The Effect of Exposure Duration on Perceived Similarity in Simultaneous Lineups

Peter S. Grieve

School of Psychology, University of Adelaide

Word Count: 8361

Author Note

This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of Psychological Science (Honours), University of Adelaide, 2018.

The research reported in this thesis was supported by grant DP160101048 from the Australian Research Council.

Correspondence concerning this article should be addressed to Peter Grieve, Adelaide University School of Psychology, Adelaide, SA, 5000.

Table of Contents

List of Figures

List of Tables

Abstract

Quantifying reliable and accurate eyewitness identification procedures which avoid wrongful convictions and give confidence to justice systems as to the accuracy of suspect guilt continues to be an area of intense research. Defining the parameters of a 'fair lineup' is central to this endeavour. Measures of similarity between lineup members have been key variables used to accurately describe what is and isn't a 'fair lineup'. To date little research has been done on how the perception of similarity may vary across groups and conditions, particularly as a result of memory encoding strength. This study aimed to understand how exposure time, a key variable for altering the encoding strength of a face, in the context of simultaneous lineups, may alter perceived similarity. Results showed that the observed data fit the Unequal Variance Signal Detection (UVSD) model well, however the predicted increases in discriminability with longer exposure duration and higher lineup similarity were not measured. Similarly no significant changes in perceived similarity were found between any of the conditions. Given observed differences in Hit (CID) and False Alarm (FA) rates between low and high similarity lineups this result suggests that judgements of perceived similarity between faces in a line up are unrelated to participants face familiarity judgements. This supports the independent observations assumption within the maximum likelihood method and indicates that overall a priori categorical classifications of lineups as wholly low or high in similarity are less important to discriminability than participants judgements about each faces familiarity to the memory of the target. This finding has implications for future research into 'fair lineup' design and measurement.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this thesis contains no materials previously published except where due reference is made. I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

Signed:

Name: Peter S. Grieve

Date: October, 2018

Acknowledgements

I'd like to sincerely thank my supervisor Dr Carolyn Semmler who's unwavering support and guidance both academic and personal allowed me to complete this work. Thanks also goes to Keith Ransom who not only provided technical support in software development, running the online components of the study and data transformations, but also personal support and encouragement. Similarly, Matthew Kaesler who developed the measurement models and provided technical advice and assistance. Also to Kym McCormack for providing stimulus material assistance and technical advice. Finally, thanks goes to my Wife and boys who's support and belief made this all possible.

Chapter 1

Introduction

1.1 Fair Lineups

The Innocence Project (2017) has identified over 350 wrongful convictions to date, of which up to 70% involve the misidentification of innocent suspects by eyewitnesses. Findings such as these have the potential to negatively influence the confidence that justice systems and the general public place in eyewitness identifications. Current research aims to produce empirical evidence which can guide justice systems as to the reliability and accuracy of the eyewitness decision. Key to this is defining which eyewitness identification procedure reliably maximises guilty and minimises innocent suspect identifications, or discriminability. Further, the statutory principle of 'innocent until proven guilty' demands that justice systems ensure that lineups be 'fair' in order to protect the a priori presumed- innocence of the suspect. As such the identification decision is required to be the result of a reliable match between the eyewitnesses' recognition and their memory of the perpetrator, and not purely as a result of the suspects' salience as the only plausible selection. A fair lineup is one in which all members of the lineup closely resemble the description of the perpetrator (Wells, Yang & Smalarz, 2015), thus ensuring that the suspect does not stand out, irrespective of whether they strongly resemble the memory of the perpetrator. Further, a number of studies (Tredoux, 2002; Clark, 2012; Fitzgerald, Oriet & Price, 2014) have suggested that there might be a level of similarity between lineup members that is optimal for producing the ideal trade-off between correct identification of perpetrators and misidentification of innocent suspects. To date however, little research has been done on how the perception of similarity may vary across groups and conditions, particularly as a result of memory encoding strength. This study aims to understand how exposure duration, a key variable for altering the encoding strength of a face, in the context of simultaneous lineups, may alter perceived similarity.

Understanding how perceived similarity varies across conditions is an important step for the development of methods for testing eyewitness memory and for developing measurement models for the eyewitness task.

1.2 The Lineup Procedure

Police eyewitness identification procedures vary depending on the nature of the crime and the jurisdiction carrying out the investigation, however, the simultaneous photo array is one of the most common methods used across Western police forces (National Research Council, 2014). The photo array consists of a number of photos of faces, usually six, presented simultaneously in a grid pattern. One of the faces will be the suspect and the others fillers. Filler selection should accord with the principles of the fair lineup rule, such that the suspect does not stand out as the only plausible selection. Under optimal conditions the witness will be told that the perpetrator may or may not be present prior to seeing the lineup, after which they will be asked if any of the lineup faces match their memory of the perpetrator. Directly following this identification procedure the witnesses' confidence in their decision should be obtained. Ideally a double blind method should be employed wherein the person conducting the procedure is not linked to the investigation and therefore cannot provide conscious or unconscious cues to the witness.

Given that the suspect may be guilty or innocent, the possible outcome from this procedure is either an identification of a guilty suspect, an identification of an innocent suspect, the selection of a known innocent filler, or the rejection of the lineup. Researchers control the base rate of these outcomes by manipulating the presence, or not, of a known guilty suspect (target) in the eyewitness task, known as a target-present (TP) lineup. A target-absent (TA) lineup is one in which it is known that none of the lineup faces have previously been seen by participants. Correct target identifications are defined as Hits and incorrect suspect identifications as False Alarms (FA). Filler selections and TP lineup rejections are

known to be wrong and classified as a Miss. TA lineup rejections are classified as Correct Rejections (see Table 1).

Table 1

Possible outcomes of simultaneous lineup procedure

	Suspect Selected	Filler Selected	Lineup Rejected
Target-present	Hit	Miss	Miss
Target-absent	FA	Miss	Correct Rejection

Under experimental conditions the presence of the target is controlled and hence innocent subject FAs are known. However, a court does not have this information. The only evidence available is knowledge of the eyewitness procedure used, the conditions under which memory encoding occurred and the confidence of the witness. These variables are generally described as either system or estimator variables (Wells, 1978).

1.3 System and Estimator Variables

System variables are those that are within the control of the Justice System and are manipulated within experiments in order to identify the procedure which optimises discriminability. The similarity of lineups, often categorised as low, moderate or high, has been treated as a system variable to date. Discriminability in this context is defined as the ability to discriminate innocent from guilty suspects when presented with either a low or high similarity lineup. Estimator variables are those which cannot be controlled within the lineup procedure and vary from one situation to another producing unique effects on the memory encoding, retrieval and recognition processes. Examples would be exposure duration, distance, lighting, weapon presence or the elapsed time between event and identification. Researchers looking at the effects of these variables on the accuracy of eyewitness

identifications are specifically measuring how reliable an identification is under those conditions, or the positive predictive value (PPV) that the identification is correct (Semmler, et al, 2018). To measure this, System variables are held constant while an Estimator variable is manipulated and the outcome analysed to quantify differences. It is essential to define the difference between discriminability and PPV, as the first helps us understand and compare procedural effectiveness at the group level, while the second allows us to predict accuracy at the individual level under varying conditions.

1.4 Measurement Models

As scientists it is essential that our theories be falsifiable and our data reproducible. The use of measurement models helps us achieve this goal by forcing us to clearly define all our constructs and by providing us with the means to test our model against observed data using model fitting techniques (Farrell & Lewandowsky, 2010). One such model is the Unequal Variance Signal Detection model (UVSD) and it has become widely accepted as a useful tool for predicting memory accuracy within the eyewitness paradigm (Wixted, Mickes, Dunn, Clark, & Wells, 2015). Procedural effectiveness at maximising memory accuracy is described as the discriminability of that process and is measured as *d'* or the distance between the means of two separate but overlapping distributions. Given a 6 person TP lineup, these distributions are assumed to be Gaussian in nature with foil familiarity modelled as 5 random draws from the innocent suspect/filler distribution (see dotted line Figure 1) and guilty suspect familiarity modelled as a single random draws from the target distribution (see solid line Figure 1). The former is assumed to have a mean of 0 and standard deviation of 1 and the later has its mean quantified by the variable *d'* and its standard deviation by *s*. A 6 person TA lineup is simply 6 random draws from the innocent suspect/filler distribution (see dotted line Figure 1). The x-axis represents the latent variable 'memory strength', or familiarity, and is generally measured by taking the eyewitnesses confidence ratings that their decision is

correct, or decision criterion *c*. This point reflects the response bias of the participant, or the level of familiarity reached when a decision to choose is made. Conservative response biases result in decisions being made only when familiarity is high, resulting in higher reported confidence levels. Liberal response biases lead to selections with lower levels of familiarity and hence reported confidence. The probability that this choice is correct (Hit) or incorrect (FA) is quantified by the overlap of the two distributions at that point.

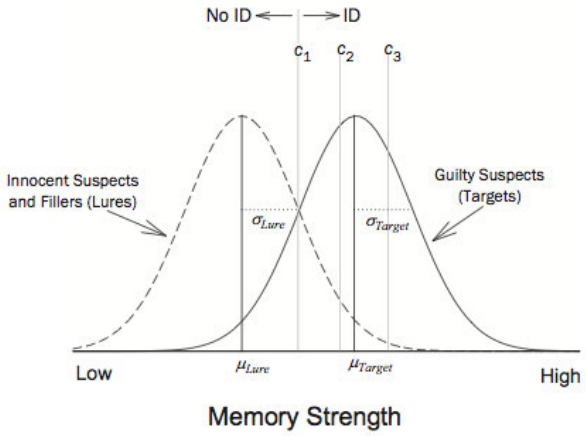


Figure 1. Basic signal detection model of lineups. The decision criterion (c_1, c_2 and c_3) correspond to participants confidence in their decision.

The cognitive processes and models which describe how an eyewitness decides which level of response bias is appropriate are beyond the scope of this thesis, however, as described by Wixted, Vul, Mickes, and Wilson (2018), the MAX decision rule is the one used in the UVSDT model. Under this rule the face with the highest familiarity above the response criterion is the one selected and if no face exceeds this threshold then the lineup is rejected.

Another helpful measurement tool is Receiver Operating Characteristic (ROC) analysis, or the plot of a binary outcomes across a given response range. For a 0.5 base rate the positive diagonal equates to chance response (50/50) for any given response bias (see Figure 2). The negative diagonal equates to discriminability between the two responses. ROC

plots describe how the diagnosticity ratio (Hits / FAs) varies as response bias moves from conservative toward liberal in nature. In eyewitness identification research we are attempting to optimise procedures in order to maximise the diagnosticity ratio across the range of response biases. This is determined by measuring which procedure produces the largest area between the discriminability plot points (ROC curve) and the chance response line, commonly referred to as the partial Area Under the Curve (pAUC) (Wixted & Mickes, 2015). Using this measure allows researchers to capture changes in procedure accuracy which may not be captured using *d'* from the UVSDT model. For example, it is possible that two conditions may produce identical *d'* values while simultaneously having variance differences in the target response distribution. This difference in variance will be observable in an ROC analysis as a skewing of the curve between the two conditions proportional to the magnitude of the effect. It is this fine grained analysis of response bias data which can provide useful PPV measures about identification accuracy and confidence under varying conditions (National Research Council, 2014).

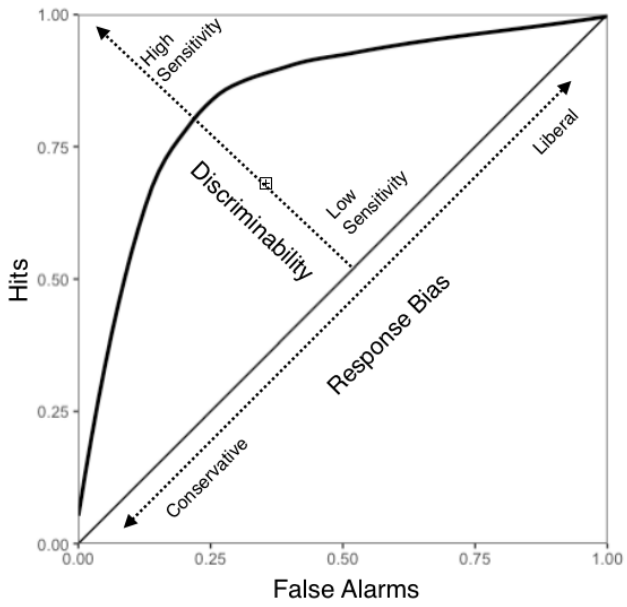


Figure 2 – Receiver Operating Characteristic (ROC) curve

1.5 The Role of Exposure Duration in Recognition Memory and Eyewitness Identification

It is well established that increased exposure duration to the face of a perpetrator is a reliable predictor of increased accuracy (Reynolds & Pezdek, 1992; Memon, Hope & Bull, 2002; Bornstein, Deffenbacher, Penrod & McGorty, 2010, Flowe & Bessemer, 2011). A meta-analysis of face recognition studies which manipulated exposure time found a consistent pattern of improved recognition accuracy with longer exposure duration across many procedural variations (Bornstein, et al, 2012). Further, the effect was found to be non-linear with the greatest improvement occurring across approximately 30 seconds. Semmler and Brewer (2006) found significant improvements in facial recognition accuracy using time scale differences as minimal as 650ms. Palmer, Weber, Brewer and Nagesh (2013) also found significant increases in eyewitness identification accuracy when manipulating exposure time from 5s to 90s. Further, their analysis of the confidence-accuracy relationship showed that higher confidence decisions correlated with high accuracy in both conditions. Exposure duration is thought to influence the cognitive processes for encoding memories such that a longer duration produces a stronger memory trace, or familiarity (Memon, Hope, & Bull, 2003). Within the UVSDT model this would be seen an increase in d' and within ROC analysis as an increase in pAUC.

How exposure duration improves memory encoding, and any subsequent relationship to improved recognition accuracy, is yet to be fully explained and is beyond the scope of this thesis. However, Reynolds and Pezdek's (1992) study suggests that the salience of facial features may be an important factor in the encoding of faces under restricted exposure durations. They found that upper-face features (hair & eyes) were found to be better remembered than lower face features (mouth, chin & nose) in low exposure conditions.

Further, recognition accuracy for all features improved as exposure duration increased. Flowe and Bessemer (2011) found increases in Hit rates from TP lineups when participants were presented with increased feature counts in one condition and increased exposure duration in another. These results suggest that the salience and number of features plays an important part in memory encoding and that this encoding process can be significantly affected by exposure duration. Further, longer exposure duration may allow for more detailed encoding of less salient features which can subsequently aid in improving later recognition accuracy.

1.6 Defining Similarity

Similarity is often referred to in psychological research without providing an exact theoretical definition. The complexities involved in measuring similarity doubtless add to this reluctance. Tversky (1977) argues that similarity judgements are made within frames of reference which provide the context upon which the judgement is made. Changes in context correspond to changes in which features are relevant for making the similarity comparison. The salience of features is then further defined by two factors: intensive and diagnostic. Intensive factors refer to measures of intensity or the signal-to-noise ratio. An example might be a skin tone or hair colour. Diagnostic factors are features upon which comparisons can be made as defined by the set of objects in question. Using these factors people can sort a larger set of objects into smaller clusters based on feature similarity. Similarity is thus described as a feature matching process where the similarity between a group of objects is a linear combination of their common and distinctive features. Tversky (1977) argues that people use these factors to sort objects into clusters in order to maximise within cluster similarity and between cluster dissimilarity. Lineups can be thought of as groups of stimuli waiting to be sorted into clusters, or sets. The memory trace of the perpetrator contains the features which define whether or not a lineup face can be categorised as belonging to the set within which it is probable that the perpetrator may be present. Low similarity lineups, in which only the

suspect is matched well to the perpetrator, make it less likely that anyone other than the suspect will be considered as part of this set. Conversely, high similarity lineups potentially contain more faces with the set of features required to be included in the set from which it is probable that the perpetrator may be present. Once the set is defined, finer grained comparisons of diagnostic features influence the similarity between the members of the set. As stated by Tversky (1977), stimuli within a set are perceived as more similar when compared to stimuli external to that set, however, when compared to each other they are also perceived as more dissimilar due to greater familiarity of the diagnostic differences within the set.

1.7 The Role of Similarity in Eyewitness Memory

The purpose of using similar looking fillers in simultaneous lineups is to attempt to prevent innocent suspects from standing out as the only plausible candidate who matches the description of the perpetrator. However, it has been argued that this must be balanced with not making the fillers so similar that the difficulty of the task prevents eyewitnesses from correctly choosing a guilty suspect when present. Finding this balance has resulted in the 'not to similar' recommendation endorsed by many researchers (Brewer & Palmer, 2010; Malpass, Tredoux, & McQuiston-Surrett, 2007; Wells et al., 1998). However, this argument rests on the assumption that increasing similarity between fillers and targets will result in a more difficult decision. An assumption for which there exists disconfirming evidence (Horry & Brewer, 2016).

Horry and Brewer (2016) used face generation software to create targets and filler siblings morphs. They then varied the similarity of these fillers across a number of decision difficulty conditions to test their theory that confidence in identifications is proportional to the difference between the target and the least similar filler of a lineup. As expected they found decreases in *d'* and confidence with increased filler similarity. However, they also

paradoxically found that choosing rates and confidence increased in lineups when all fillers had high similarity to the target. Evidence against the 'not too similar' assumption. Further, the lack of a significant effect in a 4 person lineup which retained one low similarity filler while the other fillers increased in similarity suggests that the effects of similarity on accuracy and confidence are sensitive to the perceived integrity of the lineup as a congruent whole.

A further complicating factor in perceived lineup similarity is whether or not researchers measure filler similarity to the suspect or to a general description of the perpetrator. Current recommendations in US jurisdictions are to match the foils to the description of the perpetrator (National Research Council, 2014). Either method can result in unintended biases between lineup members (Tunnicliff & Clark, 2000) which, in the UVSDT model, may violate the assumption that the familiarity of any one face can be described as a random draw from a single Gaussian distribution.

Meta-analysis of previous studies have shown that gains in discriminability for high similarity lineups occur as a result of FAs decreasing significantly more than Hits (Fitzgerald, et al , 2014; Clark, 2012). This suggests that the low similarity lineups contained a bias leading towards increased selections from TA lineups. In this study we avoid these complications by creating equal differences in similarity across all the faces in each lineup and condition – instead of focussing the manipulation of similarity on the difference between the target/innocent suspect and the fillers, we manipulate the entire similarity space of the lineup as a whole – and with reference to a normative set of similarity ratings across stimuli. Thus no face should resemble the target more than any other ensuring that the probability of one face being chosen from a TA lineup remains 1 in 6, across all conditions. Similarly, for TP lineups, the target should be equally more familiar than each of the surrounding fillers. Thus any change measured in d' or response bias between low and high similarity lineups,

within the same exposure condition, should be an unbiased result of the similarity manipulation. The former measured by the UVSDT model and the later identified via ROC analysis. This method ensures the integrity of each lineup as whole within which inter-item similarity remains constant and also provides a more ecologically valid methodology than using computer generated morphs.

Drawing on the nature of similarity as a feature comparison process, Wixted and Mickes (2014) proposed The Diagnostic Feature Detection Hypothesis in support of using unbiased simultaneous lineups. Their argument is that the simultaneous presentation of faces provides the decision maker with a finer grain comparison of common features with which to match to the memory trace, much the same as Tversky's (1977) argument. If true, increased exposure duration may amplify this effect by increasing the number of encoded features available for comparison. Tversky's (1977) findings support the view that increased feature encoding, via higher exposure, should result in better discriminability, and lower similarity, between a group of faces categorised as belonging to a common set. Further, in the high exposure high similarity condition increasing these diagnostic factors should result in a larger decrease in similarity between the target and fillers while still maintaining the integrity of the common set. Retaining the assumption that all fillers could be probable matches to the memory of the perpetrator. Whether the magnitude of this change in similarity, as a result of increased exposure duration, is great enough to have a significant effect on d' is the central question of this study.

1.8 Hypotheses

As a test of these statements the hypotheses of this study are: 1. High exposure duration will result in increased discriminability (d') over low exposure duration in a simultaneous lineup identification task; 2. Higher similarity lineups will result in increased discriminability (d') over lower similarity lineups in a simultaneous lineup identification task;

3. That the similarity between members of a lineup and the target is changed as a result of exposure duration such that: a. The reduction in perceived similarity will be higher with higher exposure duration; b. The greatest difference in perceived similarity will occur between the low exposure high similarity condition and the high exposure high similarity condition.

Chapter 2

Method

2.1 Ethics

The study was approved by The University of Adelaide School of Psychology Human Research Ethics Subcommittee (see Appendix A). Participants undertook the study anonymously and were free to withdraw any time. Consent was required from each participant before they could begin the experimental task, and participants indicated this consent by clicking 'next' on the relevant online page (see Appendix B).

2.3 Open Science Framework Pre-Registration

In accordance with the recommendations outlined by the Open Science Framework (ref). The study design, data analysis plan and procedure were pre-registered prior to any data collection (insert project link here).

2.2. Participants

A total of 1122 participants took part in the study, all recruited via the Amazon Mechanical Turk online portal. Tables 1 & 2 details the demographics for each sample. The majority of participants were citizens of the United States of America. Exclusion criteria for the study stipulated that participants must be at least 18 years of age, be proficient in English and have normal or corrected-to-normal vision. 527 participants undertook the a priori similarity rating task and were then not eligible to undertake the main study (N = 595).

2.3. Materials

2.3.1. Face Stimuli

The 90 female faces images used for the experiment came from the Adelaide University Psychology face database. These photographs were cropped from the original torso plus face size of 2850 x 4270 pixels to 1200 x 1800 pixels centred on the face and excluding as much of the torso as possible. Due to variations in the camera to subject distance

of the original photographs, some of the images required slight variations from the 1200 x 1800 resolution to create a parsimonious face size across the stimulus set.

Table 1

Participant demographics a priori similarity ratings task (n = 527)

Country	Female	Male	Unspecified	Mean Age (*SD*)	Age Range
United States	191	248	4	38.14 (*11.22*)	18 - 74
India	21	49	0	30.21 (*6.12*)	21 - 46
Other	5	8	1	35.63 (*9.83*)	19 - 59

Table 2

Participant demographics main study (n = 595)

Country	Female	Male	Unspecified	Mean Age (*SD*)	Age Range
United States	214	302	1	33.84 (*10.93*)	19 - 71
India	16	44	0	29.33 (*5.70*)	22 - 62
Other	6	12	0	33.61 (*11.06*)	23 - 68

2.3.2. Experimental Interface

Data for similarity stimulus materials and main study were collected via an online web application (computer software) coded in JavaScript. This was accessed via Amazon's Mechanical Turk framework (AMT, see https://www.mturk.com/mturk/help?helpPage=overview). AMT participants did not have direct contact with the experimenter and all participants executed the task via a web-browser. All instructions were designed to be self-contained, and were provided online via the participant's web-browser. Workers within the AMT platform conduct tasks in exchange for payment. Payment is task

based with the rate of payment set at an amount equivalent to approximately US $10.00 per hour.

2.3.3. Stimulus Development Study using Similarity Ratings

For lineup construction it was necessary to construct a distance matrix of similarity ratings for each pair of faces from human raters. The scale used for measuring similarity was a 9-point Likert scale with 2 annotations: 1 = Very High Similarity and 9 = Very Low Similarity. The same scale was used for the lineup similarity ratings task in the main study. Ratings were obtained online (see Experimental Interface). The female faces from the database were chosen over the male faces due to an a priori assessment that a lower similarity cluster of faces was more probable from the female images.

2.3.3.1 Similarity Ratings Procedure

The software presented two faces from the face database and the participant was asked to rate the similarity of the faces using the Likert scale. The number of pairs to be rated was shown at the top left hand corner of the page and indicated the progress of the participant. Each participant rated 45 pairs of faces. A total of $N = 527$ participants completed the task. When completed a message was presented thanking the participant and detailing the AMT code required to receive payment.

2.3.4 Multidimensional Scaling (MDS) and Cluster Analysis

The goal of the MDS and cluster analysis was to identify high and low similarity clusters from the pairs-ratings distance matrix (see Similarity Ratings). From these clusters high and low similarity lineups were constructing wherein each face differed from the other as equally as possible (see Appendix C for methodology). The mediod of the high similarity cluster was selected as the target (see Figure 3).

2.3.5 Lineups

From the faces identified using the cluster analysis procedure, 4 lineups were created: 1 x low similarity target-present; 1 x low similarity target-absent; 1 x high similarity target-present; 1 x high similarity target-absent.

2.3.6 Target Videos

Exposure to the target was delivered via video obtained from the Adelaide University Psychology face database. A 5 second count down prompts the participant to prepare for the video proper. This initially shows the target sitting at a desk with their back to the camera. The target then rises from the desk, walks toward the camera and out of the field of view toward the bottom right hand side of the screen. The video is shot from a camera positioned in the corner of the ceiling above, behind and to the right of the desk. The target looks up at the camera as they are walking out of the field of view of the camera. In the short exposure condition participants will be exposed to one viewing of the video with a 0.1 sec exposure of the target looking directly at the camera. In the long exposure condition participants will be exposed to one viewing of the video with a 1.5 sec exposure of the target looking directly at the camera. In the nil exposure condition the video shows just an empty room. Low short exposure video is 10 seconds and the long exposure video is 11.5 seconds in length.

2.4. Main Experiment Procedure

Participants were first presented with a Participant Information Statement detailing the purpose of the experiment, what they were required to do, study duration and contact information for questions or complaints. Consent was given when participants clicked on the 'I agree' option at the bottom of the statement. Three verification questions were asked to ensure that the participant understood the Participant Information Statement. Failure to answer the questions correctly resulted in the Participant Information Statement being displayed again. The participant could then re-read the statement and continue. After

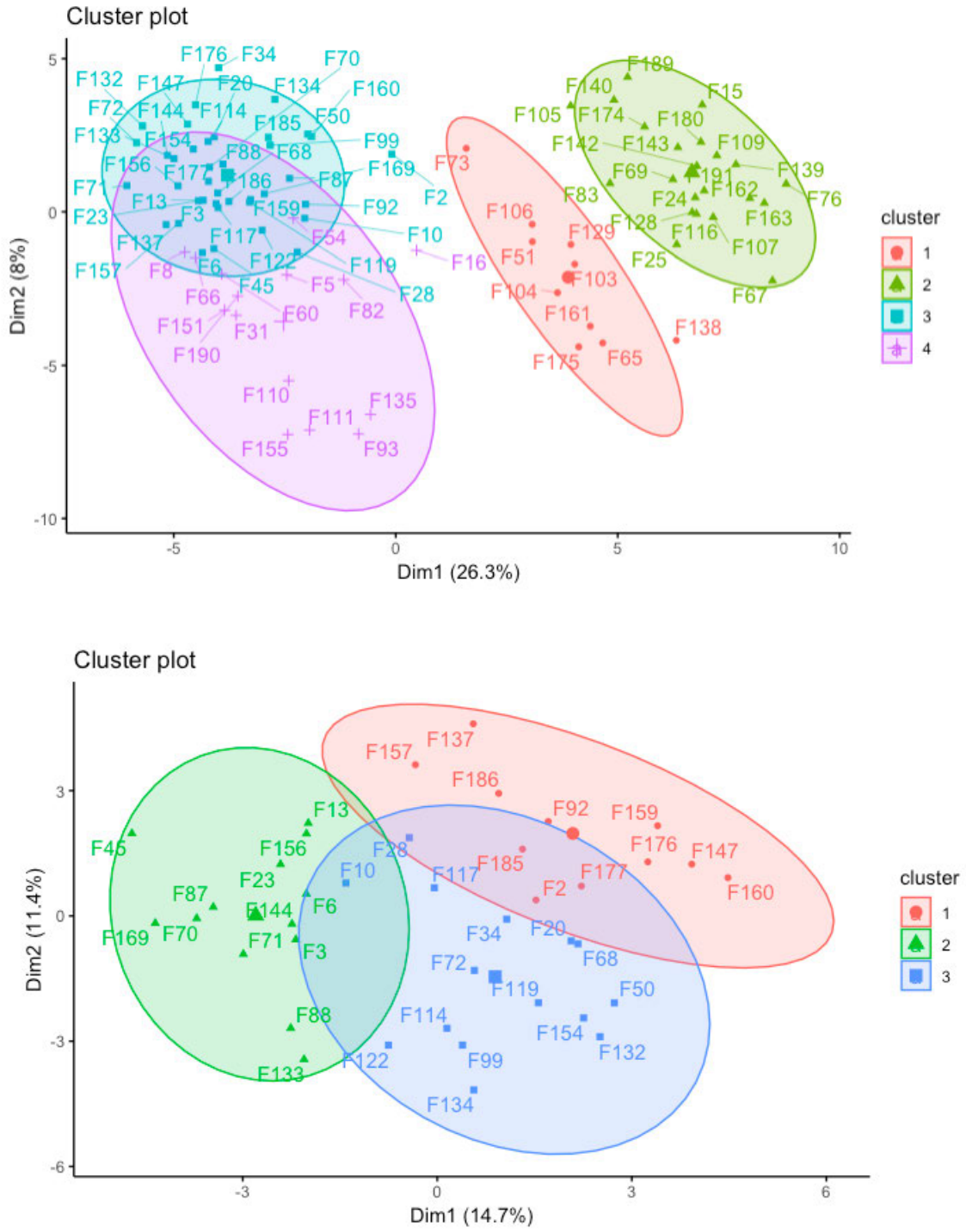


Figure 3. Top plot shows initial cluster analysis of all faces. Bottom plot shows secondary analysis of cluster 3 from initial analysis. Mediod of cluster 3 from second analysis selected as the target.

correctly answering the verification questions the experiment commenced by clicking the 'Continue' option. Demographic data of age, country and gender were obtained. Followed by a loading screen which informed the participant that the video they are going to view is being selected. When this was completed a 'play' button icon appeared to enable them to continue and start the testing phase.

The participants were randomly assigned to either the nil, short or long exposure condition. Participants viewed the appropriate length video after which they undertook a distractor task. Next they were informed that they will be asked to view a lineup which may or may not contain the target from the video. If they recognised the target they should indicate this by selecting the appropriate face in the lineup and then provide a rating between 0 to 100 % which reflects their confidence that their selection is correct. All participants then undertook the similarity ratings task. With all of the lineup faces remaining present on the screen, each combination of pairs of faces was presented and the participant rated their similarity. When completed a message was presented thanking the participant and detailing the AMT code required to receive payment.

2.5 Design

The study a between-subjects 3x2x2 design. The Independent Variables being Exposure Duration (Nil, Short, Long), Lineup Similarity (Low, High), Target (Present, Absent). The dependent variables were perceived lineup similarity, identification performance and confidence.

2.5.2 Perceived Lineup Similarity

Perceived lineup similarity is a representation of cluster cohesion, measured from the similarity pairs ratings matrix of the 7 faces of each participant in each lineup condition. The value is derived by calculating the total within sum of squares value from the post identification similarity pairs ratings.

Chapter 3

Results

Alpha was set at .05. Planned comparisons for target present lineups were :

3a No Exposure Low Similarity vs (Short Exposure Low Similarity plus Long Exposure Low Similarity)

3a.1 Short Exposure Low Similarity vs High Exposure Low Similarity

3b No Exposure High Similarity vs (Short Exposure High Similarity plus Long Exposure High Similarity)

3b.1 Short Exposure High Similarity and Long Exposure High Similarity

3c No Exposure vs (Short Exposure Low Similarity plus Long Exposure High Similarity)

3c.1 Short Exposure Low Similarity and Short Exposure High Similarity

3d No Exposure vs (Short Exposure High Similarity plus Long Exposure Low Similarity)

3d.1 Long Exposure Low Similarity and Long Exposure High Similarity

Prior to testing of hypothesis 1 and 2, the Unequal Variance Signal Detection Model (UVSD) using the max decision rule was fit to the observed data, to obtain estimates of *d'*, criteria (5 in total) and *s* (the standard deviation of the target distribution) . Table E1 details frequency counts collapsed across similarity and exposure conditions used as observed data within the model fitting process. A χ^2 goodness-of-fit statistic was computed by comparing the predicted data from the model to the observed data. The χ^2 goodness-of-fit between the predicted and observed data were subsequently minimised through the adjustment of 7 model parameters ($c_1 - c_5$, *d'*, *s*), using a maximum likelihood method (Dunn, 2010). Table 3 shows that the model fit the data well in all conditions ($p > .05$). Thus *d'* and the predicted data from the

model for each condition was found to be suitable for hypothesis testing. Figures 4 and 5 graphically display the predicted data in a signal detection format.

Table 3

Chi-Squared (χ)^2 goodness-of-fit test results and model parameters collapsed across conditions.

				Model Parameters						
						Confidence Criterion				
Condition	χ^2	(df)	*p*	*d'*	*s*	c_1	c_2	c_3	c_4	c_5
Short Exposure	0.42	(10)	1.00	2.00	0.58	1.26	1.34	1.46	1.75	2.22
Long Exposure	10.85	(10)	.37	1.81	0.61	1.19	1.24	1.30	1.54	1.93
Low Similarity	8.00	(10)	.63	2.04	0.59	1.38	1.45	1.54	1.73	2.14
High Similarity	3.61	(10)	.96	1.73	0.57	1.08	1.14	1.23	1.55	1.97

For testing hypotheses 1 and 2 the predicted Hits (CID) and FA rates at each criterion from each condition (see Table 4) were compared with a multinomial goodness-of-fit test by Monte-Carlo simulations using a log-likelihood ratio test statistic (Engels, n.d.). Hypothesis 1 predicted that longer exposure duration would result in increased discriminability (*d'*) over short exposure duration in a simultaneous lineup identification task. A multinomial goodness-of-fit log-likelihood ratio test of long exposure duration to short exposure duration predicted model data returned a non-significant result ($\chi^2(3) = -6.67$, $p = .177$), thus indicating that there is no statistical difference between the conditions. Further, the direction of the difference was opposite to that predicted by hypothesis 1, with the short exposure condition having higher discriminability (*d'*) than the long exposure condition. Figure 6 illustrates the model data plotted using Receiver Operator Curves (ROCs).

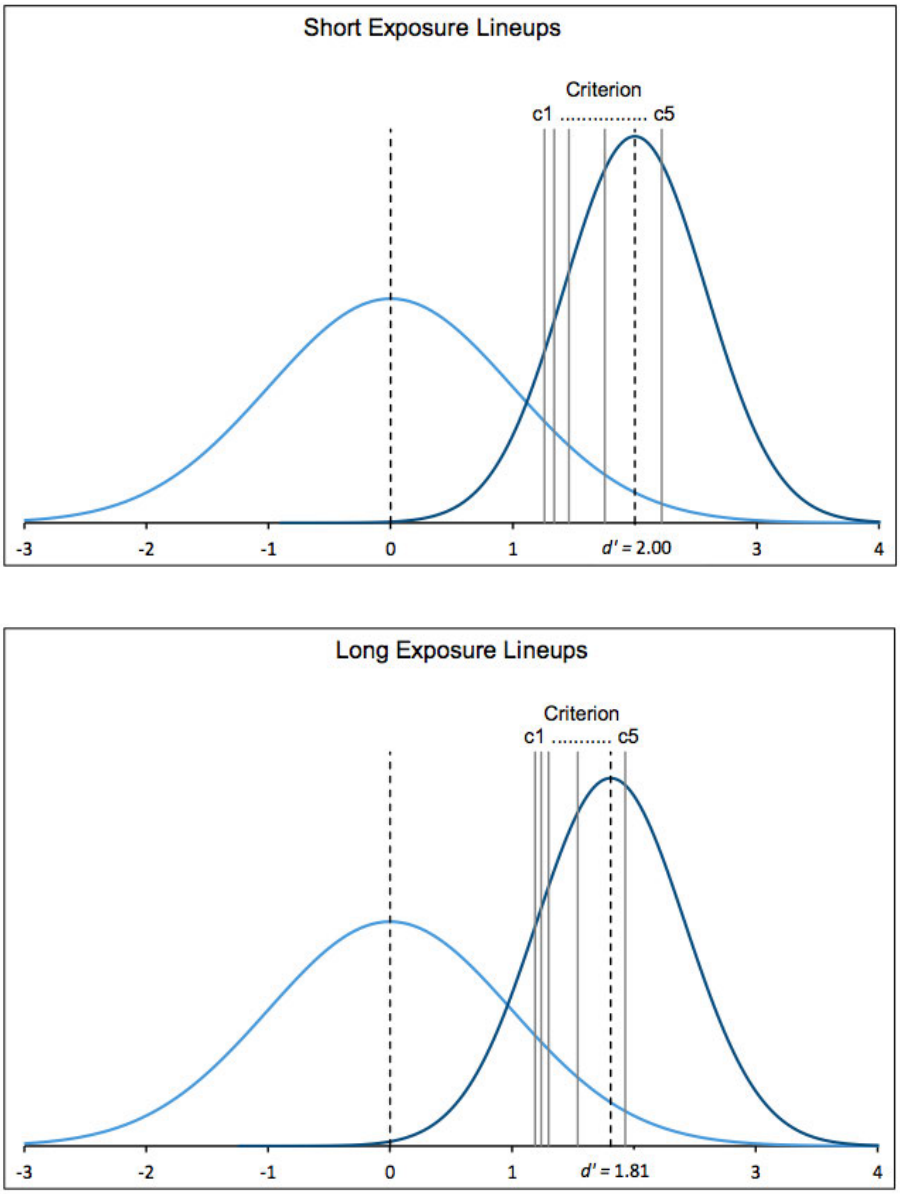


Figure 4. Plots of unequal variance signal detection model predicted data for short and long exposure lineups.

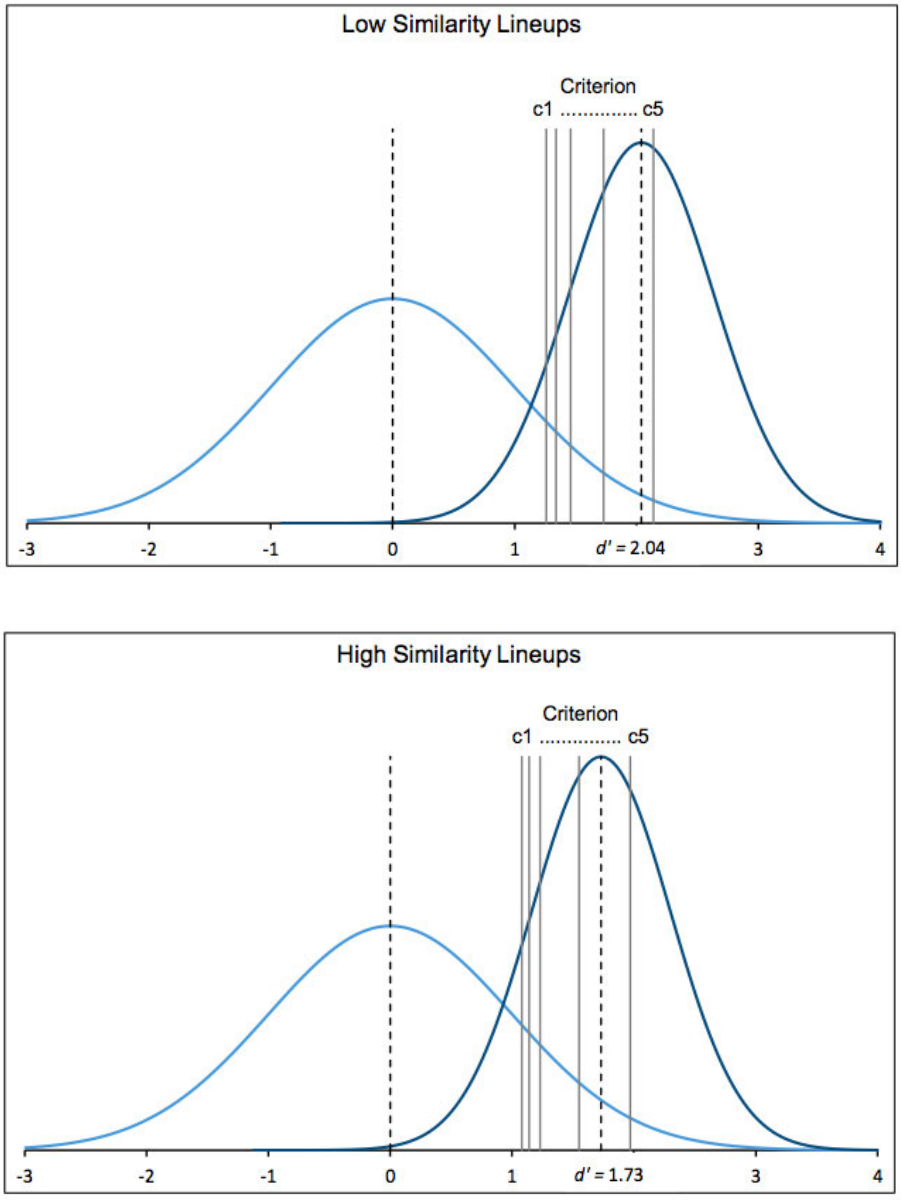


Figure 5. Plots of unequal variance signal detection model predicted data for short and long exposure lineups.

Hypothesis 2 predicted that high lineup similarity would result in increased discriminability (*d'*) over low lineup similarity. A multinomial goodness-of-fit log-likelihood ratio test comparing high lineup similarity to low lineup similarity returned a non-significant result ($\chi^2(3) = -7.35$, $p = .117$), thus indicating no statistical difference between the conditions. Equally, the direction of the effect was opposite to that predicted by hypothesis 1, with the low similarity condition having higher discriminability (*d'*) than the high similarity

condition. Figure 7 illustrates the model Receiver Operator Curves (ROCs).

Table 4

Predicted model counts by condition.

			Confidence Criteria					
Condition	Lineup		c_1	c_2	c_3	c_4	c_5	Rejections
Short Exposure	TP	CID	33.91	28.12	11.77	3.28	1.54	4.30
		TD	34.93	31.84	15.52	4.98	2.55	10.18
	TA	FA	7.27	13.51	13.46	6.86	4.46	49.43
Long Exposure	TP	CID	42.79	21.35	9.28	1.78	1.19	5.76
		TD	45.38	26.45	13.87	3.09	2.28	15.95
	TA	FA	15.30	16.68	14.21	4.08	3.20	47.53
Low Similarity	TP	CID	45.48	25.40	8.77	3.04	1.85	6.87
		TD	47.05	29.09	11.38	4.32	2.78	14.39
	TA	FA	9.30	13.17	9.45	4.95	3.86	59.26
High Similarity	TP	CID	31.41	23.06	11.78	1.99	1.16	3.82
		TD	33.30	28.33	18.11	4.07	2.90	11.28
	TA	FA	13.24	16.97	18.12	5.27	3.85	38.56

Note. TP = target present; TA = target absent; CID = correct identifications (target selection from TP lineup); TD = total identifications from TP lineup; FA = false alarm (filler selection from TA lineup).

Testing of hypothesis 3 was to consist of 4 two part planned comparisons using a 3x2x2 one way Anova model, however, Shapiro-wilk tests of normality (see Table F1) indicated that the dependent variable in most conditions was non-parametric. As a result Kruskall-Wallis H test statistics were deemed to be the appropriate analysis measures. Table F1 details descriptive statistics for each condition.

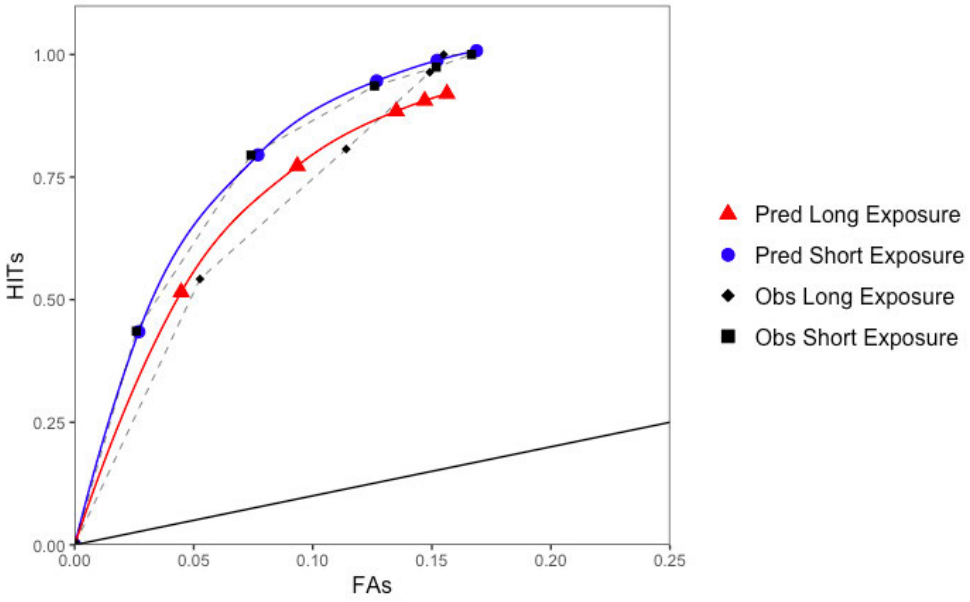


Figure 6. ROC plot of predicted (Pred) model and observed (Obs) data by exposure condition.

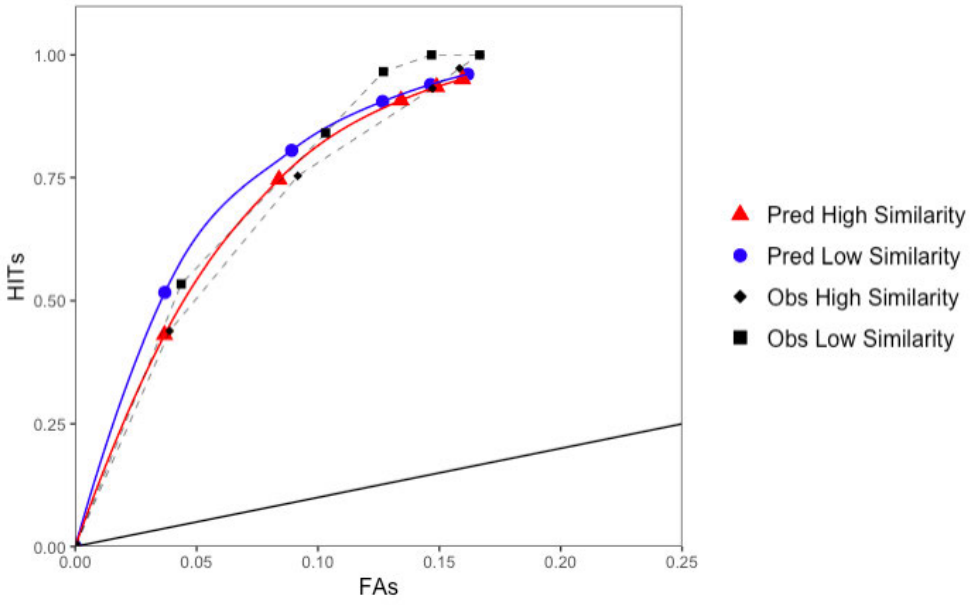


Figure 7. ROC plot of predicted (Pred) model and observed (Obs) data by exposure condition.

Hypothesis 3a predicted perceived Lineup Similarity to be significantly different between the no exposure (control) (*Mdn* = 47.07), short exposure low similarity (*Mdn* =

47.73) and long exposure low similarity (*Mdn* = 47.67) target present conditions. A Kruskall-Wallis H test indicated that there was no significant difference in perceived similarity (H(2) = 0.35, *p* = .846) between the conditions (see Figure 8).

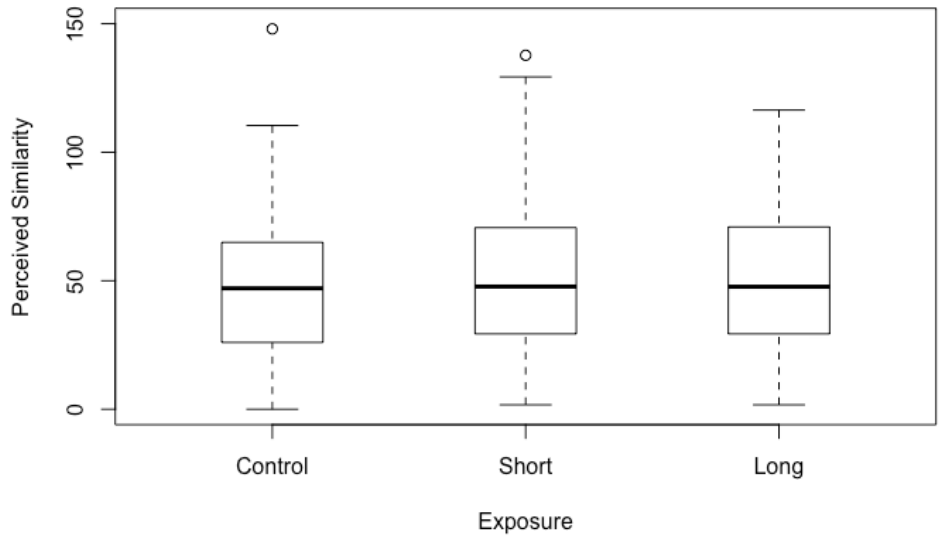


Figure 8. Perceived similarity of low similarity target present lineups by exposure condition; control = no exposure.

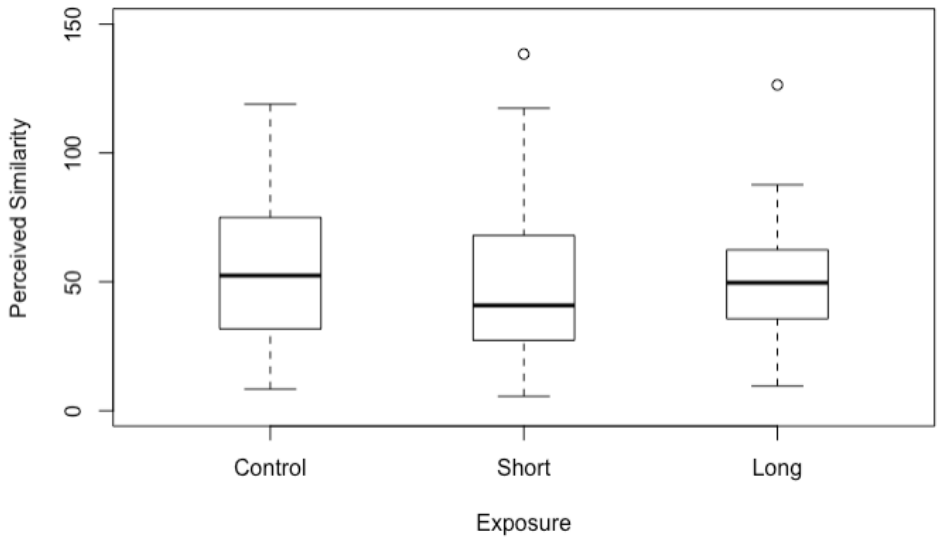


Figure 9. Perceived similarity of high similarity target present lineups by exposure condition; control = no exposure.

Hypothesis 3b predicted Perceived Lineup Similarity to be significantly different between the no exposure (control) (*Mdn* = 52.40), short exposure high similarity (*Mdn* = 40.93) and long exposure high similarity (*Mdn* = 49.60) target present conditions. A Kruskall Wallis H test indicates that there was no significant difference in perceived similarity (H(2) = 1.13, *p* = .568) between the conditions (see Figure 9).

Hypothesis 3c predicted Perceived Lineup Similarity to be significantly different between the no exposure low similarity (*Mdn* = 47.07), no exposure high similarity (*Mdn* = 52.40), short exposure low similarity (*Mdn* = 47.73), long exposure high similarity (*Mdn* = 49.60) target present conditions. A Kruskall-Wallis H test indicates that there was no significant difference in perceived similarity (H(3) = 0.983, *p* = .805) between the conditions.

Hypothesis 3d predicted Perceived Lineup Similarity to be significantly different between the no exposure low similarity (*Mdn* = 47.07), no exposure high similarity (*Mdn* = 52.40), short exposure high similarity (*Mdn* = 40.93), long exposure low similarity (*Mdn* = 47.67) target present conditions. A Kruskall Wallis H test indicates that there was no significant difference in perceived similarity (H(3) = 1.38, *p* = .711) between the conditions.

To achieve a complete analysis of the data, target absent perceived lineup similarity was visually inspected via box plots (see Figure 10). The plots show no visible differences which warrant further investigation.

Appendices F to K contain all of the R script files used to conduct the statistical comparisons and generate the plots.

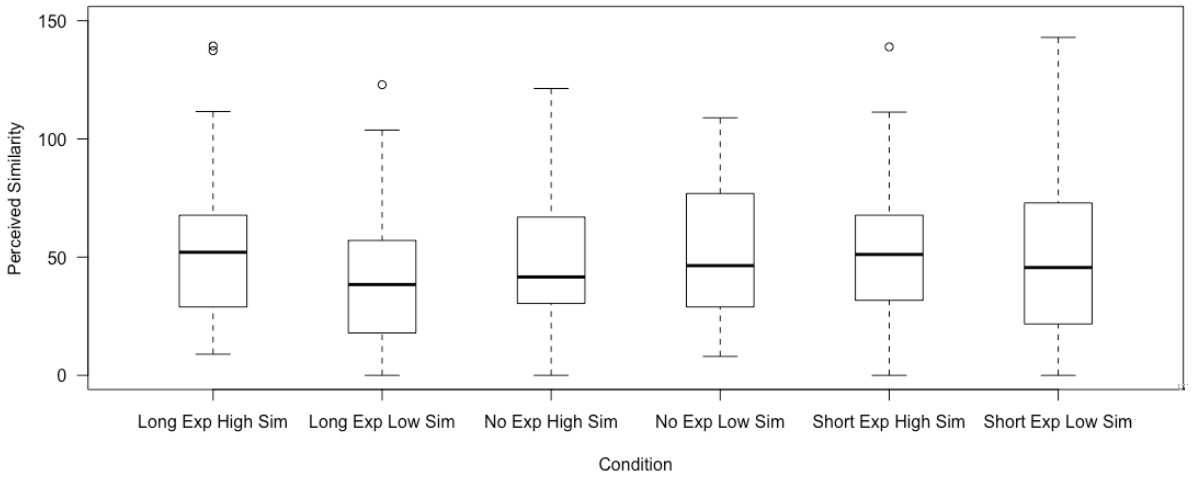


Figure 10. Perceived similarity of target absent lineups by condition.

Chapter 4
Discussion

The aim of this study was to measure changes in discriminability (*d'*) and perceived similarity between faces in a simultaneous lineup under varying exposure duration and filler similarity conditions. A key prediction was that longer exposure duration would increase the strength of memory encoding of the target and thus, improve later identification accuracy (Reynolds & Pezdek, 1992; Memon, Hope & Bull, 2002; Bornstein, Deffenbacher, Penrod & McGorty, 2010, Flowe & Bessemer, 2011). Wixted and Mickes' (2014) Diagnostic Feature Detection Hypothesis proposed that high similarity lineups should aid in target identification as a result of an increased ability to undertake finer grained comparisons of common and dissimilar features between high similarity faces. When combined with the proposal that increased feature counts from improved memory encoding could result in decreasing similarity between the target and fillers in a lineup, it was hypothesised that the effect of longer exposure duration would produce the greatest increase in perceived similarity between members of a lineup and a target in long exposure high similarity conditions.

4.1 Hypotheses 1 and 2

To test for changes in *d'* as described by hypotheses 1 and 2,the observed data, collapsed across exposure and similarity conditions, was fit to the UVSDT model. The observed data fit the model well (see Table 3) therefore *d'* was considered to be an accurate measure of the discriminability of each condition. Further, the UVSDT plots comparing short to long exposure (see Figure 4) and low to high similarity lineups (see Figure 5) gave us an accurate representation of the response criteria and distribution characteristics of each condition. It can be seen from these plots that although the response criteria contracted slightly with lower *d'* values, the change is proportional to the decrease in *d',* indicating that the averaged response criteria of participants across all the conditions remained analogous. This, along with the fact that the target distribution variance (*s)* remained almost constant

across conditions, eliminates any suggestions that response criteria or distribution variance may have diverged while *d'* remained constant between conditions. The similarity of the curves for each condition illustrated in the ROC plots (see Figures 6 & 7) confirm this finding. Finally, the result of the multinomial goodness-of-fit log-likelihood ratio tests between short and long exposure, and, low and high similarity predicted frequency counts confirmed that there was no significant difference. These non-significant findings lead to only one conclusion: that discriminability between the conditions did not significantly change as a result of either increased exposure duration or higher similarity beyond what could be expected by chance.

For the short and long exposure conditions this result may be explained by deficiencies in the short and long exposure materials (further explored in the limitations section), too weak a manipulation between conditions, or simply by an insignificant effect size. However, given the large difference in a priori similarity ratings used to create the low and high similarity lineups, the lack of any significant change in discriminability between these conditions requires a more detailed explanation. Fitzgerald, et al (2014) explained gains in discriminability for high similarity lineups as a result of FAs decreasing more significantly than Hits (CID's), mostly via increased filler selection from high similarity target absent lineups. Clark (2012) describes explanations such as this as the *no cost view* of lineup design in which innocent suspect identifications can be reduced without the cost of a simultaneous reduction in guilty suspect identifications and questions the results of studies which propose this as being possible. In our study we did not nominate an innocent suspect, rather created equally similar target absent fillers to the target, therefore filler vs nominated innocent suspect selections (FA's) cannot be directly compared. However, in our study FA's were higher in high similarity conditions, particularly within the 41-60% confidence range (from low = 6 to high = 20) (see Table E1), the opposite to Fitzgerald, et al's (2014) findings.

Further, Fitzgerald, et al's (2013) meta-analysis also noted that changes in lineup similarity had no effect on lineup rejections, however this study found a large difference in correct rejections. 58 in low compared to 36 in high similarity lineups (see Table E1). A closer look at total identifications (see Table E1) also shows that the number of choosers was higher in target absent and lower in target present high similarity lineups. Combined with the lower high similarity Hit (CID) rates this indicates that participants were both less accurate when the target was present and more willing to choose when the target was not present in high similarity lineups. This also is the opposite to the findings of Fitzgerald, et al (2014), and explains why the results differ in this study. Further, although the difference between the predicted frequency counts for each condition was not significant, these results explain why the high similarity lineups generated lower *d'* values than low similarity lineups and also supports Clark's (2012) assertion that the *no cost view* of lineup design is a false one. It is also likely that not having a nominated innocent suspect matched to the description of the target in our target absent lineups also had an impact on our results. This issue is discussed further in the lineup construction methodology section (4.4).

4.2 Hypotheses 3a, 3b, 3c and 3d

Hypotheses 3a, 3b, 3c and 3d predicted changes in perceived similarity between faces in a lineup with increasing exposure duration to the target and increasing filler similarity. Kruskall-Wallis H tests did not support any difference in perceived similarity across any of the comparisons, thus failing to support our primary hypothesis that the perceived similarity between members of a lineup and the target is changed as a result of exposure duration. Interestingly, the lack of difference in perceived similarity between the no exposure and short and long exposure conditions indicates that the presence of an encoded memory of the target, regardless of exposure duration, meant that how participants rated the similarity of the lineup faces was not measurable in this study design. Figures 8, 9 and 10 show how little variance

existed in the participants' use of the similarity measurement scale. This suggests that participants calibrated their use of the scale to fit the variance of similarity they perceived between each specific set of faces, creating a different similarity measurement outcome from the a priori similarity ratings used to construct those lineups. Given the nature of the task, that is identifying a previously seen target from a lineup, it was anticipated that the memory of the target would act as a reference point upon which the overall perceived similarity of the lineup could be judged. The fact that the results clearly show that this was not the case was unexpected. In particular, the lack of any differentiation between the low and high similarity lineups was surprising. Especially as target memory retrieval and comparison processes would have been in use during the target identification task, thus bringing the encoded memory of the target to psychological attention. In contrast to this however, the data from the low and high similarity lineups show substantive differences in Hit (CID) and FA rates (see Table E1), inferring that the similarity of the lineup faces to the memory of the target had an effect on the levels of familiarity and subsequent selections from these lineups. However, changes in judgements of perceived similarity were not measured after these processes. This provides supporting evidence for the assumption of independent observations inherent in the maximum likelihood method used in the UVSDT model. What appears to be critical is not how similar each face is to each other, and the target, but rather how many faces breach the level of familiarity required for a selection to be made. This suggests that high similarity lineups simply contain more faces with a higher likelihood of breaching this critical threshold. Something which was not captured using the perceived similarity measure.

A closer look at the Mean similarity ratings by condition (see Table F2) shows that in both target present low similarity lineups, the Mean similarity decreased when compared to the other conditions. What this possibly indicates is that perception of low similarity required the physical presence of the target within the lineup to act as a reference upon which low

similarity could be judged. When the target is absent, the memory of the target does not appear to have an effect on how the lineup similarity level is judged.

4.4 Lineup construction methodology

A secondary aim of this study was to explore a new methodology for creating unbiased experimental lineups using a priori similarity ratings, multidimensional scaling and cluster analysis. This method aimed to ensure that the psychological distance between faces, measured as similarity, were as equal as possible. This then minimises the possibility that one of the target absent faces stands out as the only plausible match to the target and solves the match-to-description / match-to-suspect confound faced in previous studies (Tunnicliff & Clark, 2000). Even though the predicted frequency counts between conditions were not significantly different the trends in the data reveal some interesting information regarding innocent suspect bias. Our study found similar increases in filler identifications to Fitzgerald, et al (2014) however there was a decrease in target absent rejections in the high similarity lineups. Fitzgerald, et al's (2013) meta-analysis noted that changes in lineup similarity had no effect on lineup rejections. The higher number of correct lineup rejections in the low similarity condition suggests that our results could be explained by the lack of innocent suspect bias in our lineup designs. That is, not having a nominated innocent suspect matched to the description of the target. Rather, our design could be seen as, by chance, having 6 innocent suspects equally matched to the description of the target, thus removing any innocent suspect bias. The higher number of correct rejections in the low similarity condition is in line with the assumptions of the MAX decision rule in which the face with the highest familiarity above the critical threshold for choosing will be the one selected (Wixted, et al, 2018). If none of the faces reach the minimum familiarity required to make a selection then the lineup is rejected. Given that no face in the target absent low similarity lineups was matched to the target, then it is logical that lower numbers of participants would perceive a

familiarity level which exceeded their critical threshold for choosing. Although not an ecologically valid methodology, as the innocent suspect will always be selected due to their similarity to the perpetrator, this methodology provides the experimenter with an opportunity to remove this bias if required in order to test other elements of the eyewitness identification process. In this study removing this bias was desirable in order to ensure that any effects of exposure duration and similarity on perceived similarity were as free from confounds as possible. Given the results detailed above, this methodology can be considered a success.

4.4 Limitations

The primary limitation of this study involved the materials available from the Adelaide University Psychology Department face database. Each video in the database had a maximum length of 11 to 12 seconds, limiting the degree of manipulation available between the short and long exposure conditions. Although previous studies found significant differences in recognition accuracy using small exposure durations (Semmler & Brewer, 2006), the results of this study indicate that a longer video may have been necessary to increase the effect size and generate a significant difference. The limited sample of faces available for selection from the database was also a limitation. The concentration of faces in the 18 to 30 age group inevitably led to lineups composed of faces from that age range. Greater differences in similarity for the low similarity lineups in particular may have been possible if a wider range of ages was available for selection.

Some limitations are also inevitable when running an online experiment through Amazon Mechanical Turk. It is not possible to control conditions such as (1) the quality and size of the images viewed by participants due to variations in monitor designs and quality, (2) the distances from which the participants viewed the images, (3) the environment within which the participant undertook the experiment and the related distractions or distortions this may have created. However, the advantage of being able to randomly sample a larger and

broader population than would otherwise be possible offsets these negatives.

Although purposefully designed into this experiment, the lack of a nominated innocent suspect in the target absent lineups could be considered as not ecologically valid and a distortion of the basic premise of the eyewitness identification procedure. However, the effects of bias toward innocent suspects in simultaneous lineups has been studied many times and is well understood. What has not been well studied is how eyewitnesses perform when that bias is not present. The author would argue that there is much yet to be discovered about understanding and measuring the cognitive processes involved in the basic process of target identification without introducing possible confounding effects such as the match to description/ match to suspect dilemma. This study has hopefully added to the knowledge in this area.

4. 5 Future Research

The results of this study add to the evidence that the UVSDT model is a useful tool for predicting eyewitness identification accuracy from simultaneous lineup. The observed data fit the model well and it's predicted values produced ROC curves which closely followed those of the observed data. As such future researchers can have increased confidence in using the UVSDT model to predict and compare discriminability rates between differing experimental conditions.

The observed lack of any relationship between discriminability and perceived lineup similarity suggests that future research into lineup similarity should focus on the nature of the relationship between similarity and familiarity to the target rather than on the similarity between faces in the lineup. The results indicate that how people perceive the similarity between lineup faces is not influenced in any way by how similar those faces are to the memory of the target. From this an interesting question is whether or not a priori categorical judgments of lineup similarity as either high or low are really very useful. Perhaps answering

the question of how many faces of a particular familiarity are required to construct a fair lineup may deliver more useful results.

Similarly, further research is warranted in exploring if the observed separation in this study between the judgements of similarity between faces in a lineup and the judgements of face familiarity and response criterion levels can be replicated under different conditions. If so, it may be that similarity is a factor not directly related to the cognitive processes of eyewitness identification. Such a finding would have important ramifications for theories about simultaneous lineup design and implications in the on-going debate over whether simultaneous, sequential or show up identification procedures protect innocent suspects the most effectively.

4.6 Conclusion

As exposure duration was not effective in producing a difference in performance, no meaningful conclusion about its effect on perceived similarity between lineup faces and a target can be drawn. Except to acknowledge that the stability of the perceived similarity measure across conditions indicates that no effect is likely even with a stronger manipulation. However, if, as suggested by the results, perceived lineup similarity between the faces in a lineup and the target has no relationship with a priori judgements of lineup similarity and lineup discriminability, then the premise of arguments around whether high similarity lineups are fairer than low similarity lineups become questionable. These arguments assume that a priori categorical judgements of lineup similarity as a whole can predict later decision making and accuracy in the eyewitness identification processes. What this study suggests is that how similar each face in a lineup is perceived to be by participants as a whole, even when the target is present, has no relationship to how accurate those participants will be when carrying out the identification task – at least within the bounds of similarity that were explored here. Rather, in support of the independent observations assumption within the

UVSD (Wixted, et al, 2018), what appears to be critical is the participants' judgement of which faces in a lineup individually exceed the level of familiarity required for a selection to be made, and which of those faces has the maximum familiarity. The fact that more faces exceed this level in a lineup containing faces equally high in similarity to the target seems self-evident and, as inferred by our results, independent of any overall a priori categorical judgement of whether the lineup as a whole is either low or high in similarity. The conclusion to be drawn is that the judgment of perceived similarity between faces in a lineup and to a target, as measured in this study, are not effected by the processes involved in eyewitness identification. Consequently, based on the results of this study, statements which conclude that high similarity simultaneous lineups, constructed using categorical a priori similarity ratings, produce higher rates of discriminability than low similarity lineups should be questioned. Further, the data from this study does not support Wixted and Mickes (2014) Diagnostic Feature Detection Hypothesis, as relative similarity judgements in lineups based on individual features appear to be unrelated to those processes used to match lineup faces to the memory of the target. The inference that these two processes appear to be separate generates questions around the basic premise of using simultaneous lineups, constructed on the basis of a priori face similarity, as a procedure designed to protect innocent suspects. It may be that the fate of an innocent suspect relies more on the luck of how familiar they are to the eyewitnesses' memory of the perpetrator than to any specific simultaneous lineup design based on similarity.

References

Brewer, N., & Palmer, M. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*(1), 77-96. doi: 10.1348/135532509X414765

Bornstein, B., Deffenbacher, K., Penrod, S., & Mcgorty, E. (2012). Effects of exposure time and cognitive operations on facial identification accuracy: A meta-analysis of two variables associated with initial memory strength. *Psychology, Crime & Law, 18*(5), 473-490. doi: 10.1080/1068316X.2010.508458

Clark, S., Erickson, E., & Breneman, M. (2011). Probative Value of Absolute and Relative Judgments in Eyewitness Identification. *Law and Human Behavior, 35*(5), 364-380. doi: 10.1007/s10979-010-9245-1

Clark, S. (2012). Costs and Benefits of Eyewitness Identification Reform: Psychological Science and Public Policy. *Perspectives on Psychological Science, 7*(3), 238-259. doi: 10.1177/1745691612439584

Engel, W., R. (2018, August 27). Re:XNomial – Exact Test For Multinomial. Retrieved from https://cran.r-project.org/web/packages/XNomial/vignettes/XNomial.html

Farrell, S., & Lewandowsky, S. (2010). Computational Models as Aids to Better Reasoning in Psychology. *Current Directions in Psychological Science, 19*(5), 329-335. doi: 10.1177/0963721410386677

Flowe, H., & Bessemer, A. (2011). The effects of target discriminability and criterion placement on accuracy rates in sequential and simultaneous target-present lineups. *Psychology, Crime & Law, 17*(7), 587-610. doi:10.1080/10683160903397540

Fitzgerald, R., Price, H., Oriet, C., Charman, S., & Lamb, Michael E. (2013). The Effect of Suspect-Filler Similarity on Eyewitness Identification Decisions: A Meta-Analysis. *Psychology, Public Policy, and Law, 19*(2), 151-164. doi: 10.1037/a0030618

Fitzgerald, R., Oriet, C., Price, H., & Kovera, Margaret Bull. (2015). Suspect Filler Similarity in Eyewitness Lineups: A Literature Review and a Novel Methodology. *Law and Human Behavior, 39*(1), 62-74. doi:10.1037/lhb0000095

Horry, R., Brewer, N., & Gauthier, Isabel. (2016). How Target–Lure Similarity Shapes Confidence Judgments in Multiple-Alternative Decision Tasks. *Journal of Experimental Psychology: General, 145*(12), 1615-1634. doi:10.1037/xge0000227

Innocence Project. (2018, April 10). Re:Exonerate the Innocent. Retrieved from https://www.innocenceproject.org/exonerate/

Malpass, Roy S, Tredoux, Colin G & McQuiston-Surrett, Dawn. (2007). Lineup construction and lineup fairness. Lindsay, Rod C. L [Ed], Ross, David F [Ed], Read, J. Don [Ed], Toglia, Michael P [Ed]. *The handbook of eyewitness psychology, Vol II: Memory for people.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, US; pp. 155-178. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc5&NEWS=N&AN=2007-04817-007.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal Of Psychology, 94*, 339-354. doi: 10.1348/000712603767876262

National Research Council. 2014. Chapter 2 Eyewitness Identification Procedures in *Identifying the Culprit: Assessing Eyewitness Identification*. (pp. 21-30). Washington, DC: The National Academies Press. https://doi.org/10.17226/18891

National Research Council. 2014. Chapter 5 Applied Eyewitness Identification Research in *Identifying the Culprit: Assessing Eyewitness Identification*. (pp. 71-101). Washington, DC: The National Academies Press. https://doi.org/10.17226/18891

Palmer, Matthew A., Brewer, Neil, Weber, Nathan, & Nagesh, Ambika. (2013). The Confidence-Accuracy Relationship for Eyewitness Identification Decisions: Effects of Exposure Duration, Retention Interval, and Divided Attention. *Journal of Experimental Psychology: Applied,19*(1), 55-71. doi: 10.1037/a0031602

Reynolds, J., & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology, 6*(4), 279-292. doi: 10.1002/acp.2350060402

Semmler, C., & Brewer, N. (2006). Postidentification feedback effects on face recognition confidence: Evidence for metacognitive influences. *Applied Cognitive Psychology, 20*(7), 895-916. doi: 10.1002/acp.1238

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. (2018). The Role of Estimator Variables in Eyewitness Identification. *Journal of Experimental Psychology: Applied 2018*, doi: 10.1037/xap0000157

Tredoux, C., & Ackerman, Phillip L. (2002). A Direct Measure of Facial Similarity and Its Relation to Human Similarity Perceptions. *Journal of Experimental Psychology: Applied, 8*(3), 180-193. doi: 10.1037/1076-898X.8.3.180

Tunnicliff, J., & Clark, L. (2000). Selecting Foils for Identification Lineups: Matching Suspects or Descriptions? *Law and Human Behavior, 24*(2), 231-258. doi: 10.1023/A:1005463020252

Tversky, A., & Estes, William K. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352. doi: 10.1037/0033-295X.84.4.327

Wells, G., & Greenwald, G. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*(12), 1546-1557. doi: 10.1037/0022-3514.36.12.1546

Wells, G., Small, M., Penrod, S., Malpass, R., Fulero, S., Brimacombe, C., & Wiener, Richard L. (1998). Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads. *Law and Human Behavior*, *22*(6), 603-647. doi: 10.1023/A:1025750605807

Wells, G., Yang, Y., Smalarz, L., & Kovera, Margaret Bull. (2015). Eyewitness Identification: Bayesian Information Gain, Base-Rate Effect Equivalency Curves, and Reasonable Suspicion. Law and Human Behavior, 39(2), 99-122. doi: 10.1037/lhb0000125

Wixted, J., Mickes, L., & Anderson, R. (2014). A Signal-Detection-Based Diagnostic-Feature-Detection Model of Eyewitness Identification. *Psychological Review, 121*(2), 262-276. doi:10.1037/a0035940

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

Wixted, J., Mickes, L., Dunn, J., Clark, S., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences,113*(2), 304-9. doi: 10.1073/pnas.1516814112

Wixted, Vul, Mickes, & Wilson. (2018). Models of lineup memory. *Cognitive Psychology, 105*, 81-114. doi: 10.1016/j.cogpsych.2018.06.001

Appendix A

Ethical Approval

THE UNIVERSITY OF ADELAIDE AUSTRALIA

School of Psychology
University of Adelaide
North Terrace, Adelaide SA 5005
Ph. 61 8 8313 5693
Fax 61 8 8313 3770

School of Psychology: Human Research Ethics Subcommittee
Approval Sheet

Dear CAROLYN

The members of the subcommittee have considered your application:

Code Number: 18/40

Title:

THE EFFECT OF EXPOSURE DURATION AND DISCRIMINABILITY IN HIGH & LOW SIMILARITY LINE-UPS

With [Student name, if applicable] PETER GRIEVE

I am writing to confirm that approval has been granted for this project to proceed. Approval is granted to 12 months from the date specified below.

Yours sincerely

[redacted]

Deputy Convenor, Human Research Ethics Subcommittee

Name: PAUL DELFABBRO

Date: 23/4/18

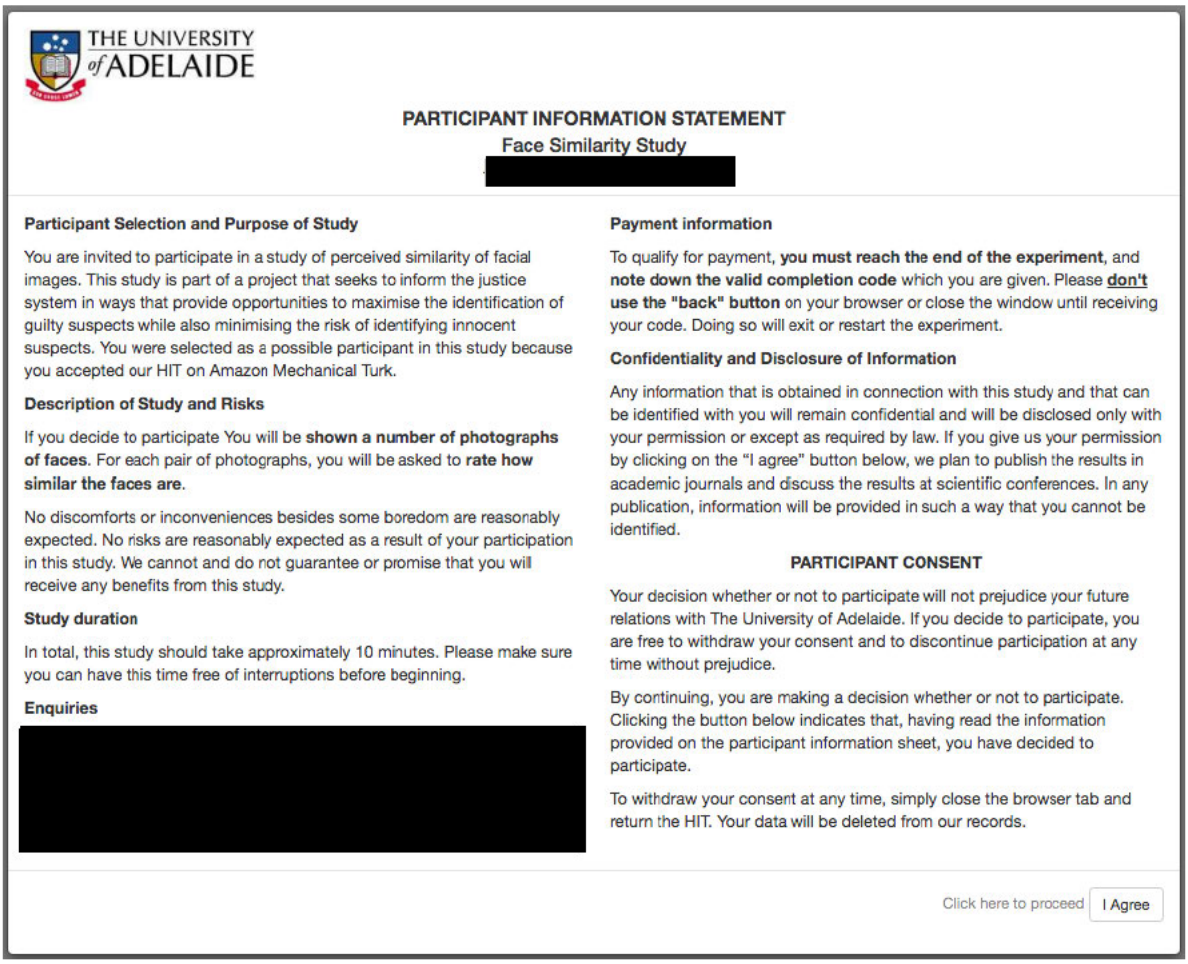
Phone Number: 8313 4736

Email: paul.delfabbro@adelaide.edu.au

NB · INFO SHEET - NOT HREC APP

Appendix B

Experimental Interface (a priori Similarity Ratings)



THE UNIVERSITY of ADELAIDE

PARTICIPANT INFORMATION STATEMENT

Face Similarity Study

[redacted]

Participant Selection and Purpose of Study

You are invited to participate in a study of perceived similarity of facial images. This study is part of a project that seeks to inform the justice system in ways that provide opportunities to maximise the identification of guilty suspects while also minimising the risk of identifying innocent suspects. You were selected as a possible participant in this study because you accepted our HIT on Amazon Mechanical Turk.

Description of Study and Risks

If you decide to participate You will be **shown a number of photographs of faces**. For each pair of photographs, you will be asked to **rate how similar the faces are**.

No discomforts or inconveniences besides some boredom are reasonably expected. No risks are reasonably expected as a result of your participation in this study. We cannot and do not guarantee or promise that you will receive any benefits from this study.

Study duration

In total, this study should take approximately 10 minutes. Please make sure you can have this time free of interruptions before beginning.

Enquiries

[redacted]

Payment information

To qualify for payment, **you must reach the end of the experiment**, and **note down the valid completion code** which you are given. Please **don't use the "back" button** on your browser or close the window until receiving your code. Doing so will exit or restart the experiment.

Confidentiality and Disclosure of Information

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or except as required by law. If you give us your permission by clicking on the "I agree" button below, we plan to publish the results in academic journals and discuss the results at scientific conferences. In any publication, information will be provided in such a way that you cannot be identified.

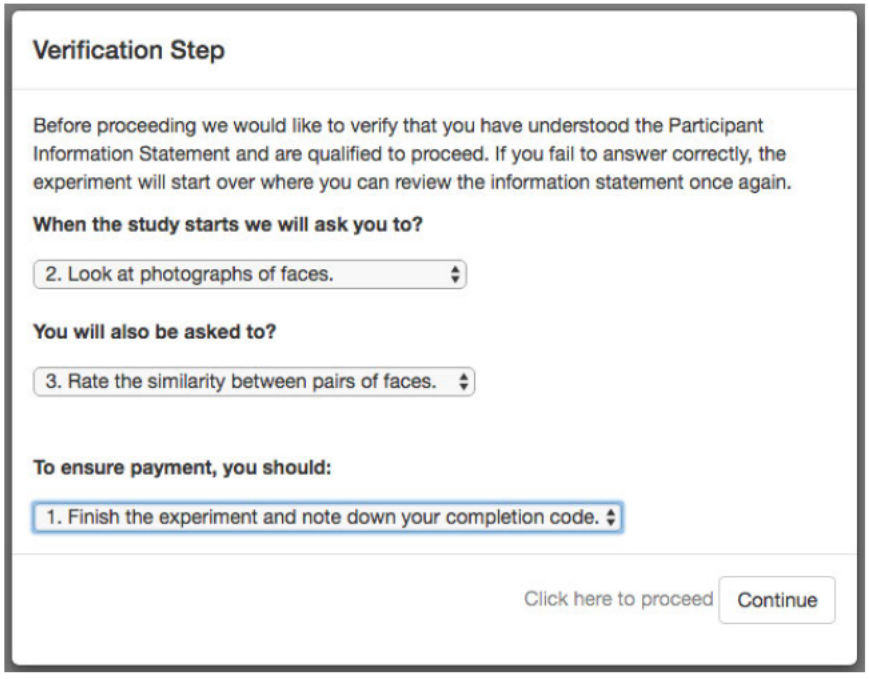
PARTICIPANT CONSENT

Your decision whether or not to participate will not prejudice your future relations with The University of Adelaide. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

By continuing, you are making a decision whether or not to participate. Clicking the button below indicates that, having read the information provided on the participant information sheet, you have decided to participate.

To withdraw your consent at any time, simply close the browser tab and return the HIT. Your data will be deleted from our records.

Click here to proceed | I Agree



Verification Step

Before proceeding we would like to verify that you have understood the Participant Information Statement and are qualified to proceed. If you fail to answer correctly, the experiment will start over where you can review the information statement once again.

When the study starts we will ask you to?

2. Look at photographs of faces.

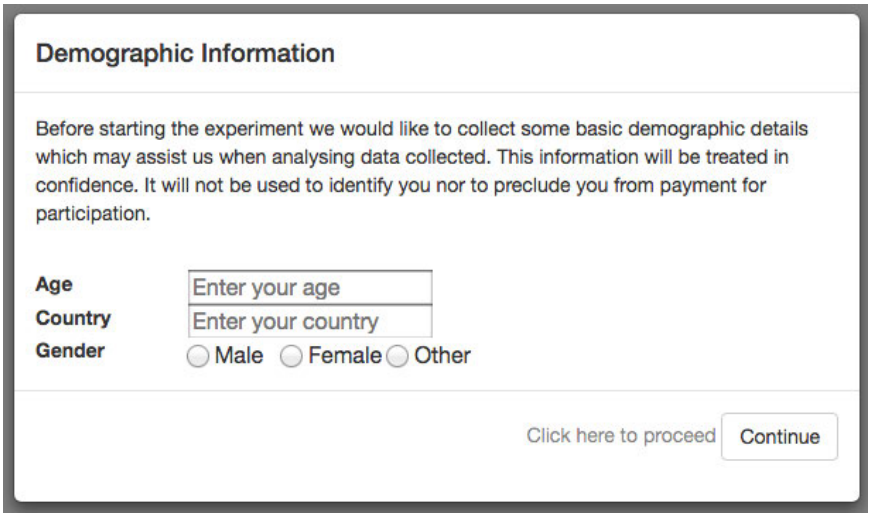
You will also be asked to?

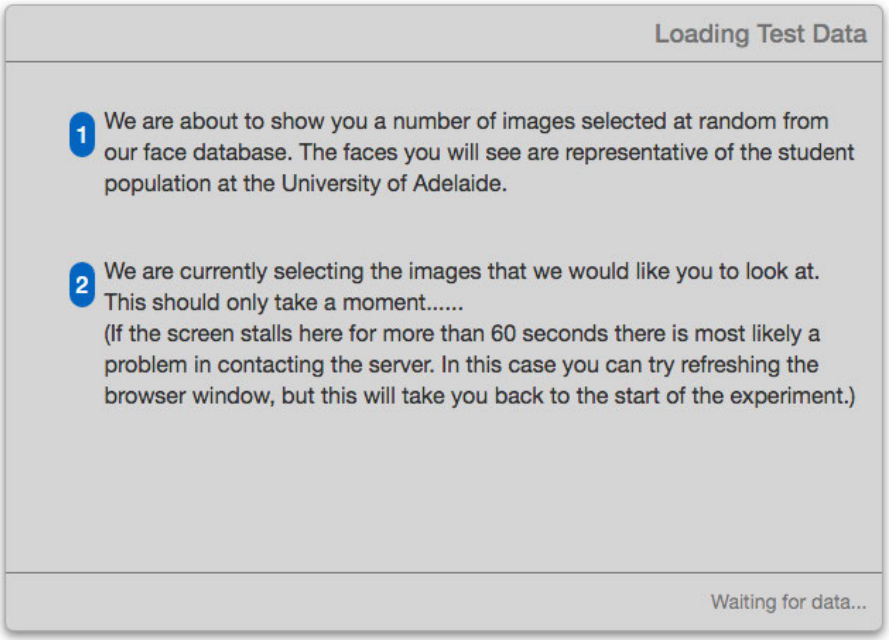
3. Rate the similarity between pairs of faces.

To ensure payment, you should:

1. Finish the experiment and note down your completion code.

Click here to proceed | Continue


Demographic Information
Before starting the experiment we would like to collect some basic demographic details which may assist us when analysing data collected. This information will be treated in confidence. It will not be used to identify you nor to preclude you from payment for participation.
Age
Enter your age
Country
Enter your country
Gender
Male
Female
Other
Click here to proceed
Continue


Loading Test Data
1
We are about to show you a number of images selected at random from our face database. The faces you will see are representative of the student population at the University of Adelaide.
2
We are currently selecting the images that we would like you to look at. This should only take a moment.....
(If the screen stalls here for more than 60 seconds there is most likely a problem in contacting the server. In this case you can try refreshing the browser window, but this will take you back to the start of the experiment.)
Waiting for data...

Ready to begin...

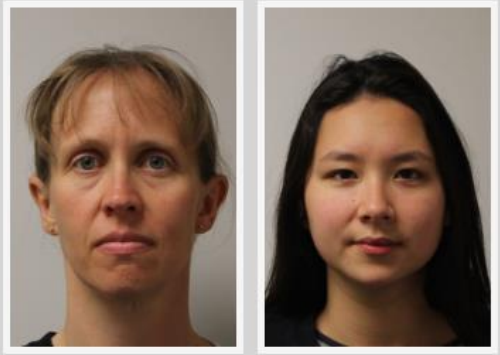
1 We are about to show you a number of images selected at random from our face database. The faces you will see are representative of the student population at the University of Adelaide.

2 We are currently selecting the images that we would like you to look at. This should only take a moment.....
(If the screen stalls here for more than 60 seconds there is most likely a problem in contacting the server. In this case you can try refreshing the browser window, but this will take you back to the start of the experiment.)

3 The images have been selected. Press the button below when you are ready to begin viewing.

Click here to continue...

Question 1 of 45 Test Phase



How similar are these two faces?

(Type the number or click the button corresponding to your choice)

Very High Similarity ○1 ○2 ○3 ○4 ○5 ○6 ○7 ○8 ○9 *Very Low Similarity*

Please select your answer...

Congratulations, you have finished.

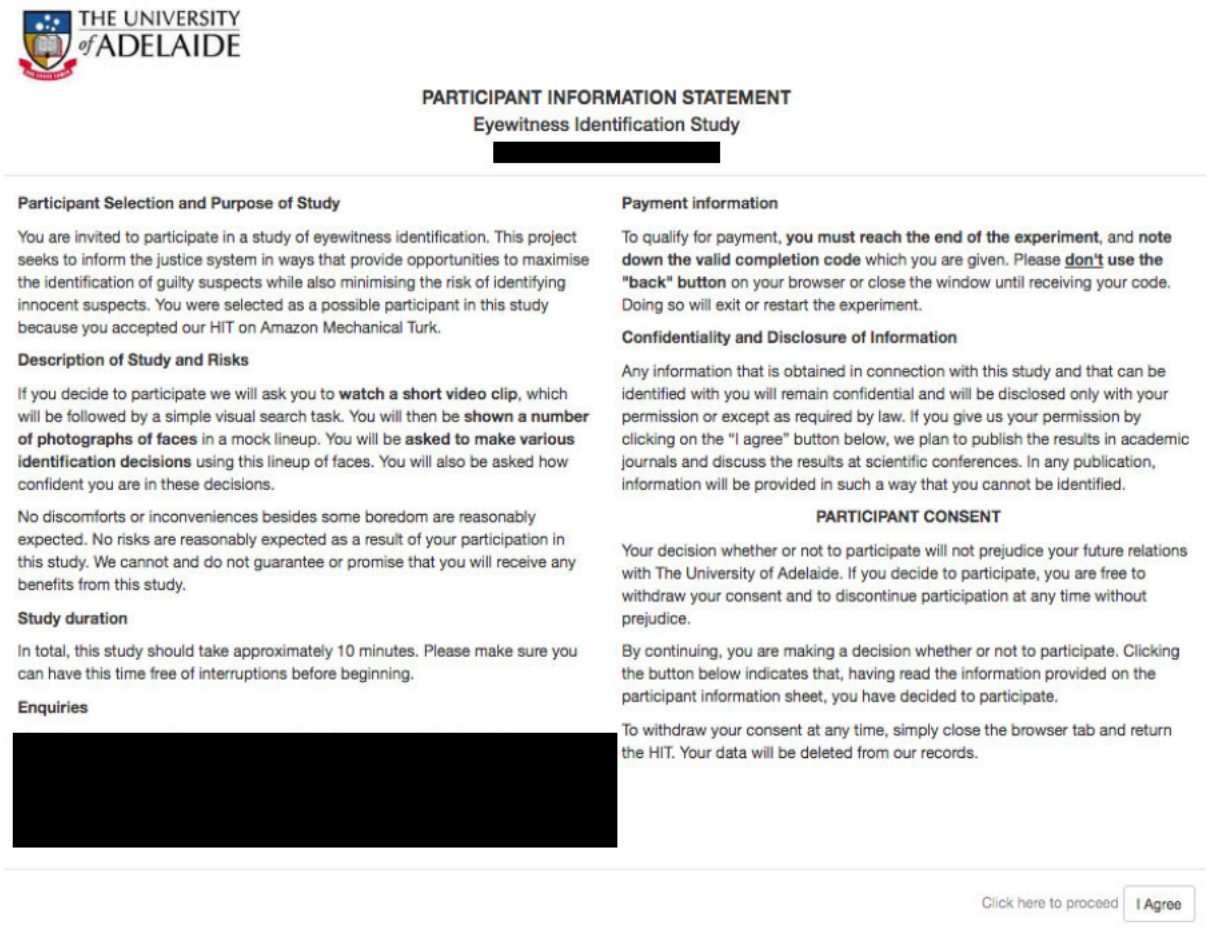
1 You have successfully completed the experiment. Thank you for your participation.

2 If you are participating via Mechanical Turk, please enter the code: **fs_1_0_5707967177424896** to complete the HIT.

Close the browser tab when done.

Appendix C

Experimental Interface (Main Study)



THE UNIVERSITY of ADELAIDE

PARTICIPANT INFORMATION STATEMENT

Eyewitness Identification Study

[redacted]

Participant Selection and Purpose of Study

You are invited to participate in a study of eyewitness identification. This project seeks to inform the justice system in ways that provide opportunities to maximise the identification of guilty suspects while also minimising the risk of identifying innocent suspects. You were selected as a possible participant in this study because you accepted our HIT on Amazon Mechanical Turk.

Description of Study and Risks

If you decide to participate we will ask you to **watch a short video clip**, which will be followed by a simple visual search task. You will then be **shown a number of photographs of faces** in a mock lineup. You will be **asked to make various identification decisions** using this lineup of faces. You will also be asked how confident you are in these decisions.

No discomforts or inconveniences besides some boredom are reasonably expected. No risks are reasonably expected as a result of your participation in this study. We cannot and do not guarantee or promise that you will receive any benefits from this study.

Study duration

In total, this study should take approximately 10 minutes. Please make sure you can have this time free of interruptions before beginning.

Enquiries

[redacted]

Payment information

To qualify for payment, **you must reach the end of the experiment**, and **note down the valid completion code** which you are given. Please **don't use the "back" button** on your browser or close the window until receiving your code. Doing so will exit or restart the experiment.

Confidentiality and Disclosure of Information

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or except as required by law. If you give us your permission by clicking on the "I agree" button below, we plan to publish the results in academic journals and discuss the results at scientific conferences. In any publication, information will be provided in such a way that you cannot be identified.

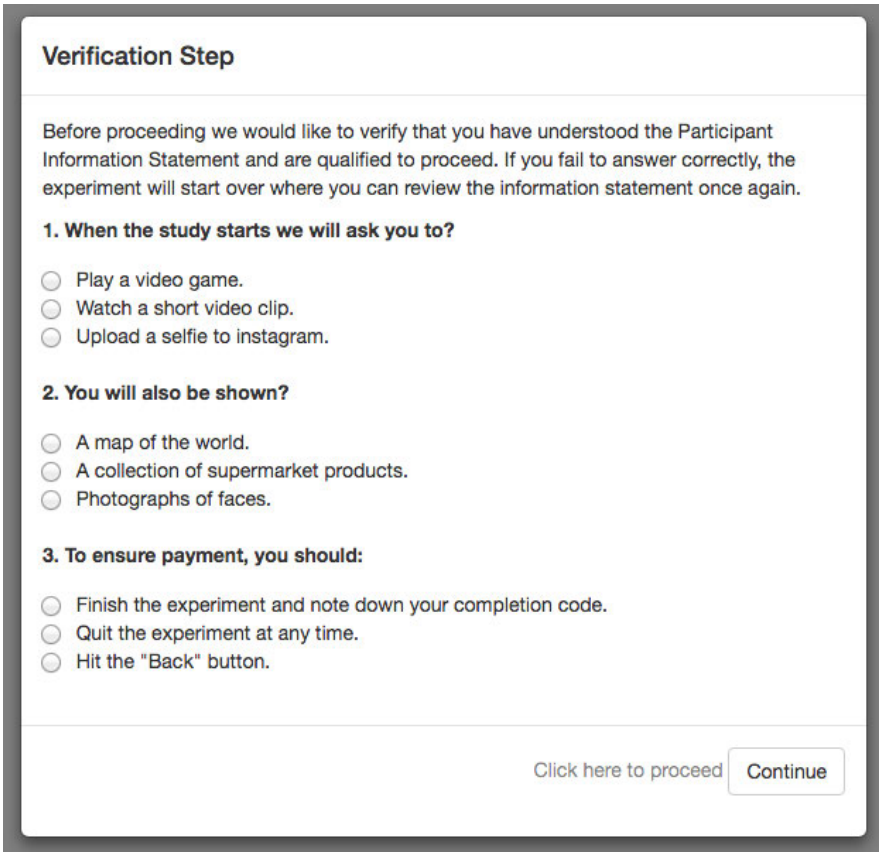
PARTICIPANT CONSENT

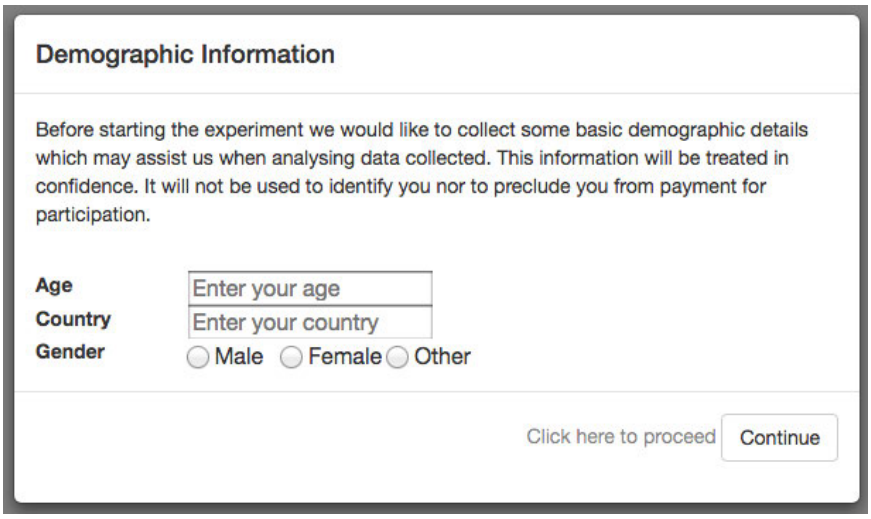
Your decision whether or not to participate will not prejudice your future relations with The University of Adelaide. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

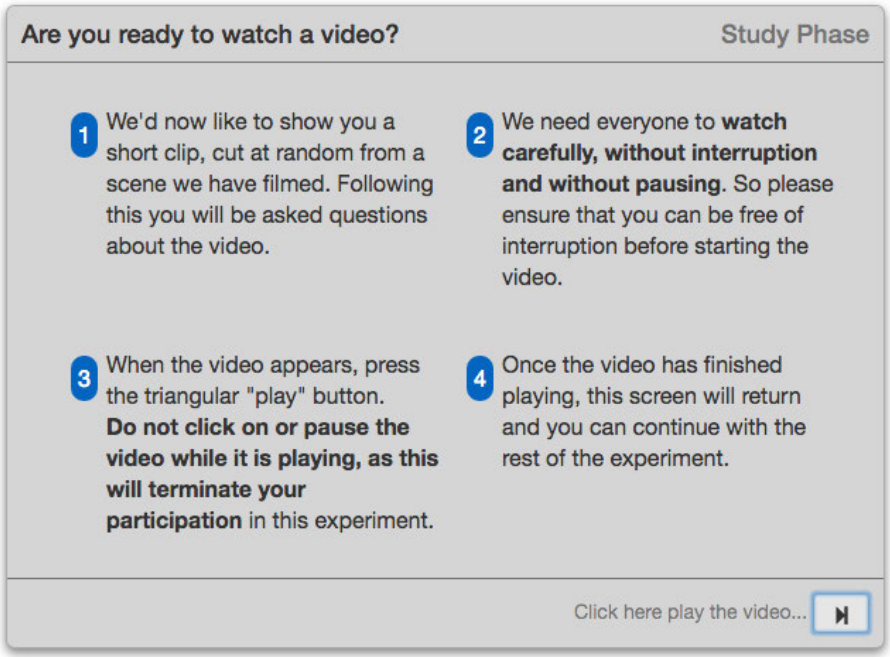
By continuing, you are making a decision whether or not to participate. Clicking the button below indicates that, having read the information provided on the participant information sheet, you have decided to participate.

To withdraw your consent at any time, simply close the browser tab and return the HIT. Your data will be deleted from our records.

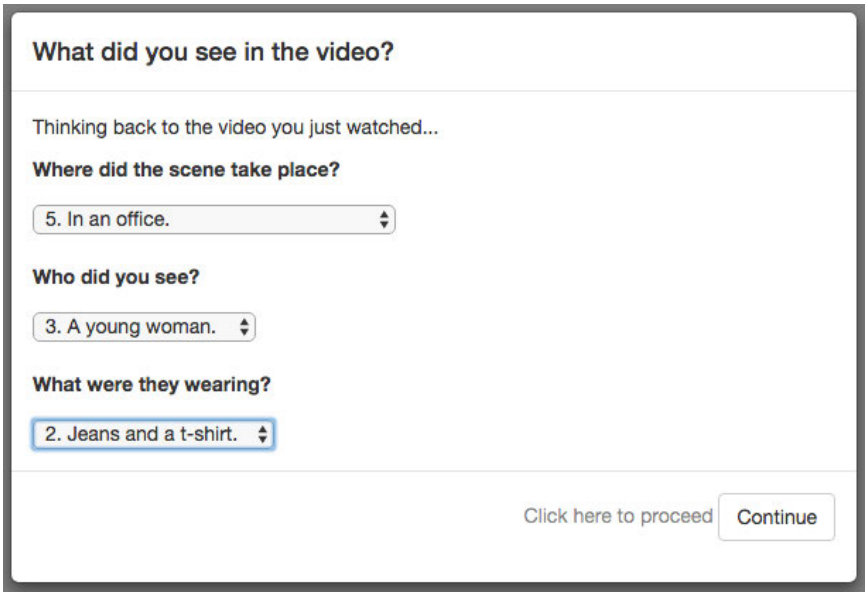
Click here to proceed | I Agree

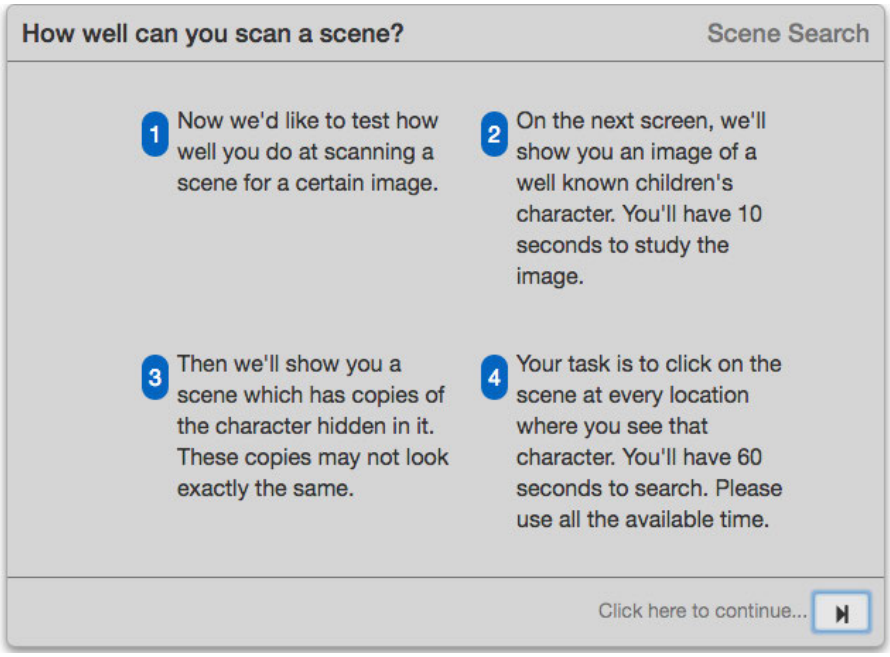

Verification Step
Before proceeding we would like to verify that you have understood the Participant Information Statement and are qualified to proceed. If you fail to answer correctly, the experiment will start over where you can review the information statement once again.
1. When the study starts we will ask you to?
Play a video game.
Watch a short video clip.
Upload a selfie to instagram.
2. You will also be shown?
A map of the world.
A collection of supermarket products.
Photographs of faces.
3. To ensure payment, you should:
Finish the experiment and note down your completion code.
Quit the experiment at any time.
Hit the "Back" button.
Click here to proceed
Continue

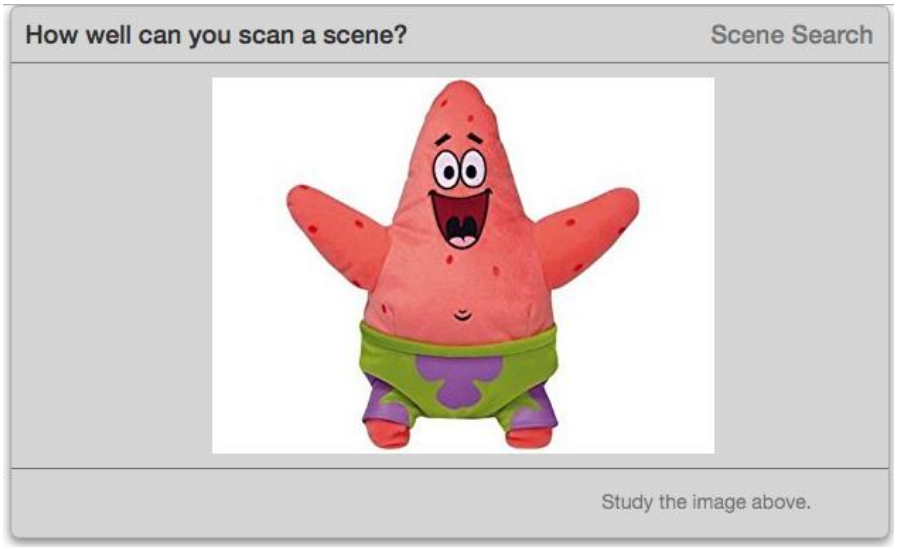

Demographic Information
Before starting the experiment we would like to collect some basic demographic details which may assist us when analysing data collected. This information will be treated in confidence. It will not be used to identify you nor to preclude you from payment for participation.
Age
Enter your age
Country
Enter your country
Gender
Male
Female
Other
Click here to proceed
Continue

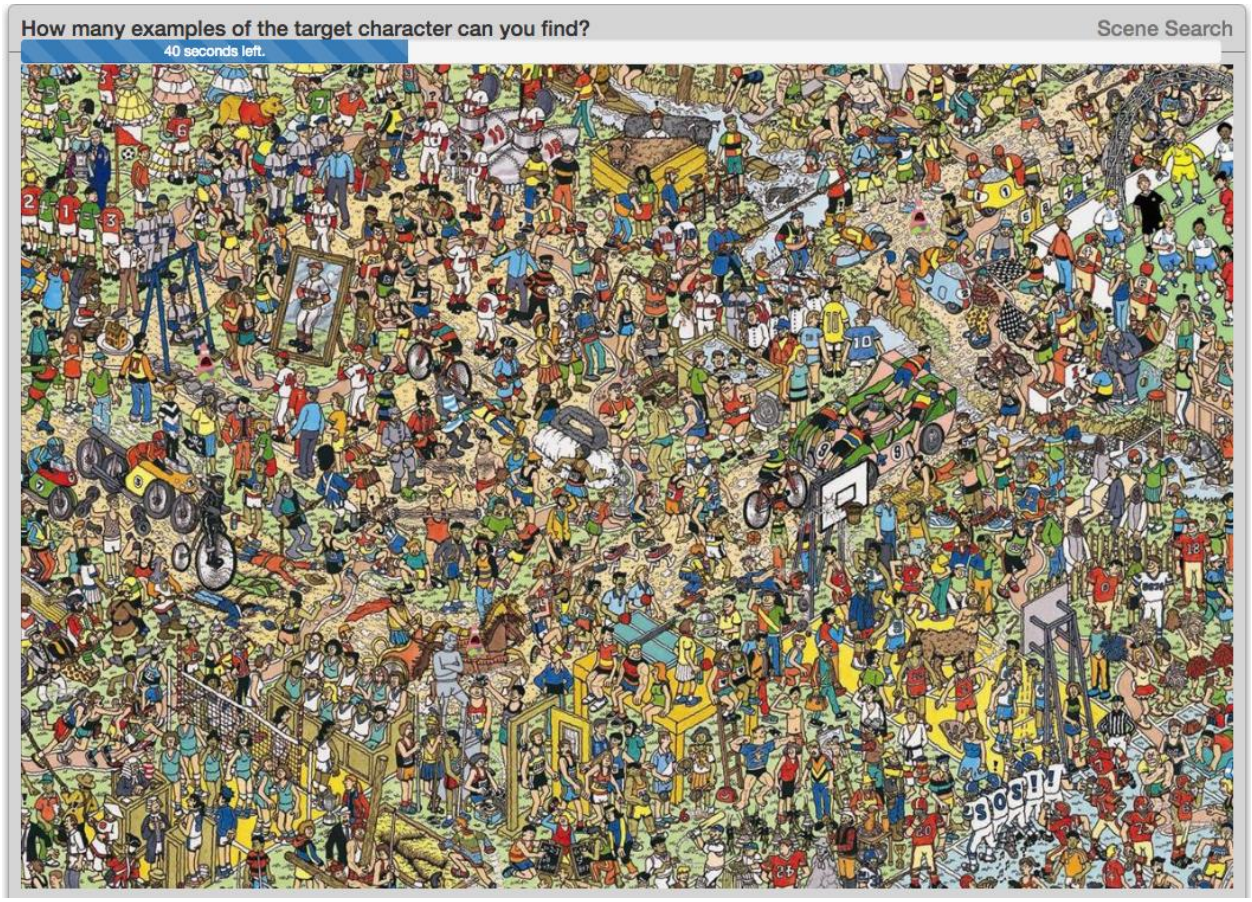

Are you ready to watch a video?
Study Phase
1 We'd now like to show you a short clip, cut at random from a scene we have filmed. Following this you will be asked questions about the video.
2 We need everyone to watch carefully, without interruption and without pausing. So please ensure that you can be free of interruption before starting the video.
3 When the video appears, press the triangular "play" button. Do not click on or pause the video while it is playing, as this will terminate your participation in this experiment.
4 Once the video has finished playing, this screen will return and you can continue with the rest of the experiment.
Click here play the video...

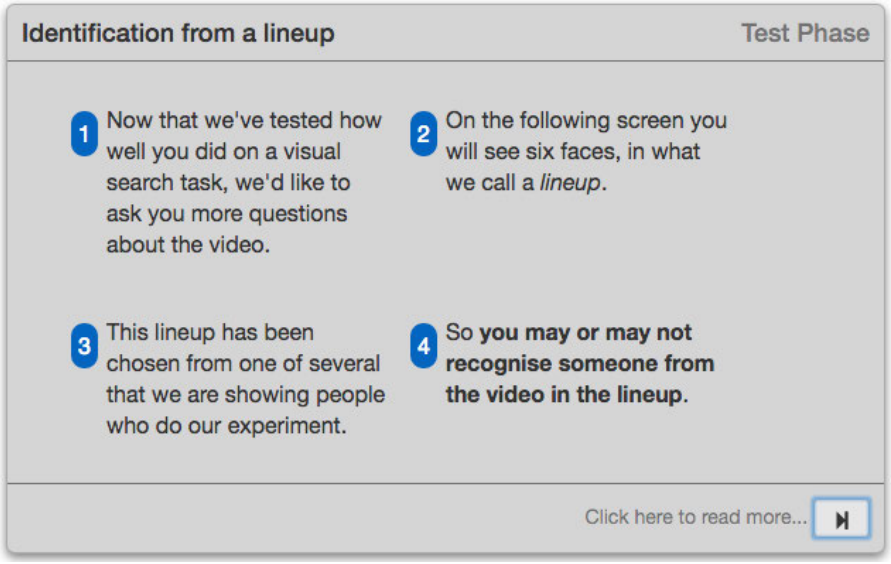

life
IMPACT

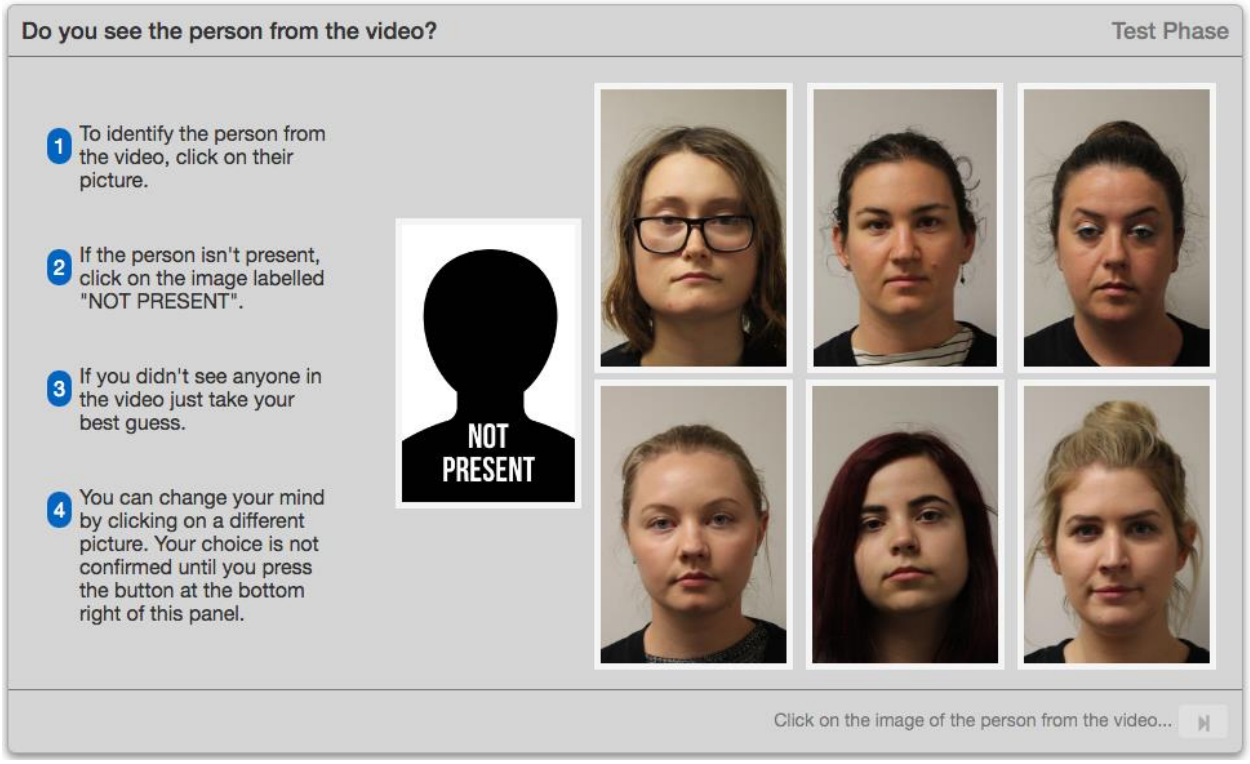

What did you see in the video?
Thinking back to the video you just watched...
Where did the scene take place?
5. In an office.
Who did you see?
3. A young woman.
What were they wearing?
2. Jeans and a t-shirt.
Click here to proceed
Continue

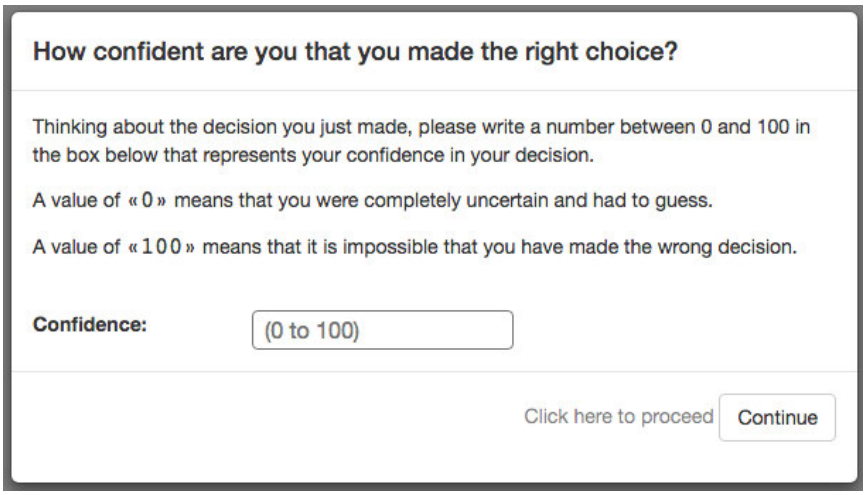

How well can you scan a scene?
Scene Search
1 Now we'd like to test how well you do at scanning a scene for a certain image.
2 On the next screen, we'll show you an image of a well known children's character. You'll have 10 seconds to study the image.
3 Then we'll show you a scene which has copies of the character hidden in it. These copies may not look exactly the same.
4 Your task is to click on the scene at every location where you see that character. You'll have 60 seconds to search. Please use all the available time.
Click here to continue...

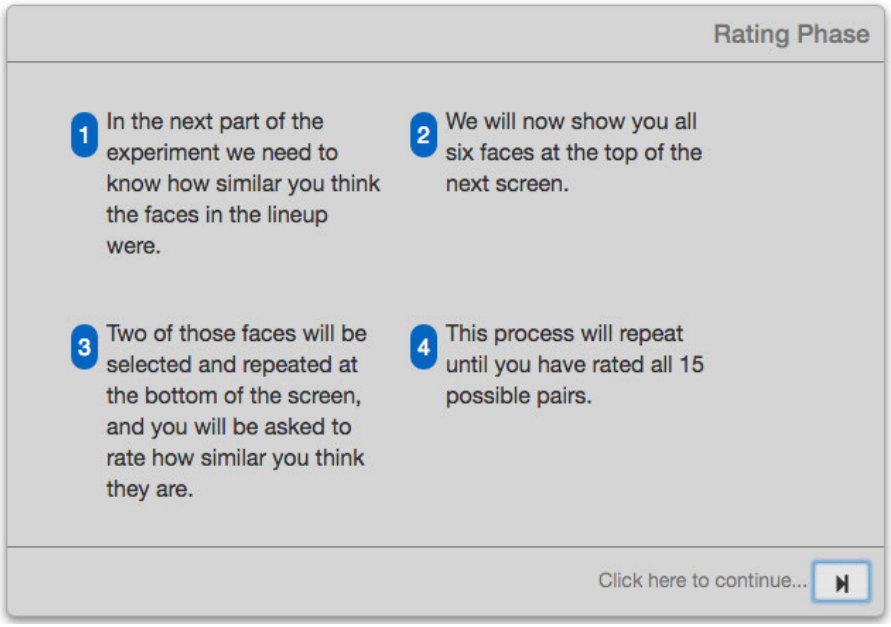

How well can you scan a scene?
Scene Search
Study the image above.

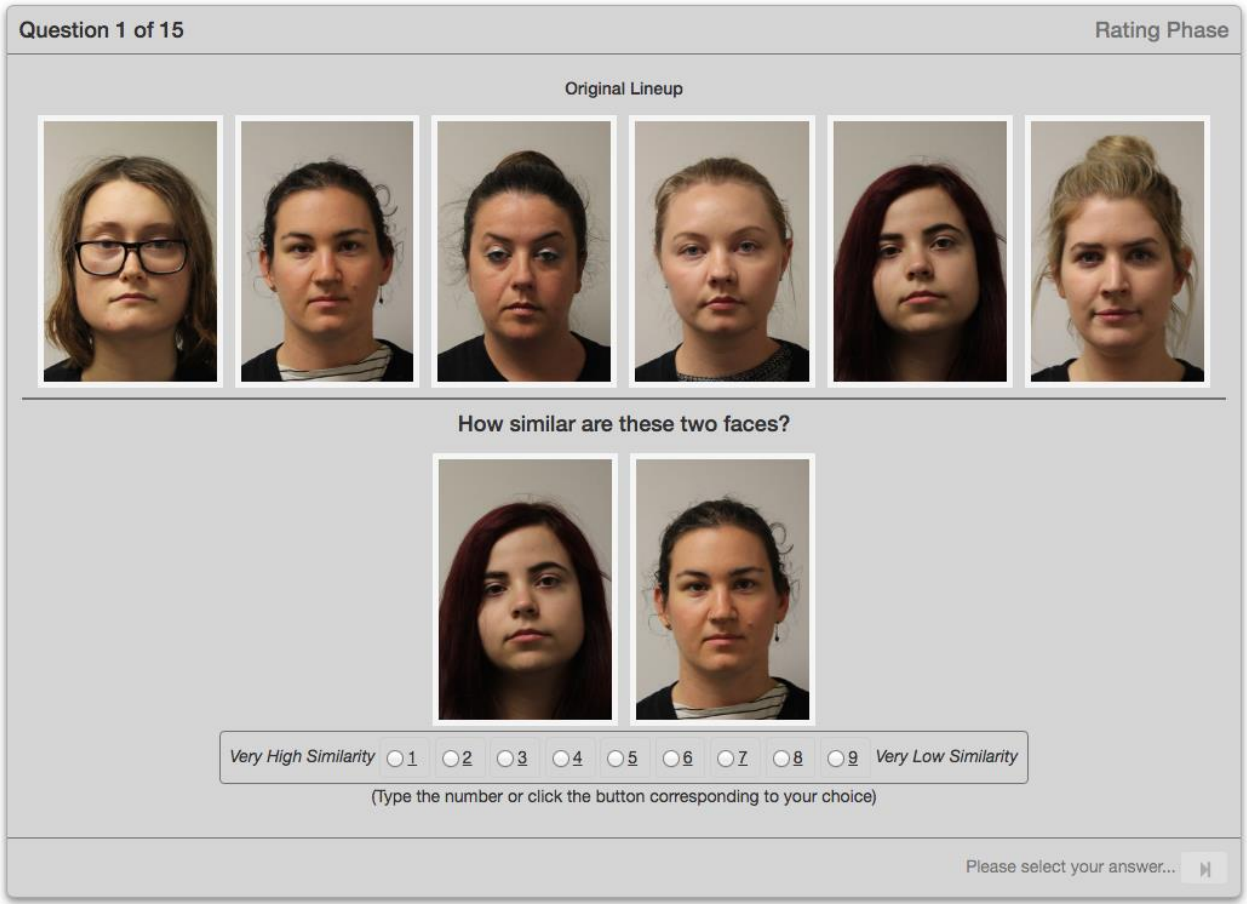

How many examples of the target character can you find?
Scene Search
40 seconds left.

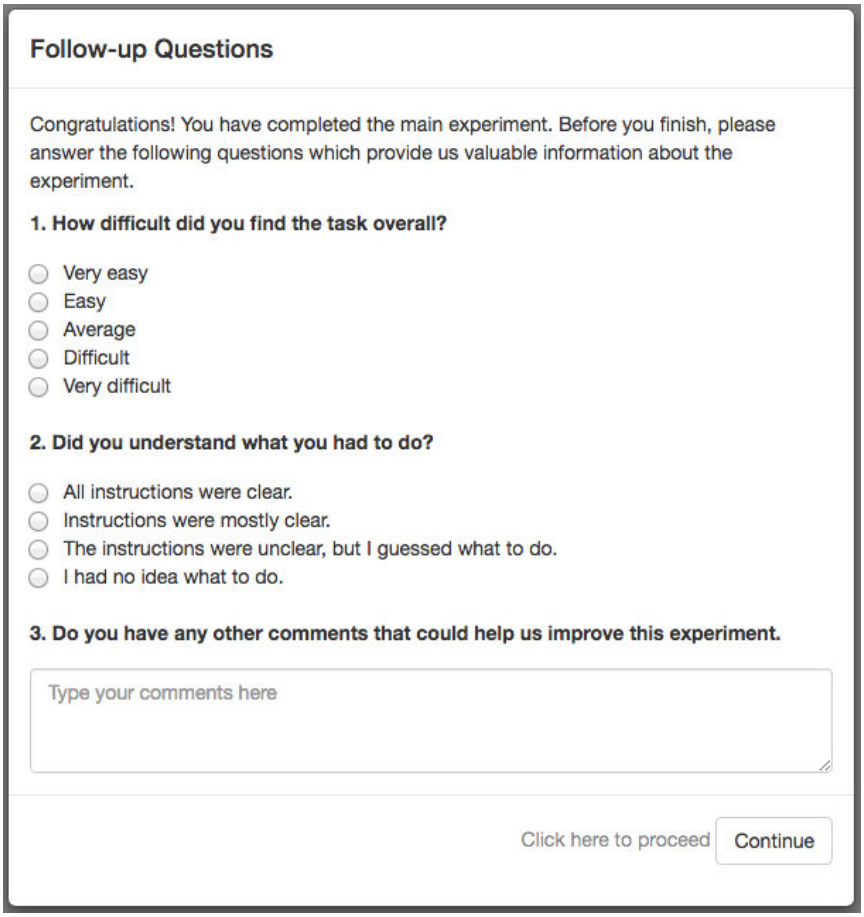

Identification from a lineup
Test Phase
1 Now that we've tested how well you did on a visual search task, we'd like to ask you more questions about the video.
2 On the following screen you will see six faces, in what we call a *lineup*.
3 This lineup has been chosen from one of several that we are showing people who do our experiment.
4 So **you may or may not recognise someone from the video in the lineup**.
Click here to read more...

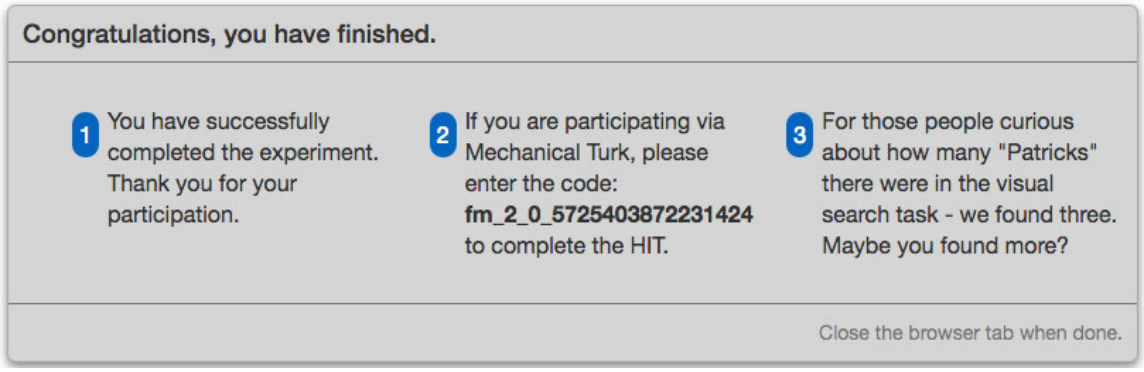

Do you see the person from the video?
Test Phase
1 To identify the person from the video, click on their picture.
2 If the person isn't present, click on the image labelled "NOT PRESENT".
3 If you didn't see anyone in the video just take your best guess.
4 You can change your mind by clicking on a different picture. Your choice is not confirmed until you press the button at the bottom right of this panel.
NOT PRESENT
Click on the image of the person from the video...


How confident are you that you made the right choice?
Thinking about the decision you just made, please write a number between 0 and 100 in the box below that represents your confidence in your decision.
A value of « 0 » means that you were completely uncertain and had to guess.
A value of « 100 » means that it is impossible that you have made the wrong decision.
Confidence:
(0 to 100)
Click here to proceed
Continue


Rating Phase
1 In the next part of the experiment we need to know how similar you think the faces in the lineup were.
2 We will now show you all six faces at the top of the next screen.
3 Two of those faces will be selected and repeated at the bottom of the screen, and you will be asked to rate how similar you think they are.
4 This process will repeat until you have rated all 15 possible pairs.
Click here to continue...


Question 1 of 15
Rating Phase
Original Lineup
How similar are these two faces?
Very High Similarity 1 2 3 4 5 6 7 8 9 Very Low Similarity
(Type the number or click the button corresponding to your choice)
Please select your answer...


Follow-up Questions
Congratulations! You have completed the main experiment. Before you finish, please answer the following questions which provide us valuable information about the experiment.
1. How difficult did you find the task overall?
Very easy
Easy
Average
Difficult
Very difficult
2. Did you understand what you had to do?
All instructions were clear.
Instructions were mostly clear.
The instructions were unclear, but I guessed what to do.
I had no idea what to do.
3. Do you have any other comments that could help us improve this experiment.
Type your comments here
Click here to proceed
Continue


Congratulations, you have finished.
1 You have successfully completed the experiment. Thank you for your participation.
2 If you are participating via Mechanical Turk, please enter the code: fm_2_0_5725403872231424 to complete the HIT.
3 For those people curious about how many "Patricks" there were in the visual search task - we found three. Maybe you found more?
Close the browser tab when done.

Appendix D

Cluster Analysis and Multidimensional Scaling

To enable unbiased target and filler selections for lineup construction cluster analysis and multidimensional scaling techniques using R were applied to a priori similarity ratings from a random sample (*N* = 600). The following R code and commentary details the process.

R code for identifying target and high similarity fillers:

```
library(cluster)
library(factoextra)
library("rgl", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")
install.packages("Hmisc")
library("Hmisc")
rm(list=ls())
Data_Dummy <- read.csv("~/Uni Files/RStudio/responseMatrix.csv", row.names = 1)
# Convert to data.frame
data_r <- as.data.frame(Data_Dummy)
# remove F113 as has nose ring, F22 & F58 glasses, F47 old
remove_face <- c("F113", "F22", "F58", "F47")
data_raw <- data_r[!(colnames(data_r) %in% remove_face)]
data_raw <- data_raw[which(rownames(data) %nin% remove_face), ]
set.seed(123)
km.res <- kmeans(data_raw, 4, nstart = 25)
print(km.res)
# Plot kmeans clusters
fviz_cluster(km.res,
        ellipse.type = "t", # Concentration ellipse
        repel = TRUE, # Avoid label overplotting (slow)
        ggtheme = theme_classic(),
        data = data_raw)
```

```
#extract cluster of interest
clusterK <- c(3) # from kmeans result - enter manually
clusterK_faces <- as.data.frame(km.res$cluster)
colnames(clusterK_faces) <- "cluster"
clusterK_faces <- subset(clusterK_faces, cluster %in% clusterK)
clusterK_faces <- row.names(clusterK_faces) # vector containing Face row names for cluster
print(clusterK_faces)
# extract kmeans cluster Faces from distance matrix
clusK_data <- data_raw[c(clusterK_faces),c(clusterK_faces)]
# Perform kmeans again to get smaller clusters within clusterK
set.seed(123)
km.res <- kmeans(clusK_data, 3, nstart = 25)
print(km.res)
# Plot 2nd kmeans clusters
fviz_cluster(km.res,
        ellipse.type = "t", # Concentration ellipse
        repel = TRUE, # Avoid label overplotting (slow)
        ggtheme = theme_classic(),
        data = clusK_data)

#extract 2nd cluster of interest
clusterK <- c(3) # from kmeans result - enter manually
clusterK_faces <- as.data.frame(km.res$cluster)
colnames(clusterK_faces) <- "cluster"
clusterK_faces <- subset(clusterK_faces, cluster %in% clusterK)
clusterK_faces <- row.names(clusterK_faces) # vector containing Face row names for cluster
print(clusterK_faces)

# extract 2nd kmeans cluster Faces from distance matrix
clusK_data <- data_raw[c(clusterK_faces),c(clusterK_faces)]
# Mediod calc using clara and cluster data set
```

```
clara.res <- clara(clusK_data, 1, samples = 50, pamLike = TRUE)
table(clara.res$clustering)
summary(clara.res)
```

```
Object of class 'clara' from call:
 clara(x = clusK_data, k = 1, samples = 50, pamLike = TRUE)
Medoids:
          F10 F114 F117 F119 F122     F132     F134 F154 F20      F28 F34 F50      F68      F72 F99
F114 5.571429    0    4    5    4 4.428571 4.857143    5 3.5 6.142857   4 4.5 5.571429 4.833333 3.6
Objective function:       7.519265
Numerical information per cluster:
     size max_diss  av_diss isolation
[1,]   15  10.2048 7.519265         0

Best sample:
 [1] F10  F114 F117 F119 F122 F132 F134 F154 F20  F28  F34  F50  F68  F72  F99
Clustering vector:
 F10 F114 F117 F119 F122 F132 F134 F154  F20  F28  F34  F50  F68  F72  F99
   1    1    1    1    1    1    1    1    1    1    1    1    1    1    1

105 dissimilarities, summarized :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 5.4807  7.7781  8.7785  8.8493  9.9173 12.4080
Metric :  euclidean
Number of objects : 15

Available components:
[1] "sample"     "medoids"    "i.med"      "clustering" "objective"  "clusinfo"   "diss"       "call"
[9] "data"
```

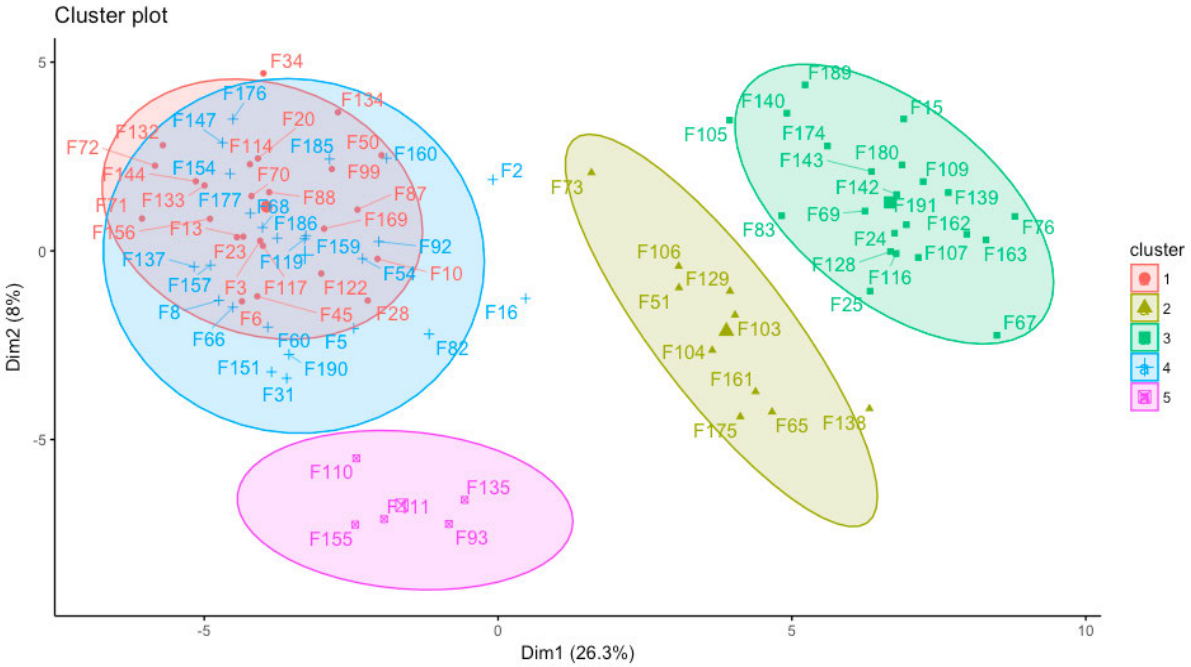
```
# check summary results to get mediod
mediod <- ("F114") # from clara results - enter manually
# create x, y, z co ords for MDS plot using cluster data set
fit <- cmdscale(clusK_data, k=3) # k is the number of dimensions (3D)
print(fit)
# transform x, y, z co-ords making mediod center of plot
mediod_data <- apply(fit, 2, function(x) x - x[mediod])
medioddf <- as.data.frame(mediod_data)
print(medioddf)
Xmax <- max(abs(medioddf$V1)) # x co ord max value
Ymax <- max(abs(medioddf$V2)) # y co ord max value
Zmax <- max(abs(medioddf$V3)) # z co ord max value
lower <- .04 # set lower radius limit
upper <- .135 # set upper radius limit
# extract data points within limits
highsim_data <- subset(medioddf, subset = (abs(V1) < Xmax*upper & abs(V1) >
Xmax*lower |
                                    abs(V2) < Ymax*upper & abs(V2) > Ymax*lower |
```

```
                                    abs(V3) < Zmax*upper & abs(V3) > Zmax*lower),select=c(V1,
V2, V3))
highsim_faces <- row.names(highsim_data) # Face labels as vector
highsim_faces <- append(highsim_faces, mediod) # add mediod
print(highsim_faces)
```

```
[1] "F117" "F132" "F134" "F20"  "F34"  "F50"  "F114"
```

R code for identifying low similarity fillers:

```
library(cluster)
library(factoextra)
library("rgl", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")
install.packages("Hmisc")
library("Hmisc")
rm(list=ls())
Data_Dummy <- read.csv("~/Uni Files/RStudio/responseMatrix.csv", row.names = 1)
# Convert to data.frame
data_r <- as.data.frame(Data_Dummy)
# remove F113 as has nose ring, F22 & F58 glasses, F47 old
remove_face <- c("F113", "F22", "F58", "F47")
data_raw <- data_r[!(colnames(data_r) %in% remove_face)]
data_raw <- data_raw[which(rownames(data) %nin% remove_face), ]
set.seed(123)
km.res <- kmeans(data_raw, 5, nstart = 25)
print(km.res)
# Plot kmeans clusters
fviz_cluster(km.res,
        ellipse.type = "t", # Concentration ellipse
        repel = TRUE, # Avoid label overplotting (slow)
        ggtheme = theme_classic(),
        data = data_raw)
```




```
#extract cluster of interest
clusterK <- c(1,4) # from kmeans result - enter manually
clusterK_faces <- as.data.frame(km.res$cluster)
colnames(clusterK_faces) <- "cluster"
clusterK_faces <- subset(clusterK_faces, cluster %in% clusterK)
clusterK_faces <- row.names(clusterK_faces) # vector containing Face row names for cluster
print(clusterK_faces)
# extract kmeans cluster Faces from distance matrix
clusK_data <- data_raw[c(clusterK_faces),c(clusterK_faces)]
# mediod determined by clara results - same as high sim face id code
mediod <- ("F114") # from clara results - enter manually
# create x, y, z co ords for MDS plot using cluster data set
fit <- cmdscale(clusK_data, k=3) # k is the number of dimensions (3D)
print(fit)
# transform x, y, z co-ords making mediod center of plot
mediod_data <- apply(fit, 2, function(x) x - x[mediod])
medioddf <- as.data.frame(mediod_data)
print(medioddf)
Xmax <- max(abs(medioddf$V1)) # x co ord max value
Ymax <- max(abs(medioddf$V2)) # y co ord max value
```

```
Zmax <- max(abs(medioddf$V3)) # z co ord max value

lower <- .7 # set lower radius limit

upper <- .78 # set upper radius limit

# extract data points within limits

lowsim_data <- subset(medioddf, subset = (abs(V1) < Xmax*upper & abs(V1) >
Xmax*lower |
                                  abs(V2) < Ymax*upper & abs(V2) > Ymax*lower |
                                  abs(V3) < Zmax*upper & abs(V3) > Zmax*lower),select=c(V1,
V2, V3))

lowsim_faces <- row.names(lowsim_data) # Face labels as vector

lowsim_faces <- append(lowsim_faces, mediod) # add mediod

print(lowsim_faces)
```

```
[1] "F156" "F157" "F160" "F186" "F5"   "F82"  "F114"
```

Appendix E

Observed Data Frequency Counts

Table E1

Frequency counts of observed data for target present and target absent lineups collapsed across similarity and exposure duration conditions

Condition	Rejections	Total ID's		Confidence				
				0-20%	21-40%	41-60%	61-80%	81-100%
Low Similarity								
Target present	15	94	CID	0	3	11	27	47
Target absent	58	42	FA	5	5	6	15	11
High Similarity								
Target present	11	87	CID	2	3	13	23	32
Target absent	36	60	FA	3	4	20	19	14
Short Exposure								
Target present	10	90	CID	2	3	11	28	34
Target absent	50	45	FA	4	7	14	13	7
Long Exposure								
Target present	16	91	CID	0	3	13	22	45
Target absent	44	57	FA	4	2	12	21	18

Note. CID = Correct identification (target selection from target present lineup); FA = False Alarm (filler selection from target absent lineup).

Appendix F

Descriptive Statistics

Table F1

Descriptive statistics and normality tests by condition – dependent variable perceived lineup similarity

Condition			Perceived Lineup Similarity			Shapiro-wilk test	
Exposure	Similarity	Lineup Type	*M*	*(SD)*	*N*	Statistic	Sig
Nil (Control)	Low	Target absent	53.22	(31.00)	51	0.93	.004
Nil (Control)	Low	Target present	48.36	(30.19)	46	0.96	.077
Nil (Control)	High	Target absent	49.80	(26.33)	45	0.85	.063
Nil (Control)	High	Target present	54.09	(30.55)	51	0.96	.094
Short	Low	Target absent	50.65	(36.25)	49	0.95	.028
Short	Low	Target present	51.91	(32.32)	51	0.95	.047
Short	High	Target absent	51.75	(28.95)	46	0.97	.330
Short	High	Target present	49.23	(30.04)	49	0.92	.002
Long	Low	Target absent	44.90	(33.38)	51	0.90	.000
Long	Low	Target present	50.95	(27.03)	58	0.97	.171
Long	High	Target absent	53.10	(29.84)	50	0.93	.008
Long	High	Target present	49.32	(22.16)	49	0.96	.077

Table F2

Descriptive statistics and normality tests by condition – lineup mean similarity rating

Condition			Mean Similarity Rating			Shapiro-wilk test	
Exposure	Similarity	Lineup Type	*M*	*(SD)*	*N*	Statistic	Sig
Nil (Control)	Low	Target absent	5.81	(1.43)	51	0.97	.253
Nil (Control)	Low	Target present	5.43	(1.81)	46	0.97	.194
Nil (Control)	High	Target absent	5.56	(1.34)	45	0.95	.066
Nil (Control)	High	Target present	5.73	(1.39)	51	0.98	.048
Short	Low	Target absent	5.71	(1.72)	49	0.95	.042
Short	Low	Target present	6.30	(1.34)	51	0.98	.382
Short	High	Target absent	5.31	(1.28)	46	0.95	.054
Short	High	Target present	5.80	(1.05)	49	0.98	.590
Long	Low	Target absent	5.75	(1.77)	51	0.93	.004
Long	Low	Target present	6.15	(1.27)	58	0.99	.883
Long	High	Target absent	5.49	(1.34)	50	0.97	.235
Long	High	Target present	5.48	(1.31)	49	0.98	.670

Appendix G

R Code for Data Transformation and Descriptive Statistics

```
---
title: "R Notebook"
output: html_notebook
---
This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook for preforming
preliminary
results processing for Peter's 2018 Honours project.
Load the required libraries:
```{r}
library(tidyverse)
library(reshape2)
library(lsr)
library(dplyr)
setwd("~/Uni Files/RStudio")
```
Read in the appropriate CSV file containing the raw results.
```{r}
rm(list=ls())
raw = read.csv("results_20180804.csv", as.is = FALSE);
```
Define a simple function to test whether a given participant has answered the video check
questions correctly.
```{r}
passedVideoCheck = function(condition, q1, q2, q3) {
```
```

```
  ifelse(startsWith(as.character(condition), "CO"),
         q1 == 5 & q2 == 1 & q3 == 1,
         q1 == 5 & q2 == 3 & q3 == 2)
}
```

```{r}
raw %>% select(- matches("Distraction"))
```

Define a simple function to transform raw data from the form exported from Google App Engine
into an initial dataframe suitable for subsequent processing.

```{r}
processRawData = function(raw)
{
    ## Drop rows corresponding to test runs etc.
    ## Add something like this for date filtering:
    ## ! startsWith(as.character(date), "YYYY-MM-DD")
    raw <- raw %>% filter(src != "admin", src != "kr", X.Appengine.Country != "AU")
    ## Add a column which captures whether a participant finished.
    raw$finished <- ! is.na(raw$experimentEndTime);
    ## The uid for any participants for whom we have reason to
    ## believe should be excluded, are captured here.
    excluded <- c(
        "Dummy"
    )
    repeaters <- c("Dummy");
```

```
repeated <- raw[raw$mtWorkerId %in% repeaters | raw$mtWorkerRepeat == "True",]
included <- raw[! (raw$mtWorkerId %in% repeaters | raw$mtWorkerRepeat == "True"),]
included <- raw %>% filter(! is.na(videoCheck_q1));
## Capture the result data for those participants who completed
## and who are not to be excluded
completed <- included[included$finished &  ! (included$uid %in% excluded),]
completed$uid <- as.character(completed$uid)
# Re-label the condition factor
completed$condition <- as.factor(completed$condition)
## Produce a filtered version that omits those people
## who failed to correctly answer the video check question
passed <- completed %>% filter(passedVideoCheck(condition,
                                    videoCheck_q1,
                                    videoCheck_q2,
                                    videoCheck_q3))

## Keep a data frame (useful for inner joins) that tracks the
## number of people in each condition.
condN =  passed %>% group_by(condition) %>% summarise(condN = n())
c1 <- passed %>%
   mutate(id_duration = Test_T1_trialEndTime - Test_T1_enterState_startLineup,
          confidence_duration = confidence_endTime_1 - confidence_startTime_1)
c2 <- c1 %>% select(uid,
                    condition,
                    total_duration = experimentEndTime,
                    id_selection = Test_T1_finalSelection,
```

```
                id_RT = id_duration,
                confidence_RT = confidence_duration,
                id_confidence = confidence_rating,
                original_order = Test_T1_suspectOrder)
c3 <- c2 %>%
  separate(original_order,
           c("Face_1","Face_2","Face_3","Face_4","Face_5","Face_6"),
           sep=":") %>%
  separate(condition,
           c("Exposure","Similarity", "Lineup_Type", "Lineup_Gender"),
            sep="_")
c3$Exposure <- factor(c3$Exposure,
                      levels = c("CO","SE", "LE"),
                      labels=c("Control", "Short", "Long"))
c3$Similarity <- factor(c3$Similarity,
                        levels = c("LS","HS"),
                        labels = c("Low","High"))
c3$Lineup_Type <- factor(c3$Lineup_Type,
                         levels = c("TA","TP"),
                         labels = c("Target Absent", "Target Present"))
faces <- passed %>%
  gather("trial", "faceId", matches("_face")) %>%
  select(uid, trial, faceId) %>%
  separate(trial, c("phase","trialId","R"), sep="_") %>%
  select(-phase,-R) %>%
  group_by(uid, trialId) %>%
```

```
      mutate(face = row_number()) %>%
      spread(face, faceId, sep="_")
    ratingResponses <- passed %>%
      select(uid, matches("Rating")) %>%
      select(uid, matches("_response"))
    ratingResponses <- passed %>%
      gather("trial", "response", matches("Rating_.*_response")) %>%
      select(uid, trial, response) %>%
      separate(trial, c("phase","trialId","R"), sep="_") %>%
      select(-phase,-R) %>%
      filter(! is.na(response))
    environment()
}
```

Process the raw data:

```{r}
e <- processRawData(raw %>% filter(src == "mt"))
```

```{r}
cohesionFunc = function(x) {
    sum((x  - mean(x)) ^ 2)
}
responseByPerson <-
  e$ratingResponses %>%
  group_by(uid) %>%
  summarise(mu = mean(response),
```

```
            sigma = sd(response),
            cohesion = cohesionFunc(response))
responseByPerson
responses <-
  e$ratingResponses %>%
  inner_join(responseByPerson, by = "uid") %>%
  mutate(zResponse = (response - mu) / sigma)
responses
df <- responses %>%
  inner_join(e$faces) %>%
  filter(! is.na(response))
df
condData <- e$c3 %>%
  #filter(Lineup_Type == "Target Present") %>%
  mutate(Correct = (id_selection == "F114")) %>%
  select(uid, Exposure, Similarity, Lineup_Type, Correct)
df <- df %>% inner_join(condData, by = "uid") %>% arrange(uid)
df %>% group_by(Similarity, Exposure, Lineup_Type) %>% summarise(mmu = mean(mu),
mcohesion = mean(cohesion)) %>% arrange(Lineup_Type)
df %>% group_by(Lineup_Type) %>% summarise(mmu = mean(mu), mcohesion =
mean(cohesion))
```

```{r}
# Individual condition descriptive stats and shapiro-wilk test
# Change variable to change conditions tested
library(psych)
```

```
dstatsData <- df %>% filter(Lineup_Type == "Target Absent", Similarity == "High",
Exposure == "Long") %>% group_by(uid) %>% summarise(mmu = mean(mu), mcohesion =
mean(cohesion))
test <- dstatsData[,-1]
describe(test)
shapiro.test(test$mmu)
shapiro.test(test$mcohesion)
```

Appendix H

R Code for Observed Data Counts and UVSDT Model Fitting by Condition

```
# Pete's lineup accuracy
# Observed data counts
# Following script filters for selected condition and then provides CID, FA and Rejection
# totals between the specified id_confidence levels (eg <= 100 & <=0 gives all totals for the
# condition).
# If you change the 'filter' settings you also have to change the 'select' settings to reflect this.
# This is how I got the frequency counts for each condition for the UVSDT scripts.
```

```{r}
confData <- e$c3 %>%
  filter(Exposure == "Long" ) %>%
  mutate(Correct = (id_selection == "F114")) %>%
  mutate(Rejection = (id_selection == "Silhouette")) %>%
  select(uid, Exposure, Similarity, Lineup_Type, id_confidence, Correct, Rejection)%>%
  arrange(id_confidence)
#confData
df <- confData %>% filter(id_confidence <= 20 & id_confidence >= 0) %>% # Bins = 100-81, 80-61, 60-41, 40-21, 20-0
  group_by(Exposure, Lineup_Type) %>%
  add_tally(Correct == "TRUE") %>%
  add_tally(Correct == "FALSE") %>%
  add_tally(Rejection == "TRUE")
df <- plyr::rename(df, c(n="Target_ID", nn="Non_Target_ID", nnn="Lineup_Rejections"))
```

```
df <- df %>% group_by( Exposure, Lineup_Type, Target_ID, Non_Target_ID,
Lineup_Rejections) %>%
  select( Exposure, Lineup_Type, Target_ID, Non_Target_ID, Lineup_Rejections)
df <- summarise(df)
df <- df %>% mutate(False_ID = abs(Non_Target_ID - Lineup_Rejections)) %>% select(
Exposure, Lineup_Type, Target_ID, Non_Target_ID, Lineup_Rejections, False_ID)
df
```

```
# Pete's model fitting script Low Sim
obs.dataLS <- matrix(data = c(47,27,11,3,0,0, # Low Sim Frequency counts from 'Petes
Lineup Accuracy' script
                              48,28,14,4,0,15,
                              11,15,6,5,5,58),
                     nrow = 3,
                     ncol = 6,
                     byrow = TRUE)
n <- 6 #lineup size
#pars <- c(1.8, 1.4, 1, 0.6, 0.2, 2, 1) c1, c2, c3, c4, c5, d, s used to simulate data above
#Likelihood functions generate predicted data.
#Given a particular set of parameters that define the likelihood surface, they give the most
likely data
#Predicted proportion of Correct IDs according to MAX model
QT <- function(c,d,s,n){
  m <- function(x) dnorm(x,mean = d, sd = s)*(pnorm(x)^(n-1))
  p <- vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
```

```
    a <- integrate(m,c[i],15)
    p[i] <- a$value
  }
  return(p)
}
#predicted proportion of total detections on TP trials MAX model
TP <- function(c,d,s,n){
  p <- vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] <- pnorm(((c[i])-d)/s)*pnorm((c[i])^(n-1))
  }
  p <- 1 - p
  return(p)
}
#predicted proportion of total detections on TA trials MAX model
TA <- function(c,n){
  p = vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] = pnorm(c[i])^n
  }
  p = 1 - p
  return(p)
}
genpred <- function(pars, obs.dataLS, n){
  c <- pars[1:(length(pars)-2)]
  d <- pars[length(pars)-1]
```

```
  s <- tail(pars,1)
  total.TP <- sum(obs.dataLS[2,])
  total.TA <- sum(obs.dataLS[3,])
  CID <- QT(c(c, -Inf),d,s,n)
  CID <- c(CID[1],diff(CID))
  TDTP <- c(TP(c,d,s,n),1)
  TDTP <- c(TDTP[1],diff(TDTP))
  TDTA <- c(TA(c,n),1)
  TDTA <- c(TDTA[1],diff(TDTA))
  CID <- CID*total.TP
  TDTP <- TDTP*total.TP
  TDTA <- TDTA*total.TA
  pred.dataLS <- rbind(CID,TDTP,TDTA)
  rownames(pred.dataLS) <- c()
  return(pred.dataLS)
}
#Chi-square
chisq <- function(pars,obs.dataLS,n){
  pred.dataLS <- genpred(pars,obs.dataLS,n)
  lastcell <- ncol(obs.dataLS)
  nc <- ncol(obs.dataLS)-1
  f <- vector(mode = "integer", length = nrow(obs.dataLS)*ncol(obs.dataLS)) #for storing and
summing chi-sq fit value
  for (i in 1:nc){
    a <- pred.dataLS[1,i] #Correct ID
    b <- obs.dataLS[1,i]
```

```
    f[1] <- f[1] + (b-a)^2/a
    a <- pred.dataLS[2,i]-pred.dataLS[1,i] #Foil ID on TP lineup
    b <- obs.dataLS[2,i]-obs.dataLS[1,i]
    f[2] = f[2] + (b-a)^2/a
    a <- pred.dataLS[3,i] #False Alarm
    b <- obs.dataLS[3,i]
    f[3] <- f[3] + (b-a)^2/a
  }
  a <- pred.dataLS[2,lastcell] #Rejection TP
  b <- obs.dataLS[2,lastcell]
  f[4] <- (b-a)^2/a
  a <- pred.dataLS[3,lastcell] #Rejection TA
  b <- obs.dataLS[3,lastcell]
  f[5] <- (b-a)^2/a
  f <- sum(f)
  return(f)
}
#optimisation
x0 = c(5,4,3,2,1,1,1) #c1, c2, c3, c4, c5, d, s
A <- cbind(c(1,0,0,0),c(-1,1,0,0),c(0,-1,1,0),c(0,0,-1,1),c(0,0,0,-1),c(0,0,0,0),c(0,0,0,0))
#added extra column for s parameter
b <- c(0,0,0,0)
#Optimize using the constraints.
outLS <- constrOptim(theta = x0, f = chisq, grad = NULL, ui = A, ci = b, mu = 1e-04,
method = "Nelder-Mead",
                     outer.iterations = 100, obs.dataLS = obs.dataLS, n = n)
```

```
#get fit statistic and parameters from model fit
chisq.modelfitLS <- outLS$value
c.modelfitLS <- outLS$par[1:(length(outLS$par)-2)]
d.modelfitLS <- outLS$par[length(outLS$par)-1]
s.modelfitLS <- tail(outLS$par,1)
pred.dataLS <- genpred(outLS$par, obs.dataLS, n)
pred.dataLS
chisq.modelfitLS
c.modelfitLS
d.modelfitLS
s.modelfitLS
rownames(pred.dataLS)<-c("CID","TD", "FA") #add row names
colnames(pred.dataLS)<-c("c1","c2","c3","c4","c5","c6") #add column names
rownames(obs.dataLS)<-c("CID","TD", "FA") #add row names
colnames(obs.dataLS)<-c("c1","c2","c3","c4","c5","c6") #add column names
pred.dataLS
obs.dataLS
obs.data1LS <- obs.dataLS[-2,]
obs.data1LS
pred.data1LS <- pred.dataLS[-2,]
pred.data1LS
obs.data1LS <- obs.data1LS[,-6]
pred.data1LS <- pred.data1LS[,-6]
fa.totalnLS <- sum(obs.data1LS["FA",])
h.totalnLS <- sum(obs.data1LS["CID",])
#This is for plotting ROCs - reshaping the data back into cumulative proportions
```

```
obs.pLS <- obs.data1LS
obs.pLS[1,] <- cumsum(obs.pLS[1,])
obs.pLS[2,] <- cumsum(obs.pLS[2,])
obs.pLS
obs.pLS[1,] <- obs.pLS[1,]/h.totalnLS
obs.pLS[2,] <- obs.pLS[2,]/fa.totalnLS
obs.pLS[2,] <- obs.pLS[2,]/6
obs.pLS
obs.pLS <- as.data.frame(obs.pLS)
c6 <- matrix(data = c(0,0),
             nrow = 2,
             ncol = 1,
             byrow = TRUE)
obs.pLS <-  cbind(obs.pLS,c6)
pred.pLS <- pred.data1LS
pred.pLS[1,] <- cumsum(pred.pLS[1,])
pred.pLS[2,] <- cumsum(pred.pLS[2,])
pred.pLS[1,] <- pred.pLS[1,]/h.totalnLS
pred.pLS[2,] <- pred.pLS[2,]/fa.totalnLS
pred.pLS[2,] <- pred.pLS[2,]/6
pred.pLS
pred.pLS <- as.data.frame(pred.pLS)
c6 <- matrix(data = c(0,0),
             nrow = 2,
             ncol = 1,
             byrow = TRUE)
```

```
pred.pLS <-  cbind(pred.pLS,c6)
obs.plotLS <- t(obs.pLS) # transform from wide to long form for ggplots
obs.plotLS <- as.data.frame(obs.plotLS)
obs.plotLS
pred.plotLS <- t(pred.pLS)
pred.plotLS <- as.data.frame(pred.plotLS)
pred.plotLS


# Pete's model fitting script High Sim
obs.dataHS <- matrix(data = c(32,23,13,3,2,0, # High Sim Frequency counts from 'Petes
Lineup Accuracy' script
                      34,28,17,5,3,11,
                      14,19,20,4,3,36),
                nrow = 3,
                ncol = 6,
                byrow = TRUE)
n <- 6 #lineup size
#pars <- c(1.8, 1.4, 1, 0.6, 0.2, 2, 1) c1, c2, c3, c4, c5, d, s used to simulate data above
#Likelihood functions generate predicted data.
#Given a particular set of parameters that define the likelihood surface, they give the most
likely data
#Predicted proportion of Correct IDs according to MAX model
QT <- function(c,d,s,n){
  m <- function(x) dnorm(x,mean = d, sd = s)*(pnorm(x)^(n-1))
  p <- vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
```

```
    a <- integrate(m,c[i],15)
    p[i] <- a$value
  }
  return(p)
}
#predicted proportion of total detections on TP trials MAX model
TP <- function(c,d,s,n){
  p <- vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] <- pnorm(((c[i])-d)/s)*pnorm((c[i])^(n-1))
  }
  p <- 1 - p
  return(p)
}
#predicted proportion of total detections on TA trials MAX model
TA <- function(c,n){
  p = vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] = pnorm(c[i])^n
  }
  p = 1 - p
  return(p)
}
genpred <- function(pars, obs.dataHS, n){
  c <- pars[1:(length(pars)-2)]
  d <- pars[length(pars)-1]
```

```
  s <- tail(pars,1)
  total.TP <- sum(obs.dataHS[2,])
  total.TA <- sum(obs.dataHS[3,])
  CID <- QT(c(c, -Inf),d,s,n)
  CID <- c(CID[1],diff(CID))
  TDTP <- c(TP(c,d,s,n),1)
  TDTP <- c(TDTP[1],diff(TDTP))
  TDTA <- c(TA(c,n),1)
  TDTA <- c(TDTA[1],diff(TDTA))
  CID <- CID*total.TP
  TDTP <- TDTP*total.TP
  TDTA <- TDTA*total.TA
  pred.dataHS <- rbind(CID,TDTP,TDTA)
  rownames(pred.dataHS) <- c()
  return(pred.dataHS)
}
#Chi-square
chisq <- function(pars,obs.dataHS,n){
  pred.dataHS <- genpred(pars,obs.dataHS,n)
  lastcell <- ncol(obs.dataHS)
  nc <- ncol(obs.dataHS)-1
  f <- vector(mode = "integer", length = nrow(obs.dataHS)*ncol(obs.dataHS)) #for storing
and summing chi-sq fit value
  for (i in 1:nc){
    a <- pred.dataHS[1,i] #Correct ID
    b <- obs.dataHS[1,i]
```

```
    f[1] <- f[1] + (b-a)^2/a
    a <- pred.dataHS[2,i]-pred.dataHS[1,i] #Foil ID on TP lineup
    b <- obs.dataHS[2,i]-obs.dataHS[1,i]
    f[2] = f[2] + (b-a)^2/a
    a <- pred.dataHS[3,i] #False Alarm
    b <- obs.dataHS[3,i]
    f[3] <- f[3] + (b-a)^2/a
  }
  a <- pred.dataHS[2,lastcell] #Rejection TP
  b <- obs.dataHS[2,lastcell]
  f[4] <- (b-a)^2/a
  a <- pred.dataHS[3,lastcell] #Rejection TA
  b <- obs.dataHS[3,lastcell]
  f[5] <- (b-a)^2/a
  f <- sum(f)
  return(f)
}
#optimisation
x0 = c(5,4,3,2,1,1,1) #c1, c2, c3, c4, c5, d, s
A <- cbind(c(1,0,0,0),c(-1,1,0,0),c(0,-1,1,0),c(0,0,-1,1),c(0,0,0,-1),c(0,0,0,0),c(0,0,0,0))
#added extra column for s parameter
b <- c(0,0,0,0)
#Optimize using the constraints.
outHS <- constrOptim(theta = x0, f = chisq, grad = NULL, ui = A, ci = b, mu = 1e-04,
method = "Nelder-Mead",
                     outer.iterations = 100, obs.dataHS = obs.dataHS, n = n)
```

```
#get fit statistic and parameters from model fit

chisq.modelfitHS <- outHS$value

c.modelfitHS <- outHS$par[1:(length(outHS$par)-2)]

d.modelfitHS <- outHS$par[length(outHS$par)-1]

s.modelfitHS <- tail(outHS$par,1)

pred.dataHS <- genpred(outHS$par, obs.dataHS, n)

pred.dataHS

chisq.modelfitHS

c.modelfitHS

d.modelfitHS

s.modelfitHS

rownames(pred.dataHS)<-c("CID","TD", "FA") #add row names

colnames(pred.dataHS)<-c("c1","c2","c3","c4","c5","c6") #add column names

rownames(obs.dataHS)<-c("CID","TD", "FA") #add row names

colnames(obs.dataHS)<-c("c1","c2","c3","c4","c5","c6") #add column names

pred.dataHS

obs.dataHS

obs.data1HS <- obs.dataHS[-2,]

obs.data1HS

pred.data1HS <- pred.dataHS[-2,]

obs.data1HS

obs.data1HS <- obs.data1HS[,-6]

pred.data1HS <- pred.data1HS[,-6]

fa.totalnHS <- sum(obs.data1HS["FA",])

h.totalnHS <- sum(obs.data1HS["CID",])

#This is for plotting ROCs - reshaping the data back into cumulative proportions
```

```
obs.pHS <- obs.data1HS
obs.pHS[1,] <- cumsum(obs.pHS[1,])
obs.pHS[2,] <- cumsum(obs.pHS[2,])
obs.pHS
obs.pHS[1,] <- obs.pHS[1,]/h.totalnHS
obs.pHS[2,] <- obs.pHS[2,]/fa.totalnHS
obs.pHS[2,] <- obs.pHS[2,]/6
obs.pHS
obs.pHS <- as.data.frame(obs.pHS)
c6 <- matrix(data = c(0,0),
             nrow = 2,
             ncol = 1,
             byrow = TRUE)
obs.pHS <-  cbind(obs.pHS,c6)
pred.pHS <- pred.data1HS
pred.pHS[1,] <- cumsum(pred.pHS[1,])
pred.pHS[2,] <- cumsum(pred.pHS[2,])
pred.pHS[1,] <- pred.pHS[1,]/h.totalnHS
pred.pHS[2,] <- pred.pHS[2,]/fa.totalnHS
pred.pHS[2,] <- pred.pHS[2,]/6
pred.pHS
pred.pHS <- as.data.frame(pred.pHS)
c6 <- matrix(data = c(0,0),
             nrow = 2,
             ncol = 1,
             byrow = TRUE)
```

```
pred.pHS <-  cbind(pred.pHS,c6)
obs.plotHS <- t(obs.pHS)
obs.plotHS <- as.data.frame(obs.plotHS)
obs.plotHS
pred.plotHS <- t(pred.pHS)
pred.plotHS <- as.data.frame(pred.plotHS)
pred.plotHS
# Pete's model fitting script Short Exp
obs.dataSE <- matrix(data = c(34,28,11,3,2,0, # Short Exp Frequency counts from 'Petes
Lineup Accuracy' script
                      35,32,15,5,3,10,
                      7,13,14,7,4,50),
                nrow = 3,
                ncol = 6,
                byrow = TRUE)
n <- 6 #lineup size
#pars <- c(1.8, 1.4, 1, 0.6, 0.2, 2, 1) c1, c2, c3, c4, c5, d, s used to simulate data above
#Likelihood functions generate predicted data.
#Given a particular set of parameters that define the likelihood surface, they give the most
likely data
#Predicted proportion of Correct IDs according to MAX model
QT <- function(c,d,s,n){
  m <- function(x) dnorm(x,mean = d, sd = s)*(pnorm(x)^(n-1))
  p <- vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    a <- integrate(m,c[i],15)
```

```
    p[i] <- a$value
  }
  return(p)
}
#predicted proportion of total detections on TP trials MAX model
TP <- function(c,d,s,n){
  p <- vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] <- pnorm(((c[i])-d)/s)*pnorm((c[i])^(n-1))
  }
  p <- 1 - p
  return(p)
}
#predicted proportion of total detections on TA trials MAX model
TA <- function(c,n){
  p = vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] = pnorm(c[i])^n
  }
  p = 1 - p
  return(p)
}
genpred <- function(pars, obs.dataSE, n){
  c <- pars[1:(length(pars)-2)]
  d <- pars[length(pars)-1]
  s <- tail(pars,1)
```

```
  total.TP <- sum(obs.dataSE[2,])
  total.TA <- sum(obs.dataSE[3,])
  CID <- QT(c(c, -Inf),d,s,n)
  CID <- c(CID[1],diff(CID))
  TDTP <- c(TP(c,d,s,n),1)
  TDTP <- c(TDTP[1],diff(TDTP))
  TDTA <- c(TA(c,n),1)
  TDTA <- c(TDTA[1],diff(TDTA))
  CID <- CID*total.TP
  TDTP <- TDTP*total.TP
  TDTA <- TDTA*total.TA
  pred.dataSE <- rbind(CID,TDTP,TDTA)
  rownames(pred.dataSE) <- c()
  return(pred.dataSE)
}
#Chi-square
chisq <- function(pars,obs.dataSE,n){
  pred.dataSE <- genpred(pars,obs.dataSE,n)
  lastcell <- ncol(obs.dataSE)
  nc <- ncol(obs.dataSE)-1
  f <- vector(mode = "integer", length = nrow(obs.dataSE)*ncol(obs.dataSE)) #for storing and
summing chi-sq fit value
  for (i in 1:nc){
    a <- pred.dataSE[1,i] #Correct ID
    b <- obs.dataSE[1,i]
    f[1] <- f[1] + (b-a)^2/a
```

```
    a <- pred.dataSE[2,i]-pred.dataSE[1,i] #Foil ID on TP lineup
    b <- obs.dataSE[2,i]-obs.dataSE[1,i]
    f[2] = f[2] + (b-a)^2/a
    a <- pred.dataSE[3,i] #False Alarm
    b <- obs.dataSE[3,i]
    f[3] <- f[3] + (b-a)^2/a
  }
  a <- pred.dataSE[2,lastcell] #Rejection TP
  b <- obs.dataSE[2,lastcell]
  f[4] <- (b-a)^2/a
  a <- pred.dataSE[3,lastcell] #Rejection TA
  b <- obs.dataSE[3,lastcell]
  f[5] <- (b-a)^2/a
  f <- sum(f)
  return(f)
}
#optimisation
x0 = c(5,4,3,2,1,1,1) #c1, c2, c3, c4, c5, d, s
A <- cbind(c(1,0,0,0),c(-1,1,0,0),c(0,-1,1,0),c(0,0,-1,1),c(0,0,0,-1),c(0,0,0,0),c(0,0,0,0))
#added extra column for s parameter
b <- c(0,0,0,0)
#Optimize using the constraints.
outSE <- constrOptim(theta = x0, f = chisq, grad = NULL, ui = A, ci = b, mu = 1e-04,
method = "Nelder-Mead",
                     outer.iterations = 100, obs.dataSE = obs.dataSE, n = n)
#get fit statistic and parameters from model fit
```

```
chisq.modelfitSE <- outSE$value

c.modelfitSE <- outSE$par[1:(length(outSE$par)-2)]

d.modelfitSE <- outSE$par[length(outSE$par)-1]

s.modelfitSE <- tail(outSE$par,1)

pred.dataSE <- genpred(outSE$par, obs.dataSE, n)

pred.dataSE

chisq.modelfitSE

c.modelfitSE

d.modelfitSE

s.modelfitSE

rownames(pred.dataSE)<-c("CID","TD", "FA") #add row names

colnames(pred.dataSE)<-c("c1","c2","c3","c4","c5","c6") #add column names

rownames(obs.dataSE)<-c("CID","TD", "FA") #add row names

colnames(obs.dataSE)<-c("c1","c2","c3","c4","c5","c6") #add column names

pred.dataSE

obs.dataSE

obs.dataSE1 <- obs.dataSE[-2,]

obs.dataSE1

pred.dataSE1 <- pred.dataSE[-2,]

pred.dataSE1

obs.dataSE1 <- obs.dataSE1[,-6]

pred.dataSE1 <- pred.dataSE1[,-6]

fa.totalnSE <- sum(obs.dataSE1["FA",])

h.totalnSE <- sum(obs.dataSE1["CID",])

obs.dataSE1

#This is for plotting ROCs - reshaping the data back into cumulative proportions
```

```
obs.pSE <- obs.dataSE1
obs.pSE[1,] <- cumsum(obs.pSE[1,])
obs.pSE[2,] <- cumsum(obs.pSE[2,])
obs.pSE
obs.pSE[1,] <- obs.pSE[1,]/h.totalnSE
obs.pSE[2,] <- obs.pSE[2,]/fa.totalnSE
obs.pSE[2,] <- obs.pSE[2,]/6
obs.pSE
obs.pSE <- as.data.frame(obs.pSE)
c6 <- matrix(data = c(0,0),
             nrow = 2,
             ncol = 1,
             byrow = TRUE)
obs.pSE <-  cbind(obs.pSE,c6)
pred.pSE <- pred.dataSE1
pred.pSE[1,] <- cumsum(pred.pSE[1,])
pred.pSE[2,] <- cumsum(pred.pSE[2,])
pred.pSE[1,] <- pred.pSE[1,]/h.totalnSE
pred.pSE[2,] <- pred.pSE[2,]/fa.totalnSE
pred.pSE[2,] <- pred.pSE[2,]/6
pred.pSE
pred.pSE <- as.data.frame(pred.pSE)
c6 <- matrix(data = c(0,0),
             nrow = 2,
             ncol = 1,
             byrow = TRUE)
```

```
pred.pSE <-  cbind(pred.pSE,c6)

obs.plotSE <- t(obs.pSE)

obs.plotSE <- as.data.frame(obs.plotSE)

obs.plotSE

pred.plotSE <- t(pred.pSE)

pred.plotSE <- as.data.frame(pred.plotSE)

pred.plotSE


# Pete's model fitting script Long Exp

obs.dataLE <- matrix(data = c(45,22,13,3,0,0,# Long Exp Frequency counts from 'Petes

Lineup Accuracy' script

                      47,24,16,4,0,16,

                      18,21,12,2,4,44),

               nrow = 3,

               ncol = 6,

               byrow = TRUE)

n <- 6 #lineup size

#pars <- c(1.8, 1.4, 1, 0.6, 0.2, 2, 1) c1, c2, c3, c4, c5, d, s used to simulate data above

#Likelihood functions generate predicted data.

#Given a particular set of parameters that define the likelihood surface, they give the most

likely data

#Predicted proportion of Correct IDs according to MAX model

QT <- function(c,d,s,n){

  m <- function(x) dnorm(x,mean = d, sd = s)*(pnorm(x)^(n-1))

  p <- vector(mode = "integer", length = length(c))

  for (i in 1:length(c)){
```

```
    a <- integrate(m,c[i],15)
    p[i] <- a$value}
  return(p)}
#predicted proportion of total detections on TP trials MAX model
TP <- function(c,d,s,n){
  p <- vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] <- pnorm(((c[i])-d)/s)*pnorm((c[i])^(n-1))}
  p <- 1 - p
  return(p)}
#predicted proportion of total detections on TA trials MAX model
TA <- function(c,n){
  p = vector(mode = "integer", length = length(c))
  for (i in 1:length(c)){
    p[i] = pnorm(c[i])^n}
  p = 1 - p
  return(p)}
genpred <- function(pars, obs.dataLE, n){
  c <- pars[1:(length(pars)-2)]
  d <- pars[length(pars)-1]
  s <- tail(pars,1)
  total.TP <- sum(obs.dataLE[2,])
  total.TA <- sum(obs.dataLE[3,])
  CID <- QT(c(c, -Inf),d,s,n)
  CID <- c(CID[1],diff(CID))
  TDTP <- c(TP(c,d,s,n),1)
```

```
  TDTP <- c(TDTP[1],diff(TDTP))
  TDTA <- c(TA(c,n),1)
  TDTA <- c(TDTA[1],diff(TDTA))
  CID <- CID*total.TP
  TDTP <- TDTP*total.TP
  TDTA <- TDTA*total.TA
  pred.dataLE <- rbind(CID,TDTP,TDTA)
  rownames(pred.dataLE) <- c()
  return(pred.dataLE)}
#Chi-square
chisq <- function(pars,obs.dataLE,n){
  pred.dataLE <- genpred(pars,obs.dataLE,n)
  lastcell <- ncol(obs.dataLE)
  nc <- ncol(obs.dataLE)-1
  f <- vector(mode = "integer", length = nrow(obs.dataLE)*ncol(obs.dataLE)) #for storing
and summing chi-sq fit value
  for (i in 1:nc){
    a <- pred.dataLE[1,i] #Correct ID
    b <- obs.dataLE[1,i]
    f[1] <- f[1] + (b-a)^2/a
    a <- pred.dataLE[2,i]-pred.dataLE[1,i] #Foil ID on TP lineup
    b <- obs.dataLE[2,i]-obs.dataLE[1,i]
    f[2] = f[2] + (b-a)^2/a
    a <- pred.dataLE[3,i] #False Alarm
    b <- obs.dataLE[3,i]
    f[3] <- f[3] + (b-a)^2/a}
```

```
  a <- pred.dataLE[2,lastcell] #Rejection TP
  b <- obs.dataLE[2,lastcell]
  f[4] <- (b-a)^2/a
  a <- pred.dataLE[3,lastcell] #Rejection TA
  b <- obs.dataLE[3,lastcell]
  f[5] <- (b-a)^2/a
  f <- sum(f)
  return(f)  }
#optimisation
x0 = c(5,4,3,2,0,1,1) #c1, c2, c3, c4, c5, d, s
A <- cbind(c(1,0,0,0),c(-1,1,0,0),c(0,-1,1,0),c(0,0,-1,1),c(0,0,0,-1),c(0,0,0,0),c(0,0,0,0))
#added extra column for s parameter
b <- c(0,0,0,0)
#Optimize using the constraints.
outLE <- constrOptim(theta = x0, f = chisq, grad = NULL, ui = A, ci = b, mu = 1e-04,
method = "Nelder-Mead",
                outer.iterations = 100, obs.dataLE = obs.dataLE, n = n)
#get fit statistic and parameters from model fit
chisq.modelfitLE <- outLE$value
c.modelfitLE <- outLE$par[1:(length(outLE$par)-2)]
d.modelfitLE <- outLE$par[length(outLE$par)-1]
s.modelfitLE <- tail(outLE$par,1)
pred.dataLE <- genpred(outLE$par, obs.dataLE, n)
pred.dataLE
chisq.modelfitLE
c.modelfitLE
```

```
d.modelfitLE

s.modelfitLE

rownames(pred.dataLE)<-c("CID","TD", "FA") #add row names

colnames(pred.dataLE)<-c("c1","c2","c3","c4","c5","c6") #add column names

rownames(obs.dataLE)<-c("CID","TD", "FA") #add row names

colnames(obs.dataLE)<-c("c1","c2","c3","c4","c5","c6") #add column names

pred.dataLE

obs.dataLE

obs.dataLE1 <- obs.dataLE[-2,]

obs.dataLE1

pred.dataLE1 <- pred.dataLE[-2,]

pred.dataLE1

obs.dataLE1 <- obs.dataLE1[,-6]

obs.dataLE1

pred.dataLE1 <- pred.dataLE1[,-6]

fa.totalnLE <- sum(obs.dataLE1["FA",])

h.totalnLE <- sum(obs.dataLE1["CID",])

fa.totalnLE

h.totalnLE

#This is for plotting ROCs - reshaping the data back into cumulative proportions

obs.pLE <- obs.dataLE1

obs.pLE

obs.pLE[1,] <- cumsum(obs.pLE[1,])

obs.pLE[2,] <- cumsum(obs.pLE[2,])

obs.pLE

obs.pLE[1,] <- obs.pLE[1,]/h.totalnLE
```

```
obs.pLE[2,] <- obs.pLE[2,]/fa.totalnLE

obs.pLE[2,] <- obs.pLE[2,]/6

obs.pLE

obs.pLE <- as.data.frame(obs.pLE)

c6 <- matrix(data = c(0,0),

        nrow = 2,

        ncol = 1,

        byrow = TRUE)

obs.pLE <-  cbind(obs.pLE,c6)

obs.pLE <- as.data.frame(obs.pLE)

obs.pLE

pred.pLE <- pred.dataLE1

pred.pLE

pred.pLE[1,] <- cumsum(pred.pLE[1,])

pred.pLE[2,] <- cumsum(pred.pLE[2,])

pred.pLE[1,] <- pred.pLE[1,]/h.totalnLE

pred.pLE[2,] <- pred.pLE[2,]/fa.totalnLE

pred.pLE[2,] <- pred.pLE[2,]/6

pred.pLE

pred.pLE <- as.data.frame(pred.pLE)

c6 <- matrix(data = c(0,0),

        nrow = 2,

        ncol = 1,

        byrow = TRUE)

pred.pLE <-  cbind(pred.pLE,c6)

pred.pLE <- as.data.frame(pred.pLE)
```

```
obs.plotLE <- t(obs.pLE)

obs.plotLE <- as.data.frame(obs.plotLE)

obs.plotLE

pred.plotLE <- t(pred.pLE)

pred.plotLE <- as.data.frame(pred.plotLE)

pred.plotLE
```

Appendix I

R Code for Significance Testing Hypotheses 1 and 2

```{r}
# Significance testing predicted data from UVSDT model
# change variable name to change conditions tested
library("XNomial")
Exp <- xmonte(dfSE,
    dfLE,
    statName = "LLR",
    detail = 2,)
Exp$pLLR
Exp$observedLLR
```

Appendix J

R Code for Significance Testing Hypotheses 3a, 3b, 3c and 3d

```{r}
# Planned contrasts using mean(cohesion)
# Change variable to change conditions tested
# Includes box plot output
modelData <- df %>% filter(Lineup_Type == "Target Present", Similarity == "Low") %>%
group_by(uid, Exposure) %>% summarise(mmu = mean(mu), mcohesion = mean(cohesion))
hist(modelData$mcohesion)
shapiro.test(modelData$mcohesion)
boxplot(mcohesion ~ Exposure,
    data = modelData,
    ylab="Perceived Similarity",
    xlab="Exposure",boxwex=.4,
    ylim=c(0,150))
kruskal.test(mcohesion ~ Exposure, data = modelData)
median <- modelData %>% filter(Exposure == "Long")
summary(median)
# Select just Short and Long
modelData <- df %>% filter(Lineup_Type == "Target Present", Similarity == "High",
Exposure != "Control") %>% group_by(uid, Exposure) %>% summarise(mmu = mean(mu),
mcohesion = mean(cohesion))
hist(modelData$mcohesion)
shapiro.test(modelData$mcohesion)
kruskal.test(mcohesion ~ Exposure, data = modelData)
```

Appendix K

R Code for ROC Plots

````
---
title: "R Notebook"
output: html_notebook
---

Pete's ROC Plots

```{r}
library(tidyverse)
library(reshape2)
library(lsr)
library(varhandle)
library(ggrepel)
```

```{r}
# labels for bins if required
labels <-c("80% - 100%", "60% - 100%", "40% - 100%", "20% - 100%", "0% - 100%", " ")
```

```{r}
# ggplot APA theme function
theme_apa <- function(legend.pos = "right", legend.use.title = FALSE,
                      legend.font.size = 12, x.font.size = 12, y.font.size = 12,
                      facet.title.size = 12, remove.y.gridlines = TRUE,
                      remove.x.gridlines = TRUE) {
  # Specifying parameters, using theme_bw() as starting point
  plot <- ggplot2::theme_bw() + ggplot2::theme(
````
````

```
  plot.title = ggplot2::element_text(face = "bold", size = 14),
  axis.title.x = ggplot2::element_text(size = x.font.size),
  axis.title.y = ggplot2::element_text(size = y.font.size,
                                       angle = 90),
  legend.text = ggplot2::element_text(size = legend.font.size),
  legend.key.size = ggplot2::unit(1.5, "lines"),
  # switch off the rectangle around symbols
  legend.key = ggplot2::element_blank(),
  legend.key.width = grid::unit(2, "lines"),
  strip.text.x = ggplot2::element_text(size = facet.title.size), # facet labs
  strip.text.y = ggplot2::element_text(size = facet.title.size),
  # facet titles
  strip.background = ggplot2::element_rect(colour = "white", fill = "white"),
  complete = TRUE
)
# Choose legend position. APA figures generally include legends that
# are embedded on the plane, so there is no efficient way to have it
# automatically placed correctly
if (legend.pos == "topleft") {
  # manually position the legend (numbers being from 0,0 at bottom left of
  # whole plot to 1,1 at top right)
  plot <- plot + ggplot2::theme(legend.position = c(.05, .95),
                                legend.justification = c(.05, .95))
} else if (legend.pos == "topright") {
  plot <- plot + ggplot2::theme(legend.position = c(.95, .95),
                                legend.justification = c(.95, .95))
```

```
  } else if (legend.pos == "topmiddle") {
    plot <- plot + ggplot2::theme(legend.position = c(.50, .95),
                                  legend.justification = c(.50, .95))
  } else if (legend.pos == "bottomleft") {
    plot <- plot + ggplot2::theme(legend.position = c(.05, .05),
                                  legend.justification = c(.05, .05))
  } else if (legend.pos == "bottomright") {
    plot <- plot + ggplot2::theme(legend.position = c(.95, .05),
                                  legend.justification = c(.95, .05))
  } else if (legend.pos == "bottommiddle") {
    plot <- plot + ggplot2::theme(legend.position = c(.50, .05),
                                  legend.justification = c(.50, .05))
  } else if (legend.pos == "none") {
    plot <- plot + ggplot2::theme(legend.position = "none")
  } else {
    plot <- plot + ggplot2::theme(legend.position = legend.pos)
  }
  # Should legend have title? If so, format it correctly
  if (legend.use.title == FALSE) {
    # switch off the legend title
    plot <- plot +
      ggplot2::theme(legend.title = ggplot2::element_blank())
  } else {
    plot <- plot +
      ggplot2::theme(legend.title =
                       ggplot2::element_text(size = 12, face = "bold"))
```

```
  }
  if (remove.y.gridlines == TRUE) {
    plot <- plot + drop_y_gridlines()
  } else {
    plot <- plot + add_y_gridlines()
  }
  if (remove.x.gridlines == TRUE) {
    plot <- plot + drop_x_gridlines()
  } else {
    plot <- plot + add_x_gridlines()
  }
  return(plot)
}
#' @title Add and remove gridlines
#'
#' @description These are convenience wrappers for editing [ggplot2::theme()]'s
#'  `panel.grid.major` and `panel.grid.minor` parameters with sensible
#'  defaults.
#'
#' @param x Apply changes to the x axis?
#' @param y Apply changes to the y axis?
#' @param minor Add minor gridlines in addition to major?
#' @param minor.only Remove only the minor gridlines?
#'
#' @importFrom ggplot2 theme element_line
#' @export
```

```
#' @rdname gridlines
add_gridlines <- function(x = TRUE, y = TRUE, minor = TRUE) {
  plot <- theme()
  if (y == TRUE) {
    plot <- plot + theme(panel.grid.major.y = element_line(colour = "grey92"))
    if (minor == TRUE) {
      plot <-
        plot + theme(panel.grid.minor.y = element_line(colour = "grey92",
                                                       size = .25))
    }
  }
  if (x == TRUE) {
    plot <- plot + theme(panel.grid.major.x = element_line(colour = "grey92"))
    if (minor == TRUE) {
      plot <-
        plot + theme(panel.grid.minor.x = element_line(colour = "grey92",
                                                       size = .25))
    }
  }
  return(plot)
}
#' @export
#' @rdname gridlines
add_x_gridlines <- function(minor = TRUE) {
  add_gridlines(x = TRUE, y = FALSE, minor = minor)
}
```

```
#' @export
#' @rdname gridlines
add_y_gridlines <- function(minor = TRUE) {
  add_gridlines(x = FALSE, y = TRUE, minor = minor)
}
#' @export
#' @rdname gridlines
drop_gridlines <- function(x = TRUE, y = TRUE, minor.only = FALSE) {
  plot <- ggplot2::theme()
  if (y == TRUE) {
    plot <- plot + ggplot2::theme(panel.grid.minor.y = ggplot2::element_blank())
    if (minor.only == FALSE) {
      plot <-
        plot + ggplot2::theme(panel.grid.major.y = ggplot2::element_blank())
    }
  }
  if (x == TRUE) {
    plot <- plot + ggplot2::theme(panel.grid.minor.x = ggplot2::element_blank())
    if (minor.only == FALSE) {
      plot <-
        plot + ggplot2::theme(panel.grid.major.x = ggplot2::element_blank())
    }
  }
  return(plot)
}
#' @export
```

```
#' @rdname gridlines
drop_x_gridlines <- function(minor.only = FALSE) {
  drop_gridlines(x = TRUE, y = FALSE, minor.only = minor.only)
}
#' @export
#' @rdname gridlines
drop_y_gridlines <- function(minor.only = FALSE) {
  drop_gridlines(x = FALSE, y = TRUE, minor.only = minor.only)
}
```

```{r}
# Combine Short and Long Exposure plot data
obs.plotLE # df from Petes SDT Long Exp script
obs.plotLE_ROC <- obs.plotLE %>% mutate(Condition = "Obs Long Exposure")
obs.plotSE_ROC <- obs.plotSE %>% mutate(Condition = "Obs Short Exposure")
pred.plotLE_ROC <- pred.plotLE %>% mutate(Condition = "Pred Long Exposure")
pred.plotSE_ROC <- pred.plotSE %>% mutate(Condition = "Pred Short Exposure")
obs.plotLE_ROC
obs.plotSE_ROC
pred.plotLE_ROC
pred.plotSE_ROC
ROCdfEXP <- rbind(pred.plotLE_ROC, pred.plotSE_ROC, obs.plotLE_ROC,
obs.plotSE_ROC)
ROCdfEXP
```

```{r}
```

```
#Short and Long ROC plot
ggplot () +
  theme_apa() +
  xlab('FAs') +
  ylab('HITs') +
  geom_point(data = filter(ROCdfEXP, Condition == "Pred Short Exposure"),
             aes(x = FA, y = CID, color=Condition),
             size = 3, shape = 16) +
  geom_point(data = filter(ROCdfEXP, Condition == "Pred Long Exposure"),
             aes(x = FA, y = CID, color=Condition),
              size = 3, shape = 17) +
  geom_smooth(data = filter(ROCdfEXP, Condition == "Pred Short Exposure"),
              aes(x = FA, y = CID, color=Condition),
              color = "blue", size=0.5, method='loess', se = FALSE) +
  geom_smooth(data = filter(ROCdfEXP, Condition == "Pred Long Exposure"),
              aes(x = FA, y = CID, color=Condition),
              color = "red", size=0.5 ,method='loess', se = FALSE) +
  geom_point(data = filter(ROCdfEXP, Condition == "Obs Short Exposure"),
             aes(x = FA, y = CID, color=Condition),
             size = 2, shape = 15) +
  geom_point(data = filter(ROCdfEXP, Condition == "Obs Long Exposure"),
             aes(x = FA, y = CID, color=Condition),
              size = 2, shape = 18) +
  geom_line(data = filter(ROCdfEXP, Condition == "Obs Short Exposure"),
              aes(x = FA, y = CID, color=Condition), linetype = 2,
              color = "black", size=0.2, method='loess', se = FALSE) +
```

```
  geom_line(data = filter(ROCdfEXP, Condition == "Obs Long Exposure"),
            aes(x = FA, y = CID, color=Condition), linetype = 2,
            color = "black", size=0.2, method='loess', se = FALSE) +
   guides(shape = FALSE,
         colour = guide_legend(override.aes = list(shape = c(17, 16, 18, 15)))) +
  scale_color_manual("",
                     breaks = c("Pred Long Exposure", "Pred Short Exposure",
                                "Obs Long Exposure", "Obs Short Exposure"),
                     values = c("black", "black", "red", "blue")) +
  geom_abline() +
  scale_x_continuous(expand = c(0, 0), limits = c(0,.25)) +
  scale_y_continuous(expand = c(0, 0), limits = c(0,1.1))
```

```{r}
# Combine Low and High Sim plot data new
obs.plotLS # df from Petes SDT Long Exp script
obs.plotLS_ROC <- obs.plotLS %>% mutate(Condition = "Obs Low Similarity")
obs.plotHS_ROC <- obs.plotHS %>% mutate(Condition = "Obs High Similarity")
pred.plotLS_ROC <- pred.plotLS %>% mutate(Condition = "Pred Low Similarity")
pred.plotHS_ROC <- pred.plotHS %>% mutate(Condition = "Pred High Similarity")
obs.plotLS_ROC
obs.plotHS_ROC
pred.plotLS_ROC
pred.plotHS_ROC
ROCdfEXP <- rbind(pred.plotLS_ROC, pred.plotHS_ROC, obs.plotLS_ROC,
obs.plotHS_ROC)
```

ROCdfEXP

```

```{r}

#Low and High ROC plot New

ggplot () +

 theme_apa() +

 xlab('FAs') +

 ylab('HITs') +

 geom_point(data = filter(ROCdfEXP, Condition == "Pred Low Similarity"),

 aes(x = FA, y = CID, color=Condition),

 size = 3, shape = 16) +

 geom_point(data = filter(ROCdfEXP, Condition == "Pred High Similarity"),

 aes(x = FA, y = CID, color=Condition),

 size = 3, shape = 17) +

 geom_smooth(data = filter(ROCdfEXP, Condition == "Pred Low Similarity"),

 aes(x = FA, y = CID, color=Condition),

 color = "blue", size=0.5, method='loess', se = FALSE) +

 geom_smooth(data = filter(ROCdfEXP, Condition == "Pred High Similarity"),

 aes(x = FA, y = CID, color=Condition),

 color = "red", size=0.5 ,method='loess', se = FALSE) +

 geom_point(data = filter(ROCdfEXP, Condition == "Obs Low Similarity"),

 aes(x = FA, y = CID, color=Condition),

 size = 2, shape = 15) +

 geom_point(data = filter(ROCdfEXP, Condition == "Obs High Similarity"),

 aes(x = FA, y = CID, color=Condition),

 size = 2, shape = 18) +

```
 geom_line(data = filter(ROCdfEXP, Condition == "Obs Low Similarity"),
           aes(x = FA, y = CID, color=Condition), linetype = 2,
           color = "black", size=0.2, method='loess', se = FALSE) +
 geom_line(data = filter(ROCdfEXP, Condition == "Obs High Similarity"),
           aes(x = FA, y = CID, color=Condition), linetype = 2,
           color = "black", size=0.2, method='loess', se = FALSE) +
  guides(shape = FALSE,
       colour = guide_legend(override.aes = list(shape = c(17, 16, 18, 15)))) +
 scale_color_manual("",
                    breaks = c("Pred High Similarity", "Pred Low Similarity",
                               "Obs High Similarity", "Obs Low Similarity"),
                    values = c("black", "black", "red", "blue")) +
 geom_abline() +
 scale_x_continuous(expand = c(0, 0), limits = c(0,.25)) +
 scale_y_continuous(expand = c(0, 0), limits = c(-.000001,1.1))
```