Journal of Experimental Psychology: General
2015, Vol. 144, No. 3, 664–673

0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000069

Simultaneous Perceptual and Response Biases on Sequential Face Attractiveness Judgments

Teresa K. Pegors, Marcelo G. Mattar, Peter B. Bryan, and Russell A. Epstein
University of Pennsylvania

Face attractiveness is a social characteristic that we often use to make first-pass judgments about the people around us. However, these judgments are highly influenced by our surrounding social world, and researchers still understand little about the mechanisms underlying these influences. In a series of 3 experiments, we use a novel sequential rating paradigm that enables us to measure biases in attractiveness judgments from the previous face and the previous rating. Our results reveal 2 simultaneous and opposing influences on face attractiveness judgments that arise from past experience of faces: a response bias in which attractiveness ratings shift toward a previously given rating and a stimulus bias in which attractiveness ratings shift away from the mean attractiveness of the previous face. Further, we provide evidence that the contrastive stimulus bias (but not the assimilative response bias) is strengthened by increasing the duration of the previous stimulus, suggesting an underlying perceptual mechanism. These results demonstrate that judgments of face attractiveness are influenced by information from our evaluative and perceptual history and that these influences have measurable behavioral effects over the course of just a few seconds.

Keywords: sequence effects, subjective judgments, decision-making, perception, attractiveness

Human faces are rich sources of information about our social world. Face attractiveness, in particular, is a holistic visual trait that we often use to make first-pass assessments of people as we associate this feature with romantic viability, sociability, and health (for reviews, see Rhodes, 2006; Zebrowitz & Montepare, 2008). Interestingly, our judgments of the attractiveness of an individual face are not based on that face alone: They are highly influenced by other faces observed in the surrounding context. For example, a person is considered more attractive if seen with an unattractive stranger (Kernis & Wheeler, 1981) or a very attractive partner or friend (the *radiation* effect; Kernis & Wheeler, 1981; Strane & Watts, 1977) or by merely appearing within a larger group of people (the *cheerleader* effect; Walker & Vul, 2014). Moreover, even faces viewed in isolation are still often judged to be more or less attractive on the basis of faces that have been viewed in the recent past (Cogan, Parker, & Zellner, 2013; Kondo, Takahashi, & Watanabe, 2012; Wedell, Parducci, & Geiselman, 1987). Surprisingly, the nature of this sequential attractiveness bias remains unclear, because the results in the literature up to this point have been, at first glance, contradictory. Whereas some studies report a contrastive effect (i.e., if the previous face was very attractive, the current face will be rated as less attractive than usual; Cogan et al., 2013; Wedell et al., 1987), other studies report an assimilative effect (i.e., if the previous face was very attractive, the current face will be rated as more attractive than usual; Kondo et al., 2012; Kondo, Takahashi, & Watanabe, 2013). In this article, we make use of a novel experimental paradigm to resolve this apparent contradiction.

Contrastive Sequential Biases

One of the first studies to show the influence of recent visual history on current ratings of attractiveness had experimenter confederates interrupt male undergraduates who were watching *Charlie's Angels* to ask them to rate the attractiveness of a girl in a photograph (who was described as a potential date). Male undergraduates who were watching *Charlie's Angels*, a TV show with three beautiful women as the main characters, rated the girl in the photograph as less attractive than did other male respondents who were watching another show (Kenrick & Gutierres, 1980). Follow-up studies in laboratory-controlled settings provided further evidence for this sequential contrast effect: Faces tended to be rated as less attractive when a beautiful face had been previously viewed, and vice versa (Cogan et al., 2013; Kenrick & Gutierres, 1980; Wedell et al., 1987).

Interestingly, this sequential contrast bias occurs for other kinds of judgments as well, including both hedonic and nonhedonic judgments (Kamenetzy, 1959; Parker, Bascom, Rabinovitz, & Zellner, 2008; Schifferstein & Frijters, 1992; Schifferstein &

This article was published Online First April 13, 2015.

Teresa K. Pegors, Marcelo G. Mattar, Peter B. Bryan, and Russell A. Epstein, Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania.

This work was supported by National Institutes of Health/National Eye Institute Grant EY-022751 to Russell A. Epstein. Teresa K. Pegors was supported by National Science Foundation Graduate Research Fellowship Grant DEG-0822. We are grateful to Steven Marchette for being part of many discussions that provoked this line of research and to Maria Olkkonen and Geoffrey Aguirre for their helpful comments on a draft of the article. Teresa K. Pegors and Marcelo G. Mattar made equal contributions to this article.

Correspondence concerning this article should be addressed to Teresa K. Pegors, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104. E-mail: tpegors@sas.upenn.edu

Kuiper, 1997; Zellner, Rohm, Bassetti, & Parker, 2003). For example, a study originally conducted for military taste testing showed that foods were rated as tasting worse when sampled after a good-quality food than when sampled after a poor-quality food (Kamenetzy, 1959). In another study, musical excerpts were given higher ratings when played after a low-rated excerpt than when played after a high-rated excerpt (Parker et al., 2008). Studies on magnitude estimates from the psychophysics literature have even demonstrated sequential contrast biases for estimates of loudness, light intensity, and size (e.g., DeCarlo & Cross, 1990; Jesteadt, Luce, & Green, 1977; Ward, 1990). The fact that contrastive biasing occurs for such a variety of stimulus types raises the question of whether the sequential contrast bias in face attractiveness judgments is mediated by the same mechanism that underlies contrastive sequential biases in other hedonic and psychophysical domains.

Assimilative Sequential Biases

In distinction to the just-described results, Kondo et al. (2012) reported an assimilative bias for sequential face attractiveness judgments. In this study, subjects made sequential attractiveness judgments of faces using a 1–7 Likert-type scale. The results showed a significant assimilative sequential bias: If the previous face was rated as very attractive, the current face would be rated as a little more attractive than usual, and vice versa. The authors attributed this bias to the previous response, not to the previous stimulus itself. Like the contrast effect, this response bias has been reported to occur for more than just attractiveness judgments of faces: This type of bias is broadly known in the decision-making literature as the *anchoring* effect. Tversky and Kahneman (1974) originally described this effect as one in which a person's current decision will be biased toward a previously given value onto which he or she will "anchor and adjust." Studies in psychophysics have also reported similar assimilative biases in magnitude judgments, and varying theories have been put forward as to the nature of such a bias (DeCarlo & Cross, 1990; Ward & Lockhead, 1971). If it is the case that face attractiveness judgments are influenced by a type of assimilative response bias, as the Kondo et al. (2012) results suggest, then the contrastive and assimilative results reported in the literature are not necessarily in conflict but may arise from two separate sources—the stimulus and the response.

Simultaneous Stimulus and Response Biases

We set out to investigate the question of whether attractiveness ratings are biased by the previous stimulus, the previous response, or both. Sequential rating paradigms, in general, render this question extremely difficult to answer, because ratings will naturally be highly correlated with the stimulus characteristic they are meant to judge. Whether a judgment bias is driven by the previous stimulus or the previous response, therefore, is very hard to determine. Although modeling solutions have been proposed in the psychophysics literature to determine the presence of biases arising from the previous stimulus and response (DeCarlo & Cross, 1990; Jesteadt et al., 1977; Matthews & Stewart, 2009; Ward & Lockhead, 1971), these solutions are necessarily limited in their ability to accurately detect and estimate effect sizes in the presence of this multicollinearity (Neter, Wasserman, & Kutner, 1989). A more effective, yet untested, method is to decorrelate these potential sources of bias in the experimental design itself.

In the following experiments, we used a novel sequential rating design to measure the effects of the previous stimulus and the previous response on face attractiveness judgments. In this design, we alternated the type of judgment on every other trial, which allowed us to obtain estimates of the bias attributable to the attractiveness of the previous face while also estimating the bias attributable to the previous, orthogonal, response. To anticipate, our results showed that opposing biases attributable to both the previous stimulus and response are indeed simultaneously present during sequential face attractiveness judgments. We then used this paradigm to test the extent of these biases across categories and to test whether the bias from the previous stimulus is perceptual or cognitive in nature. We found that both effects are category specific and that only the contrastive (stimulus) bias is dependent on exposure duration, suggesting that these effects are likely mediated by two fundamentally different mechanisms. Not only do our results reconcile the opposing effects seen in the face attractiveness literature, but these experiments demonstrate the usefulness of our experimental paradigm for testing a range of questions about judgment biases.

Experiment 1

The goal of our first experiment was to test whether face attractiveness judgments made in sequence are biased by the attractiveness of the preceding face, the rating given to the preceding face, or both. To answer this question, we asked subjects to make attractiveness judgments and hair darkness judgments of faces on alternating trials. Because all attractiveness trials were preceded by hair darkness trials, and attractiveness and hair darkness are only weakly correlated, this design allowed us to separately measure the effect of the attractiveness of the preceding stimulus and the response to the preceding stimulus on face attractiveness judgments. To determine the generality of the effects, we also measured whether hair darkness judgments were affected by the preceding stimulus and/or response during attractiveness trials.

Method

Stimuli. Two-hundred and forty-two female face images were selected to span a wide range of attractiveness and hair darkness. These came from the Glasgow Unfamiliar Face Database (http://homepages.abdn.ac.uk/m.burton/pages/gfmt/Glasgow%20Face%20Recognition%20Group.html), the Radboud Faces Database (Langner et al., 2010), the Center for Vital Longevity Face Database (Minear & Park, 2004), the CVL Face Database (created by Peter Peer [http://www.lrv.fri.uni-lj.si/facedb.html]), the Diana Theater Face Database (courtesy of Robert Schultz at the Center for Autism Research, Philadelphia, Pennsylvania), and online searches. Faces were all Caucasian, had a neutral to pleasant expression, and were forward facing. They were cropped such that the hair did not extend well below the chin, resized to a height of 400 pixels, and placed on 400 × 400 pixel backgrounds consisting of phase-scrambled variations of a single scene image (for example stimuli, see Figure 1). From the set of 242 images, 10 were chosen to form a practice set of trial images to be used across all

Figure 1. Alternating experimental design. Subjects rated either the attractiveness or the hair darkness of each female face on a 1–8 Likert-type scale. See the online article for the color version of this figure.

subjects, and the experimental trial (212 images) and memory task foil images (10 images) were randomly drawn for each subject from the remaining 232 images.

We acquired attractiveness ratings from 28 subjects not participating in our main experiments to calculate an attractiveness score for each face. Each rater made 1–8 Likert-type scale ratings of 543 male and female faces (including the 242 female faces just described). In a separate block, ratings were given to place images (see Experiment 2). Within each block, image order was randomized, and attractiveness ratings were averaged across raters for each item to determine its attractiveness score. In the current experiment, these attractiveness scores served as our stimulus values, which were considered to be independent of the stimulus history or the task.

Subjects. Our a priori sample size was set at 30 on the basis of the number of subjects used in an earlier experiment that showed a well-powered assimilation effect of one-back ratings on current ratings of attractiveness using a sequential (but nonalternating) design (Kondo et al., 2012). Thirty-two total University of Pennsylvania undergraduates were recruited and given class credit for their participation. Two subjects were excluded for not following instructions, leaving us with a total of 30 subjects (21 female).

Procedure. Subjects made a total of 106 hair darkness judgments and 106 attractiveness judgments on a 1–8 Likert-type scale. Importantly, these judgments alternated such that all attractiveness judgments were preceded by hair darkness judgments, and vice versa. Faces were presented on the screen for 4 s each, and between face presentations, a fixation cross appeared on the screen for a randomized interstimulus interval (ISI) of 0.0–0.5 s. Faces were displayed in the center of the screen, and buttons indicating the numbers between 1 and 8 were displayed at the bottom of the screen (see Figure 1). Subjects were instructed to place eight fingers on the keyboard row of numerical keys so that ratings could be made easily and quickly. To make the task easier, the current judgment type (attractiveness or hair darkness rating) was cued on the screen by the color of an outline around the face and buttons as well as by the button labels at the anchors of the scale. When a subject made a judgment, the corresponding outline of the button turned white to reinforce his or her selection.

A unique face was shown on each of the 212 trials, without any face repetitions. Faces were randomly ordered and randomly assigned to one of the two judgment types for each subject. The first judgment type that subjects made was counterbalanced between subjects. In an attempt to ensure that subjects attended to the entire face (and did not, e.g., just focus on the hair), we asked them to remember each face for a postexperiment memory test.

To acclimate subjects to the range of attractiveness in the experiment, they were trained beforehand on the alternating task with 10 faces that were not used in the main experiment. Faces for the practice were chosen to span the ranges of attractiveness and hair darkness. Participants were instructed to spread their ratings during the main experiment across the full scale on the basis of the range of faces they had seen during practice.

After the main experiment, subjects were shown a random subset of 20 images from the experiment (10 from the hair darkness trials, 10 from the attractiveness trials) and 20 novel images. These images were randomly intermixed, and subjects used a mouse to click a *Y* button on the screen if they had previously seen the image or an *N* button if they had not. Subjects completed the memory task at their own pace.

To acquire hair darkness ratings outside of the context of the alternating task, after the experiment and at their own pace, subjects rated hair darkness on the full set of female faces (242 images). Subjects also made hair darkness ratings on a separate block of male faces (not used in our subsequent analyses). Faces were presented in a different randomized order for each subject. Whether subjects rated male faces or female faces first was counterbalanced across subjects. The resulting hair darkness ratings were compiled from 28 of the subjects (two subjects' ratings were not acquired due to technical errors) and averaged across subjects to create a mean hair darkness score for each face.

Results and Discussion

For all analyses, we excluded trials on which the reaction time (RT) was $\leq$0.2 s, because a short RT might indicate an anticipation error or a rating attributable to the previous trial. There was no correlation between attractiveness ratings and RT (Pearson's r = .01), $t(29) = 0.42$ $p = .68$, and an extremely small but trending negative correlation between hair darkness ratings and RT (Pearson's $r = -.04$), $t(29) = -1.93$ $p = .06$, indicating that hair darkness was rated slightly more quickly for faces with darker hair. In the postexperiment memory test, subjects correctly rated faces seen during attractiveness judgments as familiar (M = 7.1 out of 10), $t(29) = 5.69$, $p < .001$. However, they were at chance when rating faces seen during hair darkness judgments (M = 5.23 out of 10), $t(29) = 0.68$, $p = .50$, suggesting that they paid less attention to face identity when making these judgments. The unfamiliar foils were correctly rejected ($M = 14.6$ out of 20), $t(29) = 6.76$, $p < .001$.

In our first analysis of the attractiveness ratings, we used a time-series regression analysis to determine whether these ratings were significantly influenced by either the attractiveness of the previous face or the previous response. We created a separate model for each subject by regressing individual attractiveness ratings against the mean attractiveness of the previous face and that subject's hair darkness response given to the previous face. We also included the mean attractiveness of the current face as a predictor to account for attractiveness variance not due to sequential biasing. The model used is summarized by the following equation:

$$R_t = \beta_0 + \beta_1 S_t + \beta_2 R_{t-1} + \beta_3 S_{t-1} + \varepsilon, \quad (1)$$

where R is the response, S is the average attractiveness of the face (the stimulus value), t is the trial index, and ε is the error term. (Note that R_t in this first model is a judgment of attractiveness, and R_{t-1} is a judgment of hair darkness, whereas S_t and S_{t-1} are both attractiveness values.) The dependent variable and all predictors were standardized (z scored) for each subject for the resulting beta estimates to be comparable across subjects. There was only a weak correlation between the previous response R_{t-1} (hair darkness judgment) and the previous stimulus value S_{t-1} (attractiveness judgment; mean Pearson's $r = -.135$), suggesting that these two variables are dissociable. We formally tested for multicollinearity by examining the variance inflation factor (VIF) of each of the independent variables. This number gives an estimated severity of multicollinearity—the higher the number, the more severe, with a lower bound of 1. Each of our independent variables had a very low VIF ($S_t = 1.02$, $R_{t-1} = 1.04$, $S_{t-1} = 1.04$), suggesting that multicollinearity was not a concern. (A VIF of 1.02 means that the variance of the coefficient is 0.02% larger than it would be if that predictor were uncorrelated with all other predictors). As an additional control for the nonzero correlation between the previous response and the previous stimulus value, we also ran our models with a $\beta_2 R_{t-1} \times \beta_2 S_{t-1}$ interaction term (discussed in the following paragraph).

Beta estimates of the previous stimulus and response predictors were extracted for each subject-specific regression model. Results from testing these betas against zero revealed that the response given during hair darkness trials had a significant and positive effect on subsequent attractiveness ratings (β_2: $t[29] = 2.73$, $p = .011$), whereas the attractiveness of the preceding stimulus had a significant but negative influence on current judgments of face attractiveness (β_3: $t[29] = -4.92$, $p < .001$; see Figure 2). That is, the effect of the preceding response was assimilative, whereas the effect of the preceding stimulus was contrastive. Adding the $\beta_2 R_{t-1} \times \beta_2 S_{t-1}$ interaction term to the model increased our estimate of a bias as a result of the previous response (β_2: $t[29] = 2.84$, $p = .008$) and maintained our estimate of a bias attributable to the previous stimulus (β_3: $t[29] = -4.91$, $p < .001$). The interaction term itself was not significant, $t(29) = -1.38$, $p = .18$.

The first result (β_2) parallels the assimilative bias seen by Kondo et al. (2012) but extends it by showing that this bias can be linked to the previous response rather than to the attractiveness of the previous face. Notably, this response bias occurred across judgment types: hair darkness ratings influenced attractiveness ratings. This cross-judgment influence echoes results from the decision-making literature, in which seemingly unrelated numerical values influence subsequent decisions (Critcher & Gilovich, 2008; Tversky & Kahneman, 1974). The second result (β_3) parallels other study results that have shown a contrastive effect for sequential ratings of face attractiveness and other stimulus qualities (Cogan et al., 2013; Parker et al., 2008; Wedell et al., 1987). Moreover, our design directly links this contrastive effect to the attractiveness of the previously viewed face rather than the previous response, even though the subject was attending primarily to the hair rather than the face (as evidenced by the memory results, in which subjects did not remember the faces in the hair darkness trials significantly above chance).

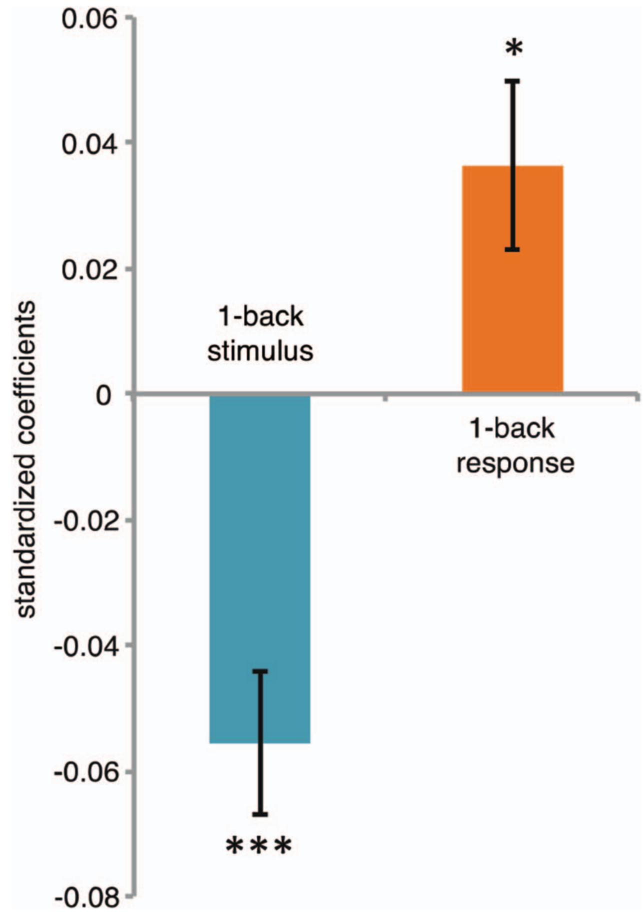

Figure 2. Regression results from the alternating face attractiveness and hair darkness design. Face attractiveness judgments were regressed against the previous response (to hair darkness) and the previous stimulus (the attractiveness of the face, an averaged score from independent subjects). Both elements of the previous trial significantly predicted current attractiveness judgments. The previous response positively predicted attractiveness judgments (an assimilative effect), and the previous stimulus negatively predicted attractiveness judgments (a contrastive effect). Error bars represent the standard error of the mean. * $p < .05$. *** $p < .001$. See the online article for the color version of this figure.

In our second regression analysis, we sought to determine whether hair darkness ratings showed the same sensitivity to stimulus and response biases. We used the same model as for the attractiveness ratings but now regressed hair darkness ratings on the hair darkness of the previous face and the attractiveness rating of the previous face (Pearson's *r* between previous stimulus and response = −.11, VIF for S_t = 1.02, VIF for R_{t-1} = 1.04, VIF for S_{t-1} = 1.03). Here, we saw a similar pattern of stimulus results in that there was a significant contrastive influence of the previous stimulus (β_3: $t[29] = -5.50$, $p < .001$). That is, faces were judged as having darker hair if they were preceded by faces with lighter hair, and vice versa. The strength of our estimate was maintained after adding the interaction term (β_3: $t[29] = -5.45$, $p < .001$). We also observed a marginal trend toward an assimilative influence from the previous response (β_2: $t[29] = 1.84$, $p = .08$), but this effect was reduced with the addition of the interaction term (β_2: $t[29] = 1.64$, $p = .11$). The interaction term itself was not significant, $t(29) = 0.74$, $p = .47$.

To get an estimate of the size of these effects in terms of raw ratings scores, we reran our original regressions using non-*z*-scored regressors. For the attractiveness ratings model, the averaged beta weight across subjects was −0.08 for the stimulus effect (β_3 range: −0.30–0.13) and 0.03 for the response effect (β_2 range: −0.09–0.19). This means that, for the stimulus effect, holding all other variables constant, if the previous face is one rating unit (on the Likert-type scale) more attractive than the sample mean, the current face will tend to be rated as 0.08 rating units less attractive than it would have been on average. For the response effect, conversely, if the previous face is rated one unit more attractive than the mean, the current face will tend to be rated as 0.03 rating units more attractive than it would have been on average. In this case, the overall effect on a rating score is contrastive as the contrastive effect of the preceding stimulus is larger on average than the assimilative effect of the preceding response. Given that these values estimate the shift that would occur with only a distance of one rating unit from the mean, and the fact that even greater variations of face attractiveness occur in the natural world, it is likely that the true effect of one face on another would be even larger and possibly quite noticeable. The hair darkness model showed effects on the same order of magnitude: The averaged beta weight across subjects was −0.05 for the stimulus effect (β_3 range: −0.20–0.04) and 0.01 for the response effect (β_2 range: −0.08–0.08).

In summary, in Experiment 1, we created an experimental design that effectively separated out possible biases resulting from the previous stimulus and the previous response. Data obtained from this alternating design revealed that there are indeed significant biases on attractiveness judgments that occur simultaneously and in opposite directions. Further, a more general mechanism appears to be driving this effect: The contrastive bias resulting from the previous stimulus was not to unique to attractiveness judgments—hair darkness ratings were biased in the same manner. In addition, face attractiveness ratings were assimilated toward hair darkness ratings, suggesting that the response effect also is able to cross task boundaries (though the fact that attractiveness ratings did not robustly affect hair darkness ratings renders the interpretation of this effect less clear).

Experiment 2

Although our first experiment showed that both the previous stimulus and the previous response biased attractiveness judgments even across task, we wanted to explore the boundary conditions of these two effects. In our second experiment, we alternated not only the task but also the category of the stimulus by alternating between face and place images. This allowed us to increase both the perceptual and the conceptual distance between trials by testing the effects of place beauty and an orthogonal place rating on face attractiveness judgments.

Method

Stimuli. The same 242 female face images were used as in Experiment 1. In addition, 373 natural scene images were selected from online sources to span a range of scene types (e.g., forests, beaches, mountains). These were cropped to 400 × 400 pixels to match the size of the face images. Place attractiveness ratings were acquired from the same 28 independent raters used to acquire face attractiveness ratings (see Experiment 1). Seven face and seven place images were used for all subjects as practice images, and the experimental trial images (106 female faces, 106 places) and memory task foil images (20 faces, 20 places) were randomly drawn for each subject from the remaining images. Each subject, therefore, saw a unique (though overlapping) set of images.

Subjects. We again set our a priori sample size to 30, to match the number of subjects in Experiment 1. We ran a total of 31 University of Pennsylvania undergraduates and excluded one due to a technical error, leaving us with 30 subjects (18 female). Subjects received course credit for their participation.

Procedure. During the experiment, face and place trials were alternated, with participants rating the temperature of the place images on a scale from 1–8 (an orthogonal judgment to place beauty) and the attractiveness of face images on a scale from 1–8. The design and procedure were similar to those used in Experiment 1, with a few key changes. The place trials were cued with the word *temperature* above the image and the words *cold* and *hot* at the scale anchors. In a practice session, subjects completed the alternating task on 14 images (seven face, seven place) that were not shown in the main experiment and that were chosen to span the range of the attractiveness and temperature. After the practice, subjects were instructed to spread their ratings across the scale on the basis of the images they had just seen. They were also instructed to remember all of the images for a postexperiment memory test. The memory test included 20 place and 20 face images seen during the main experiment and 20 place and 20 face foils.

Results and Discussion

Any trial on which the RT was ≤0.2 s was excluded. There was no correlation between RT and face attractiveness (Pearson's r = .01), $t(29) = 0.45$ $p = .66$, or between RT and place temperature (Pearson's $r = .01$), $t(29) = 0.56$ $p = .58$. In the postexperiment memory test, subjects correctly reported images seen during the experiment as familiar for places ($M = 14.9$ out of 20), $t(29) =$ 29.45, $p < .001$, and faces ($M = 14.93$ out of 20), $t(29) = 32.66$, $p < .001$, and correctly rejected the unfamiliar foils ($M = 29.9$), $t(29) = 45.93$, $p < .001$. There was no significant difference between the number of faces and places remembered, $t(29) = -0.06$, $p = .95$.

To test whether face attractiveness judgments were influenced by either the attractiveness of the preceding place stimulus or the previous response, we used the same analysis approach as in Experiment 1. That is, we regressed subject-specific face attractiveness ratings against the mean attractiveness of the face, the mean attractiveness of the preceding place, and the subjects' previous response regarding place temperature (see Equation 1). There was a very low correlation between the temperature judgments and place attractiveness (Pearson's $r = -.03$, averaged across subjects), suggesting that our design successfully decoupled the potential effects from the previous stimulus and response. Our test for multicollinearity using the VIF on each of the predictors showed low numbers, similar to Experiment 1, indicating that multicollinearity was not a concern (VIF for $S_t = 1.02$, VIF for $R_{t-1} = 1.04$, VIF for $S_{t-1} = 1.04$). For completeness, we also ran the model with an additional $\beta_2 R_{t-1} \times \beta_2 S_{t-1}$ interaction term to control for any effect of the nonzero correlation between the previous response and the previous stimulus value.

Notably, our regression analysis did not reveal biases in face attractiveness ratings resulting from either the preceding place stimulus or the previous place temperature rating, as evidenced by beta weights across subjects in t tests against zero for previous temperature rating (β_2: $t[29] = 1.64$, $p = .11$) and previous place attractiveness (β_3: $t[29] = 1.47$, $p = .15$). Thus, in contrast to the previous experiment, in which we observed both stimulus and responses effects of the preceding face on judgments of the current face, here we observed no effects of the preceding place on judgments of the current face. The results continued to be nonsignificant after adding the interaction term (β_2: $t[29] = 0.89$, $p = .38$; β_3: $t[29] = 0.74$, $p = .47$). The interaction term was also not significant, $t(29) = -0.40$, $p = .69$. (Although it would also be interesting to measure the reverse influence of previous-trial face attractiveness on place attractiveness judgments, our design only included place temperature judgments, so we could not address this additional question with the data.)

To show that the absence of effects was a result of the change of stimulus category rather than other factors, we ran another regression analysis modeling the previous four trials rather than just the previous trial. This allowed us to look for any significant effects from previous face trials (trials that were two back and four back) on the current face trial. (It is important to note, of course, that the stimulus and response predictors for these trials were highly correlated because they both measured attractiveness. Although significant results are meaningful, the true strength of the effect cannot be characterized.) Using this model, we observed a significant assimilative influence of both the two-back and four-back face attractiveness responses on the participants' current face attractiveness judgment (two-back rating: $t[29] = 3.67$, $p < .001$; four-back rating: $t[29] = 2.80$, $p = .009$) and a significant contrastive influence of both the two-back and four-back faces (two-back face attractiveness effect: $t[29] = -6.94$, $p < .001$; four-back face attractiveness effect: $t[29] = -4.00$, $p < .001$; see Figure 3).

These results have a number of implications. First, a large enough dissimilarity between trial types serves to abolish both the contrast and the assimilative biases. Although evidence from our first experiment indicated that a previous numerical response biased face attractiveness ratings across task type, there was no effect when the distance between the trials spanned two separate semantic/perceptual categories. Second, because we saw significant biases arising from the two- and four-back trials, our lack of significant weighting on the one-back place trials cannot be explained by a lack of power to find a sequential effect. Third, these results make clear that these sequential bias effects are modulated by factors other than time, because the strength of the influence is modulated in an alternating fashion by the trial type.

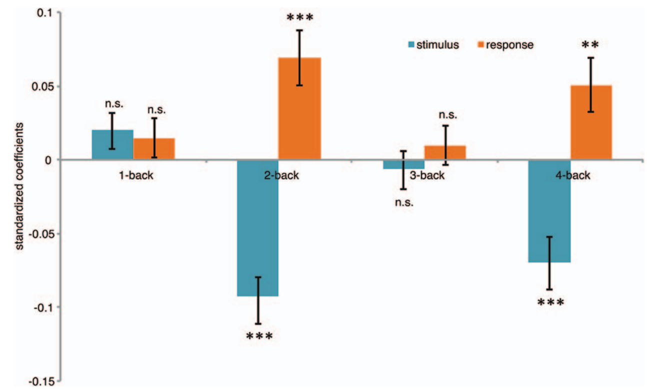

Figure 3. Regression results from the alternating face attractiveness and place temperature design. Face attractiveness ratings were regressed against the previous stimuli and the previous responses of the four preceding trials. Because of the alternating design, trials one back and three back were always place trials on which subjects judged place temperature, and trials two back and four back were always face trials on which subjects judged face attractiveness. Neither the response to place temperature nor the underlying attractiveness of places significantly predicted current face attractiveness ratings, but face trials even four trials back showed predictive power related to both the subject's response and the mean attractiveness of the face. Error bars represent the standard error of the mean. ** $p < .01$. *** $p < .001$. See the online article for the color version of this figure.

Experiment 3

In our third experiment, we attempted to better understand the nature of the contrast effect we observed for both face attractiveness and hair darkness. Sequential contrastive biases have been reported for several other types of stimuli (Kamenetzy, 1959; Parker et al., 2008; Schifferstein & Frijters, 1992; Schifferstein & Kuiper, 1997; Zellner et al., 2003). A typical interpretation of the contrastive effect is the well-known perceptual phenomenon of visual aftereffects, which has been used for over a century to describe perceptual contrast effects in motion, color, and shape (Addams, 1834; Gibson, 1933; McCollough, 1965). Perceptual aftereffects have also been observed for complex facial features such as identity, gender, ethnicity, and emotion (Leopold, O'Toole, Vetter, & Blanz, 2001; Webster, Kaping, Mizokami, & Duhamel, 2004). Aftereffects are thought to occur as a result of our visual system constantly adapting to incoming stimulus information and influencing our perception of subsequent input. According to this interpretation, subjects in our experiments would perceive faces differently on the basis of the previous face. An alternative interpretation for the contrastive bias is that a type of cognitive remapping takes place between faces and the scale itself. Under this explanation, subjects would not perceive the next face differently but, rather, would consciously or subconsciously remap certain faces or face types to the rating scale numbers on a trial-by-trial basis (e.g., "Now that I've seen face X, I'd consider face Y to be a 7, not 6").

To distinguish between these two possibilities, we modified the alternating design of Experiment 1 by varying the length of time that each face was on the screen. Previous research on face aftereffects has shown that contrastive effects are strengthened logarithmically with the length of stimulus presentation and that this effect decays exponentially over the test duration (Leopold, Rhodes, Müller, & Jeffery, 2005; Rhodes, Jeffery, Clifford, & Leopold, 2007). Therefore, if the contrast effect is perceptual in nature, trials preceded by briefly presented faces should show a weaker contrastive effect than trials preceded by faces presented for a longer duration. Conversely, if the contrast effect is a result of cognitive remapping, we would not expect stimulus duration to have a selective effect on the contrastive bias. In addition, the assimilative bias resulting from the previous response should not be modulated by the previous stimulus duration, because there is no reason to think that the response bias (due to the previous numerical rating) is perceptual in nature.

Method

Subjects. We conducted the experiment online using Amazon Mechanical Turk (mTurk [https://www.mturk.com/]) in conjunction with custom online scripts. Our move to an online version of the experiment allowed us to run the larger number of subjects needed to sufficiently power our ability to observe a modulatory effect of trial duration. A liberal sample of 415 subjects was recruited to account for both the greater variability in the online subject pool and the fact that the stimulus duration manipulation cut in half the number of trials used for each model. Our selection criteria for online subjects were that they lived in the United States, were approved in at least 95% of their mTurk task submissions, and had previously completed at least 500 approved tasks.

Stimuli. We used the same face stimuli as in Experiment 1 to make up our stimulus set for the current experiment. To obtain average attractiveness and hair darkness ratings that best reflected individual ratings from our mTurk experimental subjects, we acquired 50 attractiveness and 50 hair darkness ratings from mTurk subjects, who each rated the faces in a randomized order. These averaged ratings served as our stimulus values, which were considered to be independent of the stimulus history.

Procedure. The task and procedure were very similar to those in Experiment 1, with some key differences related to the online nature of the task and the varying face presentation times. During the experimental run, 200 faces in total were presented to each subject, and subjects alternated their judgments between face attractiveness and hair darkness. For a pseudorandom half of these trials, faces were presented for a 1,000-ms duration with a 3,000-ms ISI, and for the other half of these trials, faces were presented for a 3,750-ms duration with a 250-ms ISI. The total length of each trial did not change—only the duration of the face presentation. In total, four trial types of equal numbers were randomized throughout the experiment: 50 short- and 50 long-duration attractiveness trials and 50 short- and 50 long-duration hair darkness trials. To make clear to the subject the consistency and duration of each trial (all totaling 4 s in length), we added a progress bar to the top of the screen that visually represented the remaining trial duration during each trial.

As in the previous studies, subjects also underwent 10 warm-up trials before the experimental trials, and they were given a memory task after the main experiment in which they indicated with a simple *Y* or *N* response whether they had seen the face in the experiment. For the memory task, we intermixed 20 face images seen during the main experiment (five from each condition) with 20 face foils. All instructions throughout the course of the experiment were presented by text.

Results and Discussion

To ensure accurate modeling of stimulus features, subjects were excluded from the attractiveness analyses if their individual attractiveness ratings had a correlation score of less than .50 with the norm ratings, and subjects were excluded from the hair darkness analyses if their individual hair darkness ratings had a correlation score of less than .50 with the norm ratings. Subjects were also excluded if they missed more than five of the 200 trials. This left us with 347 subjects for our attractiveness judgment model and 391 subjects for our hair darkness judgment model. Any trial on which the RT was ≤0.2 s was also excluded. In the attractiveness judgments model, there was a significant yet small positive correlation between attractiveness ratings and RT (Pearson's $r = .11$), $t(346) = 10.58$, $p < .001$, meaning that subjects' RTs increased as face attractiveness increased. In the hair darkness judgments model, there was no correlation between hair darkness ratings and RT (Pearson's $r = -.01$), $t(390) = -1.31$, $p = .19$.

Subjects correctly reported faces seen during attractiveness judgments as familiar in both the attractiveness model ($M = 6.52$ out of 10,) $t(346) = 67.66$, $p < .001$, and the hair darkness model ($M = 6.48$ out of 10), $t(390) = 69.67$, $p < .001$, and they correctly reported faces seen during hair darkness trials as familiar in the attractiveness model ($M = 5.30$ out of 10), $t(346) = 49.65$, $p < .001$, and the hair darkness model ($M = 5.37$ out of 10), $t(390) = 53.06$, $p < .001$. Subjects correctly rejected unfamiliar faces in the attractiveness model ($M = 13.40$ out of 20), $t(346) = 81.02$, $p < .001$, and the hair darkness model ($M = 13.12$ out of 20), $t(390) = 80.12$, $p < .001$. Subjects remembered faces during attractiveness trials better than during hair darkness trials in both the attractiveness model, $t(346) = -9.92$, $p < .001$, and the hair darkness model, $t(390) = -9.47$, $p < .001$.

Before examining the effects of presentation time, we first sought to determine whether our online version of the task replicated the basic contrast and assimilation effects we found with undergraduates in the laboratory. To do this, we collapsed across the two stimulus duration conditions, using the same regression model as in Experiment 1 (see Equation 1). For both attractiveness and hair darkness judgments, we observed very robust contrastive biases as a result of the previous stimulus for the face attractiveness judgments model, $t(346) = -10.07$, $p < .001$, and the hair darkness judgments model, $t(390) = -11.48$, $p < .001$. We also observed robust assimilative biases as a result of the previous rating for the face attractiveness judgments model as an effect of the previous hair darkness rating, $t(346) = 3.03$, $p = .003$, and for the hair darkness judgments model as an effect of the previous face attractiveness rating, $t(390) = 6.22$, $p < .001$. Similar to our first experiment, there was a low negative correlation in our face attractiveness model between the two one-back predictor variables of hair darkness judgments and face attractiveness (Pearson's $r = -.11$, averaged across subjects) and nearly no correlation in our hair darkness model between the two one-back predictor variables (Pearson's $r = -.06$, averaged across subjects). The average VIF across subjects was low for each of the three predictors in both models (attractiveness model: VIF for $S_t = 1.02$, VIF for $R_{t-1} = 1.04$, VIF for $S_{t-1} = 1.04$; hair darkness model: VIF for $S_t = 1.02$, VIF for $R_{t-1} = 1.05$, VIF for $S_{t-1} = 1.05$). Finally, when we ran the models with the additional $\beta_2 R_{t-1} \times \beta_2 S_{t-1}$ interaction term to control for any shared variance between the two predictors, all terms were still significant (attractiveness model, previous stimulus: $t[346] = -9.90$, $p < .001$; attractiveness model, previous response: $t[346] = 3.38$, $p < .001$; hair darkness model, previous stimulus: $t[390] = -4.54$, $p < .001$; hair darkness model, previous response: $t[390] = 2.48$, $p = .01$).

We then turned to the main goal of our experiment: to test the hypothesis that longer stimulus durations in the preceding trial would strengthen the contrast bias but not the assimilation bias, a pattern of results that would be in line with a perceptual interpretation of the contrastive effect. To do this, we subdivided trials into those preceded by short-duration trials and those preceded by long-duration trials. For each trial type (e.g., attractiveness judgment trials preceded by short-duration trials), this resulted in 50 trials to model the contrast and assimilation biases using Equation 1. For attractiveness judgments, contrastive effects from the previous stimulus were highly significant in both trials preceded by short-duration trials, $t(346) = -5.31$, $p < .001$, and trials preceded by long-duration trials, $t(346) = -9.27$, $p < .001$; notably, though, judgments preceded by faces shown for a long duration showed a significantly greater contrast effect that judgments preceded by faces shown for a short duration, $t(346) = -3.30$, $p = .001$ (see Figure 4). In addition, although assimilative biases as a result of the previous rating were also significant on trials preceded by short-duration trials, $t(346) = 2.00$, $p = .047$, and trials preceded by long-duration trials, $t(346) = 2.56$, $p = .011$, there was no difference in the strength of the effect between the conditions, $t(346) = 0.41$, $p = .68$ (see Figure 4). Further, we also observed that the difference in contrast strength between short and long presentations was significantly greater than the difference in assimilation strength between short and long presentations (the "difference between differences"), $t(346) = -2.92$, $p = .0037$ (repeated measures t test on differences scores). Further, the same pattern of results was present when we measured these effects for hair darkness judgments preceded by short- and long-duration trials: Contrastive biases from the previous stimulus were highly significant in both trials preceded by short-duration trials, $t(390) = -5.37$, $p < .001$, and trials preceded by long-duration trials, $t(390) = -10.82$, $p < .001$; however, trials preceded by faces shown for a long duration showed a significantly greater contrast effect than those trials preceded by faces shown for a short duration, $t(390) = -4.44$, $p < .001$. There was no difference between the strength of the assimilative biases for each trial type, $t(390) = -0.80$, $p = .42$, even though assimilative biases as a result of the previous rating were significant in both trials preceded by short-duration trials, $t(390) = 4.95$, $p < .001$, and trials preceded by long-duration trials, $t(390) = 4.37$, $p < .001$. In addition, we also observed for hair darkness judgments that the difference in contrast strength between short and long presentations was significantly greater than the difference in assimilation strength between short and long presentations, $t(390) = -2.57$, $p = .011$.

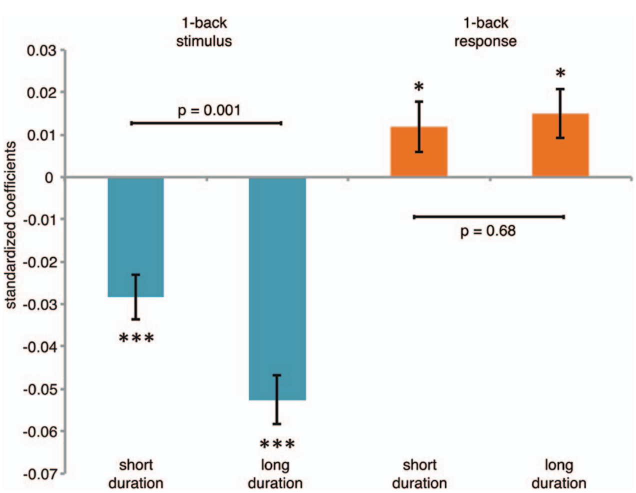

Figure 4. Results from two separate regression models comparing the effects of stimulus duration on assimilative and contrastive biases. Face attractiveness judgments preceded by short-duration trials were regressed against the previous response (to hair darkness) and the previous stimulus (the attractiveness of the face, an averaged score from independent subjects). Face attractiveness judgments preceded by long-duration trials were similarly regressed against the previous response and stimulus. Although the assimilative effect of the previous response was not modulated by duration, the contrastive effect of the previous stimulus was significantly stronger when the previous face was presented for a longer duration. Error bars represent the standard error of the mean. * $p < .05$. *** $p < .001$. See the online article for the color version of this figure.

These results are consistent with a perceptual interpretation of the contrast bias given that both attractiveness judgments and hair darkness judgments that were preceded by longer stimulus durations showed a stronger contrastive bias. Conversely, the assimilative bias did not show such a modulation, suggesting a different, nonperceptual mechanism. These results also demonstrate the robustness of our design to reveal sequential biases even in the presence of an additional condition manipulation (which decreased the number of trials per condition per subject and added a level of subjective complexity).

General Discussion

To navigate the social world, it is important to be able to evaluate face attractiveness, but such judgments are always made in relation to a larger social and environmental context. In this article, we have introduced a novel sequential rating paradigm that we used to pinpoint the source of at least two opposing contextual influences on face attractiveness judgments. First, we showed that our judgment of a face will tend to assimilate toward the response that we gave on the previous trial: If we rated the previous face as very attractive, we will tend to rate the next face as slightly more attractive than we would normally. Second, the stimulus qualities of the face that we viewed on the previous trial have a contrastive effect on our current judgment of a face: We will rate a face to be slightly less attractive if we have just seen an extremely beautiful face. Further, we used a modified sequential rating paradigm to provide evidence that this contrastive bias may be attributable to a more general perceptual aftereffect in which we perceive the next face to be slightly less attractive if we have just seen an extremely beautiful face.

Sequential Contrast Bias

In all of our experiments, we observed that the attractiveness of previous faces (as judged by an independent set of observers) negatively predicted subsequent face attractiveness ratings. This effect parallels attractiveness contrast effects seen previously in the social psychology literature (Cogan et al., 2013; Kenrick & Gutierres, 1980; Wedell et al., 1987). We also observed a strong contrast effect for hair darkness in Experiments 1 and 3; that is, faces were judged to have darker hair when preceded by faces with lighter hair, and vice versa. However, we did not observe a contrastive effect across categories in Experiment 2: Faces were not judged to be more attractive when preceded by attractive places. The fact that we observed contrastive biases for two separate perceptual characteristics (attractiveness and hair darkness) but did not observe a contrastive bias across different perceptual categories can be explained with either a cognitive or a perceptual interpretation.

According to a cognitive interpretation, it may have been the case that subjects were remapping facial features to the ratings scale on a trial-by-trial basis, finely adjusting what rating they would give to what type of face on the basis of the recent attractiveness history. Alternatively, according to a perceptual interpretation, the contrastive bias may have reflected a subtle change in how a subject perceptually processed a face, in a similar manner to a perceptual aftereffect thought be driven by neural adaptation in visual cortex. In our third experiment, we made use of the fact that perceptual aftereffects are strengthened by the duration of the adapting stimulus to show that longer face presentations on the previous trial strengthened the contrast bias in the subsequent judgment. One could argue that memory may show greater decay when the previous stimulus is presented briefly (resulting in a longer ISI), therefore weakening the stimulus effect. However, if this were the case, we might also expect the assimilative bias to be weaker when preceded by a short-duration stimulus. Significantly, this was not the case: Our results showed an equally strong assimilative bias for both presentation durations.

In fact, a perceptual interpretation is often used to explain contextual biases in judgments of high-level face attributes (Leopold et al., 2005; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Webster et al., 2004). For example, Rhodes et al. (2003) used perceptual aftereffects as an explanation for their results showing that attractiveness norms are influenced by perceptual history in the context of a blocked adaptation paradigm using compressed and expanded faces as adaptors. Thus, our results are consistent with a number of previous interpretations of contextual effects. Admittedly, however, whether a given phenomenon can be interpreted as perceptual remains a contentious topic in the literature (Morgan, Melmoth, & Solomon, 2013).

Sequential Assimilation Bias

Our experiments also revealed that previous ratings given to faces positively predict current attractiveness ratings. These results replicated the assimilative effect on face attractiveness seen by Kondo et al. (2012), but we extended their findings by linking the effect directly to the previous rating. It may be the case that Kondo et al. (2012) observed an overall assimilative effect as a result of the fact that their brief image presentations created a weaker stimulus bias relative to the response bias.

In the psychophysics literature, one interpretation of the assimilative relationship between past and current judgments is that it is a reflection of the previous judgment acting as a reference point for comparison. For example, DeCarlo and Cross (1990) provided evidence for this relative judgment model by showing that the assimilation effect on loudness estimates was decreased when subjects were instructed to make their judgments relative to a single reference loudness, presumably meaning that subjects shifted their reference away from the previous trial. Our results are consistent with this interpretation.

Our results showing no assimilative effect of responses to place temperature on face attractiveness ratings differed from the “anchor and adjust” account in the decision-making literature, in which previous values can be completely unrelated to the current judgment yet still have an assimilative influence (Tversky & Kahneman, 1974). Conversely, we did show cross-task assimilation within the same stimulus category, suggesting some level of generality to the effect. Future studies, therefore, should address exactly the set of conditions under which this bias is present.

Relevance to Sequential Tasks

Our results revealed at least two bias-inducing mechanisms that reinforce researchers’ motivation to randomize trial order for each subject when acquiring mean estimates of stimulus characteristics. Because randomization is already common practice, our results do not necessarily invalidate the many studies that use sequential rating designs. Rather, having an awareness of these potential biases may help researchers when considering appropriate experimental designs and analyses by allowing them to take into account the fact that both previous subject responses and previous stimulus presentations may affect behavior in a measurable way on subsequent trials.

Conclusion

Sequential biases during judgment tasks are ubiquitous throughout the psychological literature, and it is therefore important to understand the nature and magnitude of such effects. In this article, we introduced a novel method that uses alternating judgments to quantify simultaneous biases resulting from the previous stimulus and the previous response on current judgments. Our results demonstrated the usefulness of this technique to answer questions within the domain of face attractiveness judgments; we hope that in the future, this method will be used to answer other questions related to sequential biases in subjective judgments more generally.

References

Addams, R. (1834). An account of a peculiar optical phenomenon seen after having looked at a moving body. *The London and Edinburgh Philosophical Magazine and Journal of Science, 5,* 373–374.

Cogan, E., Parker, S., & Zellner, D. A. (2013). Beauty beyond compare: Effects of context extremity and categorization on hedonic contrast.

Journal of Experimental Psychology: Human Perception and Performance, 39, 16–22. http://dx.doi.org/10.1037/a0031020

Critcher, C., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making, 251,* 241–251. http://dx.doi.org/10.1002/bdm

DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General, 119,* 375–396. http://dx.doi.org/10.1037/0096-3445.119.4.375

Gibson, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology, 16,* 1–31. http://dx.doi.org/10.1037/h0074626

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 92–104. http://dx.doi.org/10.1037/0096-1523.3.1.92

Kamenetzy, J. (1959). Contrast and convergence effects in ratings of foods. *Journal of Applied Psychology, 43,* 47–52. http://dx.doi.org/10.1037/h0041264

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,* 131–140. http://dx.doi.org/10.1037/0022-3514.38.1.131

Kernis, M. H., & Wheeler, L. (1981). Beautiful friends and ugly strangers: Radiation and contrast effects in perception of same-sex partners. *Personality and Social Psychology Bulletin, 7,* 617–620. http://dx.doi.org/10.1177/014616728174017

Kondo, A., Takahashi, K., & Watanabe, K. (2012). Sequential effects in face-attractiveness judgment. *Perception, 41,* 43–49. http://dx.doi.org/10.1068/p7116

Kondo, A., Takahashi, K., & Watanabe, K. (2013). Influence of gender membership on sequential decisions of face attractiveness. *Attention, Perception, & Psychophysics, 75,* 1347–1352. http://dx.doi.org/10.3758/s13414-013-0533-y

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24,* 1377–1388. http://dx.doi.org/10.1080/02699930903485076

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4,* 89–94. http://dx.doi.org/10.1038/82947

Leopold, D. A., Rhodes, G., Müller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences, 272,* 897–904. http://dx.doi.org/10.1098/rspb.2004.3022

Matthews, W. J., & Stewart, N. (2009). Psychophysics and the judgment of price: Judging complex objects on a non-physical dimension elicits sequential effects like those in perceptual tasks. *Judgment and Decision Making, 4,* 64–81.

McCollough, C. (1965, September 3). Color adaptation of edge-detectors in the human visual system. *Science, 149,* 1115–1116. http://dx.doi.org/10.1126/science.149.3688.1115

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36,* 630–633.

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. *Visual Neuroscience, 30,* 197–206. http://dx.doi.org/10.1017/S095252381300045X

Neter, J., Wasserman, W., & Kutner, M. H. (1989). Multicollinearity and its effects. In *Applied Linear Regression Models* (2nd ed., pp. 295–305). Homewood, IL: Irwin.

Parker, S., Bascom, J., Rabinovitz, B., & Zellner, D. (2008). Positive and negative hedonic contrast with musical stimuli. *Psychology of Aesthetics, Creativity, and the Arts, 2,* 171–174. http://dx.doi.org/10.1037/1931-3896.2.3.171

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57,* 199–226. http://dx.doi.org/10.1146/annurev.psych.57.102904.190208

Rhodes, G., Jeffery, L., Clifford, C. W. G., & Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. *Vision Research, 47,* 2291–2296. http://dx.doi.org/10.1016/j.visres.2007.05.012

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science, 14,* 558–566. http://dx.doi.org/10.1046/j.0956-7976.2003.psci_1465.x

Schifferstein, H. N. J., & Frijters, J. E. (1992). Contextual and sequential effects on judgments of sweetness intensity. *Perception & Psychophysics, 52,* 243–255. http://dx.doi.org/10.3758/BF03209142

Schifferstein, H. N. J., & Kuiper, W. E. (1997). Sequence effects in hedonic judgments of taste stimuli. *Perception & Psychophysics, 59,* 900–912. http://dx.doi.org/10.3758/BF03205507

Strane, K., & Watts, C. (1977). Females judged by attractiveness of partner. *Perceptual and Motor Skills, 45,* 225–226. http://dx.doi.org/10.2466/pms.1977.45.1.225

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under Uncertainty: Heuristics and Biases. *Science, 185,* 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

Walker, D., & Vul, E. (2014). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science, 25,* 230–235. http://dx.doi.org/10.1177/0956797613497969

Ward, L. M. (1990). Mixed-method mixed-modality psychophysical scaling. *Perception & Psychophysics, 48,* 571–582. http://dx.doi.org/10.3758/BF03211603

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics, 9,* 73–78.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004, April 1). Adaptation to natural facial categories. *Nature, 428,* 557–561. http://dx.doi.org/10.1038/nature02420

Wedell, D., Parducci, A., & Geiselman, E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology, 23,* 230–249. http://dx.doi.org/10.1016/0022-1031(87)90034-5

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2,* 1497–1517. http://dx.doi.org/10.1111/j.1751-9004.2008.00109.x

Zellner, D. A., Rohm, E. A., Bassetti, T. L., & Parker, S. (2003). Compared to what? Effects of categorization on hedonic contrast. *Psychonomic Bulletin & Review, 10,* 468–473. http://dx.doi.org/10.3758/BF03196508

Received June 5, 2014
Revision received January 8, 2015
Accepted February 12, 2015 ■