THE ROLE OF WITHIN-PERSON AND CONTEXT VARIABILITY ON FACE LEARNING

An Honours Thesis

Submitted in Partial Fulfillment of the Requirements

for the Degree of Bachelor of Science (Honours) in Psychology

Supervised by Dr. Chris Oriet

Luther College

By

Rebekah L. Corpuz

Regina, Saskatchewan

April 2018

Abstract

We investigated whether increased variability leads to better retention of a target face. Previous research suggests that increasing variability had no effect on memory for a target presented a few minutes later. In the real world, however, the time between when we first encounter a face and when we need to be able to identify it can be much longer. We hypothesized that increasing variability, then, might help to establish a more durable representation of a face. To test this, participants were randomly assigned to one of four conditions: a) no variability (still image), b) low variability video (little to no change in target's appearance and context), c) high variability video (several changes in the target's appearance and context), d) control (no video). Participants completed a sorting task 15 minutes and 5 days after viewing the target identity. Hit rates were similar in the no variability and high variability conditions, and higher than in the low variability condition. False alarms were rare, and similar across conditions. Recognition accuracy did not decrease over 5 days. The results suggest two factors influenced face learning in this task: 1) variability in context; and, 2) opportunity to attend to the target's invariant facial features.

Keywords: variability, face learning, face recognition

Acknowledgments

Thank you to Dr. Chris Oriet for being an exceptional supervisor. Your mentorship and guidance has allowed me to grow academically and accomplish more than I ever thought possible. I will be forever grateful for this experience and the investment you have made in my life.

Dedication

I would like to express my gratitude to my family and friends for their unwavering support throughout my undergraduate studies. Your unending love and motivation is what has allowed me to be successful in not only academics, but in every area of my life.

The Role of Within-Person and Context Variability on Face Learning

Statistical Summary Representations (SSRs) have been shown to compress a large amount of detail into summary representations that encode statistical properties of sets such as mean size or variability in colour of visual items in memory (Ariely, 2001). Most of this research uses simple stimuli, such as different sized circles or oriented line segments, thus there is little research exploring whether SSRs also contribute to the processing of more complex stimuli. de Fockert and Wolfenstein (2009) investigated whether SSRs play a role in face learning in the same way that they have been repeatedly shown to play a role in the processing of sets of simple stimuli like circles and lines. The present study expanded on this work by investigating whether higher variability in the face presented would lead to more durable face learning.

de Fockert and Wolfenstein (2009) presented participants with four faces of different identities, followed by a test face. The test faces presented could be one of four types: 1) a matching member of the preceding set (matching member), 2) a member of another set (nonmatching member), 3) the morphed average identity of the preceding set (matching morph), or 4) the morphed average identity of a different set (nonmatching morph). Participants were then asked to indicate whether the test face had been present in the set of four faces previously viewed. The researchers predicted that if the identities of the set of four faces were represented as an average in memory, then there would be a high false recognition rate for test faces that were a morphed average of the identities shown in the preceding set. The results of the study showed that participants identified matching member test faces as present in the preceding set more frequently than test faces that were not of the matching member type. Morphed test faces also received present responses more than member test faces, suggesting that participants had a response bias for morphed faces. Despite this, test faces of the matching morph type received more present responses than test faces of the nonmatching morph type. This suggests that observers can extract the mean identity from a set of faces with different identities. It was the first study to show evidence that SSRs can be computed not only for simple stimuli but also for faces of different individuals (de Fockert & Wolfenstein, 2009).

De Fockert and Wolfenstein's (2009) work demonstrates that SSRs can be computed over multiple identities, but did not explore whether they are also used to summarize multiple views of the same identity. Kramer, Ritchie, and Burton (2015) directly investigated the contribution of within-person variability in familiarizing individuals with a target image. In previous experiments, faces were presented either simultaneously (de Fockert & Wolfenstein, 2009; Neumann, Schweinberger, & Burton, 2013), or sequentially (Haberman & Whitney, 2009; Leib, Fischer, Liu, Qiu, Robertson, & Whitney, 2014). In this experiment, the researchers presented faces using both methods. This experiment also differed from previous studies by using ambient colour images (Jenkins, White, Van Montfort, & Burton, 2011), which reflect real world images and would require a high level of encoding. The researchers also chose to present familiar faces (Hollywood celebrities) in addition to unfamiliar faces (Australian celebrities) to participants. This experiment replicated the method used by de Fockert and Wolfenstein (2009), but instead of showing four different identities, four photos of only one identity were shown per trial. The results suggest that participants are equally likely to say that they have seen the average of a set of photos of the same individual as they are to say that they have seen a photo that was actually shown in the display set (Kramer, et al., 2015). The results were nearly identical for both familiar and unfamiliar faces. For familiar faces, subjects were at chance at discriminating unseen exemplars and averages. However, for familiar faces subjects were about 70% accurate in discriminating unseen exemplars and averages, with no difference in performance for these two image types. Participants showed good memory for exemplars with both sequential and simultaneous presentation, with no difference between these conditions. This led to good performance when the test image was actually in the display set, but poor performance when the test image was similar to the average of the images shown or to the average of unseen photos of the same individual. The results showed that an average representation of the four faces presented was encoded in memory, and that it appears that subjects compare the test image to this average representation rather than to the faces that were presented.

Variability

Increased variability in the faces to be learned in an experiment might facilitate learning by allowing for a more descriptive average representation of the individual's identity to be established. Much of the research on face learning manipulates systematic variability by using images that vary in pose and illumination (Liu, Bhiuyan, Ward, & Sui, 2009; Longmore, Liu, &Young, 2008). Ritchie and Burton (2017) investigated unsystematic variability, the ways in which one face can vary that differ from how another face can vary (Burton, 2013; Jenkins et al., 2011). The researchers hypothesized that learning a new face would be better supported by exposure to high unsystematic variability images than low unsystematic variability images. In the first experiment the researchers used a face/name association technique. During the learning phase, participants were shown both low and high variability photographs that were paired with that individual’s name. For the test phase, the researchers used a speeded name verification task. Participants were shown a name, followed by a new photograph, and were then asked to indicate whether the photograph matched the name shown. For half of the trials, the face in the photograph shown matched the name presented. The results showed that participants were more accurate at matching names to identities they had learned with high variability than those they had learned with low variability. The researchers noted that this advantage may be due to the similarities between the test images and some of the high variability learning photographs. To show that the advantage was not due to this similarity, a second procedure was added that was identical to the first, except in the testing phase the researchers used previously unseen low variability photographs. This was done to see if the similarity between the test images and the high variability learning images was the explanation for the results in the first experiment. If this was true, it was expected that more accurate responses would be given to identities learned with low variability than with high variability when tested using low variability images. The results showed that participants were no more accurate in identifying identities they had learned with low variability than they were with identities learned with high variability, although correct responses were faster for photographs shown with low variability. What the researchers noted as surprising was that even though the test images and the high variability learning images were quite different, participants were no worse in identifying the individuals in the pictures than those with the low variability learning images. The research showed that high unsystematic variability photos allowed for a better established representation of the individual to be formed, which aided recognition later on.

To further assess the effect of variability on face learning, Ritchie and Burton (2017) conducted a second experiment that used a different dependent variable. Participants learned identities with either low or high variability, and then completed a face matching task which used new images. The researchers expected to find more accurate face matching for identities learned with high variability. During the face matching test phase, the participants were presented with two images and asked to indicate whether the images were of the same individual or of two different individuals. The participants were also presented with new identities in order to establish a baseline measure for face matching performance. The results showed that face learning with high unsystematic variability images led to improved face matching. Ritchie and Burton (2017) showed that learning from high variability photographs produces better learning than learning from low variability photographs. The findings also showed that learning faces with high unsystematic variability leads to better face/name matching, and face matching, than learning with low variability. Seeing photographs of an individual's face with high unsystematic variability allows people to establish a better representation of that individual over time.

Baker, Laurence, and Mondloch (2017) also investigated how the recognition of a newly encountered face makes the transition from unfamiliar to familiar. Participants first watched a 10 minute video of a woman reading a story. The woman was filmed on either a single day (low variability), or over three days (high variability). To increase variability, the woman's appearance and filming conditions varied in the video filmed over three days. Adults and children performed two training tasks. First, they completed a child-friendly sorting task in one of the three following conditions: no training, low variability training, or high variability training. All participants then completed a training task. The participants were then presented with a toy house with a picture of the woman reading the story (target image) displayed on it. Participants then were instructed to put all the photos of the target image in the house, while keeping any other pictures out. The researchers hypothesized that 1) performance would improve with age; 2) children in the training conditions would perform more accurately than children in the no training control group, and 3) children in the low variability condition would show less of a benefit of training than children in the high variability condition. Their results suggest that children are less able than adults to form a familiar representation of a newly encountered face with low variability in appearance. The researchers also found no evidence that the effect of training was moderated by age. Although comparing the groups directly was not a goal of the study, children and adults showed different patterns in their results. After watching a high variability video both adults and children showed evidence of becoming familiar with a new identity, but after watching a low variability video, only adults showed evidence of learning. Moreover, adults in the training conditions performed better than those in the no training control condition. However, training with videos with high variability in the actor's appearance yielded only a small benefit in accuracy relative to training with low variability videos. This was surprising due to evidence in previous studies that increasing variability facilitates learning (Ritchie & Burton, 2017).

Purpose

The results of Baker et al. (2017) are inconsistent with the results of previous research on the role of variability on face learning. Previous research has supported the hypothesis that increased unsystematic variability in the face viewed will lead to a more representative average stored in long term memory (Kramer et al., 2015; Ritchie & Burton, 2017). Based on this hypothesis, one would think that the greater the variability in the face viewed, the more accurate recognition would be later on, but the results of Baker et al. (2017) suggest that this may not be the case. Previous research has only tested participants a few minutes after first viewing a new face, but in the real world, the time between when we first encounter a face and when we need to be able to identify it can be much longer. Increasing unsystematic variability, then, might help to establish a more durable representation of a face. In the present study, we replicated the method used by Baker et al., but focused on matching an image from memory instead of an actual target image that was displayed. Participants were randomly assigned to one of three groups: 1) no variability (still image showing one viewpoint with invariant context), 2) low unsystematic variability (video showing multiple viewpoints with invariant context), or 3) high unsystematic variability (video showing multiple viewpoints and multiple contexts). Participants then completed a sorting task 15 minutes after viewing the target identity, and then again five days after viewing the target identity. We hypothesized that those in the low and high variability groups will show greater accuracy in recognizing the target identity than those in the no variability group, and that those in the high variability group will show even greater accuracy in recognizing the target identity. We also hypothesized that this process would be counterintuitive to participants, and that those in the low variability group would have a higher confidence in being able to recognize the target image than those in the high variability group.

Method

Participants

88 subjects, consisting of 60 females and 28 males (M = 22.36 years), from the University of Regina Department of Psychology Pool of Research Participants completed this experiment. All participants were unfamiliar with the identities used in the experiment. Participants were excluded from the analyses if they misidentified the target as someone they know. Before the start of the experiment, written consent was obtained from all participants. Following the completion of the experiment all participants were compensated for their time with two psychology course participation credits.

Measures

The target identity for the present study was either Javiera Mena (female) or Sonu Nigam (male). Images and videos of each target identity were taken from Google and Youtube. The no variability images of the targets were obtained from a Google search. Both the low and high variability videos were obtained from Youtube. The low variability videos show little to no changes in the target identity's appearance. The high variability videos show many changes to both the target identity's appearance and the background the target identity is presented in. Distractor identities (identities similar but not matching the target identity) were also taken from Google. The sorting task contained 32 cards, including: 16 novel images of the target identity (male or female) and 4 images of 4 similar looking distractors (similar age, hair color, face shape). These cards were 1 inch by 2 inches and the images were scaled to clearly display the faces of the targets and distractors. Two sets of cards were used so that participants were tested with different images at both 15 minutes and 5 days. The distractor task nature video was a 15 minute portion of a National Geographic Documentary from Youtube, that showed no human faces. To test the participants' confidence in their performance, two self-report questions asked participants to indicate how confident they were, on a scale of one to seven (with 1 being *not confident at all I will remember* and 7 being *certain I will remember)*, that they would be able to recognize the target image after 15 minutes and after 5 days.

Procedure

As illustrated in Figure 1, the present study closely replicated the method used by Baker et al. (2017), but with a focus on matching an image from memory instead of an actual target image that is displayed. Before beginning the experiment, participants were shown an array of 10 images, including the target image, and asked to indicate any that they recognize. Any participant that indicated that they recognize a target image did not continue in the experiment. Participants were randomly assigned to one of four conditions (high variability training, low variability training, still image training, and control) and within each condition, they were assigned to one of two target identities (male or female).

Control condition. The participants in the control condition were given 32 cards on which the test stimuli, including 4 distractor identities, were presented. The participants were asked to create a pile of cards for each different identity they thought was in the deck of cards. Participants then returned to the lab five days after the initial date and completed the sorting task a second time.

Experimental conditions. All participants in the still image condition were informed that they would be viewing a still image of a man or woman (target image) for one minute. All participants in the high and low variability training conditions were informed that they would be watching a one minute video of a man or woman (target image). After completing the training task, participants were given a short survey that asked how well they think they would be able to recognize the target image after 15 minutes and after five days. Participants were then asked to watch a short nature video (distractor task) for 15 minutes. The participants were then given 32 cards on which the test stimuli, including 4 distractor identities, were presented. Participants were asked to indicate if the test images presented belonged to the target identity they had previously encountered during the training phase. The participants were asked to create one pile of cards that matched and one pile of cards that did not match the target identity. Participants then returned to the lab five days after the initial test and completed the facial recognition sorting task a second time.

Results

Control Condition

When asked to identify how many different individuals were shown in the stack of cards, those in the control condition created between one and twenty-one different piles ($M = 11.0$), and of those piles one to twelve ($M = 4.79$) contained the target identity. This suggested that the sorting task was not trivial, and participants in the experimental condition would not be able to perform well on the sorting task without having seen the target, and would instead have to rely on their memory of the target identity in order to sort the cards.

Experimental Conditions

Sorting Task. To test whether increased variability leads to more durable face learning a 2 (Delay: 15 minutes vs. 5 days) x 2 (Stimuli: Female vs. Male) x 3 (Condition: control, low variability training, high variability training) mixed model analysis of variance (ANOVA), with recognition accuracy (d') as the dependent variable was conducted. The results, summarized in Figure 2, show that there was a significant within-subject main effect for Time, $F(1, 60) = 8.62$, $p < .005$, $\eta_p^2 = .125$, and a significant between-subject main effect for Condition, $F(2, 60) = 3.65$, $p < .032$, $\eta_p^2 = .109$. There was no significant interaction of Time by Condition, $F(2, 60) = 1.35$, $p > .267$, $\eta_p^2 = .043$, Time by Stimuli, $F(1, 60) = .022$, $p > .883$, $\eta_p^2 = .001$, or Condition by Stimuli, $F(2, 60) = .333$, $p >.718$, $\eta_p^2 = .011$. Fisher's LSD post-hoc tests ($p < .05$) revealed that both the high variability ($M = 2.39$) and still image conditions ($M = 2.23$) performed more accurately than the low variability condition ($M = 1.42$) after 5 days.

Hit Rate. To test whether increased variability leads to more accurate discrimination between the target and distractors, a 2 (Delay: 15 minutes vs. 5 days) x 3 (Condition: control, low variability training, high variability training) mixed model ANOVA, with hit rate as the dependent variable was conducted. The results showed that there was a significant within-subject main effect for Time, $F(1, 63) = 14.0$, $p < .001$, $\eta_p^2 = .182$, and a significant between-subject main effect for Condition, $F(1, 63) = 5.94$, $p < .004$, $\eta_p^2 = .159$. There was also a significant interaction of Time by Condition, $F(2, 63) = .5.25$, $p < .008$, $\eta_p^2 = .143$. To examine if there were any differences in hit rate between the two time delays, a one-way ANOVA was conducted for both the 15 minute and 5 day time delay. The analysis of hit rate showed that there was a significant difference between the means at both the 15 minute, $F(2, 63) = 6.62$, $p < .002$, $\eta_p^2 = .174$, and 5 day time delays, $F(2, 63) = 5.24$, $p < .008$, $\eta_p^2 = .143$. Fisher's LSD post-hoc tests ($p < .05$) revealed that the high variability condition ($M = .644$) performed more accurately than the low variability condition ($M = .336$) after 15 minutes, and both the high variability ($M = .646$) and still image conditions ($M = .681$) performed more accurately than the low variability condition ($M = .409$) at 5 days.

False Alarm Rate. To further test whether increased variability leads to more accurate discrimination between the target and distractors, a 2 (Delay: 15 minutes vs. 5 days) x 3 (Condition: control, low variability training, high variability training) mixed model ANOVA, with false alarm rate as the dependent variable was also conducted. The results showed that there was not a significant within-subject main effect for Time, $F(1, 63) = 8.27$, $p > .367$, $\eta_p^2 = .013$, and not a significant between-subject main effect for Condition, $F(2, 63) = .356$, $p > .702$, $\eta_p^2 = .011$. There was also not a significant interaction of Time by Condition, $F(2, 63) = 1.08$, $p > .347$, $\eta_p^2 = .033$.

Confidence Ratings. To analyze participants' ability to intuit their performance, a 2 (Delay: 15 minutes vs. 5 days) x 3 (Condition: control, low variability training, high variability training) mixed model ANOVA, with confidence ratings as the dependent variable was also conducted. The results, summarized in Figure 3, showed that there was a significant within-subject main effect for Time, $F(1, 58) = 60.8$, $p < .001$, $\eta_p^2 = .512$, and no significant between-subject main effect for Condition, $F(2, 58) = 1.29$, $p > .284$, $\eta_p^2 = .042$. There was also no significant interaction of Time by Condition, $F(2, 58) = .438$, $p > .647$, $\eta_p^2 = .015$. Participants correctly predicted that their accuracy would decrease from the 15 minute to the 5 day time delay, but the confidence ratings provided by subjects in the low variability condition were not lower than those provided by participants in the high variability or still image conditions, even though their actual performance was worse at both the 15 minute and 5 day time delay.

Discussion

The results of this study did not support the hypothesis that higher variability would lead to more durable face learning. The results also did not support the results previously found by either Ritchie and Burton (2017) or Baker et al. (2017). The main effect of condition showed that both those in the high variability and still image conditions performed more accurately than those in the low variability condition. To further investigate if higher variability would lead to more durable face learning, an analysis of both hit rate and false alarm rate was conducted. If higher variability indeed led to more durable face learning, we would expect that participants in the high variability condition would have a higher hit rate and lower false alarm rate than those in the still image condition because they would be able to more accurately discriminate between the target and distractors. The analysis of hit rate yielded significant main effects for time, condition, and the interaction between time and condition, but there were no significant effects in the analysis of false alarm rate. Further analysis of hit rate showed that accuracy was higher in both the high variability and still image conditions than those in the low variability conditions at both time delays.

Although it appears that there was also a main effect of time, which would suggest that participants showed improvement at the 5 day delay, this result should be interpreted with caution because the cards used in the sorting task were not counterbalanced across the two time delays. Based on performance in the Control condition, it would appear that the sets of cards used at the 5 day delay for both types of stimuli were easier for participants to sort than those used at the 15 minute delay. The analysis of participants' confidence ratings showed that participants in the low variability condition did not rate their confidence as lower than those in the high variability or still image conditions, even though their actual performance was worse at both time delays. This suggests that those in the lower variability condition did not intuit that they would perform poorly based on the training stimuli they viewed.

Previous research had suggested that increasing unsystematic variability facilitates learning (Ritchie & Burton, 2017), however, results of Baker et al. (2017) showed that training with videos with high variability yielded only a small benefit in accuracy relative to training with low variability videos. The method used in the two studies differed in both the training and test phases. Participants in Ritchie and Burton's (2017) study were trained with low or high variability photos and tested using either a speeded name verification task or a matching task, while those in the Baker et al. (2017) study trained using a low or high variability video and were tested using a matching task. The present study closely replicated the method used by Baker et al., but with a focus on matching a face from memory instead of a photograph that was actually presented to them. It is possible that differences in the methodology of the present study from these previous studies could account for the different outcomes observed across the three studies.

Although the results of the present study did not support the findings of previous research, this project has suggested that there may be 2 factors that influenced face learning in this study: a) attention to invariant characteristics, and b) high variability in the context in which the stimuli are presented. The results of this study suggest that those in the still image condition were able to focus on the invariant characteristics of the target and that those in the high variability condition showed a benefit from the numerous backgrounds that the target was presented in, which both led to better retention of the newly encountered face.

Limitations and Future Directions

As with any research, the present study had limitations due to aspects of the research design. The sample of participants may have been unrepresentative, as they were taken from the University of Regina Department of Psychology Pool of Research Participants. This sample was primarily Caucasian, female, and has received at least some post-secondary schooling. However, there is no a priori reason to assume that this subset of the population would differ in their learning of faces as a function of increasing variability than the population as a whole. Another possible limitation of the study could be the target identities themselves. The target identities used were both famous musicians from another country. There may be characteristics about famous individuals in general that make them more memorable to viewers. The photographs and videos of the target identities used as stimuli were mostly professional shots. These shots do not accurately mirror real life faces, but were used because the professional shots of famous individuals are of high quality and easily accessible online.

Another limitation of the present study was that the distractor identities were chosen based on a Google search of the target's face, which often brought up individuals that Google had determined looked similar to the target. This procedure lacked a measure of participants' perceived similarity between the target and distractors. Future research should collect similarity ratings prior to conducting the study, and then ensure that the stimuli used at both time delays have the same average similarity to the target.

The target identities were also of a different ethnicity than the sample population. The female target identity was Chilean and the male target identity was Indian. These identities were chosen specifically so that participants would not have any previous knowledge of who the target identities are and would ensure that recognition is due to the learning occurring in the study. It was beyond the scope of this study to examine the effect of the above limitations, but future research should investigate differences in learning of faces of familiar and unfamiliar ethnicities. Previous research has shown that individuals are better at recognizing those that are of the same ethnicity, known as the own-race bias (ORB: Johnson & Fredrickson, 2005). As previously mentioned, in the present study, both targets were of a different ethnicity than the majority of the sample tested. Thus, the accuracy in recognizing a target may be affected by the ethnicity of the participant. If the target was of the same ethnicity as the participant, it may lead to better recognition because the target is more familiar to the participant. Recent research by Kramer, Young, and Burton (2018) suggests that the binary categorization of familiar and unfamiliar faces is not sufficient for the study of face learning. They argue that there is a spectrum of faces that would be familiar to participants, which ranges from family members to interactions with media. Future research should be conducted to investigate the complex categorization of familiar faces and the effect this has on face learning.

Conclusion

This project adds to the growing research base that shows how SSRs contribute to face learning. The goal of this study was to clarify the mixed findings in previous research which investigated whether increased variability led to better face learning, and in particular, whether increased variability facilitates longer-term retention of a learned identity. These results suggest that although high variability does show an advantage over low variability, it may not be the only mechanism that contributes to learning a new face.

References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162.

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19-30.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*(8), 1467–1485.

de Fockert, J, & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces, *The Quarterly Journal of Experimental Psychology*, *62*(9), 1716-1722. doi: 10.1080/17470210902811249

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 718–734.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313–323.

Johnson, K. J., & Fredrickson, B. L. (2005). "We All Look the Same to Me": Positive Emotions Eliminate the Own-Race Bias in Face Recognition. *Psychological Science*, *16*(11), 875–881. http://doi.org/10.1111/j.1467-9280.2005.01631.x

Kramer, R. S., Ritchie, K. L., Burton, A. M. (2015) Viewers extract the mean from images of the same person: A route to face learning. *Journal of Vision 15*(1).

Kramer, R. S., Young, A. W., Burton, A. M. (2018) . "Understanding Face Familiarity." *Cognition*. *172*, 46–58. doi:10.1016/j.cognition.2017.12.005.

Lamont, A. C., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. *Memory & Cognition*, 33, 1017-1024.

Leib, A. Y., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision*, *14*(8), 1–13. doi:10.1167/14.8.26

Liu, C. H., Bhuiyan, A., Ward, J., & Sui, J. (2009). Transfer between pose and illumination training in face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 939–947.

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 77–100.

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*, 56–63.

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *The Quarterly Journal of Experimental Psychology*, *70*(5), 897-905.

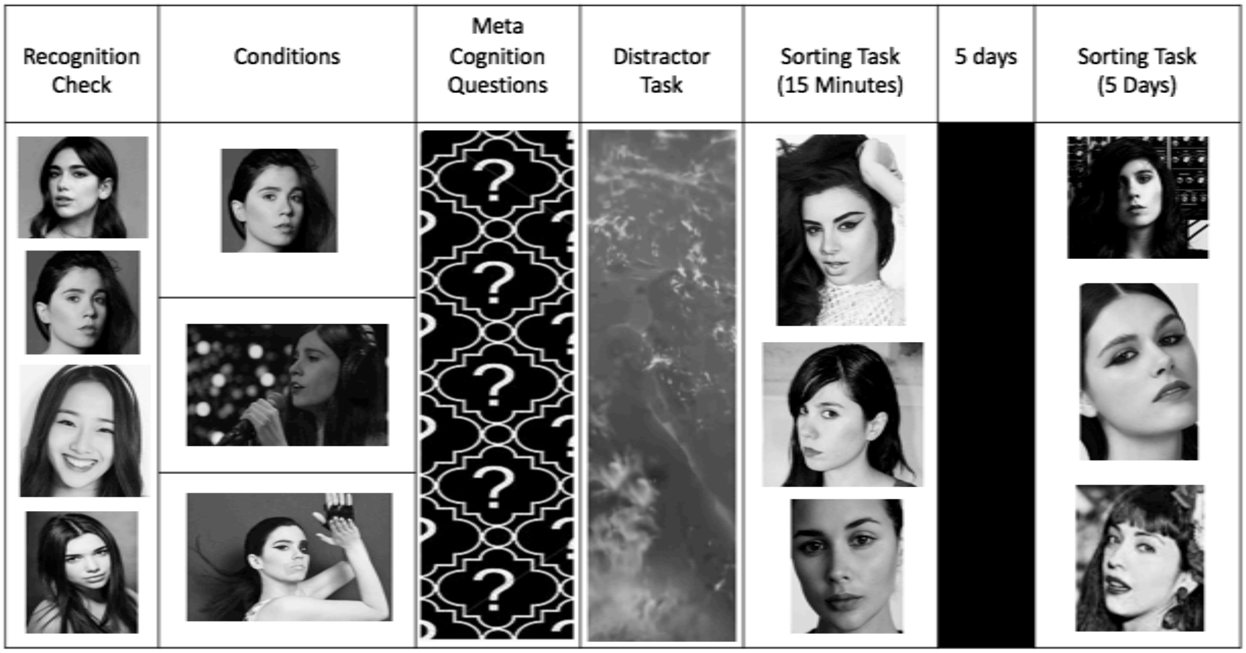

Figure 1. Procedure for experimental conditions. This figure illustrates the method used in the present study and closely follows the method used by Baker, Laurence, and Mondloch (2017).

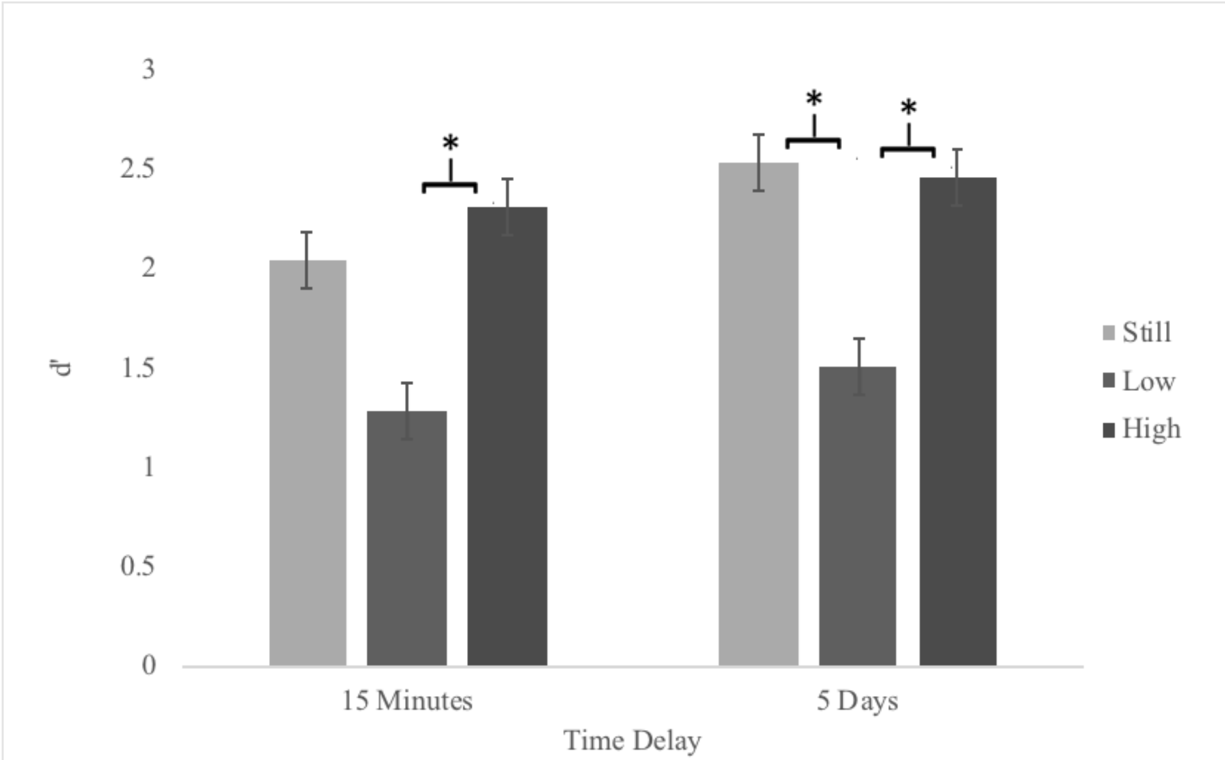

Figure 2. Results of Sorting Task Measuring Accuracy (d'). This graph illustrates the results of a mixed model ANOVA showing that those in the high variability and still image conditions performed more accurately than those in the low variability condition.

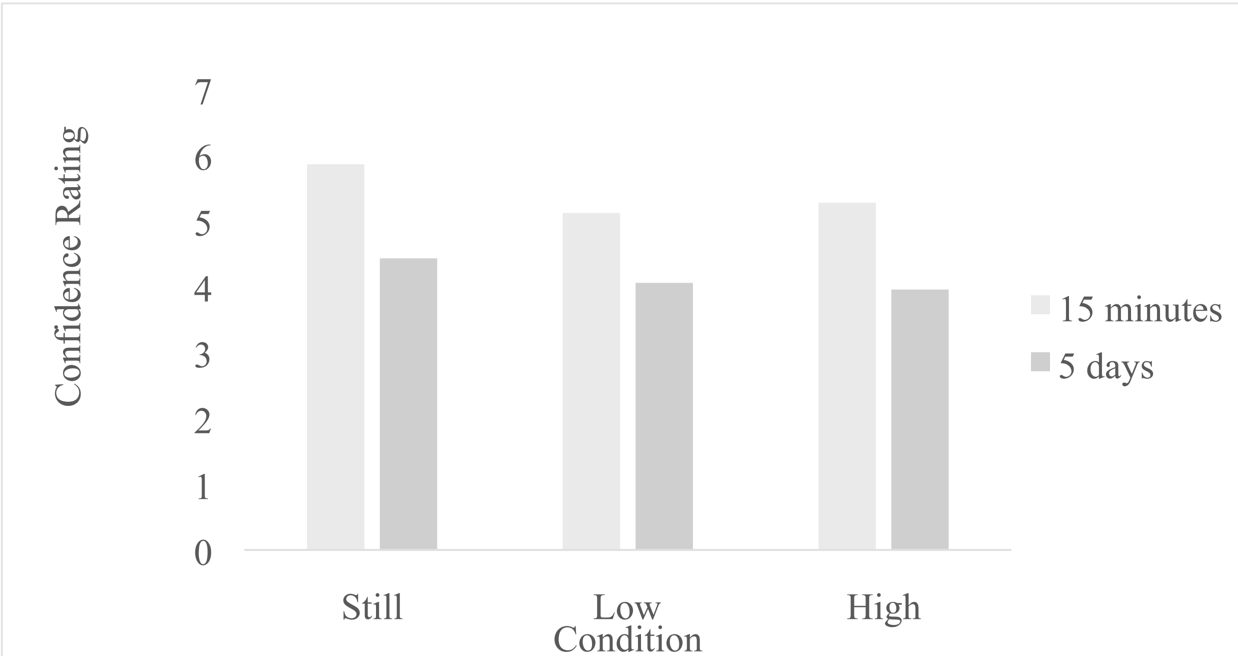

Figure 3. Results of Confidence Ratings. This graph illustrates the confidence ratings provided by subjects in the low variability condition were not lower than those provided by participants in the high variability or still image conditions, even though their actual performance was worse at both the 15 minute and 5 day time delay.