

Simultaneous perceptual and response biases on sequential face attractiveness judgments

Teresa K. Pegors*, Marcelo G. Mattar*, Peter B. Bryan, Russell A. Epstein

Department of Psychology and Center for Cognitive Neuroscience at the University of
Pennsylvania, Philadelphia, Pennsylvania, 19104

*equal contribution

This article may not exactly replicate the final version published in the APA journal. It is
not the copy of record. The final copy can be found at:
psycnet.apa.org/doi/10.1037/xge0000069

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

The authors declare no competing financial interests.

Acknowledgements: The authors are grateful to Steven Marchette for being part of many
discussions that provoked this line of research, Maria Olkkonen and Geoffrey Aguirre for
their helpful comments on an earlier draft of the manuscript, and the three anonymous
reviewers whose comments significantly improved this manuscript. This work was
supported by NIH/NEI Grant EY-022751 to RAE. TKP was supported by the National
Science Foundation Graduate Research Fellowship under Grant No. DEG-0822.

Abstract

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 three experiments, we used a novel sequential rating paradigm that enabled us to measure biases on attractiveness judgments from the previous face and the previous rating. Our results revealed two simultaneous and opposing influences on face attractiveness judgments that arise from our past experience of faces: a response bias in which attractiveness ratings shift *towards* a previously given rating, and a stimulus bias in which attractiveness ratings shift *away* from the mean attractiveness of the previous face. Furthermore, 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.

Introduction

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 solely 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), 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 based on 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 paper, 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 undergraduate males 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). Males who were watching *Charlie’s Angels*, a show with 3 beautiful women as the main characters, rated the girl in the photograph as less attractive than did other males who were watching another TV show (Kenrick & Gutierrez, 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 & Gutierrez, 1980; Wedell et al., 1987).

Interestingly, this sequential contrast bias occurs for other kinds of judgments as well, including both hedonic and non-hedonic judgments (Kamenetzky, 1959; Parker, Bascom, Rabinovitz, & Zellner, 2008; Schifferstein & Frijters, 1992; Schifferstein & 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 (Kamenetzky, 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, or 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 previous results, Kondo et al. have reported an *assimilative* bias for sequential face attractiveness judgments (2012). In this study, subjects made sequential attractiveness judgments of faces using a 1-7 Likert scale. Their 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 originally described this effect as one in which a person’s current decision will be biased towards a previously given value onto which they “anchor and adjust” (Tversky & Kahneman, 1974). 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 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. While 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 effects in the presence of this multicollinearity (Neter, Wasserman, & Kutner, 1989). A more effective, yet un-tested, 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 alternate the type of judgment on every other trial, which allows us to obtain estimates of the bias attributable to the attractiveness of the previous face while also estimating the effect attributable to the previous, orthogonal, response. To anticipate, our results show that opposing biases due to both the previous stimulus and response are indeed simultaneously present during sequential face attractiveness judgments. We then use this paradigm to test the extent of these biases across categories

and to test whether the bias due to the previous stimulus is perceptual or cognitive in nature. We find 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 study 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.

Methods

Stimuli

242 female face images were selected to span a wide range of attractiveness and hair darkness. These came from the Glasgow Unfamiliar Face Database, Radboud Database

(Langner et al., 2010), the Center for Vital Longevity Face Database (Minear & Park, 2004), CVL Face Database (Peter Peer, <http://www.lrv.fri.uni-lj.si/facedb.html>), Diana Theater Face Database (courtesy of Dr. Robert Schultz at the Center for Autism Research), 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 400x400 pixel backgrounds consisting of phase-scrambled variations of a single scene image (See Figure 1 for example stimuli). From this set of 242 images, 10 were chosen to form a practice set of trial images used across all 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 scale ratings of 543 male and female faces (including the 244 female faces describe above). 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, which was based on the number of subjects used in an earlier experiment that showed a well-powered assimilation effect of 1-back ratings on current ratings of attractiveness using a sequential (but non-alternating) design (Kondo et al., 2012). 32 total Penn undergraduates were recruited and given class credit for their participation. 2 subjects were excluded for not following instructions, leaving us with a total of 30 participants (21 female).

Procedure

Subjects made a total of 106 hair darkness judgments and 106 attractiveness judgments in on a 1-8 Likert 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 seconds each, and between face presentations, a fixation cross appeared on the screen for a randomized interstimulus-interval length of 0-0.5 seconds. 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 8 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 the subject made a judgment, the corresponding outline of the button turned white to reinforce their selection.

A unique face was shown in 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 participants attended to the entire face (and did not just focus on the hair, for example) we asked participants to remember each face for a post-experiment memory test.

To acclimate participants to the range of attractiveness in the experiment, participants 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 range of attractiveness and hair darkness. Participants were instructed to spread their ratings during the main experiment across the full scale based on the range of faces they had seen during the 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 and the "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, subjects rated at their own pace 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 (2 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 for which the reaction time (RT) was less than or equal to 0.2 seconds, as 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 $r = 0.01$, $t(29)=0.42$ $p=0.68$), and an extremely small but trending negative correlation between hair darkness ratings and RT (Pearson's $r = -0.04$, $t(29)= -1.93$ $p=0.06$), indicating that that hair darkness was rated slightly more quickly for faces with darker hair. In the post-experiment memory test subjects correctly rated faces seen during attractiveness judgments as familiar (mean=7.1 out of 10, $t(29)=5.69$, $p<0.001$). However, they were at chance when rating faces seen during hair darkness judgments (mean=5.23 out of 10, $t(29)=0.68$, $p=0.5$), suggesting that subjects paid less attention to face identity when making these judgments. The unfamiliar foils were correctly rejected (mean=14.6 out of 20, $t(29)=6.76$, $p<0.001$).

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 in order 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) (mean Pearson's $r = -0.135$), 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 us 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 non-zero correlation between the

previous response and the previous stimulus value, we also ran our models with an additional $\beta_2 R_{t-1} * \beta_2 S_{t-1}$ interaction term (see below).

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 : $t(29)=2.73$, $p=0.011$), whereas the attractiveness of the preceding stimulus had a significant but negative influence on current judgments of face attractiveness (β_3 : $t(29)=-4.92$, $p<0.001$) (See Figure 2). That is, the effect of the preceding response was assimilative, while the effect of the preceding stimulus was contrastive. Adding the $\beta_2 R_{t-1} * \beta_2 S_{t-1}$ interaction term to the model increased our estimate of a bias due to the previous response (β_2 : $t(29)=2.84$, $p=0.008$) and maintained our estimate of a bias due to the previous stimulus (β_3 : $t(29)=-4.91$, $p<0.001$). The interaction term itself was not significant ($t(29)=-1.38$, $p=0.18$).

The first result (β_2) parallels the assimilative bias seen by Kondo et al. (2012; 2013), 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 occurs across judgment types: hair darkness ratings influence 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).

In our second regression analysis, we sought to determine whether hair darkness ratings also 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: -0.11 , VIF for $S_t = 1.02$, $R_{t-1} = 1.04$, $S_{t-1} = 1.03$). Here, we saw a similar pattern of stimulus results in that there was a significant contrastive influence due to the previous stimulus (β_3 : $t(29) = -5.5$, $p < 0.001$). 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 additional interaction term (β_3 : $t(29) = -5.45$, $p < 0.001$). We also observed a marginal trend towards an assimilative influence from the previous response (β_2 : $t(29) = 1.84$, $p = 0.08$), but this effect was reduced with the addition of the interaction term (β_2 : $t(29) = 1.64$, $p = 0.11$). The interaction term itself was not significant ($t(29) = 0.74$, $p = 0.47$).

To get an estimate of the size of these effects in terms of raw ratings scores, we re-ran 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 to 0.13) and 0.03 for the response effect (β_2 range: -0.09 to 0.19). This means that, for the stimulus effect, holding all other variables constant, if the previous face is 1 “rating unit” (on the Likert 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, on the other hand, if the previous face is rated 1 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 1 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 to 0.04) and 0.01 for the response effect (β_2 range: -0.08 to 0.08).

In summary, in study 1, we created an experimental design that effectively separated out possible biases due to 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. Furthermore, a more general mechanism appears to be driving this effect: the contrastive bias due to the previous stimulus was not to unique to attractiveness judgments, as hair

darkness ratings were biased in the same manner. Additionally, face attractiveness ratings were assimilated towards hair darkness ratings, suggesting that the response effect as well 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

While our first study showed that both the previous stimulus and response biased attractiveness judgments even across task, we wanted to explore the boundary conditions of these two effects. In our second study, we not only alternated 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 conceptual distance between trials by testing the effects of *place* beauty and an orthogonal place rating on face attractiveness judgments.

Methods

Stimuli

The same 242 female face images were used from experiment 1. Additionally, 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 400x400 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). 7 face and 7 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.

Procedure

We again set our *a priori* N to 30 to match the number of participants used in study 1. We ran a total of 31 Penn undergraduate participants, and excluded one participant due to a technical error, leaving us with 30 participants (18 females). Participants received course credit for their participation.

During the experiment, face and place trials were alternated, with participants rating the temperature of the place images on a scale of 1-8 (an orthogonal judgment to place beauty), and the attractiveness of face images on a scale of 1-8. The design and procedure were similar to that 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 (7 place, 7 face) that were not shown in the main experiment and which were chosen to span the range of the attractiveness and temperature. After the practice, subjects were instructed to spread their ratings across the scale based on the images they had just seen. They were also instructed to remember all of the images for a post-experiment 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 where the reaction time (RT) was less than or equal to 0.2 seconds was excluded. There was no correlation between RT and face attractiveness (Pearson's $r = 0.01$, $t(29)=0.45$ $p=0.66$) nor between RT and place temperature (Pearson's $r = 0.01$, $t(29)=0.56$ $p=0.58$). In the post-experiment memory test, subjects correctly reported images seen during the experiment as familiar (places: mean=14.9 out of 20, $t(29)=29.45$, $p<0.001$; faces: mean=14.93 out of 20, $t(29)=32.66$, $p<0.001$) and correctly rejected the unfamiliar foils (mean=29.9, $t(29)=45.93$, $p<0.001$). There was no significant difference between the number of faces and places remembered ($t(29)=-0.06$, $p=0.95$).

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 to place temperature (see Equation 1). There was a very low correlation between the temperature judgments and place attractiveness (Pearson's $r = -0.03$, averaged across subjects), suggesting that our design successfully decoupled the potential effects from the previous stimulus and response. Our test for multicollinearity using the variance inflation factor (VIF) on each of the predictors showed low numbers similar to experiment 1, indicating that multicollinearity was not a concern (VIF for $S_t = 1.02$, $R_{t-1} = 1.04$, $S_{t-1} = 1.04$). For completeness, we also ran the model with an additional $\beta_2 R_{t-1} * \beta_2 S_{t-1}$ interaction term, to control for any effect of the non-zero correlation between the previous response and the previous stimulus value.

Notably, our regression analysis did *not* reveal biases on face attractiveness ratings due to either the preceding place stimulus or the previous place temperature rating (beta weights across subjects in t-test against zero: previous temperature rating (β_2): $t(29)=1.64$, $p=0.11$; previous place attractiveness (β_3): $t(29)=1.47$, $p=0.15$). Thus, in contrast to the previous experiment, where 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 non-significant after adding the interaction term (β_2 : $t(29)=0.89$, $p=0.38$; β_3 : $t(29)=0.74$, $p=0.47$). The interaction term was also not significant ($t(29)=-0.4$, $p=0.69$). (While 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 due to the change of stimulus category rather than to other factors, we ran another regression analysis modeling the previous 4 trials, rather than just the previous trial. This allowed us to look for any significant effects from previous face trials (trials which were 2-back and 4-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, since they both measured attractiveness. While significant results are meaningful, the true strength of the effect cannot be characterized.) Using this model, we showed a significant assimilative influence of both the 2-back and 4-back face attractiveness responses on the participants' current face attractiveness judgment (2-back

rating: $t(29)=3.67$, $p<0.001$; 4-back rating: $t(29)=2.8$, $p=0.009$), and a significant contrastive influence due to both the 2-back and 4-back faces (2-back face attractiveness effect: $t(29)=-6.94$, $p<0.001$; 4-back face attractiveness effect: $t(29)=-4.0$, $p<0.001$) (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 assimilative biases. While evidence from our first study 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 2- and 4-back trials, our lack of significant weighting on the 1-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, as the strength of the influence is modulated in an alternating fashion by the trial type.

Experiment 3

In our third study, 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 (Kamenetzky, 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 based on the previous face. An alternative interpretation for the contrastive bias is that a type of cognitive re-mapping takes place between faces and the scale itself. Under this alternative explanation, subjects would not perceive the next face differently, but rather, consciously or subconsciously re-map certain faces or face types to the numbers on ratings scale on a trial-by-trial basis (*"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 so that the length of time that each face was on the screen varied in duration. Previous research on face aftereffects has shown that contrastive effects are strengthened logarithmically with the length of stimulus presentation, and 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. On the other hand, if the contrast effect is due to cognitive "re-mapping," we would not expect stimulus duration to have a selective effect upon the contrastive bias. Additionally, the assimilative bias due to the previous *response* should not be modulated by the previous stimulus

duration, because there is no reason to think the response bias (due to the previous numerical rating) is perceptual in nature.

Methods

Subjects

We conducted the experiment online using Amazon Mechanical Turk (mTurk) 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 affect of trial duration. A liberal sample of 415 subjects were recruited to account for both the greater variability in the online subject pool and to account for the fact that the stimulus duration manipulation cut in half the number of trials used for each model. Our selection criteria for online subjects was that they lived in the United States, they were approved in at least 95% of their mTurk task submissions, and that they 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 was very similar to 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 judgments. For a pseudo-random half of these trials, faces were presented for a 1000 ms duration with a 3000 ms ISI, and for the other half of these trials, faces were presented a 3750 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, 4 trials 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 seconds 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 where 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 (5 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 model if their individual attractiveness ratings had a correlation score of less than 0.5 with the norm ratings, and subjects were excluded from the hair darkness model if their individual hair darkness ratings had a correlation score of less than 0.5 with the norm ratings. Subjects were also excluded if they missed more than 5 of the 200 trials. This left us with 347 subjects for our attractiveness judgment models and 391 subjects in our hair darkness judgment models. Any trial where the reaction time (RT) was less than or equal to 0.2 seconds was also excluded. In the attractiveness judgments model, there was a significant yet small positive correlation between attractiveness ratings and RT (Pearson's $r = 0.11$, $t(346) = 10.58$, $p < .00001$), meaning that subjects' response time increased as face attractiveness increased. In the hair darkness judgments model, there was no correlation between hair darkness ratings and RT (Pearson's $r = -0.01$, $t(390) = -1.31$, $p = 0.19$).

Subjects correctly reported faces seen during attractiveness judgments as familiar (subjects used in the attractiveness model: mean=6.52 out of 10, $t(346) = 67.66$, $p < .001$; subjects used in the hair darkness model: mean=6.48, $t(390) = 69.67$, $p < .001$), and correctly reported faces seen during hair darkness trials as familiar (subjects used in the attractiveness model: mean=5.30 out of 10, $t(346) = 49.65$, $p < .001$; subjects used in the hair darkness model: mean=5.37 out of 10, $t(390) = 53.06$, $p < 0.001$), and correctly rejected unfamiliar faces (subjects used in the attractiveness model: mean=13.40 out of 20, $t(346) = 81.02$, $p < .001$; subjects used in the hair darkness model: mean=13.12 out of 20, $t(390) = 80.12$, $p < 0.001$). Subjects remembered faces during attractiveness trials better

than during hair darkness trials (subjects used in the attractiveness model: $t(346)=-9.92$, $p<0.001$; subjects used in the hair darkness model: $t(390)=-9.47$, $p<0.001$).

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 showed very robust contrastive biases due to the previous stimulus (attractiveness judgments model: effect of previous face attractiveness: $t(346)=-10.07$, $p<0.001$; hair darkness judgments model: effect of previous trial hair darkness: $t(390)=-11.48$, $p<0.001$). We also showed robust assimilative biases in both models due to the previous rating (attractiveness trials model: effect of previous hair darkness rating: $t(346)=3.03$, $p=0.003$; hair darkness trials model: effect of previous attractiveness rating: $t(390)=6.22$, $p<0.001$). Similar to our first experiment, there was a low negative correlation in our attractiveness model between the two 1-back predictor variables of hair darkness judgments and face attractiveness (Pearson's $r = -0.11$, averaged across subjects), and nearly no correlation in our hair darkness model between the two 1-back predictor variables (Pearson's $r = -0.06$, averaged across subjects). The average variance inflation factor (VIF) across subjects was low for each of the three predictors in both models (attractiveness model VIFs for $S_t = 1.02$, $R_{t-1}=1.04$, $S_{t-1}=1.04$; hair darkness model VIFs for $S_t = 1.02$, $R_{t-1}=1.05$, $S_{t-1}=1.05$). Finally, when we ran the models with the additional $\beta_2 R_{t-1} * \beta_2 S_{t-1}$ interaction term to control for any shared variance between the two predictors, all terms were still

significant (attractiveness model: previous stimulus: $t(346) = -9.90$, $p < 0.00001$, previous response: $t(346) = 3.38$, $p < 0.0001$; hair darkness model: previous stimulus: $t(390) = -4.54$, $p < 0.00001$, previous response: $t(390) = 2.48$, $p = 0.01$).

We then turned to the main goal of our experiment: to test the hypothesis that longer stimulus duration in the preceding trial would strengthen the contrast bias but not the assimilation bias, a pattern of results which 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 due to the previous stimulus were highly significant in both duration conditions (trials preceded by short duration trials: $t(346) = -5.31$, $p < 0.001$; trials preceded by long duration trials: $t(346) = -9.27$, $p < 0.001$), but notably, judgments preceded by faces shown for a long duration showed a significantly greater contrast effect than judgments preceded by faces shown for a short duration ($t(346) = -3.30$, $p = 0.001$) (See Figure 4). Additionally, while assimilative biases due to the previous rating were also significant in both conditions (trials preceded by short duration trials: $t(346) = 2.0$, $p = 0.047$; trials preceded by long duration trials: $t(346) = 2.56$, $p = 0.011$), there was no difference in the strength of the effect between the conditions ($t(346) = 0.41$, $p = 0.68$) (See Figure 4). Furthermore, we also showed that the difference in contrast strength between short/long presentations was significantly greater than the difference in assimilation strength between short/long presentations (the “difference of differences”) ($t(346) = -2.92$,

$p=0.0037$, repeated measures t-test on differences scores). Furthermore, 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 due to the previous stimulus were highly significant in both duration conditions (trials preceded by short duration trials: $t(390)=-5.37$, $p<0.001$; trials preceded by long duration trials: $t(390)=-10.82$, $p<0.001$), yet 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 ($t(390)=-4.44$, $p<0.001$). There was no difference between the strength of the assimilative biases for each trial type ($t(390)=-0.80$, $p=0.42$), even while assimilative biases due to the previous rating were significant in both conditions (trials preceded by short duration trials: $t(390)=4.95$, $p<0.001$; trials preceded by long duration trials: $t(390)=4.37$, $p<0.0001$). In addition, we also showed for hair darkness judgments that the difference in contrast strength between short/long presentations was significantly greater than the difference in assimilation strength between short/long presentations ($t(390)=-2.57$, $p=0.011$).

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. On the other hand, the assimilative bias did not show such a modulation, suggesting a different, non-perceptual mechanism. These results also demonstrate the robustness of our design to reveal sequential biases even in the presence of an additional condition manipulation (which decreases the number of trials per condition per subject, and adds a level of subjective complexity for participants).

General Discussion

To navigate the social world, it is important to be able to evaluate face attractiveness, but these judgments are always made in relation to a larger social and environmental context. In this paper, we introduce a novel sequential rating paradigm that we use to pinpoint the source of at least two opposing contextual influences on face attractiveness judgments. First, we show that our judgment of a face will tend to assimilate towards the response that we gave in 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 view in 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. Furthermore, we used a modified sequential rating paradigm to provide evidence that this contrastive bias may be due 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 & Gutierrez, 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 showed contrastive biases for two separate perceptual characteristics (attractiveness and hair darkness) but did not show a contrastive bias across different perceptual categories can be explained with either a cognitive or perceptual interpretation.

According to a cognitive interpretation, it may have been the case that subjects were re-mapping facial features to the ratings scale on a trial-by-trial basis, finely adjusting what rating they will give to what type of face, based upon the recent attractiveness history. Alternatively, according to a perceptual interpretation, the contrastive bias may have reflected a subtle change in how the subject perceptually processes the face, in a similar manner to a perceptual aftereffect thought to be driven by neural adaptation in visual cortex. In our third study, we made use of the fact that perceptual aftereffects are strengthened by the duration of the “adapting” stimulus to show that longer face presentations in the previous trial strengthened the contrast bias on 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 show an equally strong assimilative bias for both presentation durations.

In fact, a perceptual interpretation is often used to explain contextual biases on 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. used perceptual aftereffects as an explanation of 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 (2003). 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 reveal that previous ratings given to faces positively predict current attractiveness ratings. These results replicate the assimilative effect on face attractiveness seen by Kondo et al. (2012, 2013), but we extend their findings by linking the effect directly to the previous rating. It may be the case that Kondo et al. observed an overall assimilative effect due to 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) provide 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 differ 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). On the other hand, we did show *cross-task* assimilation within the same stimulus category, suggesting some level of generality to the effect. Future studies should address, therefore, exactly the set of conditions under which this bias is present.

Relevance to Sequential Tasks

Our results reveal at least two bias-inducing mechanisms that reinforce researchers’ motivation to randomize trial order for each subject when acquiring mean estimates of stimuli. Since 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 other appropriate experimental designs and analyses, by taking into account the fact that both previous subject responses and 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 paper, we introduce a novel method that uses alternating judgments to quantify simultaneous biases due to the previous stimulus and the previous response on current judgments. Our results demonstrate 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 on 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(1), 16–22.
doi:10.1037/a0031020
- Critcher, C., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making*, 251(October 2007), 241–251. doi:10.1002/bdm
- Decarlo, L. T., & Cross, D. V. (1990). Sequential Effects in Magnitude Scaling : Models and Theory. *Journal of Experimental Psychology: General*, 119(4), 375–396.

- Gibson, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, 16(1), 1–31.
- Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology. Human Perception and Performance*, 3(1), 92–104.
- Kamenetzky, J. (1959). Contrast and convergence effects in ratings of foods. *Journal of Applied Psychology*, (1), 47–52.
- Kenrick, D. T., & Gutierrez, S. E. (1980). Contrast Effects and Judgments of Physical Attractiveness : When Beauty Becomes a Social Problem. *Journal Personality and Social Psychology*, 38(1), 131–140.
- 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(4), 617–620.
- Kondo, A., Takahashi, K., & Watanabe, K. (2012). Sequential effects in face-attractiveness judgment. *Perception*, 41, 43–49. doi:10.1068/p7116
- Kondo, A., Takahashi, K., & Watanabe, K. (2013). Influence of gender membership on sequential decisions of face attractiveness. *Attention, Perception & Psychophysics*. doi: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 & Emotion*, 24(8), 1377–1388.

doi:10.1080/02699930903485076

Leopold, D. a, Rhodes, G., Müller, K.-M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings. Biological Sciences / The Royal Society*, 272(1566), 897–904. doi:10.1098/rspb.2004.3022

Leopold, D., O’Toole, A., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4(1), 89–94. doi:10.1038/82947

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(1), 64–81.

McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science*, 149(3688), 1115–1116.

Miner, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, 36(4), 630–3.

Morgan, M. J., Melmoth, D., & Solomon, J. a. (2013). Linking hypotheses underlying Class A and Class B methods. *Visual Neuroscience*, 30(5-6), 197–206.

doi: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(3), 171–174. doi:10.1037/1931-3896.2.3.171
- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, 57, 199–226. doi: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(17), 2291–6. doi: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(6), 558–566. doi:10.1046/j.0956-7976.2003.psci
- Schiffstein, H. N. J., & Frijters, J. A. N. E. R. (1992). Contextual and sequential effects on judgments of sweetness intensity. *Perception & Psychophysics*, 52(3), 243–255.
- Schiffstein, H. N. J., & Kuiper, W. E. (1997). Sequence effects in hedonic judgments of taste stimuli. *Perception & Psychophysics*, 59(6), 900–912.
- Strane, K., & Watts, C. (1977). Females judged by attractiveness of partner. *Perceptual and Motor Skills*, 45, 225–226.

- Tversky, a, & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–31. doi:10.1126/science.185.4157.1124
- Walker, D., & Vul, E. (2014). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science*, 25(1), 230–5. doi:10.1177/0956797613497969
- Ward, L. (1990). Mixed-method mixed-modality psychophysical scaling. *Perception & Psychophysics*, 48(6), 571–582.
- Ward, L., & Lockhead, G. (1971). Response system processes in absolute judgment. *Perception & Psychophysics*, 9(18).
- Webster, M., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428(April), 357–360. doi:10.1038/nature02361.1.
- 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.
- Zebrowitz, L. a, & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, 2(3), 1497. doi: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(2), 468–73.

Figure 1: “Alternating” experimental design. Subjects rated either the attractiveness or hair darkness of each female face on a Likert scale of 1-8.

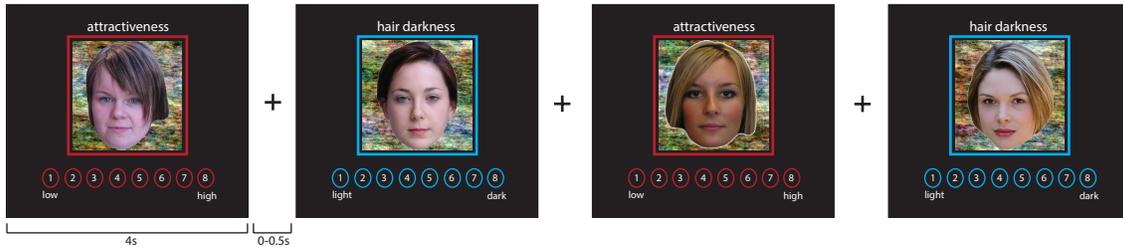


Figure 2: Regression results from the alternating face attractiveness/hair darkness design. Face attractiveness judgments were regressed against the previous response (to hair darkness) and the previous stimulus (the attractiveness of the face, an averaged score from independent subjects). Both elements 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).

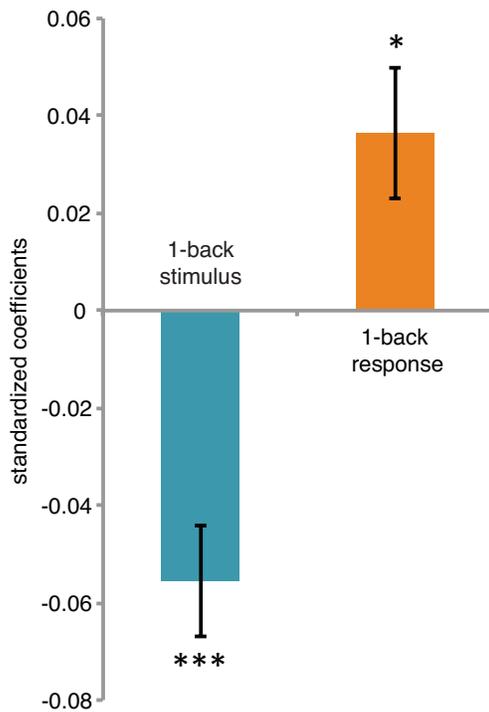


Figure 3: Regression results from the face/place alternating design. Face attractiveness ratings were regressed against the previous rating and the previous response of the four preceding trials. Because of the alternating design, trials 1-back and 3-back were always place trials in which subjects judged “place temperature”, and trials 2-back and 4-back were always face trials in 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 4 trials back showed predictive power related both to the subjects’ response and the mean attractiveness of the face.

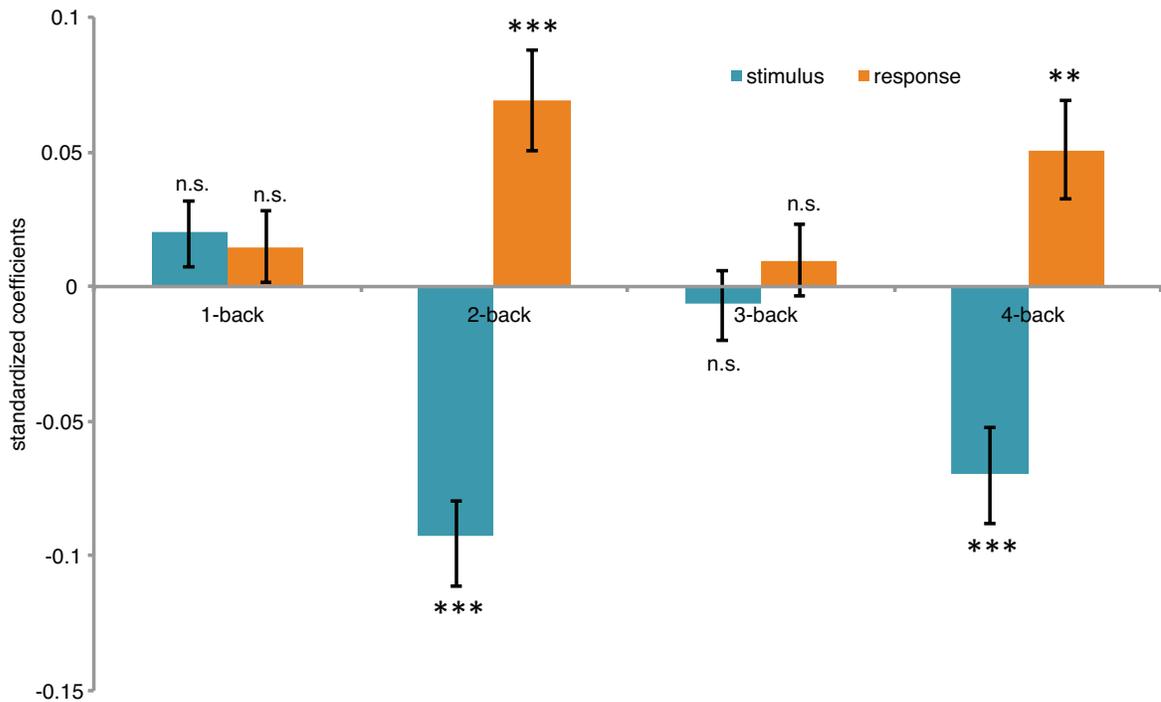


Figure 4: Results from two separate regression models comparing the effects of stimulus duration on the assimilative and contrastive biases. Face attractiveness judgments preceded by short-duration trials were regressed against the previous response (to hair darkness) and the previous stimulus (the attractiveness 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. While the assimilative effect of the previous response was not modulation by duration, the contrastive effect of the previous stimulus was significantly stronger when the previous face was presented for a longer duration.

