Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

2020, Vol. 46, No. 12, 1476–1489
http://dx.doi.org/10.1037/xhp0000869

Sequential Effects in Facial Attractiveness Judgments Using Cross-Classified Models: Investigating Perceptual and Response Biases

Robin S. S. Kramer
University of Lincoln

Alex L. Jones
Swansea University

When evaluating items in a sequence, the current judgment is influenced by the previous item and decision. These sequential biases take the form of assimilation (shifting toward the previous item/decision) or contrast (shifting away). Previous research investigating facial attractiveness evaluations provides mixed results while using analytical techniques that fail to address the dependencies in the data or acknowledge that the images represent only a subset of the population. Here, we utilized cross-classified linear mixed-effects modeling across 5 experiments. We found compelling evidence of multicollinearity in our models, which may explain apparent contradictions in the literature. Our results demonstrated that the previous image's rating positively influenced current ratings, and this was also the case for the previous image's baseline value, although only when that image remained onscreen during the current trial. Further, we found no influence of the next face on current judgments when this was visible. In our final experiment, the response bias due to the previous trial remained present even when accounts involving motor effort were addressed. Taken together, these findings provide a clear framework in which to incorporate current and past results regarding the biases apparent in sequential judgments, along with an appropriate method for investigating these biases.

Public Significance Statement

This study demonstrates the biases we show when evaluating sequences, and how they are best investigated. When considering the attractiveness of faces in sequences, this study explores the biases that are present and the contexts in which they may be absent.

Keywords: sequential effects, facial attractiveness, linear mixed-effects models, perceptual bias, response bias

Supplemental materials: http://dx.doi.org/10.1037/xhp0000869.supp

When presented with items in a sequence and asked to provide judgments of the items, people show predictable biases. For example, research has demonstrated that appearing first or last in a sequence can improve evaluations of those particular items in real-world contexts (Bruine de Bruin, 2005; Mantonakis, Rodero, Lesschaeve, & Hastie, 2009; Miller & Krosnick, 1998). In addition to particular positions in a sequence, studies have also identified an effect of direct comparison, whereby the evaluation of the current item is influenced by the evaluation of the previous item.

These sequential effects can take one of two forms: assimilation, where the current judgment is drawn toward the previous judgment, and contrast, where the opposite pattern is found. Such effects have been identified in a variety of situations, including judgments of singers on the "Idol" TV series (Page & Page, 2010) as well as both synchronized divers and gymnasts at the Olympics (Damisch, Mussweiler, & Plessner, 2006; Kramer, 2017).

Sequential Biases in Facial Attractiveness

A number of studies in recent years have focused on evaluative judgments of face photographs, and specifically, their perceived attractiveness. Findings have consistently demonstrated the presence of sequential biases in the form of an assimilation effect, whereby the current face was considered more attractive when preceded by an attractive (vs. an unattractive) face (Kok, Taubert, Van der Burg, Rhodes, & Alais, 2017; Kondo, Takahashi, & Watanabe, 2012, 2013; Kramer, Jones, & Sharma, 2013; Taubert & Alais, 2016). While these repeated demonstrations have shown how reliable this effect appears to be, its cause remains the subject of investigation.

This article was published Online First September 24, 2020.

Robin S. S. Kramer, School of Psychology, University of Lincoln; Alex L. Jones, Department of Psychology, Swansea University.

We thank Jerrica Mulgrew, Hannah Johnson, Ellen Wheeler, Lauren Hill, and Amy Haigh for their help with data collection. We would also like to thank Michael Reynolds for his input during the early stages of the project, and Chris Stride and Michael Mireku for their advice regarding data analysis.

Correspondence concerning this article should be addressed to Robin S. S. Kramer, School of Psychology, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, United Kingdom. E-mail: remarknibor@gmail.com

Two mechanisms have been proposed to explain biases when responding to faces in a sequence (e.g., Huang et al., 2018; Pegors, Mattar, Bryan, & Epstein, 2015; Taubert, Van der Burg, & Alais, 2016). The first can be classified as perceptual, whereby low-level mechanisms (e.g., sensory adaptation) result in a comparison between the attractiveness of the current and previous face. This explanation incorporates the notion that the attractiveness of the preceding face itself is influencing the present evaluation. In comparison, the second mechanism can be thought of as a higher-order response bias. Simply, the response given to the previous face is the factor influencing the current response. This might be apparent, for example, in situations where the current face is difficult to evaluate, and so participants default to repeating their previous response. Teasing apart these two mechanisms has been the goal of several studies in recent years.

In psychophysics, researchers have increasingly considered the influences of the value of the current stimulus, the value of the previous stimulus, and the judgment of the previous stimulus. The typical pattern of results suggests that the current judgment assimilates to the previous judgment while simultaneously contrasting away from the previous item's value (e.g., Jesteadt, Luce, & Green, 1977). This pattern is also evident for more complex, real-world decision making (e.g., prices of items; Matthews & Stewart, 2009). Problematically when considering facial attractiveness, there is no objective, "true" value that can be used in predictive models. While raters show a high level of agreement in their judgments, there is also a large influence of personal taste (e.g., Kramer, Mileva, & Ritchie, 2018).

One approach that researchers have taken is to focus on the nature of the perceptual bias by removing the possibility of a response bias. Chang, Kim, and Cho (2017) showed two-image sequences of artistic photographs (mostly scenes), with participants rating only the second image. In this way, the assimilation effect that was found could not have been because of a response bias because no response was given to the previous image. In addition, and using a similar approach, this same result was evident with face images (Xia, Leib, & Whitney, 2016).

Another approach in attempting to disentangle perceptual and response biases is to quantify the current face's value, along with that of the previous face, by collecting measures of baseline attractiveness. By asking a different set of participants to pre-rate the images, presenting the faces in a random order each time, and subsequently averaging ratings across participants, each face receives an attractiveness value. These values can be considered independent of any particular task or stimulus history, and represent a proxy for their actual value when predicting ratings given during the experiments (Huang et al., 2018; Pegors et al., 2015). Using this method, results have shown that evaluations of the current face were assimilated toward those given to the previous face but contrasted away from the value of the previous face. Assimilation to the previous response was present both across task boundaries (rating the attractiveness of a face after rating the hair darkness of the previous face; Pegors et al., 2015) and when responses were given orally (Huang et al., 2018). However, a contrast with the value of the previous stimulus was absent across task boundaries (rating the attractiveness of a face after rating the agreeableness of a ringtone; Huang et al., 2018).

Taken together, there remains some uncertainty in the pattern of biases shown. By investigating both perceptual and response biases in attractiveness judgments within the same design, evidence suggests that evaluations assimilate toward the previous response but contrast away from the previous image's attractiveness (Huang et al., 2018; Pegors et al., 2015). This general pattern appears robust within the literature, applying to both low-level visual and more complex, real-world decisions (e.g., Jesteadt et al., 1977; Matthews & Stewart, 2009). Indeed, this mirrors psychophysical evidence that perception is repelled away from the previous item while postperceptual decisions are attracted toward that item (Fritsche, Mostert, & de Lange, 2017). In contrast, in designs where attractiveness response biases were addressed in other ways (e.g., no response was given to the previous trial, Xia et al., 2016; upright faces were interleaved with inverted faces, Taubert et al., 2016), evaluations assimilated toward the previous image's value. Therefore, further investigation is necessary to provide clarity on this issue.

Reducing Sequential Biases

While studies have continued to identify sequential effects in judgments of attractiveness, several factors have been shown to influence their presence and/or strength. For example, assimilation to the previous response decreased when the current face was presented at a different orientation to the previous face (Taubert et al., 2016). Reduced assimilation has also been found when the current and previous face differed in sex or race (Kondo, Takahashi, & Watanabe, 2013; Kramer et al., 2013). These results support the idea that comparison with the previous face may be dependent on the two faces being sufficiently similar, for example, falling into the same perceived category. Indeed, the degree of perceived similarity between two consecutive items may determine whether assimilation (high perceived similarity) or contrast (low) takes place (Fritsche et al., 2017; Mussweiler, 2003; Mussweiler, Rüter, & Epstude, 2004).

Researchers have also suggested that timing may be an important factor when determining the nature of sequential effects (Taubert, Alais, & Burr, 2016). For example, very brief presentations have been associated with assimilation while longer durations resulted in contrast effects (Kanai & Verstraten, 2005). However, assimilation has been found for longer presentation durations and when responses were self-paced (Kondo et al., 2012; Kramer et al., 2013; Pegors et al., 2015), although longer intervals between adjacent ratings may decrease the magnitude of the effect (Attali, 2011; Xia et al., 2016). Indeed, it is possible that timing affects perceptual and response biases differently, with the suggestion that perceptual biases may be assimilative for short presentation durations and temporal intervals, while decreasing and/or becoming contrastive at longer durations and intervals (Xia et al., 2016). However, response biases may simply decrease with an increased interval between responses.

To our knowledge, all studies investigating sequential effects have presented items individually. That is, the previous face was replaced onscreen by the current face (e.g., Kramer et al., 2013). As a result, the perceptual comparison that participants make has been to the *memory* of the previous face rather than the face itself. Could this reduce the contrast effect typically shown to the attractiveness of the previous face? One might predict a stronger (or different) influence of the previous face if it remained onscreen when the current face was presented and judged. Indeed, past research has shown that evaluations of two faces presented simultaneously were assimilated while the face pair was contrasted away from the previous pair (Wedell, Parducci, & Geiselman, 1987). As such, the continued presence of the previous face may produce assimilation of the current rating toward the previous image's value, reversing the typical pattern. In addition, research investigating the "cheerleader effect" has shown that faces may simply appear more attractive when presented together rather than in isolation (e.g., Walker & Vul, 2014). The presence/visibility of the previous item mirrors numerous real-world situations in which shoppers are browsing shelves, speed-daters are meeting potential mates, and so on.

Continuing this line of reasoning, one might also consider the effect of the next item on current judgments if it too were visible. Again, when browsing a sequence of items that are all present, one might allocate some attention to the next item while considering the current one. Although yet to respond to the next item, its value may still influence the rating given to the current item. In combination with the continued presence of the previous item, it is unclear how the two may affect the current response. It is possible that the current evaluation is assimilated toward both the next and previous images' attractiveness levels. However, having already responded to the previous item, it may be that that item shows a stronger influence on current judgments in comparison with the next item. These ideas have yet to be tested and provide the main focus for the current set of experiments.

Regarding the influence of the previous response itself, which is typically assimilative (Huang et al., 2018; Pegors et al., 2015), the effect appears to be surprisingly robust. As mentioned above, even when responses were given orally, participants continued to show a response bias in their attractiveness ratings (Huang et al., 2018). Although some form of action repetition might be expected to underlie keyboard-based response biases, this explanation seems less applicable to mouth movements. For example, while finger position could mean less effort by repeating a keypress, saying out loud the same number again does not produce any apparent gains in terms of expended effort. Perhaps another way to remove the possible advantages that come with action repetition is to make sure that all motor responses are equally effortful. In our final experiment, we investigate a method through which response bias might be reduced.

Stimuli as a Random Factor

An important issue regarding previous studies investigating sequential biases involves the method of analysis used. In these experiments, samples of participants respond to samples of stimuli. That is, participants represent a subset taken from a larger population *and* the sample of stimuli presented are drawn from a larger stimulus population of interest. In statistical terms, both participants and stimuli should be thought of as random factors. Given that some amount of variance across experiments is because of the particular subset of stimuli used, the appropriate analyses must treat stimuli as a random factor to avoid statistical biases (e.g., Wells & Windschitl, 1999; Westfall, Judd, & Kenny, 2015).

Typically, researchers have followed a *by-participant* analysis, where data are collapsed across stimuli to provide a measure for each participant, with these values subsequently analyzed (e.g., Kondo et al., 2012, 2013). For example, some researchers have utilized by-participant (or "random") regression to determine the coefficients for each participant, and then conducted a one-sample *t* test to compare these to a value of zero (Huang et al., 2018; Matthews & Stewart, 2009; Pegors et al., 2015). This method implicitly treats participants, but not stimuli, as a random factor. As such, results can be generalized to other samples of participants but not to other samples of stimuli. If researchers then choose to claim the latter, this will give rise to inflated Type I error rates (Judd, Westfall, & Kenny, 2012; Murayama, Sakaki, Yan, & Smith, 2014; Westfall, Kenny, & Judd, 2014).

Less common is the *by-stimulus* analysis, where data are collapsed across participants to provide a measure for each stimulus. As above, implicitly treating stimuli, but not participants, as a random factor means that results can be generalized to other samples of stimuli but not to other samples of participants (again resulting in inflated Type I error rates; Judd et al., 2012). In some instances, researchers might even include both types of analysis, assuming that both significant by-participant and by-stimulus results permit generalization across both participants and stimuli. However, this reasoning is flawed in that the given results cannot be said to generalize to *simultaneously* new samples of both participants and stimuli because each can only be claimed when the other remains unchanged (Raaijmakers, Schrijnemakers, & Gremmen, 1999).

In recent years, advanced statistical techniques have addressed the above-mentioned issues with approaches that consider either participants *or* stimuli as a random factor. Linear mixed-effects models have been developed to effectively incorporate both random factors simultaneously by explicitly modeling the dependencies in the data. Using these methods, researchers can empirically estimate random variance components to determine to what extent effects vary randomly for each grouping factor. By taking into account the dependencies in the data (e.g., ratings given to a specific item will be more alike, as might ratings given by a specific participant), researchers are also able to allow for variation in the relationships between variables across higher level units (e.g., participants may differ in how their previous response influences their current one). Fundamentally different relationships between variables may be found when the clustering of a dataset by higher level units is controlled for. In contrast, if data points are analyzed without this clustering (e.g., Xia et al., 2016), one risks encountering "Simpson's Paradox" (Simpson, 1951), whereby a trend evident in several different groups of data disappears or reverses when the groups are combined. Surprisingly, although linear mixed-effects models have started to feature in sequential effects research (e.g., essay ratings; Zhao, Andersson, Guo, & Xin, 2017), these techniques have yet to be applied to ratings of facial attractiveness.

Current Study

There were four main aims for this study. First, previous evidence is somewhat contradictory as to the specific nature of the biases found during evaluations. For instance, whether responses assimilate toward (e.g., Xia et al., 2016) or contrast away from (e.g., Pegors et al., 2015) the attractiveness of the previous face remains unclear. The current set of experiments was able to provide additional evidence in this debate.

Second, research in this field has always considered sequential effects when viewing faces individually, one after another. To date, no studies have investigated contexts in which the previous and next item were visible during evaluation of the current item, a common occurrence in real-world responding. Therefore, the current study began to address this omission.

Third, recent research has provided evidence that participants demonstrate both assimilation and contrast effects in their responding. However, no studies to date have utilized linear mixed-effects models, considering both participants (raters) and stimuli (faces) as random factors. As a consequence, it was important to confirm this pattern of results using analyses that explicitly modeled the dependencies in the data while incorporating both random factors. The current study utilized this statistical approach to determine whether previous results were replicable under these conditions.

Fourth, previous research has failed to remove the response bias found in judgments through the use of oral responses (Huang et al., 2018). Here, the final experiment investigated an alternative method through which this type of bias might be reduced, where all motor responses require the same amount of effort, to minimize action repetition.

Experiment 1: One Image Visible

In the first experiment, we investigated whether ratings of attractiveness demonstrated both assimilation and contrast effects, in line with previous research (Huang et al., 2018; Pegors et al., 2015), when analyzed using a linear mixed-effects modeling approach. Participants viewed individual faces onscreen, one after another, and provided judgments of attractiveness, following the typical procedure in this field.

Method

Participants. Thirty-six university students (M_{age} = 19.7 years, SD_{age} = 1.0 years; 30 women; 89% self-reported as White) gave informed written consent before participating in the experiment and were verbally debriefed upon completion. Participants received course credit as compensation.

The sample sizes for this experiment and those that followed were based on the number of participants used in earlier studies showing both assimilation and contrast effects in sequential attractiveness ratings (25 to 32 participants, Huang et al., 2018; 30 participants, Pegors et al., 2015).

In addition, through simulations based on our data (SIMR package; Green & MacLeod, 2016), we calculated the power to detect fixed effects of the sizes reported by Huang et al. (2018, Experiment 1). First, we fitted a linear mixed-effects model to the data collected in this experiment (described below). Second, we replaced each of our estimated coefficients with those taken from Huang et al. (current image baseline: 1.094, previous image rating: 0.213, previous image baseline: −0.226). Finally, we used the "powerSim" function to carry out power analyses, simulating new values for the response variables using the altered model coefficients (although maintaining the same fixed and random effects structure) and then statistically testing the simulated fit. For each of the coefficients, the power to detect an effect of the size reported by Huang et al. (based on 100 Monte Carlo simulations) was high: 100%, 95% confidence interval [CI: 96.38%, 100.0%]. All experiments presented here were approved by either the University of Lincoln's School of Psychology ethics committee or Trent University's ethics committee, and were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Stimuli. One hundred female faces (all self-reported as White ethnicity) were selected from a larger database of images used in previous research (Kramer, Jones, & Ward, 2012; Scott, Kramer, Jones, & Ward, 2013, 2015). These faces were chosen to span a wide range of attractiveness levels. All images were constrained to reflect neutral expression, eyes on the camera; consistent posture, lighting, and distance to the camera; no glasses or make-up; and hair back.

Before the main experiments presented here, we collected attractiveness ratings of these faces from 41 participants (M_{age} = 30.2 years, SD_{age} = 15.4 years; 28 women; 93% self-reported as White), none of whom took part in the experiments that followed. The sample size was based on previous studies using this method of collecting baseline ratings (30 participants, Huang et al., 2018; 28 participants, Pegors et al., 2015). For each participant, the 100 images were presented online, one at a time, using the Qualtrics survey platform (www.qualtrics.com). Responses to the question "How attractive is this face?" were given on a 0 (*very unattractive*) to 9 (*very attractive*) Likert scale, and were self-paced. To minimize sequential bias effects, images were presented in a random order for each participant, and ratings were subsequently averaged across raters for each face. The resulting attractiveness scores (M = 3.35, SD = 0.80) were considered independent of any particular task or stimulus history, and served as baseline stimulus values in the experiments presented here (Huang et al., 2018; Pegors et al., 2015).

Procedure. Participants provided demographic information regarding their age, sex, and ethnicity. During the task itself, participants viewed the 100 faces (image size approximately 6.8 × 10.3 cm), presented in a random order. For each image, participants were asked "How attractive is this face?" and provided their response using a 0 (*very unattractive*) to 9 (*very attractive*) Likert scale. The task was self-paced, with images remaining onscreen until a response was given with the mouse. See Figure 1a for an example.

Rather than the current image simply being replaced with the next image after a response was given, the face was shown to move left and off the screen as the next image in the sequence moved on to the screen from the right-hand side, appearing similar to a conveyor belt. Once the face arrived in the center of the screen, participants were able to provide a rating, with the face remaining onscreen until a response was given. (This effect was included to highlight the order of the images, which was desirable for Experiments 2 and 3, where multiple images from the sequence were presented at any given time.) Viewing distance was not fixed.

Data analysis. The data were analyzed using linear mixed-effects models with crossed random effects (participants and images) because each participant rated the same series of stimuli. Therefore, participants and stimuli variance were considered at Level 2 and residual variance at Level 1. In terms of the dataset, each participant by image observation was the unit of analysis, with each row of data indicating the attractiveness rating given by that participant to that image, the image's baseline attractiveness (calculated from the prior ratings mentioned above), the attractive-

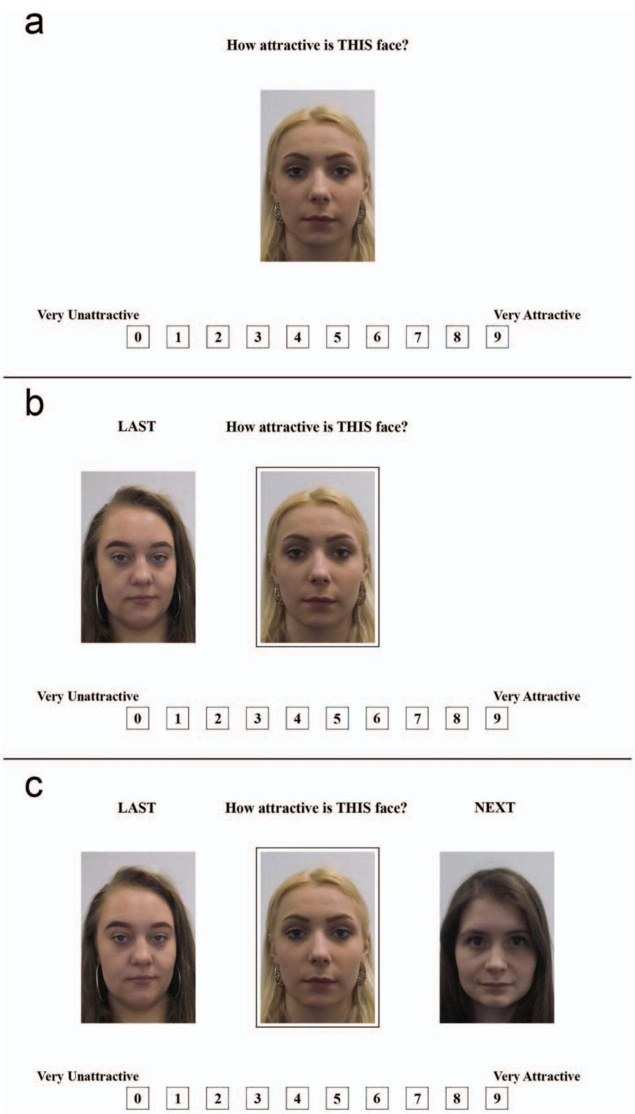

Figure 1. An illustration of the ratings task for three of our experiments. The number of images presented onscreen at a given time differed across (a) Experiment 1, (b) Experiment 2, and (c) Experiment 3. See the online article for the color version of this figure.

ness rating given by that participant to the previous image, and the previous image's baseline attractiveness (again, calculated from prior ratings).

The fixed effects were the intercept, the effect of the current image's baseline attractiveness, the effect of the rating given to the previous image, and the effect of the previous image's baseline attractiveness. Only the intercept in this model varied randomly across images, whereas the intercept and the slopes of the current image's baseline attractiveness, the rating given to the previous image, and previous image's baseline attractiveness varied randomly across participants. In addition, the covariance between random effects was set to zero. Models using more complex random effects structures were identified as singular (Barr, Levy, Scheepers, & Tily, 2013). To make the intercepts more meaningful, baseline attractiveness values were centered around their grand mean. In addition, the attractiveness ratings given by participants to the previous images were group mean centered to avoid conflating lower level (within-participant) and higher level (between-participants) variance.

Modeling was carried out in IBM SPSS Statistics v25, with multilevel analyses performed using the MIXED function. Satterthwaite approximation was applied for calculating degrees of freedom, and restricted maximum likelihood was used to fit the model.

Results

We focus on the fixed effects summarized in Table 1 (under "combined model"). (For more information regarding the fixed and random effects, see Table S1 in the online supplemental materials). Examining the tests of fixed effects, we found that ratings of the current face were, unsurprisingly, positively influenced by the current face's baseline attractiveness. That is, the more attractive the face was considered to be by a separate sample of raters, the higher the rating given to that face by the participant. In addition, the rating given to the previous image was also a positive predictor of the rating given to the current image, demonstrating the presence of an assimilation effect. Finally, the previous image's baseline attractiveness was a negative predictor of the rating given to the current image, representing a contrast effect.

In line with previous research (Wells, Baguley, Sergeant, & Dunn, 2013), we obtained estimates from an intercept-only model regarding the proportion of variance explained by different sources: 19% for stimuli, 42% for participants, and 39% residual error. These values highlighted the importance of including both participants and stimuli as random factors in the model.

As noted in previous research (Huang et al., 2018; Pegors et al., 2015), one concern with a model incorporating both the previous image's baseline *and* rated attractiveness is the presence of multicollinearity. These two predictors are expected to be correlated, given that we know that there is substantial shared taste with regard to judgments of facial attractiveness (Hönekopp, 2006; Kramer et al., 2018). If these variables are closely interrelated then quantifying the relationship between each of these predictors and the rating given to the current image is, in effect, modeling little more than random noise (e.g., in the residuals).

Therefore, we first examined the variance inflation factor (VIF) of each predictor, considering each by-participant regression separately. Across all values, the largest VIF was 2.91. While many suggest that values less than 10 are not cause for concern (e.g., Allison, 1999; Myers, 1990), others have placed this boundary at 2.5 (e.g., Chalmers et al., 2009), and indeed, recent research has demonstrated that even lower values can be associated with substantial confounding (Johnston, Jones, & Manley, 2018; Vatcheva, Lee, McCormick, & Rahbar, 2016). In addition, across participants, we found large correlations between the previous image's baseline and rated attractiveness, mean $r = .59$, 95% CI [0.54, 0.63], with prior research demonstrating that lower levels of correlation than this can be problematic for mixed-effects models (Bonate, 1999; Shieh & Fouladi, 2003). Finally, carrying out a linear mixed-effects model on all participants' data as above, but using the previous image's baseline attractiveness to predict its

Table 1
Parameter Estimates for Linear Mixed-Effects Models Predicting Ratings of Attractiveness Given to the Current Image

Fixed effects	Combined model Estimate	Combined model *SE*	Separate models Estimate	Separate models *SE*
Experiment 1				
Intercept	4.05***	0.18		
Current image baseline	0.86***	0.07	0.86***	0.07
Previous image rating$_c$	0.23***	0.02	0.18***	0.02
Previous image baseline	−0.14***	0.03	0.03	0.03
Experiment 2				
Intercept	4.23***	0.18		
Current image baseline	0.91***	0.07	0.91***	0.07
Previous image rating$_c$	0.24***	0.02	0.19***	0.02
Previous image baseline	−0.13***	0.03	0.07**	0.02
Experiment 3				
Intercept	4.08***	0.21		
Current image baseline	1.01***	0.08	1.01***	0.08
Previous image rating$_c$	0.16***	0.02	0.14***	0.02
Previous image baseline	−0.04	0.03	0.11**	0.03
Next image baseline	−0.01	0.03	−0.01	0.03
Experiment 4				
Intercept	3.46***	0.23		
Current image baseline	1.01***	0.06	1.02***	0.07
Previous image rating$_c$	0.26***	0.02	0.21***	0.02
Previous image baseline	−0.14***	0.03	0.11***	0.02
Next image baseline	0.01	0.03	0.01	0.03
Experiment 5				
Intercept	4.41***	0.18		
Current image baseline	0.91***	0.06	0.91***	0.06
Previous image rating$_c$	0.26***	0.02	0.20***	0.02
Previous image baseline	−0.17***	0.03	0.05	0.03

Note. The combined model included all fixed effects simultaneously, whereas each separate model included only one fixed effect. Subscript "c" denotes that values were mean centered at the participant level.
** $p < .01$. *** $p < .001$.

rated attractiveness, we found that the former was a strong predictor of the latter, $b = 0.87$, $SE = 0.06$, $p < .001$.

Given that we expected *and* observed that these two predictors were highly correlated, we next carried out linear mixed-effects models as before, although including only one of our three predictors in each separate model. As Table 1 (under "separate models") shows, the contrast effect because of the previous image's baseline attractiveness was absent when this variable was included as the only predictor of the current image rating. The predictive value of both the current image's baseline and the previous image's rating remained relatively unchanged.

Discussion

In line with previous research, our model including all three predictors showed that evaluations of the current face were assimilated toward those given to the previous face but contrasted away from the value of the previous face (Huang et al., 2018; Pegors et al., 2015). As such, raters simultaneously demonstrated assimilation and contrast effects in their responding. The use of a linear mixed-effects model, treating both participants and stimuli as random factors, provided evidence that this pattern of results can be generalized to other participant samples *and* other sets of stimuli.

Considering the fixed effects in Table 1, it was clear that the baseline attractiveness of the current image was the most important predictor of the response given. Indeed, there was almost a one-to-one relationship between the two variables: for a one-unit increase in baseline attractiveness on the 0–9 Likert scale, responses increased by a mean of 0.86 units along this same scale. In addition, given that all predictors were measured using the same units and scale, the estimates could be directly compared. The results suggested that the response given to the previous image produced a stronger influence on the current response (a mean increase of 0.23 units) in comparison with the previous image's baseline attractiveness (a mean decrease of 0.14 units), considering only the magnitude and not the direction of the influence, $z = 2.41$, $p = .016$ (see Paternoster, Brame, Mazerolle, & Piquero, 1998). Simply, participants' response biases were larger than their perceptual biases, although both were present.

The model also suggested significant variability across participants in the degree to which ratings given to the current image were influenced by the current image's baseline attractiveness and the ratings given to the previous image. Combined with evidence that substantial variance in responses was explained by both participants and stimuli, these results provided support for the need to incorporate both participants and stimuli as random factors in analyses.

Problematically, we found compelling evidence for multicollinearity within this model. We both predicted and demonstrated that ratings given to the previous image were highly correlated with the previous image's baseline attractiveness. As such, our model incorporating both predictors did not allow for useful interpretation of each one's influence. Further, exploration of predictive models including each variable individually suggested that the contrast effect because of the previous image's baseline may have been an artifact of this multicollinearity. Although previous researchers noted their concerns regarding this issue (Huang et al., 2018; Pegors et al., 2015), identifying low VIF values led to their failure to pursue this further.

Reanalysis of the data collected by Huang et al. (2018, Experiment 1 – available online) revealed an identical pattern of results to the one found here: a similarly large correlation between the previous image's baseline and rated attractiveness for each participant, as well as the absence of a contrast effect due to the previous image's baseline attractiveness when included as the only predictor of ratings given to the current image.

The results of this experiment may explain why some research has failed to find evidence of a contrast away from the previous image's baseline attractiveness (Taubert et al., 2016; Xia et al., 2016). While it is possible that differences in the stimulus-presentation duration could explain these differing patterns (Xia et al., 2016), we also propose that statistical confounds may have played an important role.

Experiment 2: Two Images Visible

Experiment 1 replicated previous findings of both assimilation and contrast effects when rating the attractiveness of faces in a sequence, although an investigation of multicollinearity called into question the presence of a perceptual contrast effect. As with earlier studies, participants viewed one face at a time onscreen, with each new image replacing the previous one. However, many contexts in the real world involve judging the current item while the previous item remains visible. It is unclear how (and whether) this alteration might influence sequential effects, and so the current experiment addressed this question. As such, all details of the procedure of Experiment 1 were reproduced here with the addition of the previous face remaining onscreen while the current face was being evaluated.

Method

Participants. The sample size was set to be comparable with the number of participants used in Experiment 1. Thirty-six university students ($M_{\text{age}} = 21.6$ years, $SD_{\text{age}} = 6.8$ years; 31 women; 86% self-reported as White) gave informed written consent before participating in the experiment and were verbally debriefed upon completion. Participants received course credit as compensation. There was no overlap between this participant sample and any of the other samples presented here.

Stimuli. The same 100 faces as in Experiment 1 were used here.

Procedure. The same procedure as in Experiment 1 was used here, with one exception. After providing a response for the current face, the image moved to the left but remained onscreen as the new image appeared from the right-hand side of the screen. As such, ratings of the current face were given while the previous face was still visible. The previous face was labeled "last" onscreen, and a box appeared around the current face, to make it clear to participants as to which image they were being asked to rate. As before, images remained onscreen until a response was given with the mouse. See Figure 1b for an example.

Data analysis. The data were analyzed using the same model as in Experiment 1.

Results

We focus on the fixed effects summarized in Table 1 (under combined model). (For more information regarding the fixed and random effects, see Table S2 in the online supplemental materials). Examining the tests of fixed effects, we found results identical to those of Experiment 1. Again, ratings of the current face were positively influenced by the current face's baseline attractiveness. In addition, the rating given to the current image assimilated toward the rating given to the previous image. Finally, the rating given to the current image contrasted away from the previous image's baseline attractiveness. In addition, estimates obtained from an intercept only model suggested the proportion of variance to be 20% for stimuli, 40% for participants, and 40% residual error.

As with Experiment 1, we again considered the issue of multicollinearity. Across all values for by-participant regressions, the largest VIF was 1.94. Across participants, we found large correlations between the previous image's baseline and rated attractiveness, mean $r = .57$, 95% CI [0.54, 0.61]. Finally, carrying out a linear mixed-effects model on all participants' data as above, but using the previous image's baseline attractiveness to predict its rated attractiveness, we found that the former was a strong predictor of the latter, $b = 0.91$, $SE = 0.06$, $p < .001$.

We next carried out linear mixed-effects models including only one of our three predictors in each separate model. As Table 1 (under separate models) shows, the contrast effect due to the previous image's baseline attractiveness became assimilative when this variable was included as the only predictor of the current image's rating. Again, the predictive value of both the current image's baseline and the previous image's rating remained relatively unchanged.

This assimilation due to the previous image's baseline attractiveness was not evident in Experiment 1 but inspection of the two coefficients in Table 1 (Experiment 1 – 0.03; Experiment 2 – 0.07) suggests that the difference between the two experiments may not be statistically significant. Therefore, we carried out a combined analysis of these two models by including a fixed effect of Experiment, along with an interaction between this factor and the previous image's baseline. Our results showed no effect of Experiment, $t(6{,}915) = 1.25$, $p = .212$, or previous image baseline, $t(7{,}032) = 0.38$, $p = .703$, along with a nonsignificant interaction, $t(7{,}006) = 0.71$, $p = .476$.

Discussion

The findings of this experiment replicated those of Experiment 1, again demonstrating the presence of both assimilation and contrast effects in responding when all predictors were included in the same model. In addition, the results of this experiment were almost identical to those found earlier, providing evidence that the visibility of the previous face onscreen had no effect on the nature of the sequential biases. The current response contrasted away from the baseline attractiveness of the previous face, no matter whether the previous face was visible onscreen (here) or only in memory (Experiment 1). Indeed, the estimate of the fixed effect across the two experiments was almost identical (Experiment 1: −0.14; Experiment 2: −0.13).

However, consideration of multicollinearity suggested issues with including both the previous image's baseline and rated attractiveness simultaneously. In fact, when included as the only predictor, the previous image's baseline attractiveness produced a statistically significant assimilation effect rather than the expected contrast effect. This reversal in the direction of the relationship has often been shown to be the product of multicollinearity (Mela & Kopalle, 2002). This assimilation was not evident in Experiment 1, although subsequent analysis, combining the two experiments, found no difference between them and no assimilation effect. As such, it remains unclear as to whether the continued presence onscreen of the previous image resulted in its increased salience when making a comparison with the current image. However, prior research has suggested an assimilation toward the previous image's attractiveness when both faces appeared simultaneously (Wedell et al., 1987), and so we suggest further research is needed to better clarify the pattern of results found here.

Experiment 3: Three Images Visible

Experiments 1 and 2 provided evidence that evaluations were assimilated toward the previous response while simultaneously contrasting away from the previous image's value when these predictors were included together. Multicollinearity concerns revealed that the previous image's value had no effect when considered alone and the image was offscreen, but showed (some evidence for) an assimilation effect when remaining onscreen during the current trial. In many real-world situations, if the previous item can be viewed then the next item in the sequence will also be visible. Here, both the previous and next face appeared onscreen as the current image was rated.

On the basis of previous research and the findings of Experiment 2, one might predict assimilation of the current rating toward the baseline values of both additional images (Wedell et al., 1987). However, we may also find a lesser influence of the next image in comparison with the previous image since the former has yet to be considered by participants.

Method

Participants. The sample size was set to be comparable with the number of participants used in previous experiments. Thirty-five university students (M_{age} = 24.3 years, SD_{age} = 7.1 years; 31 women; 77% self-reported as White) gave informed written consent before participating in the experiment and were verbally debriefed upon completion. Participants received course credit as compensation. There was no overlap between this participant sample and any of the other samples presented here.

Stimuli. The same 100 faces as in Experiment 1 were used here.

Procedure. The same procedure as in Experiment 2 was used here, with one exception. While providing a response for the current face, the next image in the sequence was visible to the right-hand side of the screen. (As in Experiment 2, the image that was previously rated remained visible to the left-hand side of the screen also.) As such, ratings of the current face were given while the previous face and the next face were visible. The previous and next face were labeled "last" and "next," respectively, and a box appeared around the current face, to make it clear to participants as to which image they were being asked to rate. As before, images remained onscreen until a response was given with the mouse. See Figure 1c for an example.

Data analysis. The data were analyzed using the same model as in Experiments 1 and 2, with the addition of the next image's baseline attractiveness as a fixed effect. The slope of this predictor was also allowed to vary randomly across participants. This model failed to converge as a result of the random slopes across participants for the previous image's baseline attractiveness. As such, this predictor was removed from the model.

Results

We focus on the fixed effects summarized in Table 1 (under combined model). (For more information regarding the fixed and random effects, see Table S3 in the supplemental materials). Examining the tests of fixed effects, we again found that ratings of the current face were positively and strongly influenced by the current face's baseline attractiveness. In addition, the rating given to the current image assimilated toward the rating given to the previous image (response bias). Finally, the rating given to the current image was not significantly predicted by the baseline attractiveness levels of either the previous or next images. In addition, estimates obtained from an intercept only model suggested the proportion of variance to be 19% for stimuli, 40% for participants, and 41% residual error.

As with Experiments 1 and 2, we again considered the issue of multicollinearity. Across all values for by-participant regressions, the largest VIF was 2.54. Across participants, we found large correlations between the previous image's baseline and rated attractiveness, mean r = .57, 95% CI [0.54, 0.61]. Finally, carrying out a linear mixed-effects model on all participants' data as above, but using the previous image's baseline attractiveness to predict its rated attractiveness, we found that the former was a strong predictor of the latter, b = 1.01, SE = 0.07, p < .001.

We next carried out linear mixed-effects models including only one of our four predictors in each separate model. As Table 1 (under separate models) shows, the (nonsignificant) contrast effect because of the previous image's baseline attractiveness became assimilative when this variable was included as the only predictor of the current image rating. Again, the predictive value of both the current image's baseline and the previous image's rating remained relatively unchanged. Finally, the next image's baseline continued to show no influence on the current image rating.

Discussion

In line with our earlier experiments, the rating given to the current face assimilated toward the rating given to the previous face. Further, the baseline value of the previous image showed no influence when considered in the combined model but showed an assimilative effect when analyzed separately. Finally, the value of the next face in the sequence again showed no influence on current ratings.

Previous research found that two images presented simultaneously showed assimilation in their evaluations (Wedell et al., 1987). This bias was also evident in Experiment 2 responses when we calculated separate models for our predictors. Presenting three images simultaneously resulted in a continued perceptual bias due to the previous face but no influence of the next face. The likely explanation for this difference is that, while both faces appeared onscreen while the current face was being judged, only the previous face had already been considered and rated. This distinction may be an important one with regard to sequential biases and so we sought to replicate this finding in the next experiment.

Experiment 4: All Images Visible

The results of Experiment 3 provided evidence that, when the previous and next faces were visible during consideration of the current face, evaluations were assimilated toward the rating given to the previous face and the previous image's baseline. However, we found no influence of the next image's baseline on current ratings even though this face was also visible onscreen during the trial. Here, we aimed to replicate this pattern of results, which suggested that the assimilative effect was because of the previous face being judged before the current face. In addition, we investigated a situation mirroring real-world decision-making in which all 100 face images appeared onscreen during the ratings task. This design simulated the experience provided by online dating websites in which numerous "thumbnail" profile images are displayed onscreen for ease of browsing potential partners.

Method

Participants. The sample size was set to be comparable with the number of participants used in previous experiments. Forty-one university students (M_{age} = 21.9 years, SD_{age} = 7.4 years; 32 women; 98% self-reported as White) gave informed written consent before participating in the experiment and were verbally debriefed upon completion. Participants received course credit as compensation. There was no overlap between this participant sample and any of the other samples presented here.

Stimuli. The same 100 faces as in Experiment 1 were used here.

Procedure. All 100 faces were presented onscreen throughout the ratings task. Image size was approximately 2.6 × 4.0 cm, with the order of images randomized across participants. Faces were rated from left to right along each row (20 images per row), starting with the top row and following the five rows down the screen as they progressed. The current face was highlighted by a red border and the question onscreen asked "How attractive is the highlighted face?" to make it clear to participants as to which image they were being asked to rate. After a response was given with the mouse, the border then highlighted the next face.

Data analysis. The data were analyzed using the same model as in Experiment 3.

Results

We focus on the fixed effects summarized in Table 1 (under combined model). (For more information regarding the fixed and random effects, see Table S4 in the online supplemental materials.) Examining the tests of fixed effects, we again found that ratings of the current face were positively and strongly influenced by the current face's baseline attractiveness. In addition, the rating given to the current image assimilated toward the rating given to the previous image (response bias). In contrast with Experiment 3 but in line with Experiments 1 and 2, the rating given to the current image was also significantly and negatively predicted by the baseline attractiveness level of the previous image (perceptual bias). Finally, the rating given to the current image was not significantly predicted by the baseline attractiveness level of the next image. In addition, estimates obtained from an intercept only model suggested the proportion of variance to be 16% for stimuli, 50% for participants, and 34% residual error.

Again, we considered the issue of multicollinearity. Across all values for by-participant regressions, the largest VIF was 2.39. Across participants, we found large correlations between the previous image's baseline and rated attractiveness, mean r = .58, 95% CI [0.55, 0.61]. Finally, carrying out a linear mixed-effects model on all participants' data as above, but using the previous image's baseline attractiveness to predict its rated attractiveness, we found that the former was a strong predictor of the latter, b = 1.02, SE = 0.02, $p < .001$.

We next carried out linear mixed-effects models including only one of our four predictors in each separate model. As Table 1 (under separate models) shows, the contrast effect due to the previous image's baseline attractiveness became assimilative when this variable was included as the only predictor of the current image rating. Again, the predictive value of both the current image's baseline and the previous image's rating remained relatively unchanged. Finally, the next image's baseline continued to show no influence on the current image rating.

Discussion

In line with Experiments 1–3, the rating given to the current face assimilated toward the rating given to the previous face. Further, the previous image's baseline showed a contrast effect within the combined model but an assimilative effect when analyzed separately. In line with Experiment 3, the value of the next face in the sequence had no influence on current ratings.

These findings replicated the pattern shown in Experiment 3, again demonstrating that the previous face influenced current ratings while the next face failed to do so. Here, this result was apparent when all faces in the sequence were displayed onscreen throughout the task.

Experiment 5: Circular Ratings Scale

In Experiments 1–4, we focused on the influence of the previous image's baseline attractiveness. In line with previous research (Huang et al., 2018; Pegors et al., 2015), Experiment 1 demonstrated that presenting individual images onscreen resulted in evaluations that were contrasted away from the value of the previous face in the combined model, but further investigation suggested that, in fact, the previous image baseline produced no effect on current ratings. In Experiments 2–4, we found that the previous image baseline produced an assimilation effect only when it remained onscreen during the current trial. Finally, Experiments 3 and 4 demonstrated that the next face in the sequence produced no biasing effect when displayed onscreen.

In this final experiment, we turned our attention to the assimilative influence of the previous image's rating. So far, all of our experiments identified a robust, positive effect. Here, we investigated whether this assimilation could be prevented by attempting to minimize action repetition.

Method

Participants. The sample size was set to be comparable with the number of participants used in previous experiments. Forty university students (M_{age} = 20.1 years, SD_{age} = 1.4 years; 34 women; 95% self-reported as White) gave informed written consent before participating in the experiment and were verbally debriefed upon completion. Participants received course credit as compensation. There was no overlap between this participant sample and any of the other samples presented here.

Stimuli. The same 100 faces as in Experiment 1 were used here.

Procedure. The same general procedure as in Experiment 1 was used here. However, the ratings scale was presented in a circular shape, centered on the face. As such, every response option was equidistant from the center of the image. In addition, at the beginning of each trial, the mouse cursor was relocated to this image center, where participants were then free to move it to their chosen response. Finally, after providing a response for the current face, the next image simply replaced the current one (i.e., there was no conveyor belt movement). As before, images remained onscreen until a response was given with the mouse. See Figure 2 for an example.

Data analysis. The data were analyzed using the same model as in Experiment 1.

Figure 2. An illustration of the ratings task for Experiment 5. See the online article for the color version of this figure.

Results

We focus on the fixed effects summarized in Table 1 (under combined model). (For more information regarding the fixed and random effects, see Table S5 in the online supplemental materials). Examining the tests of fixed effects, we again found that ratings of the current face were positively and strongly influenced by the current face's baseline attractiveness. In addition, the rating given to the current image was also significantly and negatively predicted by the baseline attractiveness level of the previous image (perceptual bias). Finally, and in contrast with the prediction for this experiment, the rating given to the current image assimilated toward the rating given to the previous image (response bias). In addition, estimates obtained from an intercept only model suggested the proportion of variance to be 18% for stimuli, 42% for participants, and 40% residual error.

Again, we considered the issue of multicollinearity. Across all values for by-participant regressions, the largest VIF was 2.96. Across participants, we found large correlations between the previous image's baseline and rated attractiveness, mean $r = .58$, 95% CI [0.54, 0.61]. Finally, carrying out a linear mixed-effects model on all participants' data as above, but using the previous image's baseline attractiveness to predict its rated attractiveness, we found that the former was a strong predictor of the latter, $b = 0.91$, $SE = 0.06$, $p < .001$.

We next carried out linear mixed-effects models including only one of our three predictors in each separate model. As Table 1 (under separate models) shows, the assimilation effect because of the previous image's rated attractiveness was present. Mirroring the pattern of results found in Experiment 1, we found no influence of the previous image's baseline.

Discussion

In line with Experiment 1, the rating given to the current face assimilated toward the rating given to the previous face. Indeed, the estimate of the fixed effect for the response bias across the two experiments was almost identical (for the separate models: Experiment 1 – 0.18; Experiment 5 – 0.20). This assimilation toward the previous rating was not expected to occur, given the experiment's design in which responses were provided using a circular scale, with the mouse cursor relocating to the center of that scale at the start of each trial. Explanations based on action repetition often refer to the ease with which keypresses or mouse clicks are given when the participant simply chooses not to move their hand after the previous response is submitted. Here, this account is ruled out.

General Discussion

Across five experiments, our initial approach using cross-classified linear mixed-effects modeling suggested the following pattern of results. We found a consistent response bias in all our experiments in that ratings of the current face were assimilated toward ratings given to the previous face. In addition, a perceptual bias was evident in all cases with the exception of Experiment 3, where ratings contrasted away from the baseline attractiveness value of the previous face. Finally, Experiments 3 and 4 found no evidence of a perceptual bias because of the presence of the next face onscreen during judgments.

These results generally replicate those of several previous studies exploring sequential effects with both facial attractiveness (Huang et al., 2018; Pegors et al., 2015) and other types of decisions (e.g., Jesteadt et al., 1977; Matthews & Stewart, 2009). Evaluations were assimilated toward the previous response and contrasted away from the previous item's value. This pattern is, to some extent, established in the literature but seemingly contradictory results have also been reported. For example, when participants only responded to alternate faces (each evaluated face was preceded by a face that was not evaluated), researchers found assimilation toward the previous image's value (Xia et al., 2016). Because no response was given for the previous item, this effect could not be because of a response bias. The authors argued that this difference to what has commonly been found in the literature may have been because of stimulus-presentation duration. For shorter image presentations (1 s; Xia et al., 2016), ratings may demonstrate assimilation effects. However, other studies have shown contrast effects relating to the value of the previous item for presentations as short as 4 s (Pegors et al., 2015) and 3 s (Huang et al., 2018), casting doubt on the idea that a reversal in perceptual bias takes place during the 1–3 s window. Indeed, images appeared onscreen in the current set of experiments until a response was given, demonstrating that participants showed an apparent contrast effect in self-paced designs.

Here, we propose that difficulties with identifying the nature of this perceptual bias are the result of statistical artifacts. In recent research, the baseline or "true" value of each face was determined through the prior collection and subsequent averaging of responses given by an earlier sample of raters (Huang et al., 2018; Pegors et al., 2015). Evidence that raters showed a large amount of agreement in judgments of facial attractiveness means that this is an intuitive method of quantifying the value of a face for any given viewer (Hönekopp, 2006; Kramer et al., 2018). As a direct result, we expect there to be substantial agreement between a participant's judgments and those of this prior sample. Therefore, multicollinearity will be almost inevitable. If the correlation between the previous image's baseline and rated attractiveness is substantial then including both predictors in a model will likely result in misleading outcomes (Bonate, 1999; Johnston et al., 2018; Mela & Kopalle, 2002; Shieh & Fouladi, 2003; Vatcheva et al., 2016). Here, we found across all five experiments that the influence of the previous image's baseline was altered when comparing its effect within a combined versus separate model. As a result, although we identified simultaneous assimilation and contrast effects in our combined models, we suggest that these effects are uninterpretable.

Two recent articles showed that current ratings of a face were assimilated toward ratings of the previous face while simultaneously contrasting away from the previous face's baseline attractiveness (Huang et al., 2018; Pegors et al., 2015). In both cases, the authors considered their relatively low VIF statistics to be evidence that multicollinearity was not an issue. However, researchers have demonstrated that typical "rules of thumb" were insufficient when identifying confounding (Johnston et al., 2018; Vatcheva et al., 2016). Indeed, further exploration of the data featured in Huang et al. (2018) identified large correlations between the previous image's baseline and rated attractiveness. As such, any conclusions drawn from those studies must be called into question.

Given the presence of multicollinearity, we sought to investigate how each predictor separately influenced current ratings. A clear pattern of results emerged. First, ratings of the previous image showed an assimilative effect across all five experiments. Second, we found no influence of the next face's baseline attractiveness when this image was present onscreen during judgments (Experiments 3 and 4). Third, the perceptual effect due to the previous image's baseline was always positive but only represented a statistically significant bias when the image remained onscreen during the current trial (Experiments 2–4). Although this result is less well supported, given the lack of a statistically significant difference between the patterns found in Experiments 1 and 2, it aligns with previous findings, where the perceptual effect because of the previous image was shown to be assimilative (Chang et al., 2017; Taubert et al., 2016; Wedell et al., 1987; Xia et al., 2016).

Our findings demonstrate that it may be difficult to quantify the independent contributions of the previous image's baseline attractiveness, along with the rating that that image received, when predicting the current image's rating. It is also important to note that the causal assumptions differ for these two predictors. The former represents a direct influence on the current rating—the previous image's baseline biases our perception of the current image. However, there is also an indirect component—the baseline of the previous image necessarily affects our response to that image, which in turn, is used to assess our response bias for the current image. As such, we should consider these two predictors within this causal framework. Our estimate of the perceptual bias need not be adjusted for ratings of the previous face, while what is of interest with respect to the response bias is the effect of the previous response above and beyond the influence of the previous image's baseline attractiveness. Therefore, the appropriate estimate for the response bias should be adjusted for the baseline attractiveness of the previous face. This causality may provide a clearer framework for understanding these two biases in future research.

In the set of experiments presented here, we found that assimilation toward the previous image's value was only statistically significant when both the previous and current images were presented onscreen (although the combined analysis presented a less clear account). While earlier studies support this finding with simultaneous presentation (Wedell et al., 1987), other researchers have found an assimilative effect even when the previous image was removed because of the presentation of the current image (Chang et al., 2017; Taubert et al., 2016; Xia et al., 2016). Although it remains unclear why assimilation was absent in our Experiments 1 and 5, it makes intuitive sense that the strength of the effect should be greater when the previous image is still present, allowing direct perceptual comparison with the current image. In addition, the intertrial interval may also play a role. Xia and colleagues (2016) showed assimilation toward the previous item's value for intervals of 0–6 s. Here, using a conveyor belt-style animation, approximately 1–2 s passed between trials. Evidence suggests that longer intertrial intervals produce a decay in the assimilation effect (Attali, 2011; Xia et al., 2016) and this avenue of research warrants further investigation.

While the previous image's baseline influenced current ratings under certain conditions, we found no evidence that the next image in the sequence had any effect on judgments when presented onscreen. This is a novel and interesting result, suggesting that it was important for participants to judge the previous image (although not necessarily respond to it; Chang et al., 2017) before the current image to be influenced by it. Simply presenting the next face was insufficient to produce a bias. Whether participants paid either minimal or no attention to the next image remains unclear and future studies might introduce manipulations targeted at understanding this process further.

In our final experiment, we failed to remove the response bias caused by rating the previous image. Although a robust effect in Experiments 1–4, along with numerous earlier studies (Huang et al., 2018; Pegors et al., 2015), it was presumed that this bias could at least partially be explained by action repetition or some form of inertia on the part of the judge. For instance, participants may provide a response that involves minimal effort in terms of finger or mouse movements, especially on uncertain trials. However, and surprisingly, even when responses were given orally, participants continued to show a response bias in their attractiveness ratings (Experiment 4; Huang et al., 2018). Here, participants in Experiment 5 provided responses that were equally effortful in that, on each trial, all choices were initially equidistant from the mouse cursor. However, and in line with Huang et al. (2018), the response bias continued to influence ratings. Therefore, we propose two further mechanisms that may be worth investigating: (a) the decision to select a particular response increases the likelihood of that response or similar being given on the next trial; and (b) activating motor movements orally or through hand gestures makes subsequent activation of that movement or similar more likely on the next trial. Future studies that are designed specifically to address these ideas may prove fruitful in determining the underlying cause of this bias in judgments.

The current experiments investigated sequential biases using a different method of statistical analysis. Previous findings were often the result of by-participant analytical approaches (Huang et al., 2018; Pegors et al., 2015). That is, the researchers considered participants but not images as random factors, with the result that conclusions were not generalizable to other sets of images. This is an important caveat because both participants and images should be considered subsets of larger populations. Indeed, a failure to treat images as a random effect can increase the empirical Type I error rate (Murayama et al., 2014). Using cross-classified linear mixed-effects models, researchers are now able to effectively incorporate both random factors simultaneously by explicitly modeling the dependencies in the data. This nonindependence is applicable because responses given by different participants to the same image tend to be similar. Linear mixed-effects models have now been applied to other types of sequential decisions (Zhao et al., 2017) but have yet to be used by attractiveness researchers.

All five of the current experiments demonstrated that approximately 20% of the variance in the data was because of the stimuli, although we note that the same stimuli were used in all these experiments. However, intercept only models may overestimate the importance of the random stimulus effect in the current work. That the faces' baseline attractiveness likely explained a substantial amount of the interstimulus variability in attractiveness ratings means that a more realistic estimate can be obtained from a model that includes the fixed effect of baseline attractiveness. In this case, we find that only 1–3% of the variance was because of the stimuli in our experiments. Of course, the use of cross-classified linear mixed-effects models remains the most appropriate strategy here because of the nature of the experimental design, and our recommendation is that all studies of this type should incorporate these models. Indeed, this idea has been put forward by previous researchers (e.g., Judd et al., 2012; Westfall et al., 2014; Westfall et al., 2015) and the hope is that the approach will become more common in future publications.

In all of the experiments presented here, we found that assimilation toward the previous rating (a response bias) represented a stronger influence than the assimilation toward the previous image's value (a perceptual bias). We have discussed at length how these two predictors inherently overlap because viewers show substantial agreement in their judgments of facial attractiveness (Hönekopp, 2006; Kramer et al., 2018). However, it is important to note that the influence because of the previous image's rating actually incorporates both response and perceptual components, representing both the value of the previous face in the eyes of the participant and a motor response preceding the current trial. Therefore, this may explain why the rating given to the previous face provided a stronger influence than its baseline attractiveness, with the latter representing a less direct measure of the value of the face in the eyes of the participant only. As noted, the participant's own judgment of the preceding face also provides a more accurate measure of that face's value for the participant, giving another reason why it is likely to better predict the current trial's rating.

Although the biases investigated here are considered in terms of sequential effects, it is also interesting to draw parallels with the literature on anchoring. This refers to judgments assimilating toward a previously encountered standard or value. In our experiments, participants recently encountered a value—their own previous response—and this may have served as an anchor for their current response. According to the process of "anchoring and adjustment" (Tversky & Kahneman, 1974), participants may have adjusted from this anchor toward the "correct" value (i.e., their perception of the current face's attractiveness) but stopped too soon, perhaps at the first value that seemed plausible. It appears that there are clear parallels between anchoring and sequential effects, and the extent to which these processes might share common mechanisms remains open to further research.

Investigating sequential decision making by presenting the previous and next item onscreen provides an exciting avenue for further study. Indeed, many real-world sequences are displayed visually in a way that allows access to earlier and later items, for example, browsing a shelf of items in a store. Therefore, it may be that in these circumstances, previous findings of perceptual biases are significantly altered. This design also highlights an important difference between biases in visual and other modalities—often in real-world contexts, earlier and later items may be accessible (visible) while the current item is being evaluated, but this is unlikely to be true for judging sounds or tastes in a sequence. While many studies have found commonalities in sequential biases across modalities, there could also be limits to this generalization. In terms of the underlying theory that is currently being developed regarding sequential biases, it is important to determine when to expect the presence or absence of such effects. If biases are dependent on whether items appear onscreen and/or have been previously judged, this suggests a crucial difference between comparisons with one versus multiple items during evaluative sequences.

In conclusion, across five experiments, we have shown that there is good reason to utilize a linear mixed-effects modeling approach for this field of research. Participants and images represent subsets of a larger population, and so including both of these as random factors is a necessary feature of any analysis. The results of these models have called into question previous findings of simultaneous response and perceptual biases, highlighting the issue of multicollinearity when incorporating the previous face's rated and baseline attractiveness. Instead, we have demonstrated that the current face's attractiveness evaluation was assimilated toward both the previous rating and the previous face's value, with these two predictors showing substantial overlap conceptually and statistically. In addition, we found no evidence that the next face in the sequence influenced current ratings when displayed onscreen. Finally, the response bias due to the previous face remained significant, even when explanations involving motor effort were addressed. Taken together, these results provide a clear explanation for the apparently contradictory results in the current literature, a more suitable approach for analyzing the associated data, and an initial exploration of potential real-world sequences in which the previous and next image are on display during the current judgment.

References

Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Sage.

Attali, Y. (2011). Sequential effects in essay ratings. *Educational and Psychological Measurement, 71,* 68–79. http://dx.doi.org/10.1177/0013164410387344

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bonate, P. L. (1999). The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharmaceutical Research, 16,* 709–717. http://dx.doi.org/10.1023/A:1018828709196

Bruine de Bruin, W. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychologica, 118,* 245–260. http://dx.doi.org/10.1016/j.actpsy.2004.08.005

Chalmers, J. D., Singanayagam, A., Murray, M. P., Scally, C., Fawzi, A., & Hill, A. T. (2009). Risk factors for complicated parapneumonic effusion and empyema on presentation to hospital with community-acquired pneumonia. *Thorax, 64,* 592–597. http://dx.doi.org/10.1136/thx.2008.105080

Chang, S., Kim, C.-Y., & Cho, Y. S. (2017). Sequential effects in preference decision: Prior preference assimilates current preference. *PLoS ONE, 12,* e0182442. http://dx.doi.org/10.1371/journal.pone.0182442

Damisch, L., Mussweiler, T., & Plessner, H. (2006). Olympic medals as fruits of comparison? Assimilation and contrast in sequential performance judgments. *Journal of Experimental Psychology: Applied, 12,* 166–178. http://dx.doi.org/10.1037/1076-898X.12.3.166

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology, 27,* 590–595. http://dx.doi.org/10.1016/j.cub.2017.01.006

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7,* 493–498. http://dx.doi.org/10.1111/2041-210X.12504

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 199–209. http://dx.doi.org/10.1037/0096-1523.32.2.199

Huang, J., He, X., Ma, X., Ren, Y., Zhao, T., Zeng, X., . . . Chen, Y. (2018). Sequential biases on subjective judgments: Evidence from face attractiveness and ringtone agreeableness judgment. *PLoS ONE, 13,* e0198723. http://dx.doi.org/10.1371/journal.pone.0198723

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

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity, 52,* 1957–1976. http://dx.doi.org/10.1007/s11135-017-0584-6

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103,* 54–69. http://dx.doi.org/10.1037/a0028347

Kanai, R., & Verstraten, F. A. (2005). Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision Research, 45,* 3109–3116. http://dx.doi.org/10.1016/j.visres.2005.05.014

Kok, R., Taubert, J., Van der Burg, E., Rhodes, G., & Alais, D. (2017). Face familiarity promotes stable identity recognition: Exploring face perception using serial dependence. *Royal Society Open Science, 4,* 160685. http://dx.doi.org/10.1098/rsos.160685

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

Kramer, R. S. S. (2017). Sequential effects in Olympic synchronized diving scores. *Royal Society Open Science, 4,* 160812. http://dx.doi.org/10.1098/rsos.160812

Kramer, R. S. S., Jones, A. L., & Sharma, D. (2013). Sequential effects in judgements of attractiveness: The influences of face race and sex. *PLoS ONE, 8,* e82226. http://dx.doi.org/10.1371/journal.pone.0082226

Kramer, R. S. S., Jones, A. L., & Ward, R. (2012). A lack of sexual dimorphism in width-to-height ratio in white European faces using 2D photographs, 3D scans, and anthropometry. *PLoS ONE, 7,* e42705. http://dx.doi.org/10.1371/journal.pone.0042705

Kramer, R. S. S., Mileva, M., & Ritchie, K. L. (2018). Inter-rater agreement in trait judgements from faces. *PLoS ONE, 13,* e0202655. http://dx.doi.org/10.1371/journal.pone.0202655

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science, 20,* 1309–1312. http://dx.doi.org/10.1111/j.1467-9280.2009.02453.x

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.

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis: The asymmetric effect of negative and positive correlations. *Applied Economics, 34,* 667–677. http://dx.doi.org/10.1080/00036840110058482

Miller, J. M., & Krosnick, J. A. (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly, 62,* 291–330. http://dx.doi.org/10.1086/297848

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 40, 1287–1306. http://dx.doi.org/10.1037/a0036914

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110,* 472–489. http://dx.doi.org/10.1037/0033-295X.110.3.472

Mussweiler, T., Rüter, K., & Epstude, K. (2004). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology, 87,* 832–844. http://dx.doi.org/10.1037/0022-3514.87.6.832

Myers, R. (1990). *Classical and modern regression with applications* (2nd ed.). Boston, MA: Duxbury.

Page, L., & Page, K. (2010). Last shall be first: A field study of biases in sequential performance evaluation on the Idol series. *Journal of Economic Behavior & Organization, 73,* 186–198. http://dx.doi.org/10.1016/j.jebo.2009.08.012

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology, 36,* 859–866. http://dx.doi.org/10.1111/j.1745-9125.1998.tb01268.x

Pegors, T. K., Mattar, M. G., Bryan, P. B., & Epstein, R. A. (2015). Simultaneous perceptual and response biases on sequential face attractiveness judgments. *Journal of Experimental Psychology: General, 144,* 664–673. http://dx.doi.org/10.1037/xge0000069

Raaijmakers, J. G., Schrijnemakers, J. M., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426. http://dx.doi.org/10.1006/jmla.1999.2650

Scott, N. J., Jones, A. L., Kramer, R. S. S., & Ward, R. (2015). Facial dimorphism in Autistic Quotient scores. *Clinical Psychological Science, 3,* 230–241. http://dx.doi.org/10.1177/2167702614534238

Scott, N. J., Kramer, R. S. S., Jones, A. L., & Ward, R. (2013). Facial cues to depressive symptoms and their associated personality attributions. *Psychiatry Research, 208,* 47–53. http://dx.doi.org/10.1016/j.psychres.2013.02.027

Shieh, Y.-Y., & Fouladi, R. T. (2003). The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. *Educational and Psychological Measurement, 63,* 951–985. http://dx.doi.org/10.1177/0013164403258402

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B*. Methodological, *13,* 238–241. http://dx.doi.org/10.1111/j.2517-6161.1951.tb00088.x

Taubert, J., & Alais, D. (2016). Serial dependence in face attractiveness judgements tolerates rotations around the yaw axis but not the roll axis. *Visual Cognition, 24,* 103–114. http://dx.doi.org/10.1080/13506285.2016.1196803

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports, 6,* 32239. http://dx.doi.org/10.1038/srep32239

Taubert, J., Van der Burg, E., & Alais, D. (2016). Love at second sight: Sequential dependence of facial attractiveness in an on-line dating paradigm. *Scientific Reports, 6,* 22740. http://dx.doi.org/10.1038/srep22740

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

Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology, 6,* 227. http://dx.doi.org/10.4172/2161-1165.1000227

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

Wedell, D. H., Parducci, A., & Geiselman, R. 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

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115–1125. http://dx.doi.org/10.1177/01461672992512005

Wells, T., Baguley, T., Sergeant, M., & Dunn, A. (2013). Perceptions of human attractiveness comprising face and voice cues. *Archives of Sexual Behavior, 42,* 805–811. http://dx.doi.org/10.1007/s10508-012-0054-0

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science, 10,* 390–399. http://dx.doi.org/10.1177/1745691614564879

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143,* 2020–2045. http://dx.doi.org/10.1037/xge0000014

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of attractiveness. *Journal of Vision, 16*(15), 28.

Zhao, H., Andersson, B., Guo, B., & Xin, T. (2017). Sequential effects in essay ratings: Evidence of assimilation effects using cross-classified models. *Frontiers in Psychology, 8,* 933. http://dx.doi.org/10.3389/fpsyg.2017.00933

Received April 2, 2020
Revision received June 29, 2020
Accepted July 23, 2020 ■