Attraction to the recent past in aesthetic judgments: A positive serial dependence for rating artwork

Sujin Kim
School of Psychology, The University of Sydney, New South Wales, Australia

David Burr
School of Psychology, The University of Sydney, New South Wales, Australia
Department of Neuroscience, Psychology, Pharmacology, and Child Health, University of Florence, Florence, Italy

David Alais
School of Psychology, The University of Sydney, New South Wales, Australia

Visual perception can be systematically biased towards the recent past. Many stimulus attributes—including orientation, numerosity, facial expression and attractiveness, and perceived slimness—are systematically biased towards recent past experience. This phenomenon has been termed serial dependence. In the current study, we tested whether serial dependence occurs for aesthetic ratings of artworks. A set of 100 paintings depicting scenery and still life was collected from online archives. For each participant, 40 paintings were randomly selected from the set, and presented sequentially 20 times in random order. Serial dependence was quantified for each observer by measuring how their rating response on each trial depended on the attractiveness of the previous trial. The data were pooled across participants and fitted with a Bayesian model of serial dependence. Results showed that the current painting earned significantly higher aesthetic ratings when participants viewed a more attractive painting on the previous trial, compared to when they viewed a less attractive one. The magnitude of serial dependence was greatest when the attractiveness distance between consecutive paintings was relatively close. The effect held both for 1 s exposure times, and for brief 250 ms exposures (followed by a mask). These findings show that aesthetic judgments are not sequentially independent, showing that positive serial dependencies are not limited to low-level perceptual judgments.

Introduction

When we pass through an art gallery viewing a series of paintings, do we appreciate every artwork in isolation, or is our aesthetic response to a given painting influenced by the one we just saw? Moreover, if our aesthetic judgment is not independent across paintings, is a sequence of two paintings judged to be more or less similar in attractiveness? Recent work has shown that perception is systematically biased by the recent past, an effect known as serial dependence. Although serial dependence shows that current perception is distorted—biased by the recent past and thus smeared over time—this bias can be beneficial in reducing overall error (Cicchini, Anobile, & Burr, 2014; Cicchini, Mikellidou, & Burr, 2017, 2018). The notion is that by combining current sensory input with recent input, perception is more stable and optimal, and this attribute can be achieved without great cost because the natural world tends to be very stable from moment to moment (Dong & Atick, 1995).

Assimilative serial dependence has been reported for a variety of basic sensory attributes, such as orientation, motion, and numerosity (Alais, Leung, & Van der Burg, 2017; Cicchini et al., 2014; Corbett, Fischer, & Whitney, 2011; Fischer & Whitney, 2014). This assimilation toward the recent past extends to processing higher level visual information such as body shape (Alexi et al., 2018), face gender, attractiveness, and expression (Hsu & Yang, 2013; Taubert, Alais, & Burr, 2016; Taubert, Van der Burg, & Alais, 2016; Xia, Leib, & Whitney, 2016) and preference for photographs (Chang, Kim, & Cho, 2017). Interestingly, evaluation

Citation: Kim, S., Burr, D., & Alais, D. (2019). Attraction to the recent past in aesthetic judgments: A positive serial dependence for rating artwork. *Journal of Vision*, *19*(12):19, 1–17, https://doi.org/10.1167/19.12.19.

https://doi.org/10.1167/19.12.19 Received March 29, 2019; published October 18, 2019 ISSN 1534-7362 Copyright 2019 The Authors

of these higher order visual properties could effectively be considered an aesthetic judgment based on interenal referents and values. Indeed, aesthetic judgments are very common in our everyday lives. We often judge objects, people, and places in terms of beauty and attractiveness, as we might do when inspecting an architectural design, a home decoration, a face, an outfit of clothes, or even the layout of a document or slide presentation. This behavior raises the possibility that sequential dependence may manifest in more abstract visual evaluations such as aesthetics in art appreciation.

However, with the conflicting results reported to date, it is yet to be clarified whether the nature of serial dependence in aesthetic judgment is assimilative (positive) or contrastive (negative). Unlike the above-mentioned studies, which showed attraction toward the preceding trial when rating the attractiveness of face or photographs (Chang et al., 2017; Taubert, Van der Burg, et al., 2016), there have also been studies showing that a series of aesthetic judgments will display contrastive dependence or "hedonic contrast" to the immediate past. For example, Kenrick and Gutierres (1980) found that a photo of a human face was rated as less attractive following an exposure to a highly attractive individual. Likewise, it was found that faces considered to be in the same category are subject to a contrastive serial dependence (Cogan, Parker, & Zellner, 2013). This phenomenon has been studied in a variety of aesthetic dimensions including music (Parker, Bascom, Rabinovitz, & Zellner, 2008) and visual arts (Dolese, Zellner, Vasserman, & Parker, 2005; Khaw & Freedberg, 2018).

In the current study, we investigated whether the attractiveness rating of artworks is subject to serial dependence and whether the bias is assimilative or contrastive in relation to the preceding one. To this end, we presented paintings sequentially and asked observers to rate their attractiveness. In addition, two presentation durations were compared: a brief 250 ms presentation that was immediately followed by a noise mask to curtail further processing, and a longer unmasked presentation of 1,000 ms. By using two different exposure times, we expected to elucidate the timescale of serial dependence in attractiveness rating and what factors elicit such bias. This is relevant to an ongoing debate concerning at which stage of visual processing serial dependence occurs. Whereas many previous studies have suggested that the sequential bias arises at an early perceptual level (Alais et al., 2017; Cicchini et al., 2014; Cicchini et al., 2017, 2018; Fischer & Whitney, 2014; Van der Burg, Alais, & Cass, 2015), there is also evidence advocating engagement of postperceptual processes such as working memory or decision making (Fritsche, Mostert, & de Lange, 2017; Kiyonaga, Scimeca, Bliss, & Whitney, 2017). If serial dependence reflects changes in later stages, we would expect to see a difference in bias dependent on viewing time as 1,000 ms ought to be long enough to allow feedback and cognitive input (e.g., knowledge, experience) to operate during stimulus presentation, while there would be little time for feedback loops using 250-ms presentations followed by disruptive poststimulus noise masks which curtail any postpresentation visual analysis. For data analysis, we used a Bayesian model of serial dependence to predict the bias on a given trial (Cicchini et al., 2014).

Experiment 1

Methods

Participants

Twenty-four participants (14 females, 10 males; mean age, 23 years old, ranging from 20 to 28) with normal or corrected-to-normal vision were recruited from the university student population through online advertisement. They provided written consent and were paid AU $20/hr for their participation. This research accorded with the principles of the Declaration of Helsinki and was approved by the Human Research Ethics Committee of the University of Sydney. All participants were invited to complete postexperiment questionnaires asking about their interests and knowledge in art (to examine its relationship to any serial dependence in aesthetic judgments); 16 responded, and their data were analyzed.

Stimuli and apparatus

A total of 100 paintings depicting scenery or still life were selected from online archives (http://www.artcyclopedia.com: all reproductions of original paintings; see all selected images in Supplementary File S1). We excluded portraits and figure paintings which mainly depict faces since it was previously reported that face attractiveness is positively assimilated toward the preceding one (Taubert, Van der Burg, et al., 2016). However, some scenery paintings with figures were included when an individual's face could not be identified (i.e., small size or located in the periphery; 29% of the entire set). Selected paintings varied in their styles and date of production, which ranged from the 15th to the 20th centuries. Each stimulus was presented in one of four possible sizes (100% to 85% of its pixel size in 5% steps) and at randomly jittered locations to minimize local adaptation. The average dimensions of all stimuli before the size manipulation subtended a horizontal extent of $18° \pm 1.6°$ (*SD*) and $14° \pm 2.6°$ (*SD*) vertically. The center position of the stimulus

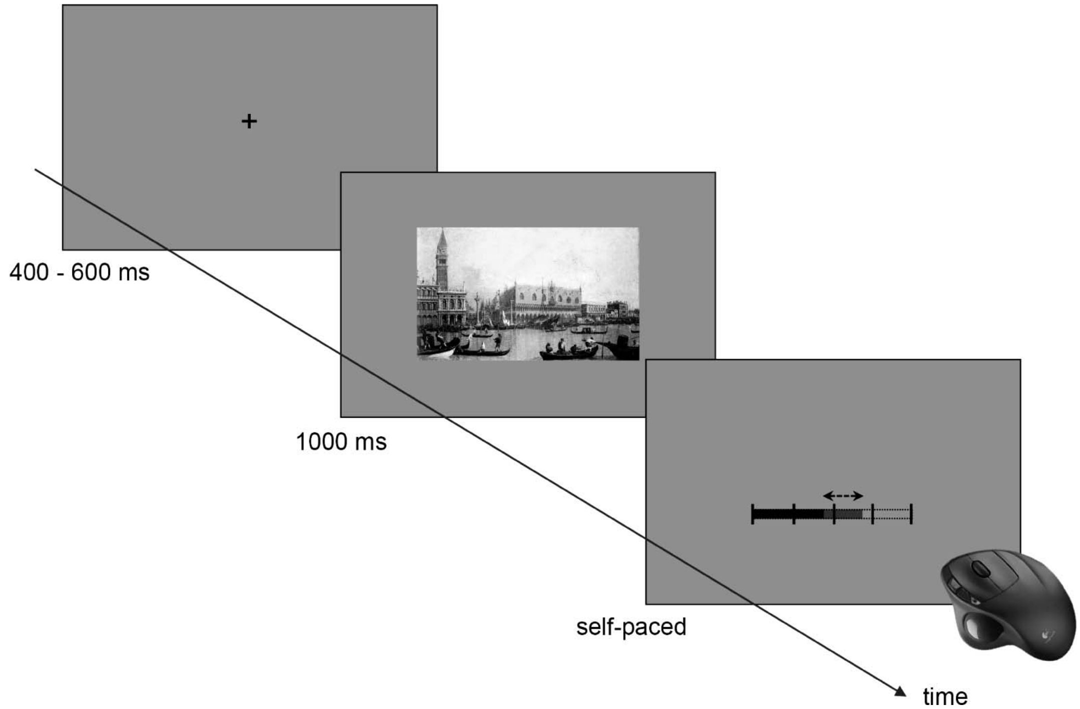

Figure 1. Experimental procedure for Experiment 1. After a variable intertrial interval, each trial began with the fixation point presented at the center of the screen followed by an image of a painting presented for 1,000 ms, after which a slide bar appeared. All paintings were presented in their original color. Participants were asked to rate the attractiveness of the painting by adjusting the slider with a trackball mouse. The slider was filled with black color to indicate the adjusted point. The procedure for Experiment 2 was identical except for two differences: The image presentation time was shortened to 250 ms, and it was followed by a static white-noise mask for 500 ms before the slide bar appeared. Sample image is *View of the Ducal Place in Venice* by Canaletto (1755, image in the public domain).

presentation was jittered on every trial by randomly offsetting the x and y coordinates of the center point within a range of $1.5° \times 1.5°$. Stimulus presentation was controlled by MATLAB (MathWorks, Natick, MA) using Psychophysics Toolbox-3 (Brainard, 1997; Pelli, 1997). All paintings were presented on a mean luminance gray background on a 13-in. CRT monitor (1280×1024 resolution, 85 Hz frame rate) viewed from a distance of 57 cm.

Design and procedure

For each participant, 40 paintings were randomly selected from the set of 100 paintings at the beginning of the experiment. Participants rated each painting 20 times throughout the experiment, completing 800 trials in two blocks separated by a 1-min break. The order of the stimulus presentation was randomized with a constraint that the same painting was never presented consecutively. Each trial consisted of a 1,000 ms stimulus presentation followed by the attractiveness rating slide bar (see Figure 1) which was adjusted by the participant with a trackball mouse. The rating bar was randomly positioned for every trial to minimize any response bias related to starting point. Participants were given instructions encouraging them to use the whole range of the slide bar when making a response. They were also told it was not a speeded task but to try to respond promptly after each image presentation. Once the attractiveness rating was adjusted with the mouse trackball, participants pressed the space bar to record their response. The next trial started after a random interval between 400 and 600 ms. There were three practice trials before the main experiment to familiarize participants with the experimental procedure.

Modelling: Cicchini et al. (2014) introduced a Kalman-filter model for serial dependence in which the current response (R_i) to the current stimulus (X_i) in a given trial is predicted by a weighted sum of the previous response and current stimulus:

$$R_i = w_{i-1}R_{i-1} + (1 - w_{i-1})X_i \quad (1)$$

where w_{i-1} is the weight given to the previous response. As paintings were not seen or rated by subjects before the main trials, we assumed that a subject's response to a painting on a given trial (R_{i-1}) would be the best estimate for predicting the effect of immediate past on the upcoming trial, as it is effectively a recursive accumulation of past and present. For the current stimulus, we used the mean value of individual's 20 ratings on each painting calculated on completion of experiment.

It can be shown (Cicchini et al., 2018) that the ideal weight w_{i-1} is determined by a combination of the relative uncertainty of the response to the current trial and the distance between the current stimulus and previous response (d):

$$w_{i-1} = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_{i-1}^2 + d^2} \quad (2)$$

where σ_i^2 and σ_{i-1}^2 represent the variance of the underlying noise in judgments of the current and the previous stimuli. As the variance in each participant's attractiveness judgments across their 40 paintings was fairly consistent, we assumed that $\sigma_i = \sigma_{i-1}$, and averaged root-variance across all paintings to obtain a more robust measure for each participant (σ).

Inserting Equation 2 into Equation 1 and rearranging, the predicted bias in judging the current stimulus is given by

$$R_i - X_i = \frac{d}{2 + (d/\sigma)^2} \quad (3)$$

When the distribution of all participants' root-variance was examined, it was highly skewed to the left due to a few outliers with extremely large values (see Figure 2B). Therefore, for the modelling of group data, we used the median root-variance of all participants. Unless otherwise specified, all reported errors represent standard deviations.

Results

Aesthetic ranking

For each of the 24 participants, we chose 40 paintings randomly. These were displayed 20 times in pseudorandom order (as described in methods), and with a slide bar, participants rated each painting for attractiveness. According to the previous studies suggesting an effect of habituation (Imamoglu, 1974; Leder, 2001; Leder, Gerger, Brieber, & Schwarz, 2014; Park, Shimojo, & Shimojo, 2010) or mere-exposure effect (Palmer, Schloss, & Sammartino, 2013; Zajonc, 1968) on liking ratings, we checked for a linear trend of ratings over repetition. A regression line was fitted to the mean ratings of the same round (i.e., *n*th rating on each painting) across 40 paintings, and the results showed a significant linear trend in 15 out of 24 subjects (mean slope of the linear regression –0.33 ± 0.44; –0.23 ± 0.38 for all subjects). However, whether the data were detrended or not did not change the results shown in the following section (see Figure A1). Therefore, the undetrended raw data was used for further analyses. Figure 2A shows the group-mean ratings averaged over all participants, for the entire set of 100 paintings, ranked by average rating. The overall rating function was a reasonable approximation to linear, slope = 0.41, $F(1, 98) = 2525$, $p < 0.001$, $R^2 = 0.96$.

To examine the degree of agreement between each individual's mean ratings to the grand averages, individual mean ratings were correlated with the average ratings of corresponding paintings calculated from the data of other individuals (i.e., group mean ratings were calculated excluding the subject whose 40 mean ratings will be correlated). Figure 2C shows the distribution of correlation coefficients, and the average of Pearson's r was 0.48 ± 0.25. Although results showed a good degree of agreement across participants in general, there were two individuals who yielded negative correlation with grand averages.

There were considerable individual differences in the response *range* used by observers (mean range of attractiveness rating for 40 paintings 69.5 ± 20). This high variance in attractiveness ranges can mislead the group results, as the distances between two successive trials cannot be represented equivalently across participants. We therefore used a minimum/maximum scaling to normalize each individual's raw ratings data with their 40 mean ratings to span a range of 0 to 100 before conducting further analysis. It should be noted, however, that normalization of the data was not essential for getting the results, although the effect size was slightly reduced due to less data points at the extremes of the intertrial distances and thus larger variances for those points.

Serial dependence in aesthetic judgments

For each observer, serial dependence was measured by plotting the difference between the current rating and their mean rating of that image as a function of the distance between the previous response and the current stimulus. The first trial of each block was necessarily excluded as it had no preceding trial. Subsequent trials were binned into ten (i.e., each bin width was 20) according to the normalized distance between the previous and current trial, which ranged from 100 to 100. Serial dependence data for each bin were averaged per individual, and these values were averaged across participants to calculate the group mean.

Figure 3A shows the results. Attractiveness ratings were clearly biased toward the responses to the preceding trials, especially when the difference between current and previous trials was not too great. That is, participants rated a painting's attractiveness higher than average when preceded by one judged as more attractive (positive distance), and lower when preceded by one judged as less attractive (negative distance).

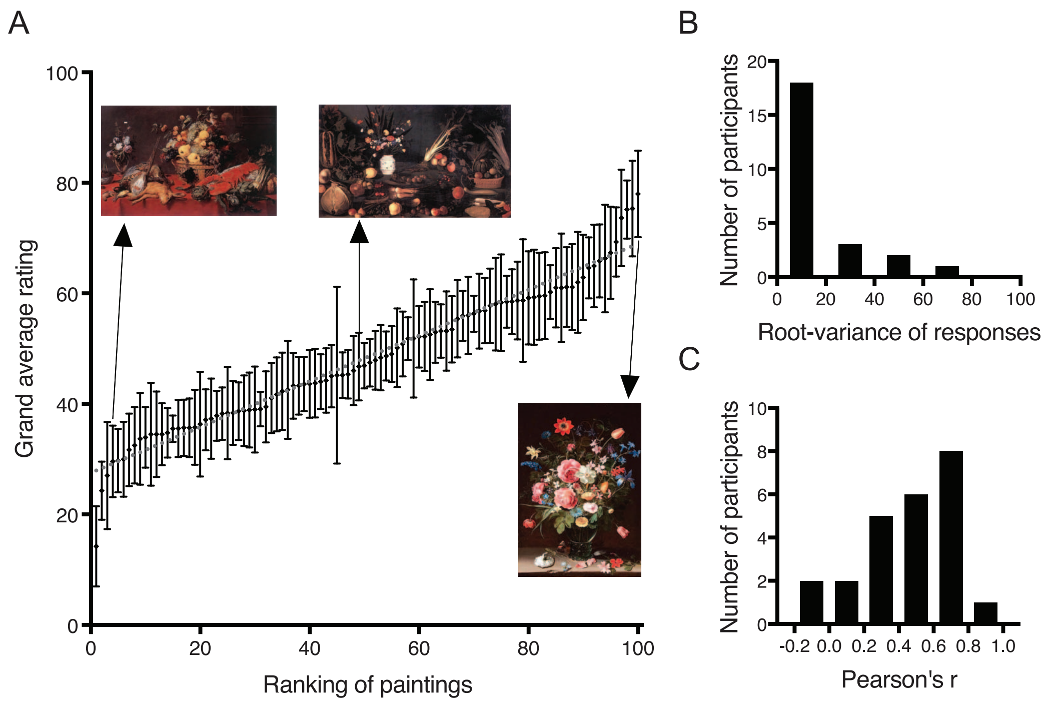

Figure 2. Mean attractiveness of all paintings across participants and its correlation with individual ratings and root-variance of all participants. (A) Grand average ratings for the entire set of stimuli in rank order. As 40 stimuli were randomly selected from a set of 100 paintings for each individual, the total number of ratings per painting in this group mean plot differs slightly (see *Design and procedure*). The ranked paintings form a consistent order that is well described by a linear fit, represented by the dotted line. Error bars represent ±1 standard error of the ratings. Sample images are *Still Life with Crab, Poultry, and Fruit* by Frans Snyders (1618), *Still Life with Flowers and Fruit* by Caravaggio (1601) and *Flower Still-Life* by Clara Peeters (*n.d.*) from left to right. All images are in the public domain. (B) Histogram of root-variance of responses across participants. (C) Histogram of Pearson's *r* for the correlation between each individual's ratings of their sample of 40 paintings and the corresponding group mean ratings.

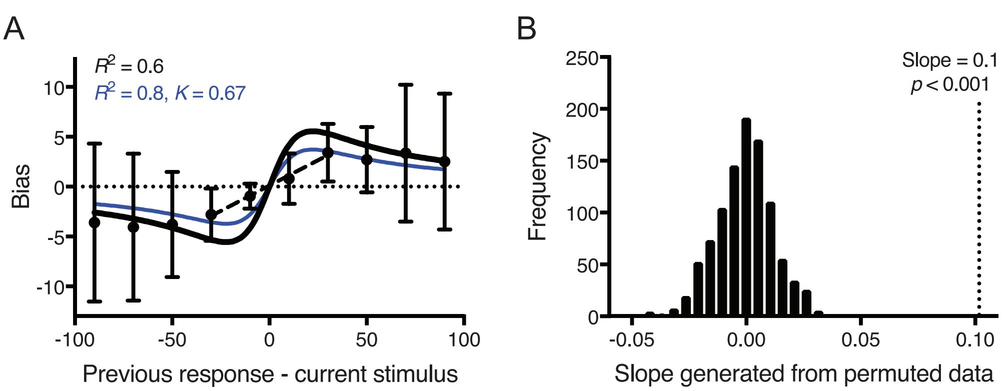

Figure 3. Data from Experiment 1 showing assimilative serial dependence in judging attractiveness of artworks for a viewing time of 1,000 ms. (A) Data points show group mean bias plotted as a function of the distance between the previous rating and the mean attractiveness of the current painting. Error bars represent standard deviations over the group. The black continuous curve shows the best fitting Kalman-filter model ($R^2 = 0.6$) described in the methods, and the blue one shows predicted bias when scale-factor (K) was introduced to the model. The model prediction shows that the peak serial dependence magnitude occurs when the current response and previous stimuli are relatively close. The dashed line represents the best linear fit to the middle four data points. (B) Results of a permutation test on the slope of the best fitting line shown at left. The best fitting slope was 0.1 (dotted line), and the test revealed it was significantly steeper than slopes generated by linear fits to 1000 permutations of the data.

Importantly, the magnitude of these effects gradually reduced as the intertrial stimulus distance became more extreme. One curious feature of the data is that the error bars are larger at the extremes. This could be due to there being less data points in these bins, which might then raise a question about whether those data validly represent a reduction in bias at the extremes of intertrial stimulus distance. However, given that we also found the same tendency to return to zero baseline in Experiment 2 (see Figure 5A), with similar numbers of data points in the extreme bins, the decline in bias seems robust. Interestingly, we did not observe high variances in the extreme bins for Experiment 2. Although the reason for this is not clear, it might be related to the longer inspection times in Experiment 1 allowing more cognitive evaluation and hence individual variability, whereas the ratings in Experiment 2 with briefer presentations might be more determined by low-level perceptual properties. It is premature to draw firm conclusions on this point, and future studies could investigate this further.

The thick continuous line is the prediction of the Kalman filter model (Equation 3). This model captures the essence of the serial dependence, including the fact that it is strongest when the difference in past response and present stimuli was not too great. The model also takes into account the relative reliability of the past and the current, and this consistency is represented as a root-variance of responses. For the group model prediction, we used the median root-variance across subjects ($\sigma = 15.72$). The fit of the model to the group mean data is $R^2 = 0.6$, very acceptable, as there were no free parameters. With one degree of freedom given to the magnitude of prediction (a scale factor K which weights the entire right-hand side of Equation 3), the scaled model gives an improved fit of $R^2 = 0.8$ when K has a value of 0.67 (blue continuous line in Figure 3A). In line with previous serial dependence studies showing assimilative patterns (Alexi et al., 2018; Bliss, Sun, & D'Esposito, 2017; Fischer & Whitney, 2014; Fritsche et al., 2017), our finding suggests that aesthetic ratings are systematically biased towards information from the recent past.

The assimilative serial dependence observed in our results and reflected in the model was further supported when we fitted a linear regression to the four data points in the central part of the response-stimulus distance range (dotted line in Figure 3A). In this central range, the previous response is of comparable attractiveness to the current one, and the slope of this line provides an effective way to measure the magnitude of the bias (Alexi et al., 2018). The slope of the regression line was 0.1, which was tested for significance by 1000 iterations of a permutation test. The results showed that the slope was significantly steeper than any of those generated by the permuted data ($N = 1000$, $p < 0.001$; see Figure 3B). For the permutation test, we shuffled the trial order of each individual's raw data and recomputed the intertrial distance on each iteration. The slope was calculated by fitting a regression line to the central-range data points of the newly generated group mean bias. Complementing this, a Bayesian linear regression analysis indicated that the slope was meaningfully different from zero ($BF_{10} = 6.51$, $R^2 = 0.99$).

Correlation between art score and serial dependence

Sixteen of 24 participants also completed questionnaires devised to measure their interest in and knowledge of art (Specker et al., 2018). Scores ranged between 0–70 for art interest and 0–26 for art knowledge. Average scores across participants were 39.2 ± 12.4 and 5 ± 3.2, respectively, for the interest and the knowledge questionnaires. To see whether these two indexes were related to the magnitude of serial dependence in aesthetic ratings, we correlated the slope of the linear fit to the middle four data points with scores for art interest and knowledge, separately. Neither of the correlations reached statistical significance, although both art scores showed a weak negative relationship to the bias magnitude: art interest, Pearson's $r(14) = -0.17$, $BF_{10} = 0.37$, $p = 0.52$; art knowledge $r(14) = -0.25$, $BF_{10} = 0.46$, $p = 0.35$ (see Figure 4). The lack of correlations could be due to the fact that our subjects were all naïve to arts and showed low art knowledge in general.

Response time

As time to make an aesthetic judgment was self-paced, we also investigated whether response time (RT) related to the serial dependence effect. Mean RT was calculated for every participant (average across participants 1.57 s ± 0.54) and then correlated with the magnitude of serial bias. Results showed that the relationship between RT and the bias magnitude is nonsignificant: $r(22) = 0.15$, $p = 0.49$ (see Figure A2). Interestingly, RT was negatively correlated with both art scores: art interest, $r(14) = -0.5$, $p = 0.05$; art knowledge, $r(14) = -0.59$, $p = 0.02$; the more one is interested in or knowledgeable about art, the less time is spent making an aesthetic judgment.

Experiment 2

Most investigations of serial dependence have used brief stimulus presentation times on the order of two to three hundred milliseconds, whereas we presented artworks for 1,000 ms in Experiment 1. Based on

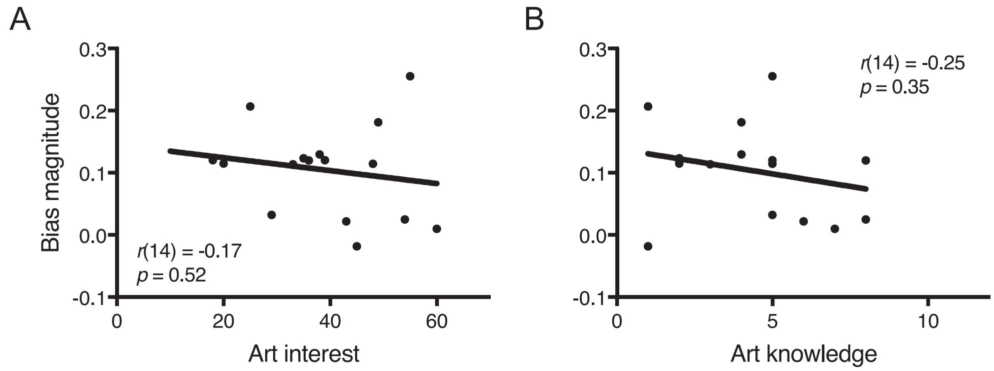

Figure 4. Results from a questionnaire on art interest and art knowledge. Sixteen out of 22 participants completed the questionnaire asking about their interest in and knowledge of art. (A) Scatterplot relating level of interest in art and the magnitude of the serial dependence effect (the slope of a linear fit to the middle four data points), for 16 observers. (B) Scatterplot for the same observers showing the relationship between the serial dependence effect and their level of art knowledge.

previous findings, this is more than long enough to make meaningful aesthetic judgments, which can be formed as quickly as 100 ms (Locher, Krupinski, Mello-Thoms, & Nodine, 2007) or even faster (Verhavert, Wagemans, & Augustin, 2018), but does this response time necessarily mean that the sequential bias of aesthetic judgments would occur quickly as well? With dissimilar exposure time to an artwork, the evidence available on which to base an aesthetic judgment would be qualitatively different, even if the same conclusion can be reached. For instance, top-down factors such as relevant knowledge and art taste would be more easily accessible when viewing time is longer. If we posit that a certain factor of aesthetic judgment is contributing to its serial dependence more than the others, the magnitude of bias would vary with the sort of evidences the current judgment is grounded on.

Experiment 2 examines this by repeating our first experiment with the stimulus exposure shortened to just 250 ms, immediately followed by a noise mask to terminate image persistence. This is shorter than an average fixation duration for art appreciation (Locher et al., 2007), which might be long enough to process basic visual features of the painting but not for higher level factors to be engaged, which may require several fixations and time to invoke knowledge and experience about the artwork. Once the image is replaced by the poststimulus mask, the image that endures in the mind's eye is of the noise mask, making further reflection and evaluation on the artwork very difficult. If such higher level factors bring about serial dependence in attractiveness ratings, we predict the positive serial dependence will be significantly reduced in Experiment 2. Alternatively, it is possible that aesthetic judgments are strongly guided by initially extracted visual information in the image and that brief exposure durations will suffice to produce a significant assimilation to the preceding aesthetic rating.

Methods

Participants

Twenty-two new participants (six males, 16 females; mean age, 25 years old, ranging from 20 to 37) with normal or corrected-to-normal vision recruited from the University student population gave informed consent and were compensated for their time.

Stimuli and apparatus

Stimuli and apparatus were identical to those in Experiment 1 except for the following details. First, we used a single set of 40 paintings as stimuli for all participants in this experiment. These were selected from the 100 paintings used in Experiment 1, and the criteria for selection were (a) they had been rated by more than six participants in the previous experiment, and (b) from one hundred paintings divided into 10 bins according to the ratings obtained in Experiment 1, four paintings were selected from each bin. Second, stimulus duration was only 250 ms (cf. 1,000 ms in Experiment 1) and every painting was immediately masked with a static Gaussian white noise image for 500 ms after stimulus presentation to minimize visual persistence (Teichner & Wagner, 1964) and color afterimages. The noise mask had the same size as the image stimulus and was created by randomly assigning pixel luminance values from a normal distribution ($\mu = 128$, $\sigma = 50$; 0–255 grayscale).

Design and procedure

Apart from the shorter stimulus duration and the introduction of the postimage noise mask, the experimental design and procedure were identical to Experiment 1.

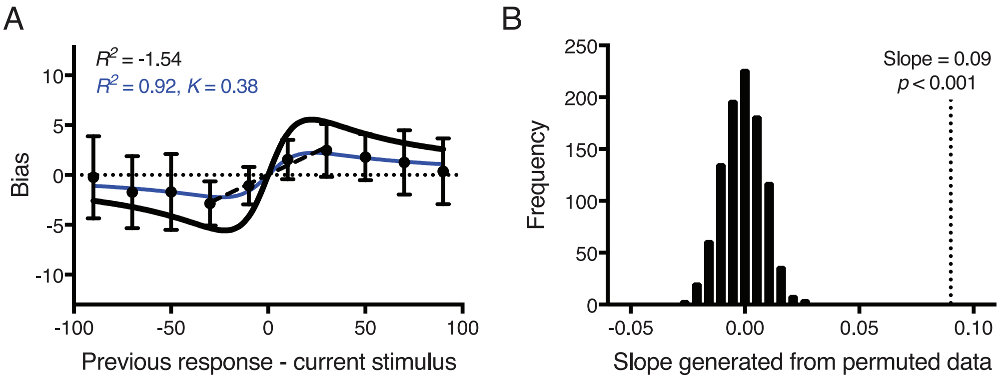

Figure 5. Data from Experiment 2 showing assimilative serial dependence in judging aesthetic attractiveness for a viewing time of 250 ms. (A) Group mean bias plotted against the distance between previous response and the current stimulus. The Kalman-filter model fit to the group mean data is represented as a continuous curve. The blue-colored curve represents the scaled model prediction ($R^2 = 0.92$, scale-factor = 0.38). Error bars represent standard deviations. The best-fitting slope to the middle four data points is represented as a dashed line. (B) The slope of best fit to the middle four data points is 0.09 (dotted line) and was found to be significant by permutation test ($N = 1000$, $p < 0.001$).

Results

Serial dependence in aesthetic judgments

We first checked the global linear trend of responses as in Experiment 1 and found that 14 out of 22 subjects showed significant trend in attractiveness rating (mean slope –0.44 ± 0.56; –0.31 ± 0.48 for all subjects). However, detrending of the data did not drive the serial dependence (see Figure A1) and hence the raw data was used for further analyses. Sequential bias of aesthetic ratings was analyzed using the same method described in the Results section of Experiment 1. As shown in Figure 5A, we again observed that attractiveness judgments of briefly presented images were assimilated toward the preceding response. The parameter-free Kalman-filter model did not provide a good fit to the group mean serial dependence data, with negative R^2 (meaning the fit was worse than the mean). However, introducing the scale factor K to the model greatly improved the predicted magnitude of bias. A scaling of 0.38 produced a fit with $R^2 = 0.92$, and the median root-variance was 16.6, which was similar to that in Experiment 1.

These results are interesting in that the current aesthetic judgment is still assimilated toward the previous response but with a reduced magnitude. Given that the bias in aesthetic judgment could be predicted fairly well without scaling in the 1,000 ms viewing condition, serial dependence of aesthetic judgments might require longer exposures to an image to develop. Despite the smaller degree of bias in the 250 ms viewing condition, the slope of the linear fit to the data points in the central range was significantly steeper than any of those generated by the permutation test ($N = 1000$, $p < 0.001$; $BF_{10} = 3.71$, $R^2 = 0.97$; see Figure 5B).

Correlation of aesthetic ratings across experiments

Grand average ratings for the 40 paintings showed a very high correlation across the two experiments (see Figure 6A); $r(38) = 0.81$, $p < 0.001$. This high level of agreement between the experiments (and between two independent groups of observers) suggests that both viewing durations were sufficient for participants to process artworks and rate their attractiveness. This is in line with previous reports showing that people are able to form a gist perception of artwork that is consistent across different presentation durations (Augustin, Defranceschi, Fuchs, Carbon, & Hutzler, 2011; Locher, 2015; Locher et al., 2007), or an aesthetic judgment in as brief as 30 ms (Verhavert et al., 2018). However, a recent study accentuated the individual differences rather than shared tastes in aesthetic ratings of artworks (Vessel, Maurer, Denker, & Starr, 2018). It is probable that correlating grand average ratings between two experiments might underestimate the variability of the ratings among subjects. Therefore, we ran another analysis in which ratings of one subject from Experiment 1 was correlated with the grand average ratings of Experiment 2. Although Pearson's r varied across subjects ranging from –0.22 to 0.86, the mean of the r values was clearly positive, indicating that in general individuals showed a fairly high agreement with the other group of subjects (median $r = 0.4 \pm 0.33$).

To investigate what leads to the high agreement among participants, we first looked at the correlation of artworks according to their styles, namely abstract versus representational (13 and 27 out of 40 paintings, respectively), because previous work reported lower agreement over observers for ratings of abstract images (Brinkmann, Commare, Leder, & Rosenberg, 2014; Leder et al., 2014; Leder, Goller, Rigotti, & Forster, 2016; Schepman, Rodway, Pullen, & Kirkham, 2015;

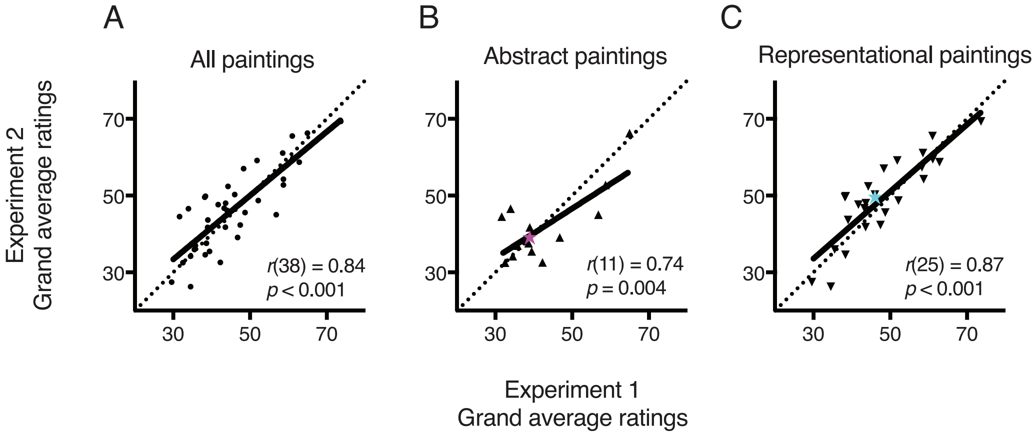

Figure 6. Grand average ratings of paintings across two experiments. (A) Group mean attractiveness for 40 paintings were highly correlated between two experiments with different viewing times (1,000 ms vs. 250 ms) and different observers. The dotted line represents the identity line. (B) and (C) Grand averages from panel (A) plotted separately for abstract and representational paintings. Dotted lines and colored markers in each plot show the regression line and the median ratings of two experiments, respectively.

Vessel & Rubin, 2010). We observed that aesthetic ratings of both abstract and representational paintings showed consistency across participants of both experiments (see Figure 6B and C). These results were again confirmed when single subject's ratings from Experiment 1 were correlated with grand averages of Experiment 2: the median value for Pearson's r across participants was 0.59 ± 0.41 and 0.47 ± 0.37, respectively, for abstract and representational paintings. However, abstract paintings tended to be rated as less attractive and in a narrow range (excepting a couple of outliers), while the representational paintings were rated across a much wider range of attractiveness and had a higher mean attractiveness rating overall. The discrepancy between our findings and previous reports regarding abstract painting may be due to semantic content. In the abovementioned studies, abstract images were synthetic or lacked semantic content whereas many of ours were abstract renditions of real-world scenes (see Supplementary File S1 for the list of stimuli) and they also report that real-world/semantic content leads to greater consistency over observers.

Based on previous findings showing regularities in image statistics of artworks (Graham & Redies, 2010; Redies, Hasenstein, & Denzler, 2007), we calculated the Fourier spectral slope of the 40 paintings in Experiment 2 to see whether it would be predictive of aesthetic rating (for detailed methods, see Schweinhart & Essock, 2013). The Fourier amplitude spectra for artworks generally follows an inverse power law, with amplitude declining with increasing spatial frequency, $1/f^{\alpha}$ (as is common for natural images). Our results showed that the mean α of our stimuli was 1.33 ± 0.17. However, the correlation with the grand average ratings collapsed across two experiments did not reach statistical significance, $r(38) = -0.17$, $p = 0.28$. When we extended the same analysis to the 100 images used in the Experiment 1, we obtained a very similar mean value of $\alpha = 1.31 \pm 0.18$ and the correlation to the grand average ratings again was not significant, $r(98) = -0.07$, $p = 0.46$ (see Figure 7A through D).

We also tested whether the color content of a painting affects its aesthetic rating. To this end, chromatic diversity and mean hue of each painting were analyzed (Li & Chen, 2009). Whereas chromatic diversity was not predictive of attractiveness rating, $r(38) = 0.22$, $p = 0.18$, mean hue of the painting showed a significant correlation with the grand average ratings, $r(38) = 0.33$, $p = 0.04$. To be more precise, participants showed a tendency to prefer paintings with cool colors (blue end of the spectrum) to warm colors (red end). The same analyses on the full image set resulted in $r(98) = 0.14$, $p = 0.15$ for the chromatic diversity and $r(98) = 0.34$, $p < 0.001$ for the mean hue (see Figure 7E).

Art interest and knowledge do not correlate with serial dependence

All participants except one completed the same questionnaires used in Experiment 1 to measure their levels of art interest and art knowledge. When averaged across participants, the mean scores for art interest and knowledge were 41.6 ± 11.6 and 4.5 ± 2.7, respectively. When individual scores for these questionnaires were plotted against the magnitude of each subject's serial dependence effect, we found that neither correlation reached statistical significance (see Figure 8).

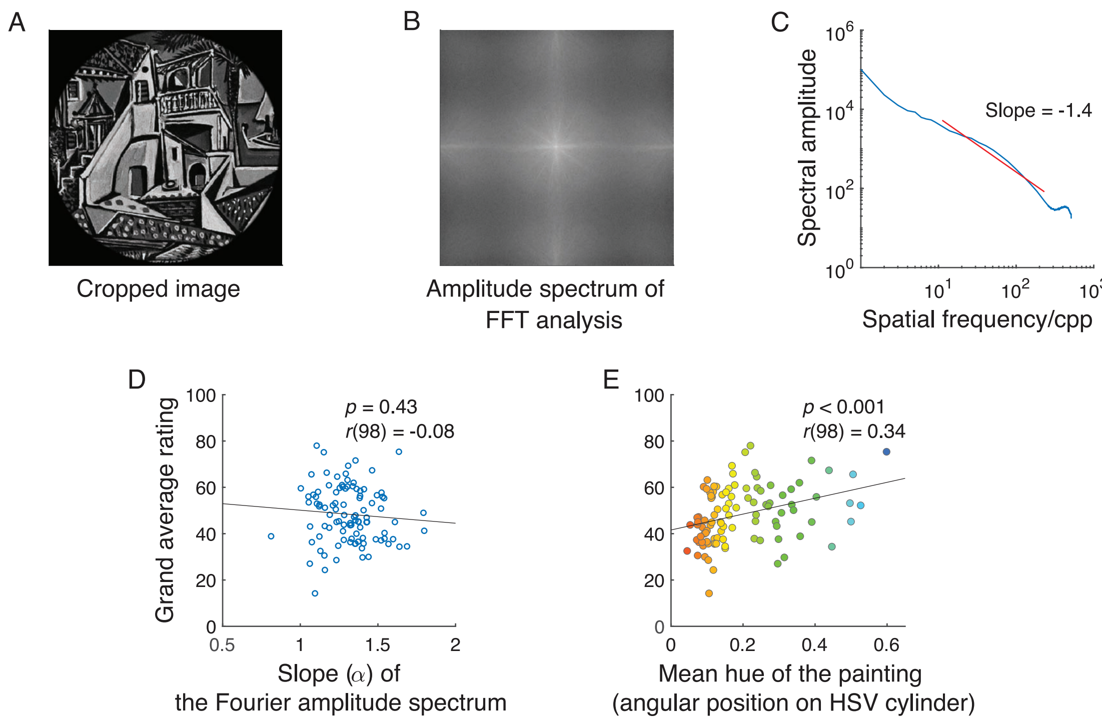

Figure 7. Image statistics of a sample painting and correlation results. (A) The grayscale image was first cropped into a circle with the edge blurred before a fast Fourier transform (FFT) analysis. (B) and (C) The amplitude spectrum of the image was plotted as a function of spatial frequency which was defined by cycles per pixel (cpp). The slope of the amplitude spectrum was calculated by fitting a regression line to the average amplitude binned according to spatial frequency between 10–256 cpp. (D) Correlation between the Fourier spectral slope and the grand average rating for 100 paintings. Results did not reach statistical significance. (E) Correlation between the mean hue of the painting and its grand average rating. Hue value was defined as an angular position on the HSV cylinder, with red primary being 0°. For calculation of the mean hue, only the pixels with saturation value greater than 0.2 and brightness between 0.15 and 0.95 were considered for each painting. The color of each data point represents the mean hue of the corresponding painting. Results showed that the paintings with cool colors (e.g., blue) tended to be preferred to those with warm colors (e.g., red).

Response time

We conducted further analyses by correlating mean RT (average across participants 1.25 s ± 0.39) with the bias magnitude and with both art questionnaire results for all participants (see Figure A2). Similar to the condition when viewing time was 1,000 ms, mean RT did not correlate with the bias magnitude, $r(19) = 0.19$, $p = 0.4$. Furthermore, it was observed that neither questionnaire score yielded a meaningful relationship with RT: art interest, $r(19) = -0.08$, $p = 0.72$; art knowledge, $r(19) = 0.04$, $p = 0.86$. Interestingly, the mean RT of subjects was significantly different depending on whether stimulus presentation time was 1,000 ms or 250 ms, $t(44) = 2.26$, $p = 0.03$, being shorter in Experiment 2 where the duration was 250 ms.

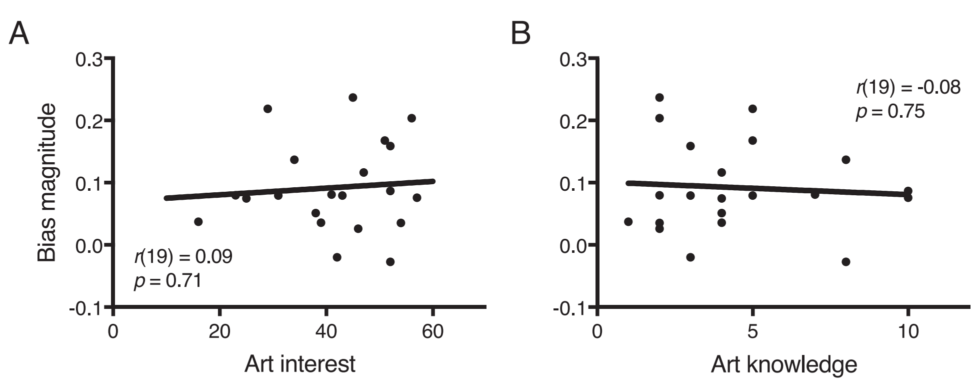

Figure 8. Questionnaire scores for (A) art interest and (B) art knowledge for 21 observers from Experiment 2. Neither art interest nor art knowledge was correlated with the magnitude of serial dependence.

Discussion

In this study, we showed that the aesthetic judgment of artworks was assimilated towards the recent past. The currently viewed painting tended to be rated as more attractive when it was preceded by one with a higher attractiveness rating and was rated as less attractive when preceded by one with a lower attractiveness. The serial dependence effect was systematically modulated by the similarity of two successive paintings, being more prominent when the previous painting was judged to be relatively similar in attractiveness as the current one. This trend was well captured by the Kalman-filter ideal-observer model (Cicchini et al., 2014; Cicchini et al., 2018), which predicts that the current percept should be a weighted sum of the past and present, where the weights are determined by the reliability (inverse variance) associated with the past and present stimuli. Here, the model successfully captures the key qualitative features of the data and provides a good quantitative fit.

Our findings add to previous studies suggesting that assimilative serial dependence occurs for higher level judgments such as face and body attractiveness (Alexi et al., 2018; Taubert, Van der Burg, et al., 2016; Xia et al., 2016), and extends the scope of assimilative serial dependences further into the domain of aesthetic judgments of artwork. It remains unanswered, however, at which stage of aesthetic judgment this positive bias occurs. Aesthetic judgments are thought to involve many levels of processing, from basic sensory to high-level cognitive processes (Leder, Belke, Oeberst, & Augustin, 2004), leaving many possibilities open for how the current percept is attracted toward the past. One possible mechanism proposed by several studies is that positive serial dependence occurs at the perceptual stage of visual processing (Cicchini et al., 2017, 2018; Liberman, Manassi, & Whitney, 2018; Manassi, Liberman, Kosovicheva, Zhang, & Whitney, 2018). For instance, Cicchini and colleagues (2017) devised an experimental task where the stimulus and response were dissociated and showed that the stimulus rather than the response is the primary driver of serial dependence. On the other hand, some researchers suggested that positive serial dependence can be attributed to a postperceptual, decision-making process (Fritsche et al., 2017; Pegors, Mattar, Bryan, & Epstein, 2015) or mnemonic process (Bliss et al., 2017). However, more recent studies using very similar experimental methods as these research groups showed that positive serial dependence cannot be solely explained by such postperceptual processes (Cicchini et al., 2017; Manassi et al., 2018).

Our experiments do not speak conclusively to the issue of whether serial dependence of aesthetic judgment acts at a perceptual or postperceptual level. However, a few findings suggest that at least a portion of positive serial dependence might be driven by early visual processing. First, we observed that the uncertainty measures of subject responses (σ) were comparable across both experiments, two-sample t-test $t(44) = 0.84$, $p = 0.4$, and the mean attractiveness ratings were highly correlated (see Figure 6A). This indicates that the visual features extracted from the artwork on a given trial were equivalently informative for reaching an aesthetic judgment in both the short- and long-duration viewing conditions.

However, while ratings and their variation were similar for both durations on individual trials, serial dependence was weaker for the shorter duration. The amplitude (K) of the Bayesian observer model predicting the serial dependence from one trial to the next had to be scaled down in the short-duration condition, from a value of $K \sim 0.7$ for the long duration to $K \sim 0.4$ for the short duration. Thus, the shorter duration appears to reduce the carryover of the previous judgment to the current one in a process that is distinct from the attractiveness judgment itself. The participant's task was the same for both durations, so there should have been no differences in decisional processes in the two experiments, while the perceptual processes were curtailed in the short-presentation experiment with a poststimulus mask. This result would seem to suggest that serial dependence occurs, at least to some extent, at the perceptual level. It is possible that the poststimulus noise mask curtailed image processing and impaired the retention of the artwork in working memory while the rating judgment was made. Alternatively, it is possible that the noise mask may have functioned as second image in the sequence, interposed between the artwork and the rating, thereby turning the serial dependence into a weaker, two-back effect.

Correlation analysis of mean attractiveness ratings showed that aesthetic judgments of paintings could be formed very quickly and consistently across different exposure times. This is not surprising given that people are able to establish a general impression within a single fixation (Locher, 2015; Locher et al., 2007) and judge the similarity of two paintings in terms of styles in 50 ms and even faster for the content (Augustin, Leder, Hutzler, & Carbon, 2008). We also found a higher agreement on attractiveness ratings across different viewing time conditions for representational paintings relative to abstract paintings, indicating that the time required to develop an aesthetic judgment of an artwork depends on its style. In addition, we add to the previous findings by showing a contribution of color to aesthetic judgment of paintings (Brachmann & Redies, 2017; Leder et al., 2004; Li & Chen, 2009; Palmer & Schloss, 2010). As suggested by Palmer and Schloss (2010), our subjects gave higher attractiveness ratings to paintings with an average hue of cooler color (e.g., blue or green) than those with warmer color (e.g., red or orange). Lastly, we examined whether the Fourier spectral content of the image affects the attractiveness. Fourier amplitude spectra of spatial frequency in natural images are known to follow an inverse power law $1/f^{\alpha}$, with α approximating 1.2 (Graham & Field, 2007; Graham & Redies, 2010; Redies, Hanisch, Blickhan, & Denzler, 2007; Redies, Hasenstein, et al., 2007; Schweinhart & Essock, 2013). This statistical regularity is also seen in paintings and other visual artworks (Mather, 2014; Spehar, Walker, & Taylor, 2016). It has been suggested that artists mimic the spectral slope of natural images to make them aesthetically pleasing to the human visual system, which has evolved to optimally encode the statistics of natural scenes (Mather, 2014). Although we found a typical α value for the mean slope of amplitude spectra of the paintings used in our study, variation in slope among the images was not predictive of the attractiveness, likely due to the small range of α values.

Art expertise has been discussed in a number of studies in relation to the processing of artworks. Pang, Nadal, Muller-Paul, Rosenberg, and Klein (2013) measured brain activities of art experts and laypersons during art appreciation using electroencephalography. Counterintuitively, their findings showed decreased activity in the late stages of visual processing among art experts, compared to laypersons. This was interpreted as reflecting greater neural efficiency with increasing expertise, and a similar finding has been observed in the music domain as well (Brattico & Pearce, 2013; Muller, Hofel, Brattico, & Jacobsen, 2010). In the same vein, it is reasonable to assume that art experts might be more efficient in individuating each work of art in the presentation sequence and compartmentalizing their evaluation. If so, one might expect a negative relationship between the magnitude of serial dependence and art expertise. However, this conjecture was not supported by the correlation between art scores and serial bias magnitude in the current study. One possible reason for these nonsignificant relationships could be the relatively small range of variability in art scores, as none of the participants in our study were art experts in a strict sense (e.g., art major students or artists). Accordingly, our sample of art knowledge scores was overwhelmingly concentrated at the low end (see Figure 4B and 7B). This is a speculation for the moment, and there is still a debate over whether practice and expertise result in neural efficiency (Else, Ellis, & Orme, 2015; Kelly & Garavan, 2005; van Paasschen, Bacci, & Melcher, 2015). A future study would need a much broader range of art expertise than our sample exhibited in order to test this, or better yet, directly contrast groups of experts and novices.

Conclusion

In conclusion, we found a serial dependence for aesthetic judgment of artwork, with aesthetic ratings of participants biased toward the recent past. The bias was stronger when the stimulus was viewed for 1 s rather than 250 ms (curtailed by a poststimulus mask). Although the serial effect was smaller for the shorter duration, this is unlikely due to cognitive factors, as there was no limit on response time and thus little reason to expect different levels of influence from cognitive sources during aesthetic judgments. We have suggested that the weaker serial effect for brief presentations could have a mnemonic origin, resulting from the reduced time to encode the image into working memory. This conjecture would need corroboration from further evidence from future studies. Future studies should also recruit participants with a greater range of art expertise and knowledge, to shed light on how these interact with serial dependence of aesthetic judgment.

Keywords: serial dependence, aesthetic judgment, artwork, Bayesian observer model

Acknowledgments

This work was supported by an Australian Research Council grant DP150101731 to DA and DB, and by the EU Horizon 2020 research and innovation programme under Grant Agreement No 832813 Spatio-temporal mechanisms of generative perception —GenPercept.

Commercial relationships: none.
Corresponding author: Sujin Kim.
Email: skim5220@uni.sydney.edu.au.
Address: School of Psychology, The University of Sydney, New South Wales, Australia.

References

Alais, D., Leung, J., & Van der Burg, E. (2017). Linear summation of repulsive and attractive serial dependencies: Orientation and motion dependencies sum in motion perception. *Journal of Neuroscience*, *37*(16), 4381–4390, https://doi.org/10.1523/JNEUROSCI.4601-15.2017.

Alexi, J., Cleary, D., Dommisse, K., Palermo, R., Kloth, N., Burr, D., & Bell, J. (2018). Past visual experiences weigh in on body size estimation. *Scientific Reports*, *8*(1), 215, https://doi.org/10.1038/s41598-017-18418-3.

Augustin, M. D., Defranceschi, B., Fuchs, H. K., Carbon, C. C., & Hutzler, F. (2011). The neural time course of art perception: An ERP study on the processing of style versus content in art. *Neuropsychologia*, *49*(7), 2071–2081, https://doi.org/10.1016/j.neuropsychologia.2011.03.038.

Augustin, M. D., Leder, H., Hutzler, F., & Carbon, C. C. (2008). Style follows content: On the microgenesis of art perception. *Acta Psychologica*, *128*(1), 127–138, https://doi.org/10.1016/j.actpsy.2007.11.006.

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), 14739, https://doi.org/10.1038/s41598-017-15199-7.

Brachmann, A., & Redies, C. (2017). Computational and experimental approaches to visual aesthetics. *Frontiers in Computer Neuroscience, 11*, 102, https://doi.org/10.3389/fncom.2017.00102.

Brainard, D. H. (1997). The Pscyhophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48–61, https://doi.org/10.1037/a0031624.

Brinkmann, H., Commare, L., Leder, H., & Rosenberg, R. (2014). Abstract art as a universal language? *Leonardo*, *47*(3), 256–257, https://doi.org/10.1162/LEON_a_00767.

Chang, S., Kim, C. Y., & Cho, Y. S. (2017). Sequential effects in preference decision: Prior preference assimilates current preference. *PLoS One*, *12*(8), e0182442, https://doi.org/10.1371/journal.pone.0182442.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences, USA*, *111*(21), 7867–7872, https://doi.org/10.1073/pnas.1402785111.

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14):6, 1–9, https://doi.org/10.1167/17.14.6. [PubMed] [Article]

Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings*. *Biological Sciences*, *285*(1890), https://doi.org/10.1098/rspb.2018.1722.

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, https://doi.org/10.1037/a0031020.

Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating stable representations: Serial dependence in vision. *PLoS One*, *6*(1), e16701, https://doi.org/10.1371/journal.pone.0016701.

Dolese, M., Zellner, D. A., Vasserman, M., & Parker, S. (2005). Categorization affects hedonic contrast in the visual arts. *Bulletin of Psychology and the Arts*, *5*, 21–25.

Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, *6*(3), 345–358, https://doi.org/10.1088/0954-898x_6_3_003.

Else, J. E., Ellis, J., & Orme, E. (2015). Art expertise modulates the emotional response to modern art, especially abstract: An ERP investigation. *Frontiers in Human Neuroscience*, *9*, 525, https://doi.org/10.3389/fnhum.2015.00525.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743, https://doi.org/10.1038/nn.3689.

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, *27*(4), 590–595, https://doi.org/10.1016/j.cub.2017.01.006.

Graham, D. J., & Field, D. J. (2007). Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, *21*(1–2), 149–164, https://doi.org/10.1163/156856807782753877.

Graham, D. J., & Redies, C. (2010). Statistical regularities in art: Relations with visual coding and perception. *Vision Research*, *50*(16), 1503–1509, https://doi.org/10.1016/j.visres.2010.05.002.

Imamoglu, E. O. (1974). Initial evaluation of stimuli as another limiting condition of the exposure effect. *British Journal of Social and Clinical Psychology*, *13*(2), 157–159, https://doi.org/10.1111/j.2044-8260.1974.tb00103.x.

Hsu, S. M., & Yang, L. X. (2013). Sequential effects in facial expression categorization. *Emotion*, *13*(3), 573–586, https://doi.org/10.1037/a0027285.

Kelly, A. M., & Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, *15*(8), 1089–1102, https://doi.org/10.1093/cercor/bhi005.

Kenrick, D. T., & Gutierres, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, *38*(1), 131–140, https://doi.org/10.1037/0022-3514.38.1.131.

Khaw, M. W., & Freedberg, D. (2018). Continuous aesthetic judgment of image sequences. *Acta Psychologica*, *188*(2018), 213–219, https://doi.org/10.1016/j.actpsy.2018.04.018.

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences*, *21*(7), 493–497, https://doi.org/10.1016/j.tics.2017.04.011.

Leder, H. (2001). Determinants of preference: When do we like what we know? *Empirical Studies of the Arts*, *19*(2), 201–211, https://doi.org/10.2190/5tae-e5cv-xjal-3885.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(Pt. 4), 489–508, https://doi.org/10.1348/0007126042369811.

Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition and Emotion*, *28*(6), 1137–1147, https://doi.org/10.1080/02699931.2013.870132.

Leder, H., Goller, J., Rigotti, T., & Forster, M. (2016). Private and shared taste in art and face appreciation. *Frontiers in Human Neuroscience, 10*, 155, https://doi.org/10.3389/fnhum.2016.00155.

Li, C., & Chen, T. (2009). Aesthetic visual quality assessment of paintings. *IEEE Journal of Selected Topics in Signal Processing*, *3*(2), 236–252, https://doi.org/10.1109/jstsp.2009.2015077.

Liberman, A., Manassi, M., & Whitney, D. (2018). Serial dependence promotes the stability of perceived emotional expression depending on face similarity. *Attention Perception & Psychophysics*, *80*(6), 1461–1473, https://doi.org/10.3758/s13414-018-1533-8.

Locher, P. (2015). The aesthetic experience with visual art "at first glance". In P. Bundgaard & F. Stjernfelt (Eds.), *Investigations Into the Phenomenology and the Ontology of the Work of Art* (pp. 75–88). Cham, Switzerland: Springer, Cham., https://doi.org/10.1007/978-3-319-14090-2_5.

Locher, P., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, *21*(1–2), 55–77, https://doi.org/10.1163/156856807782753868.

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review*, *25*(6), 2245–2253, https://doi.org/10.3758/s13423-018-1454-5.

Mather, G. (2014). Artistic adjustment of image spectral slope. *Art & Perception*, *2*(1–2), 11–22, https://doi.org/10.1163/22134913-00002018.

Muller, M., Hofel, L., Brattico, E., & Jacobsen, T. (2010). Aesthetic judgments of music in experts and laypersons—an ERP study. *International Journal of Psychophysiology*, *76*(1), 40–51, https://doi.org/10.1016/j.ijpsycho.2010.02.002.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences, USA*, *107*(19), 8877–8882, https://doi.org/10.1073/pnas.0906172107.

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, *64*, 77–107, https://doi.org/10.1146/annurev-psych-120710-100504.

Pang, C. Y., Nadal, M., Muller-Paul, J. S., Rosenberg, R., & Klein, C. (2013). Electrophysiological correlates of looking at paintings and its association with art expertise. *Biological Psychology*, *93*(1), 246–254, https://doi.org/10.1016/j.biopsycho.2012.10.013.

Park, J., Shimojo, E., & Shimojo, S. (2010). Roles of familiarity and novelty in visual preference judgments are segregated across object categories. *Proceedings of the National Academy of Sciences, USA*, *107*(33), 14552–14555, https://doi.org/10.1073/pnas.1004374107.

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, https://doi.org/10.1037/1931-3896.2.3.171.

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*(3), 664–673, https://doi.org/10.1037/xge0000069.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442, https://doi.org/10.1163/156856897x00366.

Redies, C., Hanisch, J., Blickhan, M., & Denzler, J. (2007). Artists portray human faces with the Fourier statistics of complex natural scenes. *Network*, *18*(3), 235–248, https://doi.org/10.1080/09548980701574496.

Redies, C., Hasenstein, J., & Denzler, J. (2007). Fractal-like image statistics in visual art: Similarity to natural scenes. *Spatial Vision*, *21*(1–2), 137–148, https://doi.org/10.1163/156856807782753921.

Schepman, A., Rodway, P., Pullen, S. J., & Kirkham, J.

(2015). Shared liking and association valence for representational art but not abstract art. *Journal of Vision*, *15*(5):11, 1–10, https://doi.org/10.1167/15.5.11. [PubMed] [Article]

Schweinhart, A. M., & Essock, E. A. (2013). Structural content in paintings: Artists overregularize oriented content of paintings relative to the typical natural scene bias. *Perception*, *42*(12), 1311–1332, https://doi.org/10.1068/p7345.

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2018). The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts*, https://doi.org/10.1037/aca0000205.

Spehar, B., Walker, N., & Taylor, R. P. (2016). Taxonomy of individual variations in aesthetic responses to fractal patterns. *Frontiers in Human Neuroscience, 10*, *350*, https://doi.org/10.3389/fnhum.2016.00350.

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*, 32239, https://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, https://doi.org/10.1038/srep22740.

Teichner, W. H., & Wagner, M. (1964). Visual afterimages as a source of information. *Human Factors*, *6*, 141–156, https://doi.org/10.1177/001872086400600204.

Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific Reports, 5*, 14526, https://doi.org/10.1038/srep14526.

van Paasschen, J., Bacci, F., & Melcher, D. P. (2015). The Influence of art expertise and training on emotion and preference ratings for representational and abstract artworks. *PLoS One*, *10*(8), e0134241, https://doi.org/10.1371/journal.pone.0134241.

Verhavert, S., Wagemans, J., & Augustin, M. D. (2018). Beauty in the blink of an eye: The time course of aesthetic experiences. *British Journal of Psychology*, *109*(1), 63–84, https://doi.org/10.1111/bjop.12258.

Vessel, E. A., Maurer, N., Denker, A. H., & Starr, G. G. (2018). Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition*, *179*, 121–131, https://doi.org/10.1016/j.cognition.2018.06.009.

Vessel, E. A., & Rubin, N. (2010). Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, *10*(2):18, 11–14, https://doi.org/10.1167/10.2.18. [PubMed] [Article]

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of attractiveness. Journal of Vision, *16*(15):28, 1–8, https://doi.org/10.1167/16.15.28. [PubMed] [Article]

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of Personality and Social Psychology, *9*(2, Pt.2), 1–27, https://doi.org/10.1037/h0025848.

Appendix

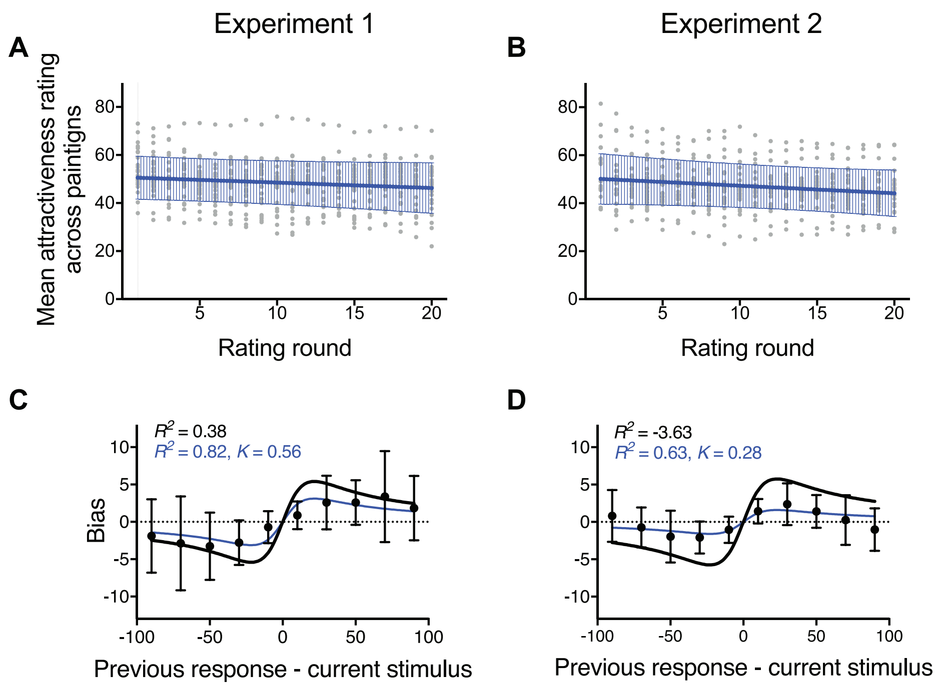

Figure A1. Global linear trend of mean attractiveness ratings across experiment. (A) Mean attractiveness ratings over repetition for all subjects from Experiment 1. Each dot represents a mean attractiveness across 40 paintings on nth round of rating within each individual. Linear trend over repeated ratings were tested by fitting a regression line to twenty mean ratings. Results showed a negative trend for 11 subjects, positive trend for four subjects and nonsignificant trend for nine subjects. The blue line with error bar (1 *SD*) represents the average of linear fits across subjects and the mean slope of the line was 0.23 ± 0.38. (B) Results from Experiment 2. Negative trend was found for 10, positive trend in four, and nonsignificant trend in eight subjects. Mean slope of the group linear fits was −0.31 ± 0.48. (C) and (D) Serial dependence results from detrended data. Raw data were detrended based on the rating round for a certain painting within subject. Same analysis as in the manuscript was conducted on the detrended data to examine serial dependence in attractiveness rating of artworks. We still observed that the aesthetic rating was assimilated toward the preceding one despite a slightly reduced magnitude in bias. Permutation test on the slope of data points in the central range confirmed that it is significantly steeper than any of the generated slopes (slope = 0.09, $p < 0.001$ for Experiment 1; slope = 0.08, $p < 0.001$ for Experiment 2).

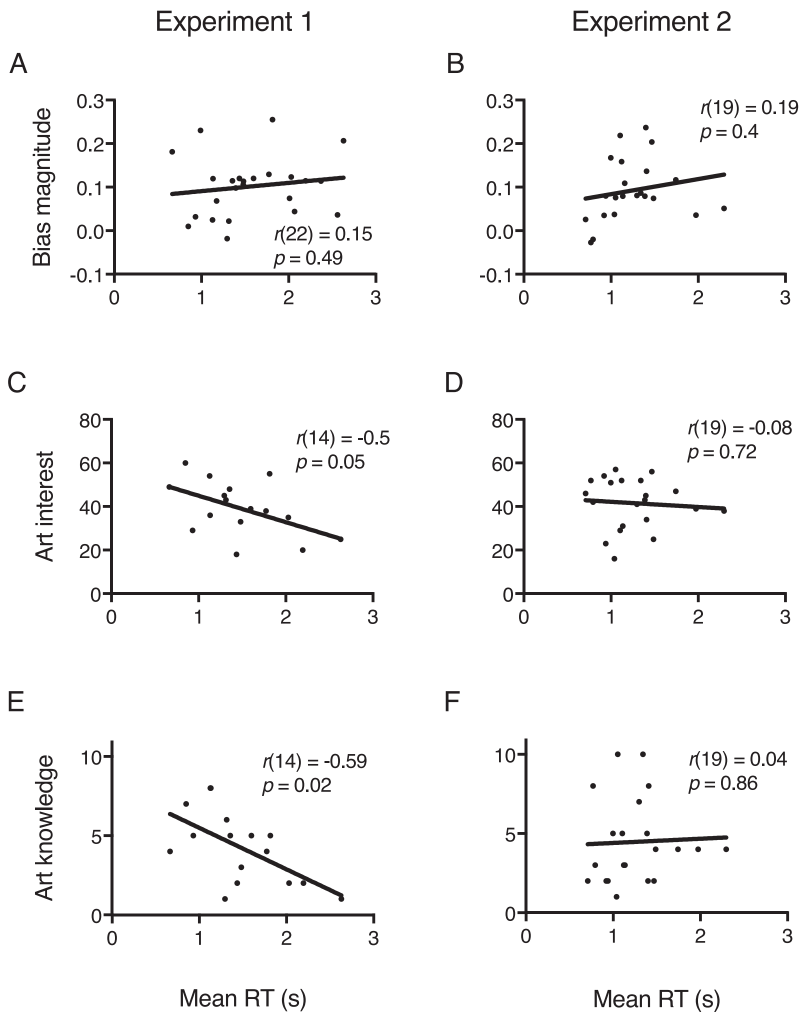

Figure A2. Mean response time to rate attractiveness and its relationship with bias magnitude and art scores. Left and right columns of figures depict results from Experiment 1 and Experiment 2, respectively. (A, B) Bias magnitude correlated with mean response time. Each data point represents each subject. Results showed that the correlation was not significant for both experiments. (C, D) Art interest scores correlated with mean response time. Mean response time was negatively correlated with the level of interest in art, only when the viewing time for artwork was 1 s. (E, F) Art knowledge score correlated with mean response time. Similar to the art interest, one tends to spend less time in making an aesthetic judgment with higher level of art knowledge.