Hypothesis One: Sequential lineups provide better discriminability than simultaneous.

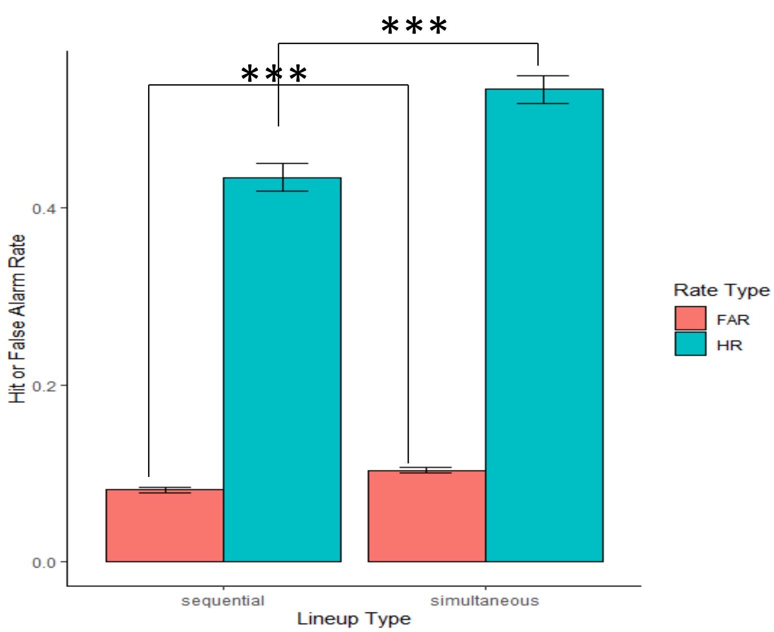
As this study did not use a designated innocent suspect, all identifications made in a target absent lineup were considered to be a false alarm. Overall false alarm rates were calculated by dividing all positive identifications made in target absent lineups for each condition first by lineup size, six for this study, then by the total number of target absent trials in each condition. Appendix 5 displays these outcome proportions, with Appendix 6 further partitioning the data by confidence ratings given, while Table 2 shows the varying hit rates and false alarm rates. When collapsing all simultaneous and sequential data across conditions, a significant difference was found between the hit rates and false alarm rates, with sequential lineups having a lower hit rate (R^2 =0.34, *F*(1,568)=20.25, *p*=8.237e-06) and a lower false alarm rate (R^2 =0.037, *F*(1,568)=21.72, *p*= 3.939e-06), as seen in Figure 3. Additionally, when computing d' by using the z transformations for HR and FAR, there was a significant difference between lineup types (R^2 =0.008, *F*(1,568)=4.78, *p*= 0.029), with simultaneous lineups having a d' score of 1.17 and sequential lineups having a lower d' score of 0.90. While these preliminary figures and results do lend support for the hypothesis that simultaneous and sequential lineups vary on response tendencies, the results indicate an advantage towards simultaneous lineups. Additionally, a conclusion on which lineup type provides the optimum outcomes cannot be established solely from comparing hit rates and false alarm rates and d' scores. This is due to it being probable that sequential lineups are yielding lower hit rates and false alarm rates due to a more conservative response bias instead of a true difference in discriminability. Furthermore, it is possible that additional differences in response tendencies are being masked across study conditions and could be further elucidated through ROC analyses.

Table 2

Hit and False Alarm Rates per Condition

	Condition	Hit Rate	False Alarm Rate
Simultaneous Lineups	A	0.50	0.11
	B	0.51	0.11
	C	0.50	0.10
	D	0.62	0.10
Sequential Lineups	A	0.46	0.08
	B	0.42	0.08
	C	0.44	0.08
	D	0.42	0.08

Note. False alarm rates were calculated in the standard practice of first dividing total identifications by lineup size, in the absence of a designated innocent suspect.



Condition A had no stopping rule and confidence ratings, Condition B had a stopping rule and confidence ratings, Condition C had a stopping and no confidence ratings, Condition D had no stopping rule and no confidence ratings given.

Figure 3: Regression comparing HR and FAR for simultaneous and sequential lineups for all possible lineups.

To further compare diagnostic accuracy of simultaneous and sequential lineups, ROC analyses were conducted for the conditions that supplied confidence ratings with each positive judgment. As ROC analyses plot hit rate as a function of false alarm rate across varying confidence ratings, data from conditions where confidence judgements were not obtained were not eligible for comparison here. For the simultaneous lineups, Figure 4 panel A shows participant responses across the two sequential stopping rule conditions and aggregated for all participants within the two conditions. A visual inspection of the figure shows that the three curves are virtually indistinguishable, with the reported AUC for each curve being 0.034. A bootstrapping analysis using the roc.test function within the pROC package in r, created by Mangul, Martin, Eskin, and Blekham (2011), with 10,000 samples indicated no significant difference in the ROCs and their corresponding AUCs for the two simultaneous lineup conditions, $p = 0.56$. Since all simultaneous data curves, regardless of sequential lineup condition, did not differ significantly, this indicates that participant responses to the other type of lineup did not impact their choices on simultaneous lineups. Regardless of if participants saw sequential lineups with or without a stopping rule, their hit and false alarm rate tendencies remained consistent for their simultaneous lineup trials.

Next, the overall simultaneous data was compared to aggregated sequential lineup data across all serial positions. As seen in Figure 4 panel B, the visual indication is that simultaneous lineups slightly outperform all aggregated sequential lineup data by displaying the higher curve tending towards the top lefthand corner of the figure along with the greatest partial AUC for the area of the x-axis covered by all graphs. Additional bootstrapping tests were conducted, each with 10,000 samples. The results indicated no significant difference between the overall sequential data and aggregated simultaneous data, p=0.17. These results indicate a slight simultaneous advantage over sequential lineups, in contrast to Hypothesis 1.

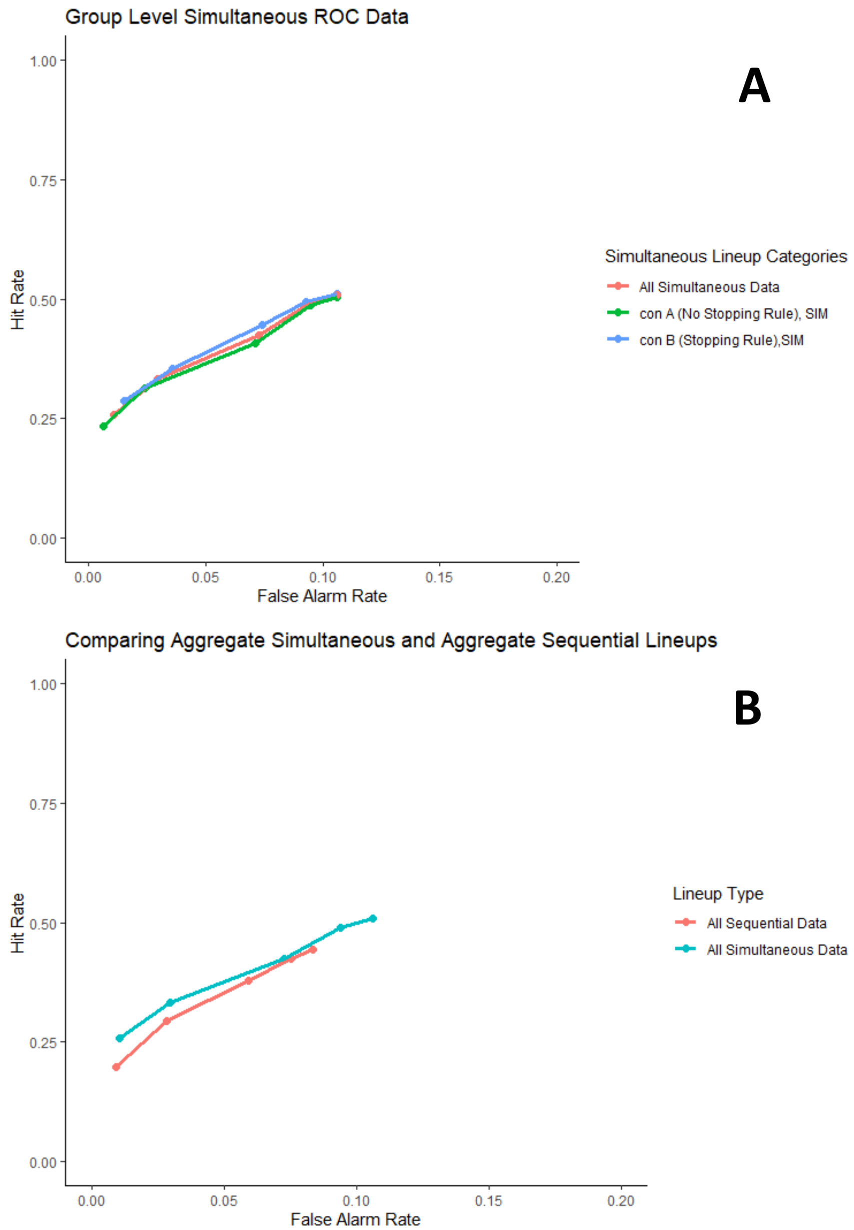


Figure 4: ROC comparisons for Simultaneous and Sequential Data. Panel A shows the data for all simultaneous lineups where confidence ratings were obtained. Panel B shows the aggregated simultaneous lineup data for both confidence conditions compared to the aggregated sequential lineup data.

In comparing aggregate data across simultaneous and sequential lineups, hypothesis one was not supported by the data observed in the experiment. The simultaneous lineup procedure produced a greater AUC than the combined sequential lineup conditions. However, it is plausible that aggregating the sequential lineup data to a single ROC curve is responsible for hiding relevant response trends. It is well established that show ups provide the worst discriminability outcomes and are nearly always outperformed by simultaneous and sequential lineups (Gronlund, Wixted, Mickes 2014; Wixted & Mickes 2014). The show up procedure is when only one face is shown to an eyewitness and the eyewitness must make their decision solely based on that face without the opportunity to make any other judgements. Sequential lineups where the designated innocent or guilty suspects are placed in position one are essentially show ups, which could lead to the erroneous conclusion that serial position one in sequential lineups would provide poorer discriminability when compared to simultaneous lineups or other serial positions. In contrast, the opposite tends to occur, with serial position one outperforming not only the remaining serial positions but also simultaneous lineups and showups (Wilson et al.,2019).

In the present study, the reported AUC for serial position one of the overall combined sequential lineup data was 0.049, a clear outperformance of all of the simultaneous lineup data which had a reported AUC of 0.034. This indicates that task dependent features and the aggregation of data in sequential lineups may inhibit the use of ROCs and AUC to determine differences in diagnostic accuracy and therefore which procedure is more favorable.

Hypothesis Two: The no stopping rule will provide greater discriminability than a stopping rule.

To first identify if response tendencies differed between the stopping rule and no stopping rule conditions, linear regressions were conducted for the hit rates and false alarm rates for all participant data. Demonstrated in Figure 5, the resulting regressions were not statistically significant for comparing hit rates (R^2=0.002, F(1,283)=0.669, *p*=0.41) nor for false alarm rates (R^2 < 0.001, F(1,283)=0.068, *p*=0.79). Furthermore, d' was calculated for each individual participant then used to determine if sequential lineup type was a significant predictor of the d' score. This regression showed that having a sequential stopping rule was not a significant predictor of d' scores (R^2 =0.002, F(1,283)=0.68, *p*=0.41), and the mean d' scores were 0.82 for the stopping rule condition and 0.97 for the no stopping rule condition. As mentioned with hypothesis one, comparing solely hit rates and false alarm rates can conflate accuracy with response bias and hide response trends pertaining to the task dependent differences, which therefore necessitates comparisons of applicable trials with ROC analyses.

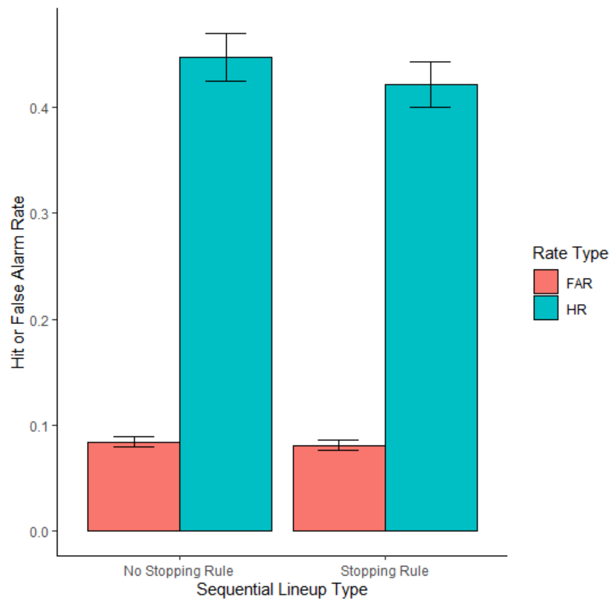


Figure 5: Regression comparing HR and FAR for the stopping rule and no stopping rule sequential lineup conditions.

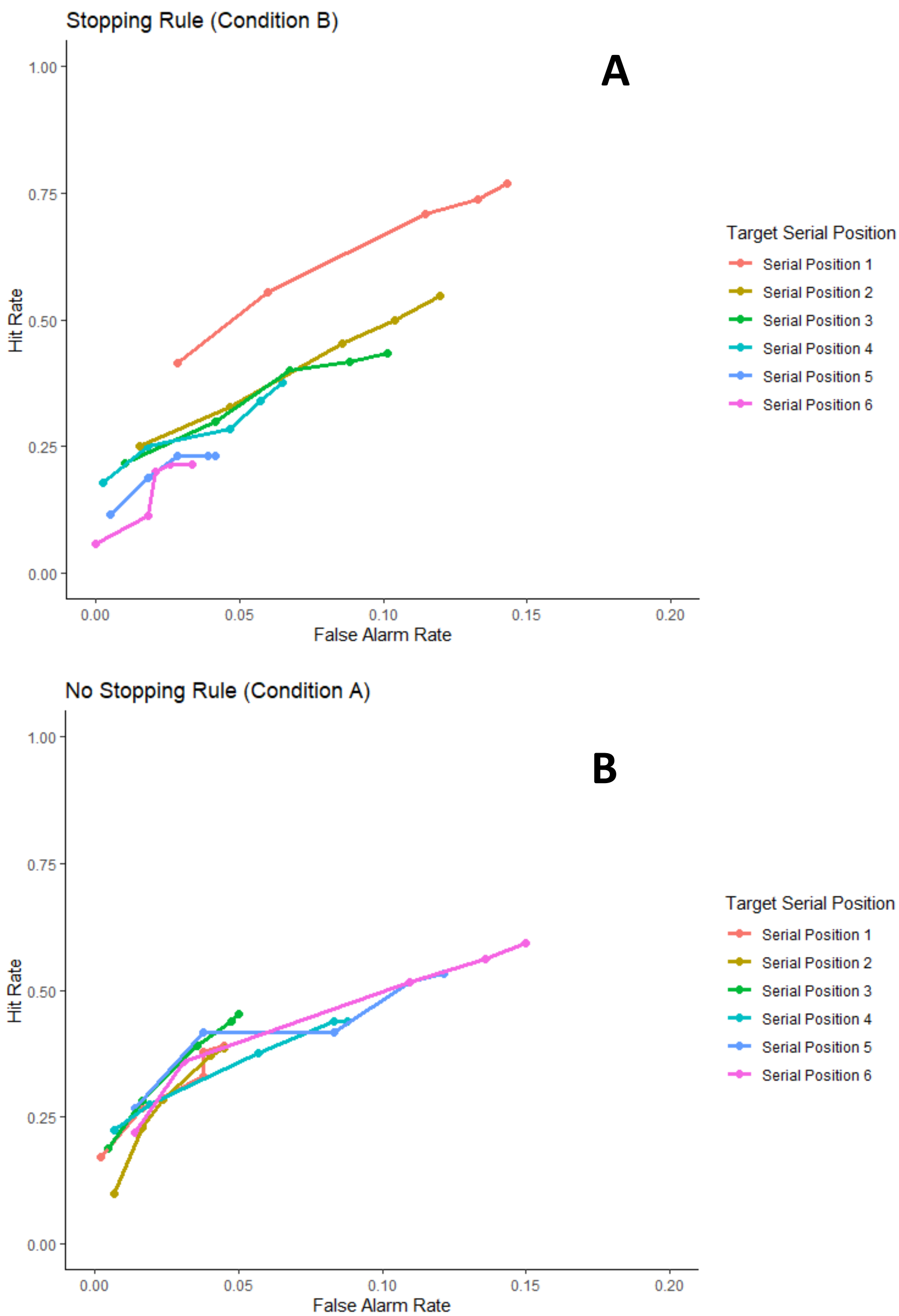
To adequately compare sequential lineup procedures and understand if considering sequential lineups at the aggregate level hides serial position effects, the sequential lineup data must be partitioned out and analyzed at the position level. This will highlight if participants respond differently to lineups with and without a stopping rule, and which procedure produces greater discriminability. To do so, Table 3 was generated to highlight response frequencies per condition and serves as the basis for the serial position ROC curves generated. Only the conditions that had participants give confidence ratings were included in table 3 as the method of ROC curve generation chosen for this study relies upon varying confidence ratings as opposed to varying response biases. The ROC curves in Figure 6 were constructed by using the bolded frequencies seen in the diagonals of the table to calculate hit rates for each serial position and using the corresponding target absent identification frequency. An example of how this was done would be for Condition A in the first part of the table, calculating the hit rate for position one as 58 divided by 82, which results in a HR of 0.71, and calculating the FAR by dividing 62 by 420, which results in 0.15.

Table 3

Response frequencies for "Yes"/"No" choices for sequential lineups for conditions A and B

	Target Position						
	One	Two	Three	Four	Five	Six	Target Absent Lineups
	Response frequencies for imposing a stopping rule on the no stopping rule condition (Condition A)						
"YES" to face in position ONE	58	12	13	13	14	16	**62**
"YES" to face in position TWO	0	30	9	14	11	10	**45**
"YES" to face in position THREE	4	0	27	8	6	7	**35**
"YES" to face in position FOUR	0	6	1	23	6	4	**24**
"YES" to face in position FIVE	1	2	1	0	14	3	**26**
"YES" to face in position SIX	2	2	0	1	0	11	**18**
Rejected Lineups	17	18	13	21	9	13	210
Total	82	70	64	80	60	64	420
	Response frequencies for the no stopping rule condition (Condition A)						
"YES" to face in position ONE	32	0	2	3	2	3	**19**
"YES" to face in position TWO	4	27	2	2	3	2	**19**
"YES" to face in position THREE	4	1	29	1	5	2	**21**
"YES" to face in position FOUR	4	7	2	35	2	3	**37**
"YES" to face in position FIVE	11	6	9	6	32	3	**51**
"YES" to face in position SIX	10	11	7	12	7	38	**63**
Rejected Lineups	17	18	13	21	9	13	210
Total	82	70	64	80	60	64	420
	Response frequencies for the stopping rule condition (Condition B)						
"YES" to face in position ONE	50	18	13	13	11	14	**55**
"YES" to face in position TWO	1	35	10	9	11	11	**46**
"YES" to face in position THREE	2	2	26	3	9	10	**39**
"YES" to face in position FOUR	1	0	1	21	7	6	**25**
"YES" to face in position FIVE	0	1	1	1	16	1	**16**
"YES" to face in position SIX	2	0	1	2	0	15	**13**
Rejected Lineups	9	8	8	7	15	13	190
Total	65	64	60	56	69	70	384

Note. Items in gray indicate correct identifications of target faces. Bolded items indicate values used to generate ROC curves based on target serial positions used to generate Figure 6.



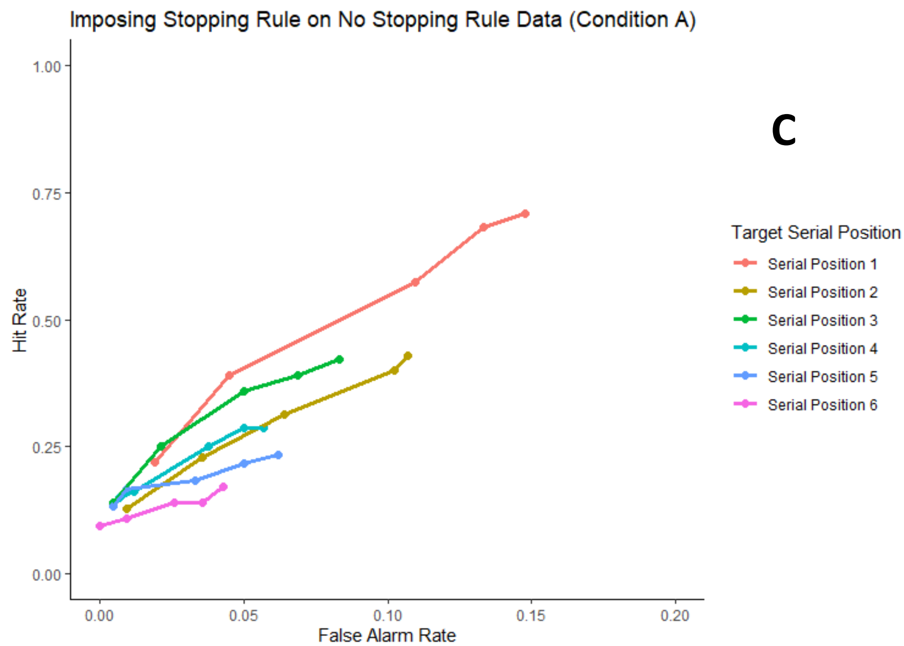


Figure 6: ROCs by Serial Position. Panel A shows data from the stopping rule (condition B) sequential lineups based on serial position of the target face. Panel B shows *ROC data for sequential lineups in the no stopping rule condition (condition A) based on serial position of the target face. Panel C demonstrates imposing a stopping rule on the no stopping rule condition (condition A). The curves plot hit rate as a function of false alarm rate for sequential lineups based on serial position of the target face.*

The serial position ROC curve generated for the stopping rule condition (Figure 6 Panel A) indicates that as serial position increases, discriminability decreases. However, the no stopping rule condition (seen in Figure 6 Panel B) does not follow this same trend. In the stopping rule condition, the pAUC values from position one to position six are: 0.064, 0.037, 0.028, 0.016, 0.006, and 0.004, respectively. A bootstrapping test with 10,000 samples indicated a significant difference between position one and position six with a p value of 0.024. For the no stopping rule condition, pAUC values from serial positions one through six range from: 0.011, 0.009, 0.013, 0.025, 0.042, and 0.053, respectively, which indicates that as serial position increases discriminability generally increases-a direct contrast to the previously mentioned ROC. A bootstrapping analysis comparing pAUC values for positions one and six did not produce a significant finding, p=0.064. Interestingly, when choosing to impose a stopping rule on the no stopping rule condition by only counting the initial identification even in the face of lineups resulting in multiple identifications, the trend for pAUC reverts back and mirrors the organic stopping rule condition, as seen in Figure 6 Panel C. In this case, pAUC for serial positions one through six are: 0.054, 0.024, 0.022, 0.010, 0.010, and 0.005, indicating that increases serial position leads to decreased discriminability, and the pAUC differed significantly from serial positions one and six, p=0.022.

To create an overall comparison across sequential lineup types, the aggregate stopping rule and no stopping rule ROCs curves were compared. These curves represent the empirical discriminability seen by these lineup types, as they encompass all lineup trials, unlike the serial position data. The partial AUC for the sequential lineups demonstrated a slight advantage for the no stopping rule condition, pAUC=0.023, followed by the stopping rule condition, pAUC=0.021, and imposing a post hoc stopping rule provided with worst pAUC at 0.019. These values did not differ significantly, and therefore did not support hypothesis two with an advantage towards the no stopping rule condition, p=0.064.

The stopping rule is a feature of sequential lineups conducted in the United States, where participants are required to make a judgement on a single presented face at a time before proceeding. In the United Kingdom sequential lineups follow a different procedure where all eyewitnesses must first view all nine lineup members in sequential order at least twice before being able to make a decision for the lineup (Palmer & Brewer 2012). The US version of the sequential lineup has demonstrated better discriminability than the UK version (Seale-Carlisle & Mickes 2016), but its inability consistently outperform simultaneous lineups led to the hypothesis that removing the stopping rule and allowing participants to make corrective judgements by only counting the last identification made could improve lineup performance. In contrast to the prior prediction, a significant difference was not found between participants that responded to sequential lineups with a stopping rule and those without a stopping rule, although, discriminability was greater in the no stopping rule condition and the associated p-value comparing the corresponding ROCs was 0.064, indicating that further comparisons may be warranted.

As shown in Appendix 7 and Appendix 8 there is a likely impact of output interference on the data collected[2]. This is seen in the data through the decrease in overall accuracy illustrated through the drop in the proportion of trials resulting in a hit as trial length increases. This finding is consistent with Criss, Malmberg, and Shiffrin (2011) which demonstrated that in word-based recognition memory tasks interference can arise due to interference from the presentation of other words, not just the target word. Due to this, it is possible that a partial explanation as to why the comparison between the sequential stopping rule and no sequential stopping rule conditions was just shy of the threshold from being considered significant could have been due to

[2] HR and FAR for simultaneous lineups that appeared as trial one were 0.697 and 0.115, respectively. For sequential lineups trial one HR was 0.513 and the FAR was 0.100. A logistic regression using test block (one through four, as indicated in Appendices G and H) indicated a significant predictor on if the trial resulted in a correct outcome (hit or correct rejection), $p<0.001$. This indicates that accuracy decreases as trial block increases.

output interference from the presentation of all other faces and the aggregation of sequential serial positions creating undue noise in the data. A way to test this theory would be to conduct an experiment with a larger participant pool in which each participant responds to a single sequential lineup, with or without a stopping rule and reanalyzing the ROC data.

Even though the overall comparisons of AUC for the different sequential conditions did not vary significantly, parsing the data out by serial condition did give insight into how participants change their individual lineup decisions based if a stopping rule was present or not. As seen in panels A and C of Figure 6, the stopping rule condition and when a post hoc stopping rule was applied to the no stopping rule condition followed the trend presented in previous research where discriminability decreases as serial position increases (Rotello & Chen, 2016; Wilson et al., 2019). However, the no stopping rule did not follow this same trend, instead showing the reverse. Excluding serial position one, from serial positions two through six AUC demonstrated slight increases as serial position progressed. Previous reasonings for why serial position may lead to an increase in discriminability relates to an eyewitness's ability to learn and distinguish diagnostic features from lineup defining features. Wilson, Donnelly, Christenfeld, and Wixted (2019) were the first to demonstrate this trend, with a comparison of AUC for serial position one yielding lower discriminability when compared to aggregate data from serial positions two through four. This may indicate that participants become more aware of differences between lineup members as more faces are presented sequentially and that making a subsequent identification serves as a correction to a previous misidentification in the absence of a stopping rule.

Hypothesis Three: Confidence judgments will not impact discriminability.

In order to test research question four on the use of confidence judgements, multinomial processing tree (MPT) modeling was implemented, as using MPT modeling allows for the use of observed participant responses to elucidate estimations of unseen processes. Figure 7 outlines the model used, with "yes" and "no" referring to the decision made by the eyewitness. Using the MPTinR package in r (Singmann & Kellen 2013), the models were fit to compare the corresponding confidence and no confidence judgment conditions to see if the probability of response outcomes change depending on if participants did or did not provide confidence judgements. Each model was able to estimate and further be constrained to hold P1 and P2 (the probabilities of trials ending in a "yes" response) equal and C1 and C2 (the probability of that "yes" response being correct) equal, if applicable. Parameter estimates[3] and corresponding significance levels can be seen in Table 4. Non-significant findings are in support of hypothesis three that the act of giving confidence ratings will not significantly change how participants respond to lineups, and no statistically significant findings are reported.

[3] P1 indicates the probability that the participant chose any face within a lineup for one of the confidence rating conditions, while P2 indicates the same finding for the non-confidence rating condition being compared. P1 and P2 are used in both target present and target absent lineups to denote a positive identification. Within target present lineups, c1 indicates that for the confidence condition being investigated, if a positive identification was given, what is the probability of that identification being correct (a hit). C2 indicates the same finding for the conditions where confidence ratings were not given.

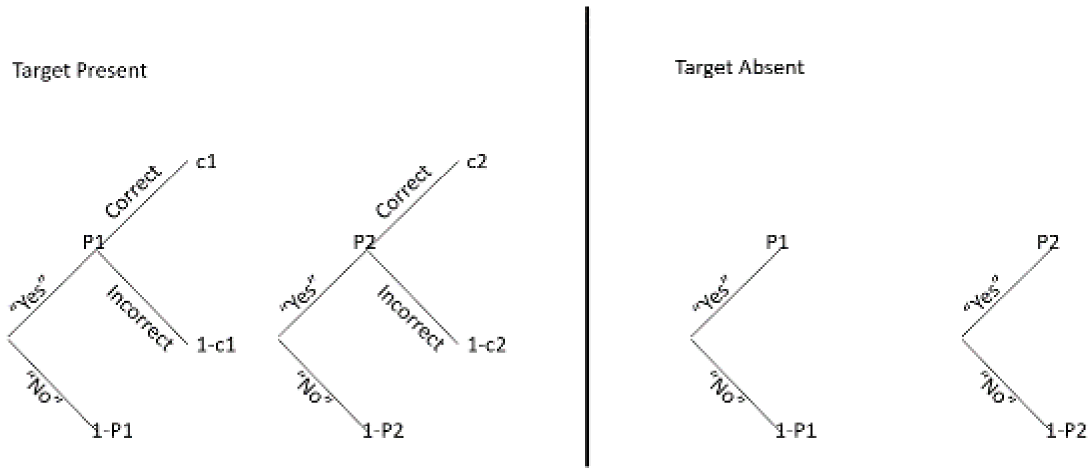


Figure 7: Multinomial Processing Tree (MPT) model. "Yes" and "No" correspond to the eyewitness's decision when presented with a face in a lineup. Correct and incorrect refer to if the eyewitness made the correct decision. In a target present lineup the correct decision is a hit while the incorrect decision is a foil ID and responding "no" is a miss. For a target absent lineup any "yes" decision is a false alarm while "no" responses produce correct rejections. In the absence of a designated innocent suspect the only response options for a target absent lineup are the binary ones listed in the model on the right of the figure. The two trees shown for each lineup type (target present vs target absent) reference the two conditions being compared.

Table 4

Reported Parameter Estimates and p-values for MPT models comparing corresponding confidence and no confidence judgement conditions.

Parameter Estimates	Simultaneous				Sequential			
	Target Present		Target Absent		Target Present		Target Absent	
	SR	NSR	SR	NSR	SR	NSR	SR	NSR
P1	0.84	0.81	0.36	0.36	0.84	0.78	0.50	0.50
P2	0.88	0.81	0.38	0.40	0.85	0.81	0.51	0.53
c1	0.61	0.63	-	-	0.50	0.58	-	-
c2	0.69	0.62	-	-	0.49	0.54	-	-
P-value for Restricted Parameter								
P2	0.06	0.88	0.55	0.15	0.71	0.31	0.55	0.36
c2	0.13	0.81	-	-	0.75	0.19	-	-

Note. SR indicates trials with a stopping rule for sequential lineup trials and NSR indicates no stopping rule for sequential lineup trials. Simultaneous SR or NSR trials indicate that the participants that responded to those simultaneous lineups also responded to sequential lineups with or without a stopping rule. P1 and c1 refer to conditions for confidence rating while P2 and c2 pertain to no confidence rating conditions.

It is particularly important to verify that the mere act of giving confidence ratings does not impact how eyewitnesses respond to lineups. Many police departments lack a structured format for how lineups are conducted, and in order to encourage the implementation of collecting confidence it is necessary to show that the proportion of hits and false alarms are not significantly changed when these judgements are given (Police Executive Reform Forum 2014). The indication that the MPT models did not show a significant decrease in favorable responses is a step in the correct direction, as previous research has shown that identifications made with higher confidence judgements are less impacted by estimator variables that are outside of the justice system's control (Semmler, Dunn, Mickes, Wixted 2017).

Hypothesis Four: The Ensemble Model Will Fit Best

Table 5 indicates the differences in decision variables, rules, and calculations for discriminability across the following SDT models. For a more detailed review of the differences between the independent observation model, integration model, and ensemble, and to see the derivation of discriminability equations, see Wixted and Mickes (2018).

Table 5

Differences in how SDT models use decision variables, rules, and equations.

SDT Model	Decision Variable	Decision Rule	Equation
IO	Raw memory strength of faces	MAX face identified if decision variable (raw memory strength) exceeds criterion.	$d'_{IG} = \mu_{target}$
INT	Summed memory strength for all lineup members	MAX face identified if decision variable (summed memory for all lineup members) exceeds criterion.	$d'_{IG} = \frac{\mu_{target}}{\sqrt{k[1+(k-1)\rho}}$
ENS	Difference between memory strength of a lineup member and average memory strength for all lineup members	MAX face identified if decision variable (difference in memory strength) exceeds criterion.	$d'_{IG} = \frac{\mu_{target}}{\sqrt{\frac{(1-\rho)k}{(k-1)}}}$
SDT-SEQ	Raw memory strength of single presented face	First face with memory strength to exceed criterion.	$h(c) = 1 - \Phi(\frac{c-\mu_{target}}{\sigma})$ $f(c) = 1 - \Phi(c)$

Note. IO refers to the Independent Observation model, INT to the Integration model, ENS to the Ensemble model and SDT-SEQ to the First Above Criterion Model. For the three simultaneous lineup models, information in this table is adapted from Wixted and Mickes (2018), and the discriminability equation refers to d'_{IG} (the ability to differentiate between innocent and guilty suspects). For the sequential SDT model, the equation refers to the hit rates and false alarm rates, relative to a decision criterion (c), where Φ is the normal cumulative distribution function. Modeling of the sequential lineup procedure is still underway and not exhaustive; therefore, discriminability can be calculated by taking the difference between the means of the target and lure distributions, further information on SDT-MAX can be seen in Kaesler, Semmler, and Dunn (2017).

Independent Observation Model

According to the independent observation model, the decision maker will identify the face that produces the MAX signal (the MAX face) if the MAX signal exceeds the given criterion. This model predicts that a memory strength signal is generated for each member of a lineup independently, and therefore correlated memory strength signals, or how similar lineup members look to each other, for these lineup members will have no impact on discriminability. Importantly, this independence of memory strength is brought into question due to research on the addition of implausible alternatives in lineups. Implausible alternatives refer to inclusions of "dud" lineup members that are implausible for a decision maker to choose to identify. According to the independent observation model, all lineup members, including implausible alternatives, should not impact the memory strength, or associated confidence value, for the surrounding faces in a lineup. However, a study on comparative judgements has indicated that implausible alternatives can increase a person's confidence in the surrounding, more plausible, items (Chambers & Windschitl, 2004).

Integration Model

The Integration Model contains a detection component reliant upon overall summed memory strength for a given lineup wherein if the summed memory strength across the lineup exceeds the given criterion, then an identification will be made. A separate identification component indicated that if this detection component is met, the lineup member that generates the MAX memory strength signal will be identified. This model predicts that as lineup members more closely approximate each other and memory signals become more correlated, discriminability will decrease and the decision maker will have increased difficulty in differentiating between innocent and guilty suspects (Wixted et al., 2018). This poses concern for a misinformed interpretation that fillers should not be chosen based on their similarity to the eyewitness's original description of the perpetrator which in turn paves the way for unfair lineups to be utilized. Recently, this model has shown concerningly poor fits to observed data (Wixted et al., 2018; Kaesler, 2021) which is especially problematic as it is the most commonly applied SDT model to eyewitness research (Palmer, Brewer, & Weber, 2010; Palmer, Brewer & Horry, 2013; Horry, Brewer, Weber, & Palmer, 2015).

Ensemble Model

The Ensemble model directly conflicts with the predictions made by the integration model due to its prediction that discriminability will increase as correlated memory signals across lineup members increase. In this model, the decision variable is the difference between the memory face of a single face and the averaged memory strength score of all lineup members, if this decision variable exceeds the given criterion a positive identification will be made. This model emphasizes the importance of focusing on diagnostic features that differ between lineup members while attending less to lineup defining features that are shared between all lineup members, and is based on Wixted and Mickes's (2014) Diagnostic Feature Detection Theory (DFDT). This theory states that eyewitnesses can learn to focus solely on the varying diagnostic features in order to make correct identifications. Recently, this model provided the best fit to simultaneous lineup data when assessed concurrently with the two previously mentioned models according to chi-squared goodness of fit results (Wixted et al., 2018).

First Above Criterion Model

The first SDT model dedicated specifically to sequential lineups was penned by Kaesler et al. (2017). This model is deemed the first above criterion model and accounts for the sequential lineup stopping rule by setting a decision variable deeming the first face to generate a memory strength signal exceeding the criterion as the one that is identified. This model was the first of its kind for sequential lineups, however, current limitations include its incapability of accounting for criteria shifts as serial positions increase along with its inability to make a prediction regarding correlated memory signals. This model has been shown to provide an adequate fit to sequential stopping rule data, while in the absence of a sequential stopping rule previous research has elected to fit a standard SDT model with free parameters that can be added to mimic an UVSD model (Wilson et al., 2019, Kaesler, Dunn, Ransom, Semmler 2020).

Results

Model fits were conducted based on Open Science Foundation code made available from the paper by Kaesler, Semmler, and Dunn (2017). Parameter estimates can be seen in Table 6. When fitting the integration model to the simultaneous data, the fit demonstrated was the worst compared to the other models, although the model did still fit the data without a significant deviation, $\chi^2(8) = 15.16$, $p = 0.06$. The ensemble model provided the best fit to the simultaneous data $\chi^2(8) = 11.23$, $p = 0.19$, and the independent observations model did provide an adequate fit between the two others, $\chi^2(8) = 12.45$, $p = 0.13$. The model fit to the sequential stopping rule data, the first above criterion model, was obtained from the same open-source code and provided an adequate fit to the data $\chi^2(8) = 14.39$, $p = 0.07$.

Table 6

Parameter Estimates for SDT Models

Parameter Estimates	Simultaneous			Sequential
	IO	INT	ENS	SDT-SEQ
μ_{target}	1.87	3.21	2.16	1.06
σ_{target}	0.90	1.64	0.87	1.02
C_1	0.91	-0.71	0.92	1.45
C_2	0.94	0.63	0.95	1.53
C_3	1.03	1.23	1.04	1.57
C_4	1.23	2.57	1.22	1.76
C_5	1.75	5.17	1.71	2.22
χ^2	12.45	15.16	11.23	14.39
df	8	8	8	8
p-value	0.13	0.06	0.19	0.07

Note. IO refers to the Independent Observation model, INT to the Integration model, ENS to the Ensemble model and SDT-SEQ to the First Above Criterion Model. None of the reported model fits were statistically significant at the critical value of 15.507, $\alpha = .05$.

In Table 6, the *C* parameters refer to the different confidence criteria for each model fit. C1 indicates a more liberal criterion, whereas C5 indicates the most conservative criterion where more certainty is required for a participant to give a positive identification. μ target refers to the mean memory strength for the target distribution; The mean for the integration model, and all other parameter estimates, cannot be directly compared to that of the remaining models as it relies on the transformed memory strength values corresponding to the summed memory strength while all other models use the raw, untransformed memory strength as the axis value. In equal variance SDT models, discriminability can be measured through μ_{target} with a higher value indicating greater discriminability. The parameter estimates in Table 6 indicate that the ensemble model provided the best theoretical discriminability over the independent observations model and in comparison to the first above criterion model. Additionally, the confidence criteria for the ensemble and independent observations model are nearly identical and do not indicate a substantial shift in response bias predictions. In comparison, the confidence criteria for the first above criterion model are higher on the memory strength axis and indicate a more conservative response bias.

In support of hypothesis four, the ensemble model did provide the best fit to the simultaneous data, relative to the independent observation and integration models. However, all three models did provide an adequate fit of the data and it is therefore not possible currently to state that any of the three models had a poor enough fit to be discounted, which is in contrast to Wixted and colleagues (2018) which showed that the integration model performed far worse than any of the opposing models. The parameter estimates did indicate that the First Above Criterion model provided a the worse fit to sequential data overall than the Independent Observations and Ensemble models did to simultaneous data, however, the confidence criteria in the First Above Criterion model indicate that response tendencies were more conservative compared to the simultaneous lineups. The Integration model relies on transformed memory strength values of the sum of all memory strength signals generated by lineup members, and therefore is not directly comparable to the remaining models which rely on untransformed memory strength signals for the criteria locations.

Interestingly, the integration model and ensemble model have contrasting predictions on how correlated memory signals impact discriminability. The first above criterion model did provide an adequate fit of the sequential lineup data, but the lack of a model that accounts for correlated memory signals, as the ensemble model does, still leaves the question of how similarity between lineup members can impact discriminability in a sequential lineup. As the Ensemble model is the mathematical representation of the DFDT, the success of the ensemble model for simultaneous data lends support for the DFDT which theorized that eyewitnesses rely on the similarity of lineup members to discount any non-diagnostic features that are shared by all lineup members and instead focus on the variations among diagnostic features among lineup members. Additionally, the results from hypothesis two indicating a visual advantage of not having a stopping rule for improving discriminability along with the success of the fit of the Ensemble model to simultaneous data indicate that the DFDT may have merit when applied to both simultaneous and sequential procedures, and that eyewitnesses can learn the difference between which features of lineup members to attend to versus which to discount. Additionally, as the Ensemble model relies on comparing the memory strength of each individual face in a lineup to the average memory strength generated by all lineup members and predicts that as correlated memory signals increase discriminability will as well, the success of this model advocates against the inclusion of "dud" lineup members, or implausible alternatives, and instead for the use of fair lineups.

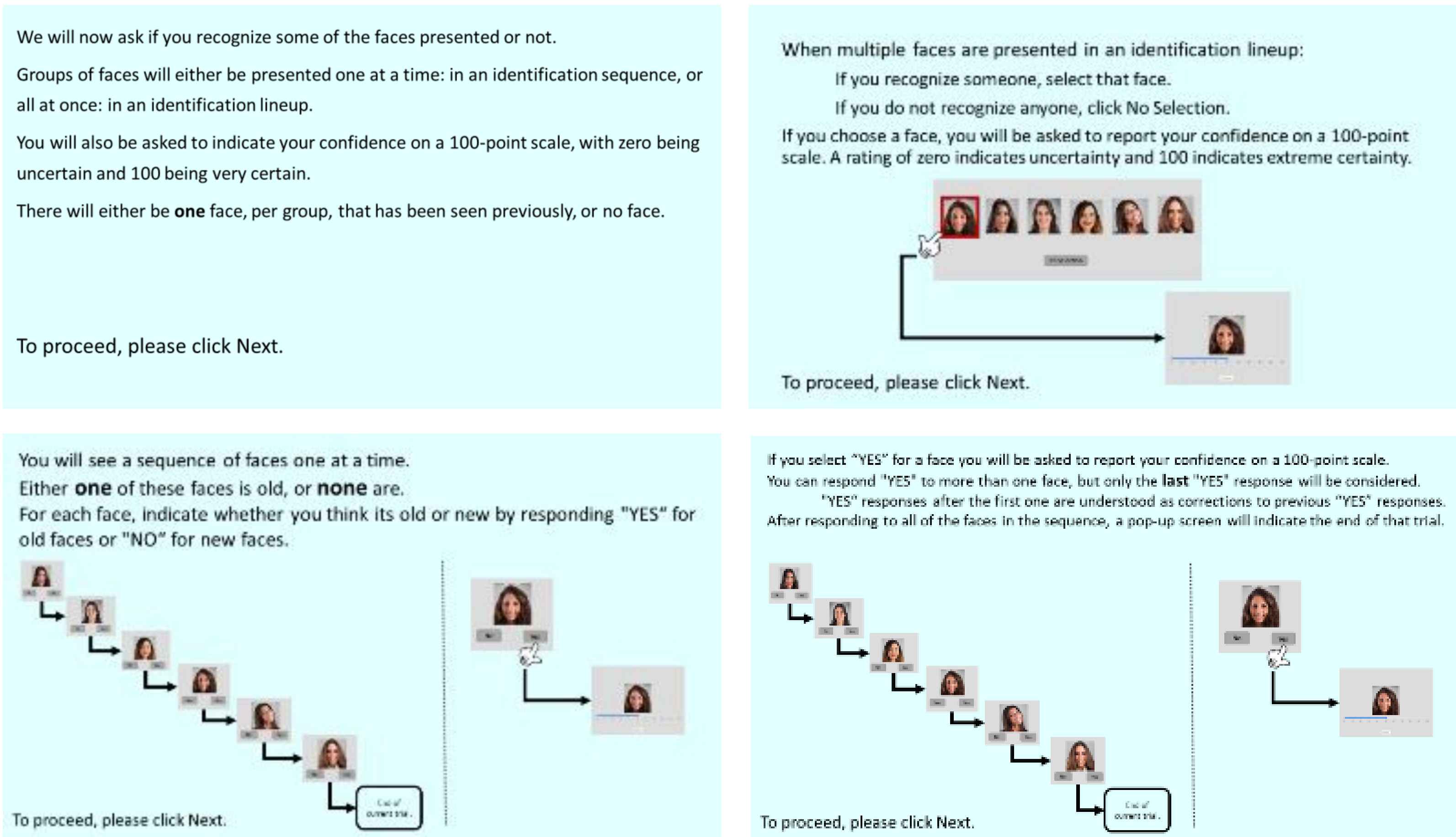
Conclusion

This study aimed to further unite eyewitness and recognition memory research by drawing from a typical recognition memory task using signal detection theory to investigate which eyewitness lineup procedure techniques should be implemented in police departments across the Unites States. Ultimately, the simultaneous lineup visually provided the best possible discriminability, compared to stopping rule and no stopping rule sequential lineups, though this difference was only significant when comparing HR, FAR, and d', not ROCs nor AUC. A significant difference did not occur in discriminability between the conditions with a sequential stopping rule and without. However, it is still possible that task dependent features of the sequential lineup, such as the individual judgments, obscure the ability of differences in target memory in eyewitnesses from being salient, indicating that further research into sequential lineups is still warranted. Furthermore, the no stopping rule condition did lend support for the DFDT by illustrating that discriminability can increase as serial position increases, a finding that has only previously been seen empirically once before (Wilson et al., 2019). This indicates that in a sequential lineup participants may be able to learn to differentiate between target and lure faces as the lineup progresses. This finding also contributes to why sequential lineups tend to outperform showups, even when sequential data is aggregated across positions, even though serial position one in a sequential lineup is essentially a showup. Further research is still needed to determine which lineup procedure should be used across police departments, and how memory effect, such as memory for target faces, differ across procedures. As the act of giving confidence ratings did not significantly impact lineup decision outcomes, these judgements should be collected by police departments as indications of an eyewitness's certainty. Model fits did not provide any distinct findings on which theories may be most consistent with an eyewitness's actions, instead, all four SDT models fit to the data provided adequate fits and therefore could not be ruled out.

The results pertaining to model fits implicate the necessity of future research into the merits of the applied SDT models. The models in this paper have varying predictions of how correlated memory signals, and therefore how similar lineup members should be, impact discriminability. The independent observations model predicts that similarity, or lack thereof, of lineup members has no impact on an eyewitness's decision, but empirically this theory is unsupported and discounted through research on implausible duds (Chambers & Windschitl, 2004). The integration model predicts that the more correlated memory signals become, the worse discriminability will be. This poses a threat to the necessity of fair lineups, where lineup members are chosen based on similarity to the description given by an eyewitness. The ensemble model predicts that correlated memory signals improve discriminability and therefore promote the use of fair lineups without any implausible duds included that are definitively not the perpetrator. The issue then arises that if all three models have strikingly different predictions on how similar lineups should be and how correlated memory signals impact discriminability, how did all three models adequately fit the empirical data? It is imperative that the advancements in SDT models be contrasted with more basic equal and unequal variance SDT models in order to validate the merits of the additional theories and free parameters. Furthermore, more weight should be given to empirical discriminability measures such as AUC of ROC curves and perhaps less to theoretical measures pertaining to model fits.

Appendix 1

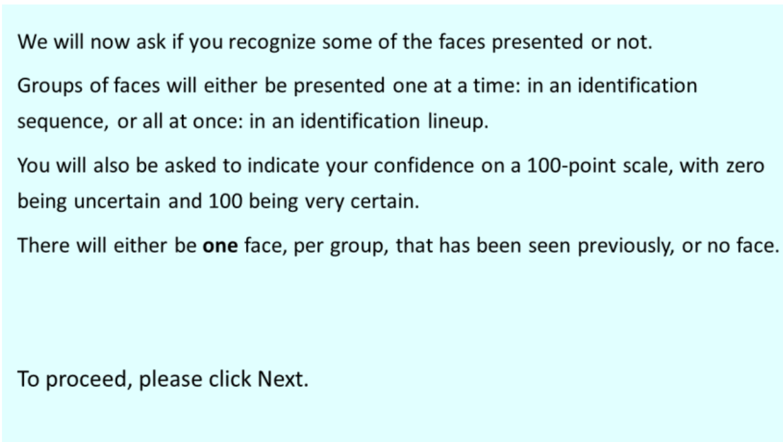
Instruction Slides for Condition A

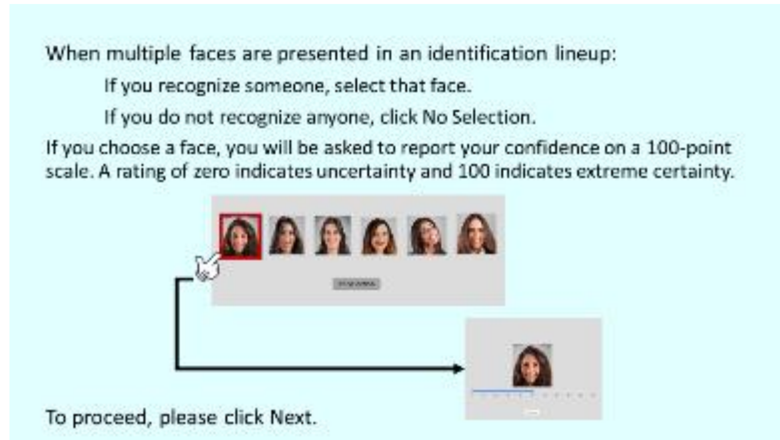


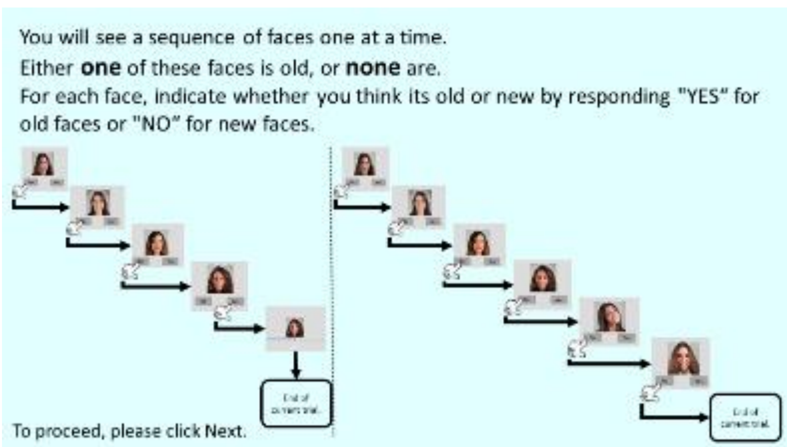
Appendix 1. The instruction slides shown to participants in condition A after the study phase. Slides went in order from one through four.

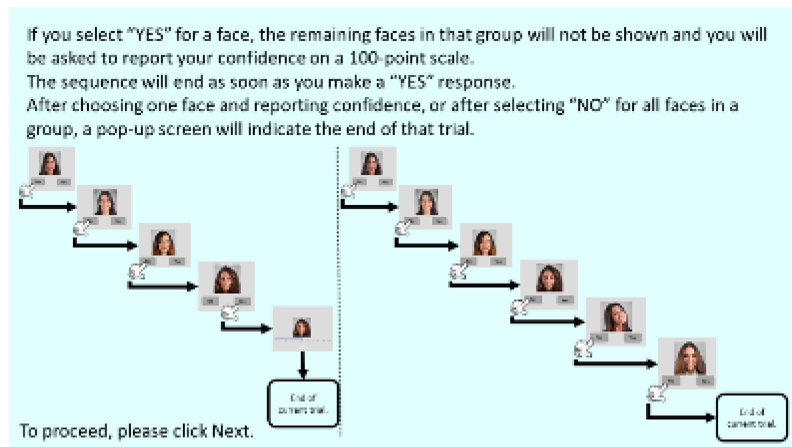
Appendix 2

Instruction Slides for Condition B





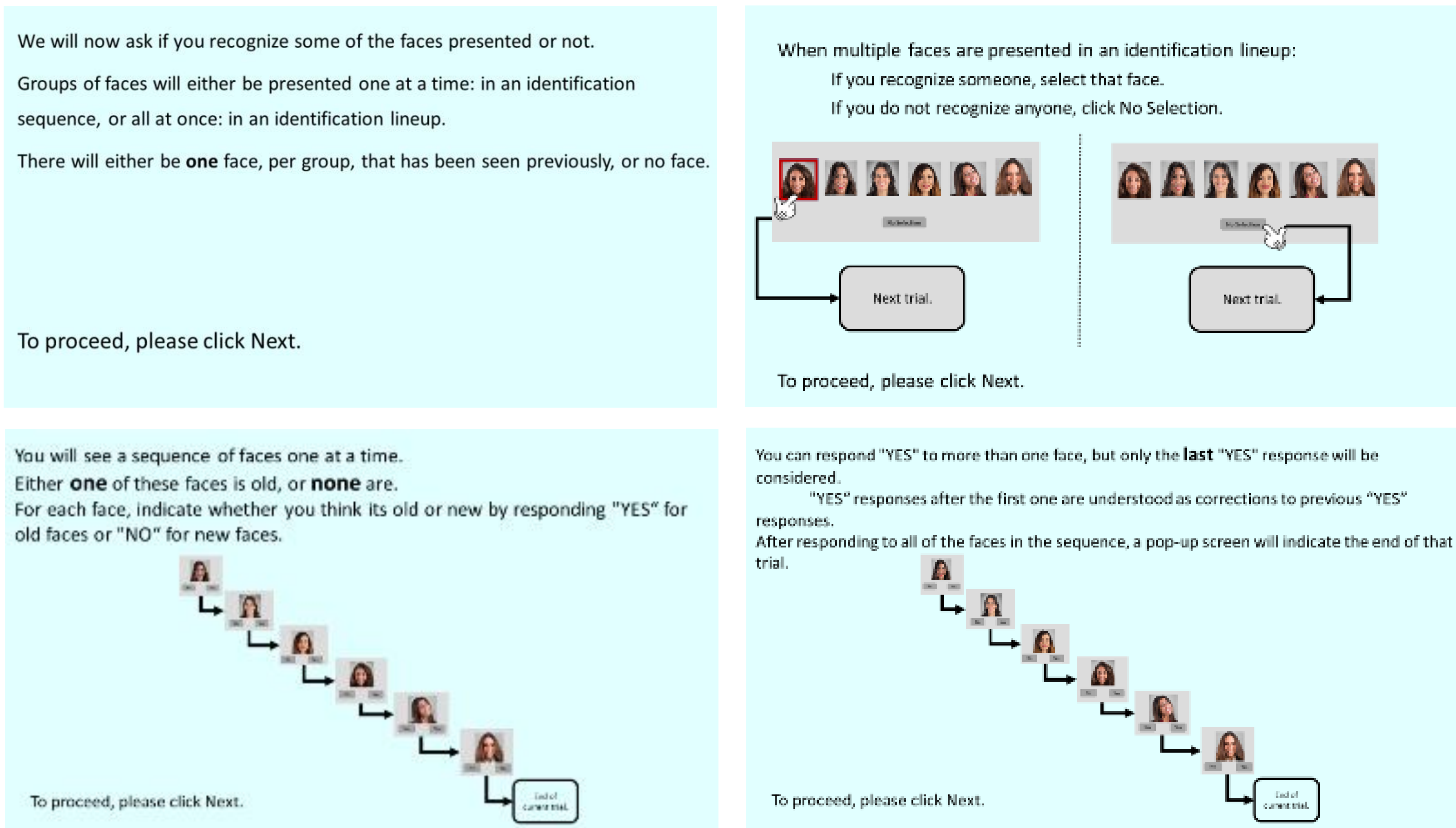




Appendix 2. The instruction slides shown to participants in condition B after the study phase. Slides went in order from one through four.

Appendix 3

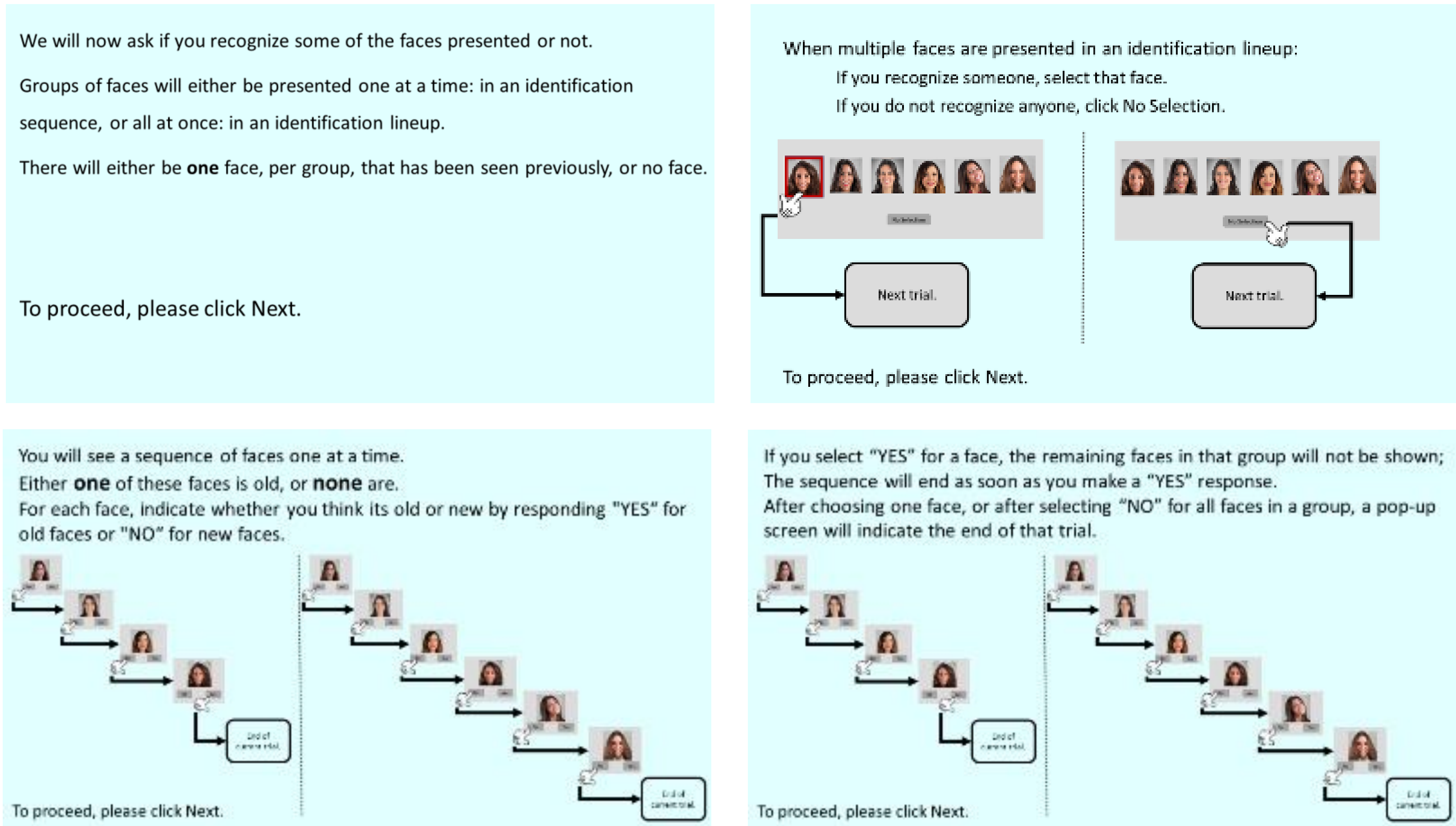
Instructions for Condition C



Appendix 3. The instruction slides shown to participants in condition C after the study phase. Slides went in order from one through four.

Appendix 4

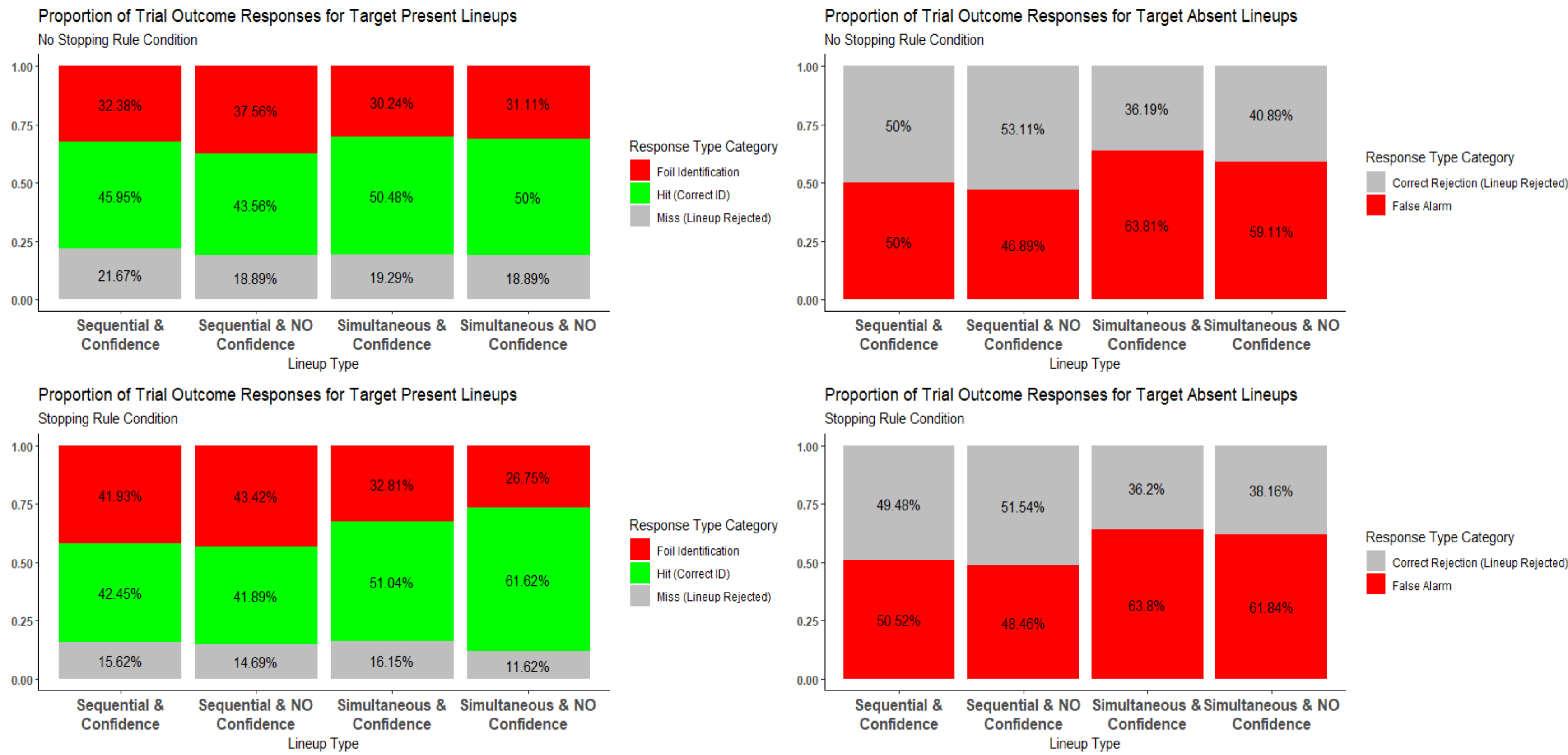
Instructions for Condition D



Appendix 4. The instruction slides shown to participants in condition D after the study phase. Slides went in order from one through four.

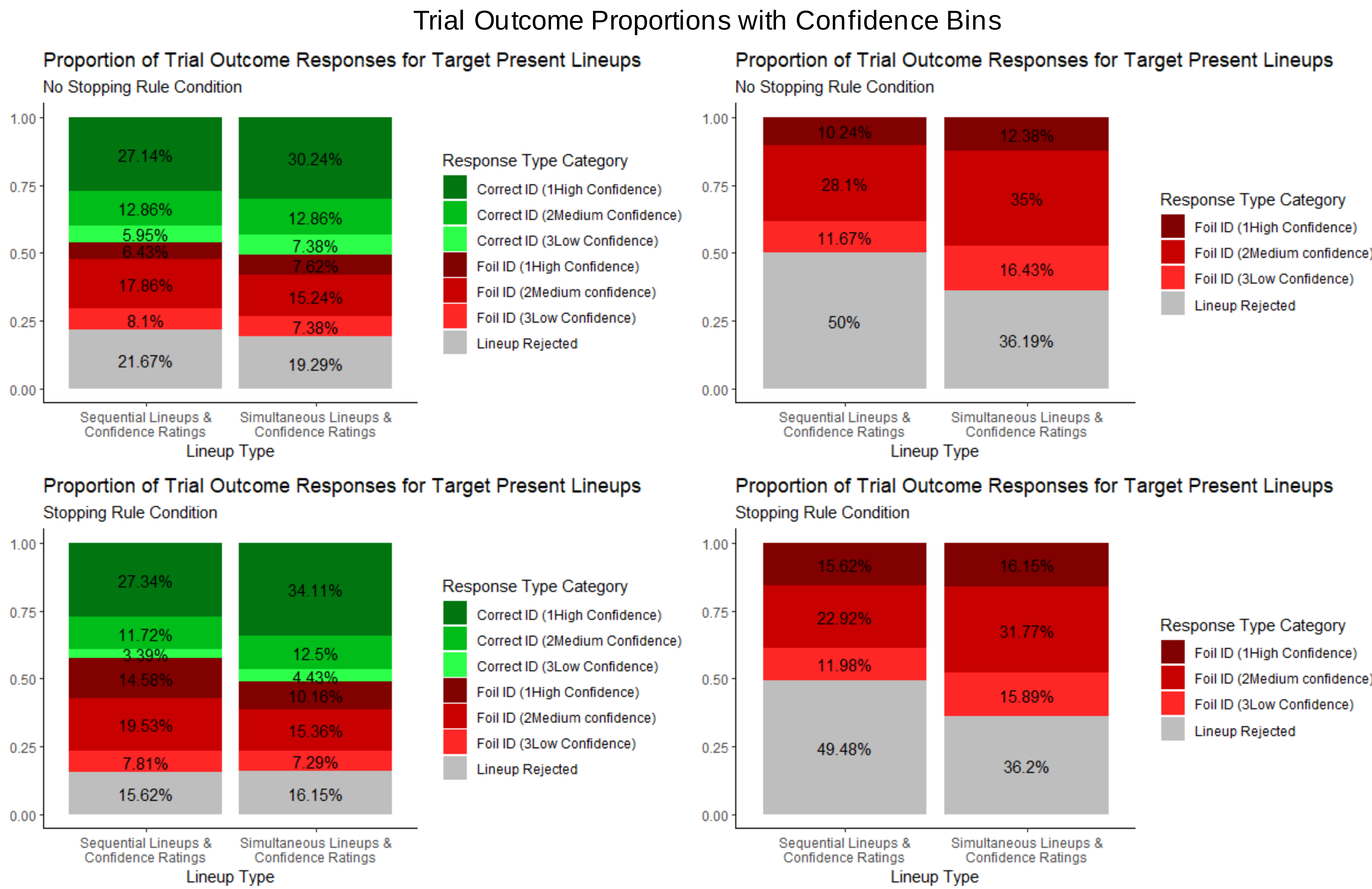
Appendix 5

Trial Outcome Proportions



Appendix 5. The top two panels show the proportion of response outcomes for the no stopping rule conditions with the left panel indicating target present trials and the right panel indicating target absent trials. The bottom two panels show the varying conditions without stopping rules, with the left panel again indicating solely target present trials and the right panel representing target absent trials. Response types are broken into five categories: hits (correct identifications of the target face), misses (failing to make an identification in a target present lineup), foil identifications (making a misidentification in a target present lineup), correct rejections (correctly not making an identification in a target absent lineup), and false alarms (making any identification in a target absent lineup).

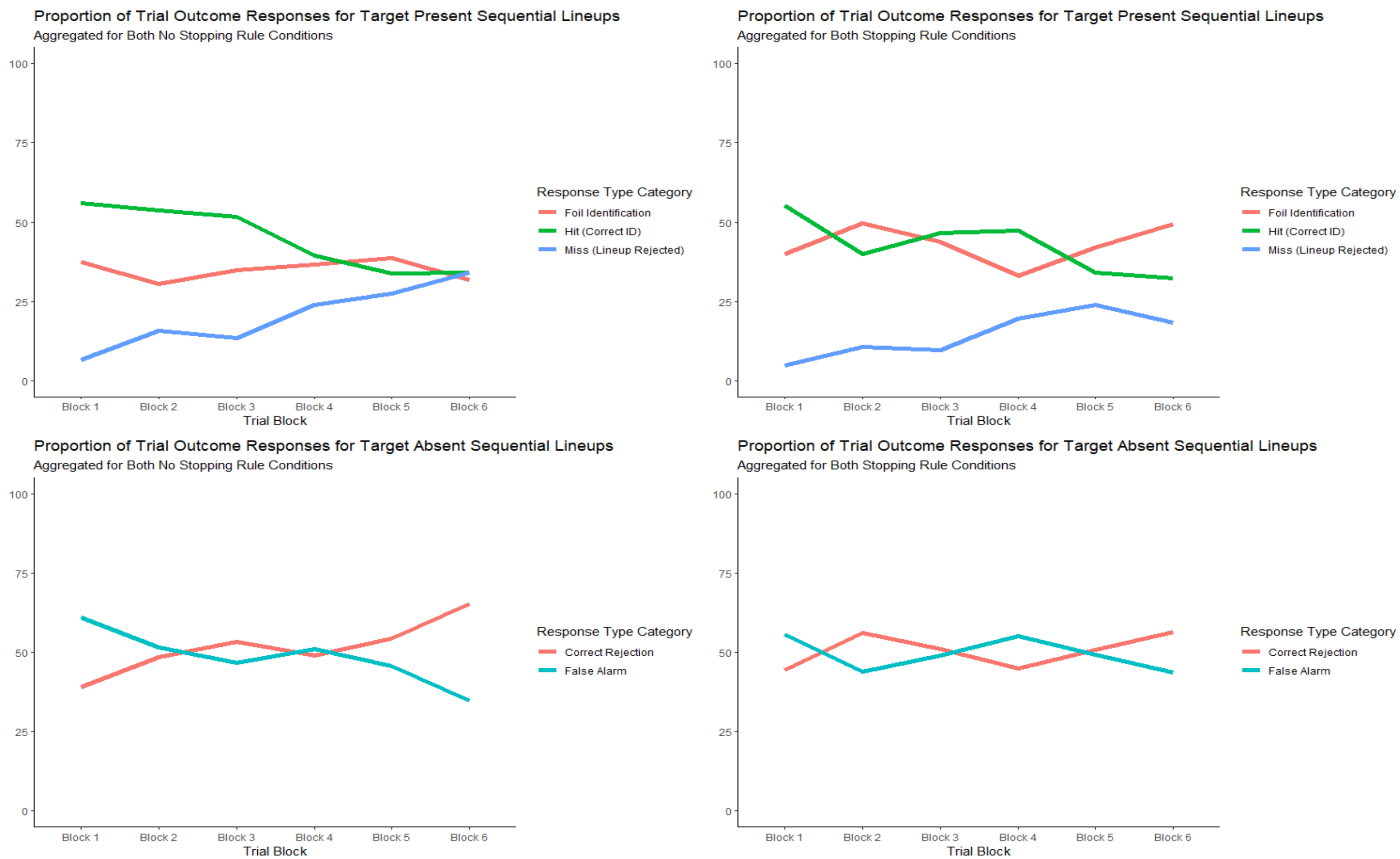
Appendix 6



Appendix 6. The conditions from Appendix A where confidence ratings were given broken up by high (67-100), medium (34-66), and low (0-33) confidence identifications, collapses across stopping rule conditions.

Appendix 7

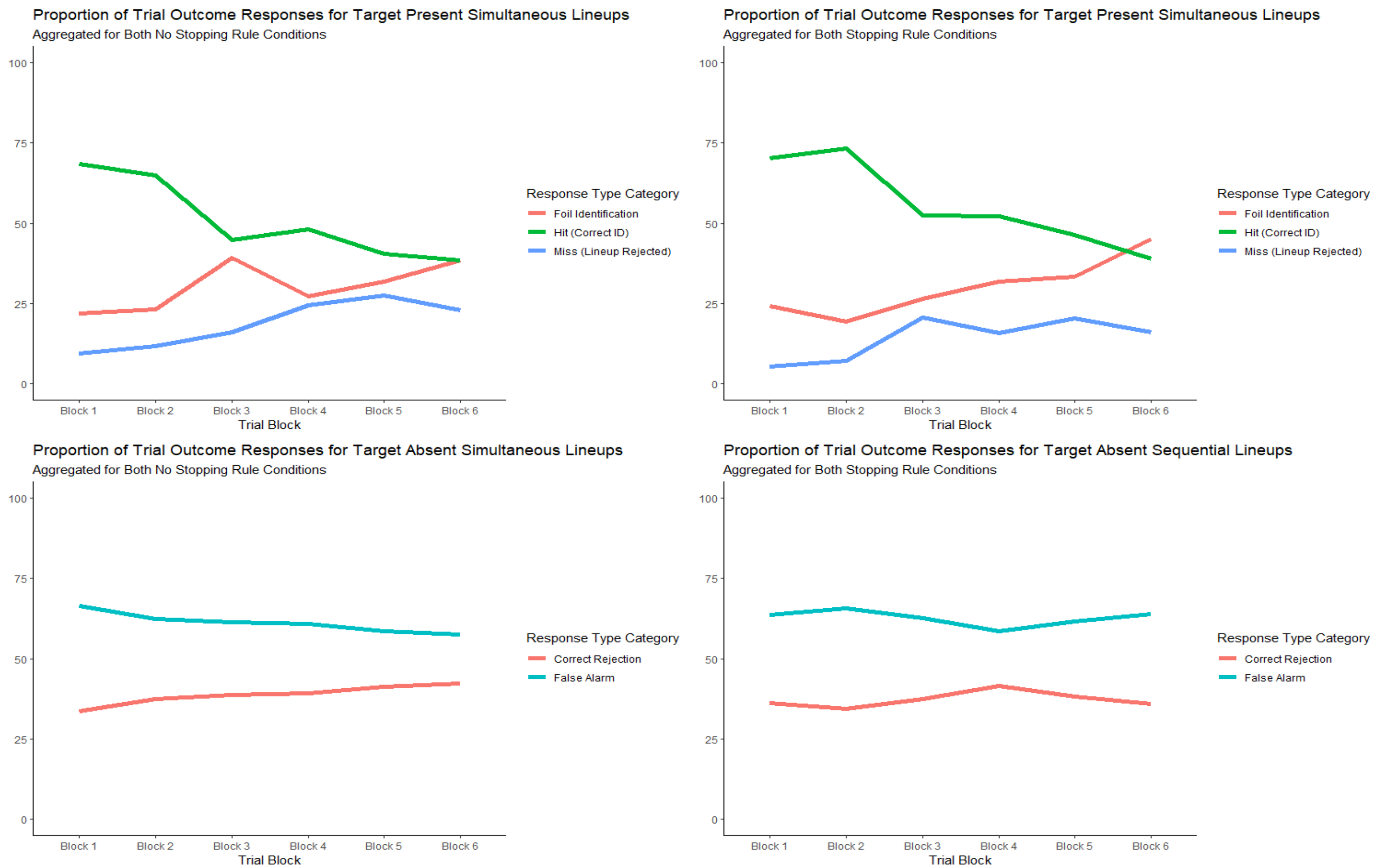
Sequential Response Outcomes by Block



Appendix 7. Sequential lineup data indicating response proportions by trial block. Each trial block indicates responses to 4 individual lineups.

Appendix 8

Simultaneous Response Outcomes by Block



Appendix 8. Simultaneous lineup data indicating response proportions by trial block. Each trial block indicates responses to 4 individual lineups.

References

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, 29, 395-424.

Criss A. H., Malmberg K. J., Shiffrin R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, 64, 316–326.

Duncan, M. (2006). A Signal Detection Model of Compound Decision Tasks. *Defense Research and Development Canada*, 1–56.

Finley, J. R., Wixted, J. T., & Roediger, H. L. (2020). Identifying the guilty word: Simultaneous versus sequential lineups for DRM word lists. *Memory & Cognition*, *48*(6), 903–919. https://doi.org/10.3758/s13421-020-01032-6

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science*, *23*(1), 3–10. https://doi.org/10.1177/0963721413498891

Gross, S. R., Shaffer, M., & Registry of Exonerations, National, Exonerations in the United States, 1989–2012 (2012). U of Michigan Public Law Working Paper No. 277, *7th Annual Conference on Empirical Legal Studies Paper*. http://dx.doi.org/10.2139/ssrn.2092195

Horry, R., Brewer, N., Weber, N., & Palmer, M. A. (2015). The effects of allowing a second sequential lineup lap on choosing and probative value. *Psychology, Public Policy, and Law*, *21*(2), 121–133. https://doi.org/10.1037/law0000041

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. Law and Human Behavior, 36(4), 257–265. https://doi.org/10.1037/h0093959

Innocence Project: DNA Exonerations, (2016). 1989-2014: Review of Data and Findings From the First 25 Years. 79 *Albany Law Review* 717

Kaesler, M., Semmler, C., & Dunn, J. (2017). Using Measurement Models to Understand Eyewitness Identification. In G. Gunzelman, A. Howes, T. Tenbrik, & E. Davelaar (Eds.), Proceedings of the 39th Annual Meeting of the Cognitive Science Society (pp. 1-7). London, UK.

Kellen, D., McAdoo, R.M. (2022) Toward a more comprehensive modeling of sequential lineups. *Cogn. Research* **7**, 65. https://doi.org/10.1186/s41235-022-00397-3

Lindsay, R. C., & Wells, G. L. (1980). What price justice? exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior*, *4*(4), 303–313. https://doi.org/10.1007/bf01040622

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. https://doi.org/10.1037/0021-9010.70.3.556

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Erlbaum.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Mangul, S., Martin, L. S., Eskin, E., & Blekham, R. (2011). Improving the accuracy if gene expression estimates in the absence of a reference genome. BMC bioinformatics, 12(1), 77. https://doi.org/10.1186/1471-2105-12-77

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*(4), 361–376. https://doi.org/10.1037/a0030609

Mickes, L., Seale-Carlisle, T. M., Wetmore, S. A., Gronlund, S. D., Clark, S. E., Carlson, C. A., Goodsell, C. A., Weatherford, D., & Wixted, J. T. (2017). ROC's in eyewitness identification: Instructions versus confidence ratings. *Applied Cognitive Psychology*, *31*(5), 467–477. https://doi.org/10.1002/acp.3344

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less biased criterion setting but does not improve discriminability. *Law and Human Behavior*, 36(3), 247–255.

Police Executive Research Forum. (2014). A National Survey of Eyewitness Identification Procedures in Law Enforcement Agencies. *National Institute of Justice* , 1–142. https://doi.org/doi.org/10.3886/ICPSR34274.v1

Rotello, C. M., & Chen, T. (2016). ROC curve analyses of eyewitness identification decisions: An analysis of the recent debate. *Cognitive Research: Principles and Implications*, *1*(10). https://doi.org/10.1186/s41235-016-0006-7

Seale-Carlisle, T. M., & Mickes, L. (2016). US line-ups outperform UK line-ups. Royal Society Open Science, 3(9), 160300.

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. (2017). The Role of Estimator Variables in Eyewitness Identification. Journal of Experimental Psychology: Applied. 24. 10.1037/xap0000157.

Singmann H, Kellen D (2013). "MPTinR: Analysis of multinomial processing tree models in R." *Behavior Research Methods*, **45**(2), 560–575. doi:10.3758/s13428-012-0259-0, https://doi.org/10.3758/s13428-012-0259-0.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the Sequential Lineup Superiority Effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*(1), 99–139. https://doi.org/10.1037/a0021650

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, *25*(5), 459–473. https://doi.org/10.1023/a:1012888715007

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

Wilson, B. M., Donnelly, K., Christenfeld, N., & Wixted, J. T. (2019). Making sense of sequential lineups: An experimental and theoretical analysis of position effects. *Journal of Memory and Language*, *104*, 108–125. https://doi.org/10.1016/j.jml.2018.10.002

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262–276. https://doi.org/10.1037/a0035940

Wixted, J. T., & Mickes, L. (2015). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition*, *4*(4), 318–323. https://doi.org/10.1016/j.jarmac.2015.08.009

Wixted, J. T., & Mickes, L. (2015). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory and Cognition*, *4*(4), 329–334. https://doi.org/10.1016/j.jarmac.2015.08.007

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive Psychology*, *105*, 81–114. https://doi.org/10.1016/j.cogpsych.2018.06.001

Resume

Gabriella Larson

Syracuse, NY

gklarson@syr.edu

Education

Syracuse University | Syracuse, New York

Cognitive Psychology | 05/2025

Cognitive Psychology PhD program

- Coursework focuses on statistical knowledge, furthering programming skills, comprehensive reading of new literature in the field of computational modeling, neuroscience, and memory and learning
- Completed First Year Project comparing procedures for eyewitness identification including simultaneous lineups and showups
- Currently working on Master's Thesis investigating sequential lineups for eyewitness identification

2021-2022 Graduate Fellowship Recipient

Clubs and Organizations:

- Syracuse Graduate Employees United (SGEU) Union Representative
- Psychology Action Committee (PAC)
    - Positions Held: Cognitive Area Representative (Fall 2021-Spring 2022) and Events Committee (Fall 2022-Spring 2023)
- Psi Chi
    - Application based international honor society in psychology

Primary Advisor: Dr. Amy Criss

Secondary Advisor: Dr. Dan Corral

Suffolk University | Boston, MA

Psychology B.S. | 05/2020

- One year studying abroad in Madrid, Spain
- Member of Honors Program
- Overall GPA: 3.477

- Major GPA: 3.533
- Key Courses: Research Method and Design (A), Behavioral Statistics (A-), Developmental Psychopathology (A-), Physiological Psychology (B-), Psychology Internship (A)
- Clubs: Forensic Science Society (treasurer)

Teaching and Mentoring

Teaching Assistant-ships:

- PSY313-Fall 2022-Spring 2023
    - Editing and optimizing lab materials used to teach three weekly labs consisting of hands-on learning of research based psychological materials.
- PSY205-Fall 2020, Spring 2021
    - Responsibilities included planning and teaching four weekly hour long recitations, executing activities to foster inclusivity and equity while allowing for comprehensive learning of course materials, and being available for one-on-one office hours as needed to support students

Instructor of Record:

- PSY205- Summer 2021, Summer 2022
    - Responsibilities included writing a syllabus, planning weekly lessons, creating and guiding varying types of class activities catering to different styles of learning, designing relevant quizzes and exams, allowing students to demonstrate their knowledge in different ways including through multiple choice quizzes, written essays, and class presentations.

Mentoring Research Assistants:

- Memory and Decision Making Lab RAs- Fall 2021-Present
    - Mentoring current research assistants through assisting with resume building, identifying and applying to graduate school programs, identifying prospective job opportunities for post graduation, teaching valuable skills including how to locate and read empirical research articles, data cleaning and coding based in R and JsPsych, obtaining research participants' informed consent, and how to work with research participants.

Research Interests and Experience

Eyewitness Identification

- Bridging the gap between current eyewitness identification and recognition memory literature in order to better understand how eyewitnesses make their decisions for varying lineup and identification procedures.
- Comparing simultaneous lineups, sequential lineups, and showups, to identify which procedure yields the highest discriminability and best allows eyewitnesses to differentiate between innocent and guilty suspects.

Signal Detection Theory

- Using the theoretical framework to guide research questions pertaining to eyewitness identification, particularly through the use of signal detection theory-based models to compare theoretical and empirical measures of discriminability for varying identification procedures.

Certificates

Health Stream Training, 27 Certificates, October 2019

- National Patient Safety Goals, HIPPA, Ethical Standards, Research Laboratory General Safety Training

Citi Training, October 2019

- Research Ethics and Compliance Training

Professional Skills

---

jsPsych, Rstudio, Matlab, Microsoft Excel, Statistical Package for the Social Sciences (SPSS), Quantitative Sensory Testing