The Temporal Cheerleader Effect: Attractiveness Judgments Depend on Surrounding Faces Through Time

Alexia Toskos Dils (alexia.toskos@purchase.edu)
Scott Niedopytalski (scott.niedopytalski@purchase.edu)
Jeffrey Arroyo (jeffrey.arroyo@purchase.edu)
Stephen J. Flusberg (stephen.flusberg@purchase.edu)
SUNY Purchase College, Department of Psychology
735 Anderson Hill Road, Purchase, NY 10577, USA

Abstract

Previous research has found that people are seen as more attractive when they appear in a group rather than in isolation. The present study asks whether faces that surround us in *time* also affect how attractive we appear to be. Participants rated the attractiveness of famous female faces presented in a sequence of three and in isolation. We found that people do integrate information about attractiveness over time, but that temporal context has the opposite effect of static context. People perceived faces as *less* attractive in a series than in isolation. We also varied the attractiveness of surrounding faces in order to examine how the serial position of contextual information might figure into people's judgments. We found that faces presented earlier in the sequence figured more heavily into people's judgment than did faces presented later in the sequence. These findings highlight the role of temporal context in perceptions of attractiveness.

Keywords: face perception, attractiveness, serial position effects, ensemble coding, cheerleader effect

Introduction

In a 5th season episode of the American comedy television series The Office, the employees of Dunder Mifflin paper company spend an entire day debating whether actress Hilary Swank is "hot" or not. The office workers are torn on the issue, battle lines are drawn, and emotions get heated. Though in part a satirical referendum on the public's dark obsession with – and objectification of – celebrity, the plotline of this episode raises an important question: What factors influence perceptions of attractiveness?

The intrinsic features of individual faces certainly contribute to the perception of attractiveness, for both evolutionary and cultural reasons. Female faces, for example, are rated as more attractive the more sexually dimorphic and prototypically "female" they are (Valenzano, Mennucci, Tartarelli, & Cellerino, 2006), and the more symmetrical they are (Perret et al., 1999). However, certain situational factors such as amount of exposure to a face (Rashidi, Pazhoohi & Hosseinchari, 2012) as well as the perceived market value of the person making the judgment (Morgan & Kisley, 2014) can also impact attractiveness judgments.

Researchers have also examined whether contextual factors can influence attractiveness ratings. Recent work, for example, has uncovered a so-called "cheerleader effect" in which people are rated as more attractive when they appear in a group than when they appear in isolation (Walker & Vul, 2013). Walker and Vul explain these findings as a sort of perceptual averaging phenomenon. The idea is that people spontaneously extract an ensemble code when viewing a group of faces, and because average faces are seen as highly attractive (Langlois, Musseman, & Roggman, 1994), attractiveness ratings for the faces that contributed to the ensemble receive a boost. This suggests that perceptions of attractiveness are in part constructed online, in the moments we experience another person's face.

Often the faces we encounter in a crowded place are processed serially, rather than all at once. For example, faces come in and out of sight as we walk down the street, scan a room, or swipe through profile pictures on social media websites. Interestingly, evidence from studies of object perception suggests that the visual system is capable of constructing average representations over *time* in addition to space. Albrecht and Scholl (2010) found that people's estimates of the average diameter of a growing or shrinking disc depended on which part of the disc received the most screen time. Participants overestimated the average when frames on the larger end of the spectrum hung on the screen longer than did frames on the smaller end of the spectrum, and vice versa.

Is information from faces spontaneously integrated over time in a similar way? The present study asks whether the cheerleader effect extends to faces that appear near one another in time in addition to space. Importantly, the study was also designed to address whether serial position influences how information about attractiveness is integrated over time. Do all faces in a sequence figure equally into the ensemble code, or is the average representation that people extract weighted more heavily by faces appearing early or late in the series?

Though research suggests the visual system computes ensemble codes of information presented close in space and time, the data is often inconsistent with a simple averaging account. For example, studies of the perception of serially presented lines have reported a recency effect in people's judgments: Estimates of line length were biased toward lines that appeared toward the end of the sequence (Weiss & Anderson, 1969). And while Walker and Vul (2013) concluded that an averaging effect best captured the data they observed in their work on the cheerleader effect, not all of their findings are consistent with this interpretation. For example, if the cheerleader effect results from simple averaging, then the more faces that contribute to the ensemble code, the more attractive the resulting average should be (i.e., increasing the number of faces in a group photograph should strengthen the cheerleader effect). Walker and Vul (2013, Experiment 4) tested this hypothesis with set sizes ranging from two to 16. Although they found a cheerleader effect within each set size, the magnitude of the effect did not increase with the number of faces in the set.

In light of these findings, the present study was designed to differentiate between different possible accounts of the *temporal* cheerleader effect we investigated. On the one hand, it is possible that faces appearing at the end of a sequence bias the perceived attractiveness of faces earlier in the sequence (i.e., a recency effect). This would be consistent with the work of Weiss and Anderson (1969) on line perception. However, it is also possible that the first face might influence the perceived attractiveness of faces appearing later in the sequence. There are at least two reasons we might expect such a primacy effect. First, there might be a *contrast effect* (Kenrick & Gutierres, 1980; Pegors, Mattar, Bryan, & Epstein, 2015), such that seeing a highly attractive face may make subsequent faces seem less attractive by contrast, and vice versa. Second, there might be an *anchoring effect*, such that subsequent faces appear to have a similar attractiveness level as a previously presented face (Pegors, Mattar, Bryan, & Epstein, 2015; Taubert, Van Der Burg, Alais, 2016; Tversky & Kahneman, 1974).

Both contrast and anchoring effects have been observed in recent studies examining the effects of the attractiveness of the previous face on online judgments of serially presented faces (Pegors et al. 2015; Taubert et al., 2016). Pegors and colleagues suggest that the contrast effect derives from perceptual components of the judgment, whereas the anchoring effect derives from a bias to respond in the same way as in the previous trial. The present study extends this work to offline judgments of facial attractiveness. Does the temporal context in which a face appears affect later memory of how attractive that face was? Offline judgments further allow us to examine any serial position effects more fully. Specifically, this design element allows us to test whether faces that follow us in time can retroactively meddle with how attractive we are judged to be.

In order to (a) determine whether or not an offline, temporal cheerleader effect exists, and (b) understand the mechanisms underlying such an effect, participants in the present study rated the attractiveness of a variety of famous female faces. On each trial, faces were presented either in isolation or in a series of three faces. Afterward, participants were cued by the name of each celebrity they saw and asked to rate how attractive the person looked in the photograph. Celebrity faces were used (as opposed to non-famous faces) to allow for offline attractiveness ratings to be collected after all faces disappeared from view. The celebrity's name uniquely picked out which face participants were being asked to rate.

The stimulus set included two versions of each celebrity: an "attractive" version where the celebrity was photographed favorably, and an "unattractive" version where the celebrity was photographed uncharitably (see Figure 1 for example stimuli). On trials where faces were presented in a series, the middle face was always an attractive face and was either preceded by or followed by an unattractive face. Therefore, in the *unattractive first* condition, participants first saw an unattractive face followed by two attractive faces. In the *unattractive last* condition, participants saw two attractive faces followed by an unattractive face.

To the extent that people do integrate information about attractiveness over time, offline judgments of a given face should depend on whether the face was presented in isolation or in the middle position of a series of three faces. If the mechanism by which information is integrated is a contrast effect, then viewing an unattractive face in the first position of the sequence should cause the middle face to seem more attractive than it does in isolation, and vice versa. Alternatively, if the mechanism is anchoring, then seeing an unattractive face in the first position of the sequence should make subsequent faces seem less attractive than they do in isolation, and vice versa. Finally, if the mechanism is simple averaging, then all faces should appear more attractive when they are presented in a sequence, since research suggests that averaged faces appear more attractive (Langlois, Musseman, & Roggman, 1994). However, since there are only three faces being averaged together here and one was specifically chosen to appear highly unattractive, it could be the case that the average of this small set would actually appear less attractive. What's more, these effects may depend on where the unattractive face appears in the sequence. If the averaging effect is subject to primacy effects, then the first face in the sequence will figure more heavily into the average. However, if the averaging effect is subject to recency effects, then later faces in the sequence will figure more heavily into the average. This study will help to rule out some of these possibilities.

Experiment

Methods

Participants 50 "master-level" participants were requested from Amazon's Mechanical Turk worker pool for this study. 68 people actually followed the link to the experiment and began the study, but 17 of those people did not complete the task. The remaining 51 participants were included in the study. There were 27 males and 23 females in the sample, with a mean age of 41.20 (*SD* = 12.67). One participant declined to provide demographic information. Each participant received $1.50 for their time spent on the study.

Stimuli and Procedure Participants rated the attractiveness of faces in three within-subjects conditions: *isolation* (a face shown by itself), *unattractive first* (an unattractive face followed by two attractive faces in sequence), *unattractive* *last* (two attractive faces followed by an unattractive face in sequence).

Figure 1. Example stimuli used in this study. On the left is the “unattractive” photograph of Debra Messing that participants saw, and on the right is the “attractive” photograph of the same celebrity.

Photographs of 75 unique female celebrities made up the stimulus set for this study. Two photographs were collected for each celebrity: an “attractive” version in which the person was made up for an event, and an “unattractive” version in which the celebrity was captured poorly (i.e., photographs found in tabloids claiming to have “shocking” pictures of the person; see Figure 1 for example stimuli). Therefore, the final stimulus set consisted of 150 celebrity photographs. Each participant saw a subset of 75 faces sampled from this full stimulus set (50 attractive faces and 25 unattractive faces). Participants saw either the attractive version or the unattractive version of each unique celebrity, and never both. Whether a celebrity was presented as attractive or unattractive was counterbalanced across participants.

The study was conducted using Qualtrics survey building software. On each trial, participants saw female celebrity faces presented either in isolation or in a series of three faces. Each face was presented along with the celebrity’s name and remained on the screen for one second. Afterward, participants were instructed to rate the attractiveness of each face they had just seen, cued only by the celebrity’s name. Participants entered their responses using a continuous sliding scale without numbers on it. This was designed to prevent participants from tracking the specific ratings they assigned to faces throughout the study. The scale ranged from “very unattractive” to “very attractive”. Participants were allowed to take as much time as they needed to enter their ratings. The order in which participants were prompted to rate each celebrity was randomized on each trial. See Figure 2 for a schematic representation of the trial structure in this task.

Each participant completed 125 trials. 75 trials consisted of faces presented in isolation, and 50 trials consisted of faces presented in a series. Of the 50 trials in which the faces were presented in a group, 25 were *unattractive first* trials, and 25 were *unattractive last* trials. The middle face in the sequence was always attractive. Which face appeared in which position in the sequence was counterbalanced across participants. All trial types were interleaved, and trials were presented to participants in random order.

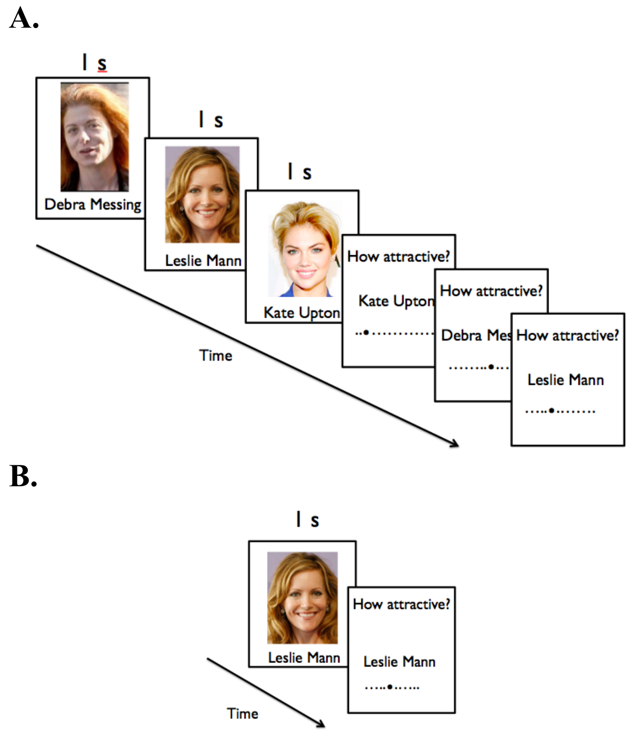

Figure 2. Trial structure for the conditions in which faces were presented in a sequence (A) and in isolation (B). Part A depicts an example *unattractive first* trial.

Participants rated each *attractive* photograph at two critical points in the study: once in isolation and once in the middle position of a series of three faces. We wanted to know whether attractiveness ratings for each photograph depended on this contextual manipulation, so we limited our analyses to the 50 attractive faces each person saw, as well as to ratings of the face presented only in the middle position on group trials.

At the end of the study, participants supplied information about their gender, race/ethnicity, and age. We also included a manipulation check designed by Oppenheimer, Meyvis, and Davidenko (2009) to ensure that all participants in our sample read task instructions carefully.

Results

All 51 participants in this study correctly answered the instructions manipulation check, so data from all participants were analyzed. One participant declined to provide demographic information and therefore was not included in analyses that examined effects of gender.

To confirm that our participants found the “attractive” faces to be more attractive than the “unattractive faces,” we compared participants’ raw attractiveness ratings between these two stimulus types. Indeed, participants rated the attractive faces more favorably on a scale from 1-100 (M = 70.91; SD = 13.12) than they did the unattractive faces, M = 34.46; SD = 16.46; $t(50) = 15.83$, $p < .001$.

For each participant, responses on analyzed trials were converted to z-scores. For each attractive face that a given subject saw, we subtracted the standardized rating of that face presented in isolation from the standardized rating of that same image shown in a group. Therefore, positive difference scores indicate that faces were rated as *more* attractive when presented in a group compared to isolation, and negative difference scores mean faces were rated as *less* attractive when presented in a group compared to isolation. Overall, participants rated celebrity faces as significantly *less* attractive when they were presented in a series than when they were presented in isolation ($M_{group\text{-}isolation}$ = -0.098, $SD_{group\text{-}isolation}$ = 0.16), $t(50) = -4.24$, $p < .05$[1].

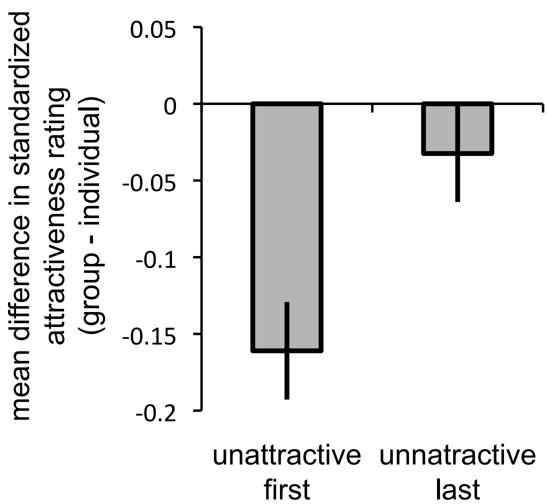

Figure 3: Shift in standardized attractiveness ratings for faces presented in a sequence from those same faces presented in isolation. Error bars represent 1 SEM.

The differences in z-scores were submitted to an ANOVA with one within-subjects factor (order of attractiveness: unattractive first vs. unattractive last) and one between-subjects factor (gender of participant: male vs. female). The ANOVA revealed a significant main effect of order of attractiveness on the attractiveness deficit for faces presented in a group relative to isolation. Specifically, the effect was stronger when participants saw an unattractive face first in the sequence (M = -0.16, SD = 0.23) than it was when participants saw an unattractive face last in the sequence (M = -0.03, SD = 0.23), $F(1, 48) = 9.78$, $p < .005$ (see Figure 3). These effects were similar in men and women: The attractiveness deficit did not depend on the gender of the participant ($F(1, 48) = 0.04$, $p > .50$), and there was also no interaction between gender and order of attractiveness, $F(1, 48) = 1.97$, $p > .05$.

Post-hoc analyses revealed that seeing an unattractive face at the beginning of the sequence caused the middle face to appear less attractive than it did in isolation ($M_{group\text{-}isolation}$ = -0.16, $SD_{group\text{-}isolation}$ = 0.23), $t(50) = -5.07$, $p < .001$. While the presence of an unattractive face at the end of a sequence produced a numerical reduction in attractiveness of the middle face relative to isolation, this shift did not reach statistical significance ($M_{group\text{-}isolation}$ = -0.03, $SD_{group\text{-}isolation}$ = 0.23), $t(50) = -1.03$, $p > .05$.

[1] Due to an error in setting up the study, one trial was lost in the unattractive last condition in one third of participants. Therefore, only 24 trials instead of 25 went into data analysis for those participants in that condition.

Discussion

The purpose of this study was to determine whether and how people integrate information from faces over time when judging the attractiveness of others. In our experiment, participants viewed a series of attractive and unattractive female celebrity faces in isolation and in a series of three images. They were then asked to rate the attractiveness of each face (cued only by their name) on a sliding scale from very unattractive to very attractive. The results indicate that people do integrate information over time when judging the attractiveness of faces. Contrary to the “cheerleader effect,” which suggests that faces are perceived as *more* attractive when they are presented in a group, the results from this experiment suggest that faces are perceived as *less* attractive when presented in a series than when they are presented in isolation. The results further show that this effect is stronger when an unattractive face is presented first in the series compared to when it is presented last.

There are a number of reasons why attractiveness ratings might be different for faces presented in a series compared to isolation. For example, seeing faces in a sequence may lead to a *contrast effect*, where seeing an attractive face causes a subsequent face to appear less attractive (and vice versa). Or there could be an *anchoring effect* where people rate subsequent faces to be similarly attractive to the initial face. Finally, there could be an *averaging effect* where all faces in the sequence contribute to an ensemble code that figures into attractiveness judgments for each face in the sequence. The results from the present study help to rule out one of these three possibilities: the contrast effect. Had our manipulation produced a contrast effect, then seeing an unattractive face first would have caused the middle face in the series to seem more attractive. Instead, our results showed the opposite pattern: viewing an unattractive face made the subsequent face appear to be less attractive than when it was viewed in isolation.

The data are partially consistent with an anchoring effect. An attractive anchor face should cause subsequent faces to seem more attractive than usual, and an unattractive anchor should cause subsequent faces to seem less attractive than usual. Indeed, unattractive faces at the beginning of the sequence caused subsequent faces to appear less attractive than usual, but the opposite pattern did not obtain for attractive faces presented early in the sequence. Therefore these data are only partly in line with the findings of Taubert et al. (2016).

Pegors and colleagues (2015) concluded from their study examining sequential attractiveness judgments that contrast effects result from stimulus (perceptual) bias, whereas anchoring effects result from response bias. The present findings are not consistent with this explanation as the study design itself rules out response bias as a possible contributor. On trials in which faces were presented in a group, participants saw all three faces before responding to any one of them. They then entered their responses in a different order than the one in which the faces were presented. We analyzed the data based on the stimulus order and not the response order, so the partial anchoring effect should not necessarily be interpreted as a result of response bias. Furthermore, the lack of a contrast effect is a departure from the findings of Pegors and colleagues, suggesting there may be a qualitative processing difference between online and offline judgments of serially presented faces that persists well beyond the presentation of the face.

The findings presented in this paper are perhaps most consistent with an averaging effect. By including an unattractive face in the sequence of three, we may have reduced the attractiveness of the ensemble code participants extracted from the series, which then biased subsequent ratings of individual faces in the sequence. Furthermore, it appears that the serial position of the unattractive face influenced how the average representation was constructed. Contrary to the findings of Weiss and Anderson (1969), our results revealed a primacy effect on perceptual averaging. Namely, unattractive faces presented at the beginning of the series figured more heavily into the average than did unattractive faces presented at the end of the series. However, because averaging faces together generally *increases* perceived attractiveness, this account depends on the possibility that averaging only a small number of faces together (three), one of which is especially unattractive, can sometimes reduce rating attractiveness (perhaps, for example, when the faces are of familiar people, as in the present study). Indeed, morphs containing familiar (celebrity) faces are rated as less attractive than are the component faces used to generate the morph (Halberstadt, Pecher, Zeelenberg, Ip Wai, & Winkielman, 2013). Therefore, using celebrity face may itself have lowered the attractiveness of the average face people constructed over time in the present study. Future work is required to confirm this possibility, however.

In order to preserve the design element in this study that no participant saw both the attractive and unattractive versions of the same celebrity, we did not include trials in which all faces in a sequence were attractive, nor did we include trials in which all faces were unattractive. Future work will examine these trial types so as to build a more complete picture of the way in which attractiveness in faces is integrated over time.

One ongoing study attempts to further differentiate between an anchoring and averaging account of our findings by replicating the present study with unattractive faces in the target (middle) position on group trials. If our findings reflect a cheerleader effect, then the results from this new study should mirror the results in the present study. Specifically, the middle face in the sequence would be rated as less attractive than a face presented in isolation, but this effect would only appear in the unattractive first condition. However, if our current findings reflect an anchoring effect, we should predict that the *unattractive last* condition in this ongoing study would produce the strongest anchoring effect – the condition in which an attractive face appears in the first position of the sequence. After all, this is the condition where the difference in attractiveness is largest between the first and middle faces. Therefore, an anchoring effect in this new version of the task predicts that the middle face would be rated as *more* attractive in a series than it is in isolation.

Our judgments about the attractiveness of others factor into a variety of important decisions we make every day. For example, whether a candidate would make a good fit for a job (Dipboye, Fromkin, & Wilback, 1975), or whether to ask a person out on a date can be influenced by how we perceive the attractiveness of that individual. This study highlights one way in which such important decisions might be influenced by the faces that happen to surround us in the moments leading up to them.

Acknowledgments

This research was supported by undergraduate research support awards from the School of Natural and Social Sciences at SUNY Purchase to SN and JA.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21*(4), 560-567.

Dipboye, R. L., Fromkin, H. L., & Wilback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic in evaluation of job applicant resumes. *Journal of Applied Psychology, 60*(1), 39-43. doi:10.1037/h0076352

Halberstadt, J., Pecher, D., Zeelenberg, R., Ip Wai, L., & Winkielman, P. (2013). Two faces of attractiveness: Making beauty in averageness appear and reverse. *Psychological Science, 24*(11), 2342-2346.

Kenrick, D. T., Gutierres, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology, 38*(1), 131-140.

Langlois, J. H., Musselman, L., Roggman, L. A. (1994). What is average and what is not average about attractive faces? *Psychological Science,5*(4), 214-220.

Morgan, L. K., & Kisley, M. A. (2014). The effects of facial attractiveness and perceiver's mate value on adaptive allocation of central processing resources. *Evolution and Human Behavior, 35*(2), 96-102. doi:10.1016/j.evolhumbehav.2013.11.002

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation check: Detecting satisficing to increase statistical power. *Journal of Experiment Social Psychology, 45*, 867-872. doi:10.1016/j.jesp.2009.03.009

Pegors, T., Mattar, M., Bryan, P., & Epstein, R. (2015). Simultaneous perceptual and response biases on sequential attractiveness judgments. *Journal of Experimental Psychology: General, 114*(3), 664-673. doi:doi:10.1037/xge0000069

Perrett, D. I., Burt, D. M., Penton-Voak, I., Lee, K. J., Rowland, D. A., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior, 20*, 295-307. doi:10.1016/S1090-5138(99)00014-8

Rashidi, M., Pazhoohi, F., & Hosseinchari, M. (2012). Effect of facial stimuli exposure time on evaluation of facial attractiveness. *Australian Journal of Psychology, 64*(3), 164-168. doi:10.1111/j.1742-9536.2011.00050.x

Taubert, J., Van Der Burg, E., Alais, D. (2016). Love at second sight: Sequential dependence of facial attractiveness in an online dating paradigm. *Scientific Reports, 6*, 22740. doi:10.1038/srep22740

Tversky, A., Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Valenzano, D. R., Mennucci, A., Tartarelli, G., & Cellerino, A. (2006). Shape analysis of female facial attractiveness. *Vision Research, 46*(8-9), 1282-1291. doi:10.1016/j.visres.2005.10.024

Walker, D., & Vul, E. (2013). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science, 25*(1), 230-235. doi:10.1177/0956797613497969

Weiss, D. J., & Anderson, N. H. (1969). Subjective averaging of length with serial presentation. *Journal of Experimental Psychology, 82*(1), 52-63. doi:10.1037/h0028028