Temporal and spatial ensemble statistics are formed by distinct mechanisms

Haojiang Ying[a,b], Edwin Burns[b,c], Amanda M Choo[b,d], and Hong Xu[b]*

[a]Department of Psychology, Soochow University, Suzhou, China

[b]Psychology, School of Social Sciences, Nanyang Technological University, Singapore

[c]Department of Psychology, University of Richmond, USA

[d]School of Biological Sciences, Nanyang Technological University, Singapore

Manuscript accepted for publication in Cognition on the 04/11/2019

*Correspondence:
Dr. Hong Xu
48 Nanyang Ave, HSS-04-06
Psychology, School of Social Sciences
Nanyang Technological University
Singapore 639818
Phone: +65 6592-1571
Fax: +65 6795-5797
Email: xuhong@ntu.edu.sg

Declarations of interest: none

Abstract word count: 312
Introduction and Discussion word count: 2990
Method and Results word count: 5623
Number of References: 76

Abstract

Our brains can extract a summary representation of the facial characteristics provided by a group of faces. To date, there has been a lack of clarity as to what calculations the brain is actually performing during this ensemble perception. For example, does ensemble processing average the fiducial points (e.g., distance between the eyes, width of the mouth) and surface characteristics (e.g., skin tone) of a set of faces in a fashion that produces what we call a 'morph average' face from the group? Or does ensemble perception extract a general 'gist average' of the face set (e.g., these faces are unattractive)? Here, we take advantage of the fact that the 'morph average' face derived from a group of faces is more attractive than the 'gist average'. If ensemble perception is performing morph averaging, then the adaptation aftereffects elicited by a morphed average face from a group should be equivalent to those elicited by the group. By contrast, if ensemble perception reflects gist averaging, then the aftereffects produced by the group should be distinct from those elicited by the more attractive morphed average face. In support of the morph averaging hypothesis, we show that the adaptation aftereffects derived via temporal ensemble perception of a group of faces are equal to those produced by the group's morphed average face. Moreover, these effects increase as a linear function of increasing attractiveness in the underlying group. We also reveal that spatial ensemble processing is not equal to temporal ensemble processing, but instead reflects the 'gist' attractiveness of the group of faces; e.g., these faces are unattractive. Finally, we show that gist averaging of a spatially presented group of faces is abolished when a temporal manipulation is additionally employed; under these circumstances, morph averaging becomes apparent again. In summary, we have shown for the first time that temporal and spatial ensemble statistics reflect qualitatively different perceptual calculations.

Keywords: rapid serial visual presentation, adaptation, ensemble statistics, face, attractiveness

Introduction

When we are presented with an array of stimuli in a scene, our brains involuntarily extract the ensemble statistics of the information that they convey (Alvarez, 2011; Haberman & Whitney, 2007, 2012; Whitney & Yamanashi Leib, 2017). For example, we can accurately report the mean emotion from a group of emotional faces (Haberman & Whitney, 2007, 2009; Whitney & Yamanashi Leib, 2017; Wolfe, Kosovicheva, Leib, Wood, & Whitney, 2015; Ying & Xu, 2017). Such averaging is considered to be a type of ensemble statistics (Alvarez, 2011; Ariely, 2001; Haberman, Brady, & Alvarez, 2015; Haberman & Whitney, 2007, 2009, 2012; Whitney & Yamanashi Leib, 2017; Ying & Xu, 2017), and can occur both spatially (i.e., multiple faces presented at once in a scene; e.g., Haberman & Whitney, 2007, 2009; Ying, Burns, Lin, & Xu, 2019) and temporally (i.e., different faces presented one at a time in rapid succession; e.g., Haberman, Harp, & Whitney 2009; Ying & Xu, 2017).

Despite researchers widely describing ensemble statistics as extracting the gist of a scene, it is still far from clear what this 'gist' represents (Alvarez, 2011; Whitney & Yamanashi Leib, 2017). For example, does ensemble coding extract a general representation of the group's mean characteristics, whereby the faces are summarized via what we call 'gist averaging'; e.g., the mean attractiveness of these unattractive faces is unattractive? Alternatively, does the brain calculate the fiducial points for each face (e.g., distance between eyes, width of the lips) with their surface characteristics (e.g., the redness of the cheeks), and then average them together to create a new mean face derived from this information? We call this latter form of ensemble coding 'morph averaging' due to the fact that it is very similar to how specialist computer morphing software creates an average face from a group of faces (Debruine & Tiddeman, 2017, Tiddeman, Burt, & Perrett, 2001).

Remarkably to date, there has been no clear evidence to support either the gist or morph averaging accounts of ensemble coding. Here, we tested these potential hypotheses by taking advantage of the well-established fact that a computer-generated average face, created by averaging the fiducial points and surface characteristics of a group of faces, is generally more attractive than the individual faces from which the average is comprised (DeBruine, Jones, Unger, & Little, 2007; Galton, 1878; Perrett, May, & Yoshikawa, 1994; Valentine, Darling, & Donnelly, 2004). This effect has been documented from the dawn of modern psychology, with Galton (1878) relaying that averaging leads to '…in every instance, a decided improvement of beauty' (Valentine et al., 2004). By requiring participants to perceive facial attractiveness in a temporal ensemble fashion, we can clearly test for the first time whether the morph average (i.e., the ensemble statistics of the group is equivalent to the morphed average face, such that a group of unattractive faces should no longer be perceived as unattractive) or the gist average (i.e., ensemble perception of the group should be less attractive than the morph average, such that a group of unattractive faces remains unattractive) hypothesis of ensemble coding is correct.

We therefore adapted participants to a group of faces presented one at a time in rapid serial visual presentation (RSVP; Potter, 1976). We chose an adaptation paradigm instead of a direct rating approach as adaptation is a powerful method that can detect perceptual effects even when explicit ratings are unable to (Liu, Montaser-Kouhsari, & Xu, 2014). After adapting to a face for a few seconds, the facial characteristics of the adapting face appear less apparent in subsequently viewed faces (Leopold, O'Toole, Vetter, & Blanz, 2001; Luo, Burns, & Xu, 2017; Rhodes & Jeffery, 2006; Webster, Kaping, Mizokami, & Duhamel, 2004; Webster & MacLeod, 2011; Xu, Dayan, Lipkin, & Qian, 2008; Ying & Xu, 2017); thus, adapting to an attractive face will lead to the subsequently viewed face as being less attractive; a powerful visual illusion known as an attractiveness adaptation aftereffect (Pegors, Mattar, Bryan, & Epstein, 2015; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Ying et al., 2019). The magnitudes of these adaptation aftereffects reflect the strength of different attributes present in the adapting face; i.e., an extremely attractive face will produce larger aftereffects than a face that is only moderately attractive (e.g., Webster et al., 2004; Ying et al., 2019). In our first experiment, we therefore compared the adaptation aftereffects produced by a group of RSVP faces, versus those elicited by their computer-generated, morph average: if they are indistinguishable from one another, then it would suggest that ensemble statistics is not a simple extraction of the group's gist (e.g., these faces are unattractive), but instead stems from a process that is consistent with morph averaging the fiducial points and surface aspects of the faces together. By contrast, if our gist averaging hypothesis is correct, the computer-generated morph average face should produce adaptation aftereffects that are distinct from the RSVP streams. This is because the computer-generated morph average face is invariably more attractive than the underlying group it is comprised of (DeBruine et al., 2007; Perrett et al., 1994; Valentine et al., 2004).

Experiment 1: Temporal ensemble statistics represent morph averaging

In our first experiment, we directly tested our morph versus gist averaging hypotheses by comparing the adaptation aftereffects produced by an RSVP stream of faces to the morphed average face derived from their group. If ensemble coding represents the morph average, then we should observe (a) similar and correlated aftereffects between the RSVP face stream and its computer-generated morph average, and (b) since this morph average will be more attractive than the individual faces in the group, the unattractive face stream may fail to generate aftereffects in the direction that we would expect from those typically induced by unattractive faces (e.g., the faces may produce no aftereffects, or even make faces presented after them seem less attractive). On the other hand, if the ensemble coding represents gist averaging, then the unattractive face stream should generate a significant aftereffect (e.g., faces presented after the stream should appear more attractive relative to no adaptation baseline) since the gist average of an unattractive face stream is still considered to be unattractive.

Experiment 1: Methods

Participants

Twenty-nine participants (14 Females; Mean Age: 22.03) with normal or corrected-to-normal vision were recruited from Nanyang Technological University. We aimed to recruit 30 participants; however, one dropped out during the experiment and was not replaced, thus leaving us with only 29 participants. We selected this sample size based upon previous face attractiveness adaptation work (*n* = 30 in Pegors et al, 2015). Written informed consent was provided by participants in all four experiments beforehand. This study was approved by the Institutional Review Board (IRB) at Nanyang Technological University, Singapore, in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving human participants.

Apparatus

Visual stimuli were presented on a 17-inch Philips CRT monitor (refresh rate 85 Hz, spatial resolution 1024 × 768 pixels; comparison between CRT and LCD monitor can be found in Zhang et al., 2018). The monitor was controlled by an iMac Intel Core i3 computer running

Matlab R2010a (Mathworks, MA, USA) via Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The experiment was conducted in a dimly lit room. During the experiment, participants rested their heads on a chin rest 75 cm in front of the monitor. Each pixel subtended 0.024° on the screen.

Visual Stimuli

Thirty-Five Chinese female faces were chosen from the N-FEE database (Yap, Chan & Christopoulos, 2016). Due to copyright restrictions we are not allowed to publicly publish these images, so we have used faces from the KDEF database for illustrative purposes (Lundqvist, Flykt, & Öhman, 1998). In this study, we only selected portrait pictures from 35 female Chinese Singaporeans with neutral expressions. All face images were grey scaled and masked so that only the facial region of each face was visible to the participants. The luminance of the face images was equalized via SHINE toolbox (Willenbockel et al., 2010). Every participant rated the attractiveness of the 35 faces at least two weeks before the main experiment in Experiment 1 (adapted from Rhodes & Jeffery, 2006; 1 for most unattractive and 7 for most attractive). Prior to rating, participants were exposed to all of the faces, each for 400ms in a randomized order, in order to gauge the range of attractiveness in the faces before rating each face. Each face was rated four times, with the mean rating for each face ranging between 2.67 and 5.00 (M = 3.53, SD = 1.31). Inter-rater reliability was high (Cronbach's alpha = .98). The adapting stimuli were selected from the four faces rated as most attractive and the four that were least attractive.

The test faces included one of the most attractive and one of the most unattractive faces from the originally rated 35 faces (excluding the adaptors), and a further five faces that were produced by morphing these two faces in equally incremental steps between them (thus giving us

7 attractiveness units ranging from the original unattractive face through to the original attractive face) via Webmorph (Debruine & Tiddeman, 2017). Therefore, there are seven test faces in total. To minimize low-level adaptation as per prior research (Burns et al., 2017; Rhodes et al., 2003; Ying & Xu, 2017; Zhao & Chubb, 2001), the adapting stimuli were displayed at $3.20^{\circ} \times 4.03^{\circ}$, which was roughly 133% of the size of the test stimuli. The adapting stimuli and the test stimuli were always presented at the same side of the central fixation cross within one trial, and their centers were roughly 3.8° away from the central fixation cross (159 pixels). Our reason for presenting the faces in the periphery was because adaptation aftereffects have been found to be greater in the visual periphery compared to the fovea (Bachy & Zaidi, 2014; Chen, Chen, Gao, Yang, & Yan, 2015; Ying & Xu, 2017). Similar to Haberman, Lee, and Whitney (2015), we are aware that the ‘attractiveness unit’ is arbitrary, and we do not mean that the (perceived) attractiveness differences between the testing faces are strictly linear. The ‘attractiveness unit’ merely represents the relative differences between these faces.

Procedure

Participants completed five blocks: baseline, RSVP unattractive, RSVP attractive, computer-generated average unattractive morph, and computer-generated average attractive morph. In the baseline condition, participants simply rated the test faces, which were presented for 400 ms, as attractive or unattractive. Each test face was presented 10 times at random giving a total of 70 trials in each block. The same test face sampling occurred in the attractive RSVP block, but this time participants viewed an RSVP stream of the four attractive adapting faces prior to viewing each test face. The temporal frequency of the RSVP sequence was 42.5 Hz, with each face displayed for 23.5 ms per face frame (with no interval between two face frames, the same as Ying & Xu, 2017). Thus, each adapting face was presented 40 times, in a random order, during the 3.764 s adaptation phase (23.5 ms × 4 faces × 40 repetitions). Figure 1 displays the trial sequence. This method was repeated for the unattractive RSVP block, except the RSVP stream comprised the unattractive adaptors. The same process occurred for the attractive morphed average block, except during adaptation when participants were simply presented with a single face that was created by morphing all of the four attractive adaptors' visual properties together. The same was true for the unattractive morphed average block, except the unattractive adaptors were used to create its adapting face morph. The blocks were presented in a random order, with instructions given beforehand. Participants were given breaks that were roughly equal in duration to an experimental block to disperse any carryover effects. Participants practiced for 5-10 trials before participating in each of the experiments reported here.

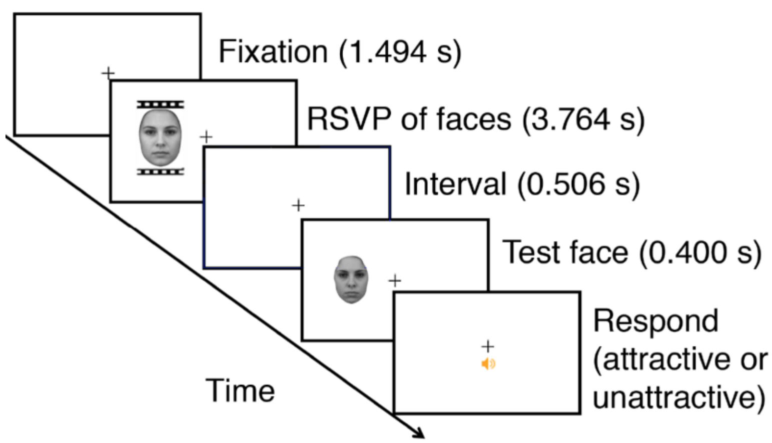


Figure 1. Example trial sequence from the RSVP adaptation condition (the demonstrated faces are AF01NES and AF34NES from the KDEF database). Participants fixated on the cross at all times. After 1.494 s, the RSVP of the faces appeared onscreen for 3.764 s. After a short interval (0.506 s), the test face appeared for 0.4 s. Then a beep sound prompted participants to judge the target face by pressing the 'A' button as attractive, or the 'S' button as unattractive.

In each trial, the test stimulus presented was one of the seven test faces selected at random. After that, a 50 ms beep sound prompted for participants to respond. Participants had to press the “A” or “S” key to express whether they found the test faces “attractive” or “unattractive” respectively. Such two-alternative forced choice (2-AFC) methods have been commonly used in adaptation experiments (e.g., Fox & Barton, 2007; Webster et al., 2004; Xu, Dayan, Lipkin, & Qian, 2008). After the participant responded in each trial, the trial would terminate, thus commencing the next trial. No feedback was given throughout. Within each block there were 70 trials, which comprised a presentation of each of the 7 test faces 10 times in a random sequence.

Analysis

Participants’ responses were sorted into proportions of ‘attractive’ responses to each test stimulus per adaptation condition. A psychometric curve was created with the x-axis indexing the test stimuli and the y-axis plotting the fractions of ‘attractive’ responses. Subsequently, the psychometric curves were fitted with a sigmoidal function $f(x) = 1/ [1 + e^{-a(x-b)}]$, where $a/4$ is the slope and b provides the test-stimulus parameter corresponding to 50% of the psychometric function, the point of subjective equality (PSE). We measured the adaptation aftereffects by comparing the difference between the PSEs of the adapting conditions and the baseline condition. Any subsidiary pairwise comparisons after the analysis of variance (ANOVA) were Bonferroni corrected. Note that goodness of fit was evaluated by coefficient of determination ($R^2 = 1$ indicates the perfect fit). The mean goodness of fit (R^2) for all experiments was > 0.89, indicating that the predicted lines fitted the observed data well.

To confirm that any non-significant results truly supported the null hypothesis, we used Bayes Factors to analyze the data (Dienes, 2014; Rouder, Speckman, Sun, Morey, & Iverson, 2009) in addition to the traditional Frequentist analyses. In brief, Bayes Factor utilizes the observed evidence for either the null or alternative hypothesis, with this weight of evidence realized as a ratio between the likelihoods of the hypotheses. For instance, '$BF_{01} = 3$' suggests that the observed data is 3 times more likely to fit the null-hypothesis compared to the alternative hypothesis. Generally, $BF_{01} > 3$ is suggested to provide evidence for the null hypothesis. All statistical analyses were conducted in JASP 0.8.6 (JASP team, 2018), R 3.4.3 (R Core Team, Vienna, Austria), Matlab R2017a (Mathworks, MA, USA) and SPSS Statistics 22 (IBM, NY, USA).

Experiment 1: Results and Discussion

The results from all the participants judging the facial attractiveness of the test faces under various conditions are shown in Figure 2A. We plotted the fraction of attractive responses as a function of the proportion of attractiveness of the test faces. The black (solid line with filled squares) psychometric curve is the baseline condition without adaptation. After adapting to the most attractive RSVP face stream, the participants judged the test faces as unattractive more frequently than baseline, and the psychometric curve (blue dashed line with open diamonds, RSVPa) shifted to the right. This is the standard facial-attractiveness aftereffect (Hsu & Young, 2004; Webster et al., 2004). The same finding occurred after adapting to the morphed average of this face stream (light blue solid line with filled diamonds, Statica). Curiously, after adapting to the most unattractive face stream (red dotted line with circles, RSPVu) or its morphed average (magenta dashed-dotted line with filled circles, Staticu), there were no adaptation aftereffects observed relative to baseline.

To determine the presence of adaptation aftereffects in our experiment, we performed paired *t*-tests between the baseline PSE and the PSEs of the adaptation conditions (Figure 2B).

As expected, both the attractive RSVP and morph average conditions produced significant aftereffects (both *ps* $< .001$), with participants reporting the test faces as unattractive more frequently in the two attractive conditions relative to the no adaptation baseline. Surprisingly, neither of the unattractive conditions produced any aftereffects (both *ps* $> .62$). Bayesian *t*-tests provided further support for the null hypothesis (RSVPu: $BF_{01} = 4.52$; Staticu: $BF_{01} = 5.06$): the unattractive conditions did not generate significant aftereffects relative to baseline. Participants did not seem to be processing either set of unattractive adaptors as unattractive. These findings contradict the outcome predicted by the gist averaging hypothesis, for if this hypothesis had been correct, then the unattractive RSVP group should have displayed aftereffects that shifted the psychometric curve in the opposite direction to those found in our attractive conditions (i.e., negative relative to baseline, where test faces were rated as attractive more frequently after adaptation).

To test whether temporal ensemble perception was indistinguishable from the computer-generated morph average, we performed a two-way repeated-measures ANOVA on the PSE shifts relative to baseline with factors of Attractiveness (attractive vs. unattractive) and Adaptor (RSVP vs. morph average). While there was a significant main effect of Attractiveness ($F(1, 28) = 49.55$, $p < .001$, $\eta_p^2 = .64$) due to the attractive conditions producing larger aftereffects than the unattractive conditions, there was no significant main effect of Adaptor ($F(1, 28) = 0.001$, $p = .99$, $\eta_p^2 < .001$) nor any interaction ($F(1, 28) = 0.46$, $p = .50$, $\eta_p^2 = .016$). Bayesian *t*-tests comparing the attractive RSVP condition versus the attractive morph average ($BF_{01} = 4.57$), and the unattractive RSVP versus the unattractive morph average ($BF_{01} = 4.50$), provided further evidence for the null hypothesis. This confirms that the RSVP streams were processed by our participants in a similar way to their morph averages. Further support for this came from the fact that the aftereffects from attractive (r = .65, p < .001; blue open diamonds with dashed line in Figure 2C) and unattractive (r = .43, p = .019; red full circles with solid line) RSVP streams were correlated with their computer-generated morphed average face counterparts.

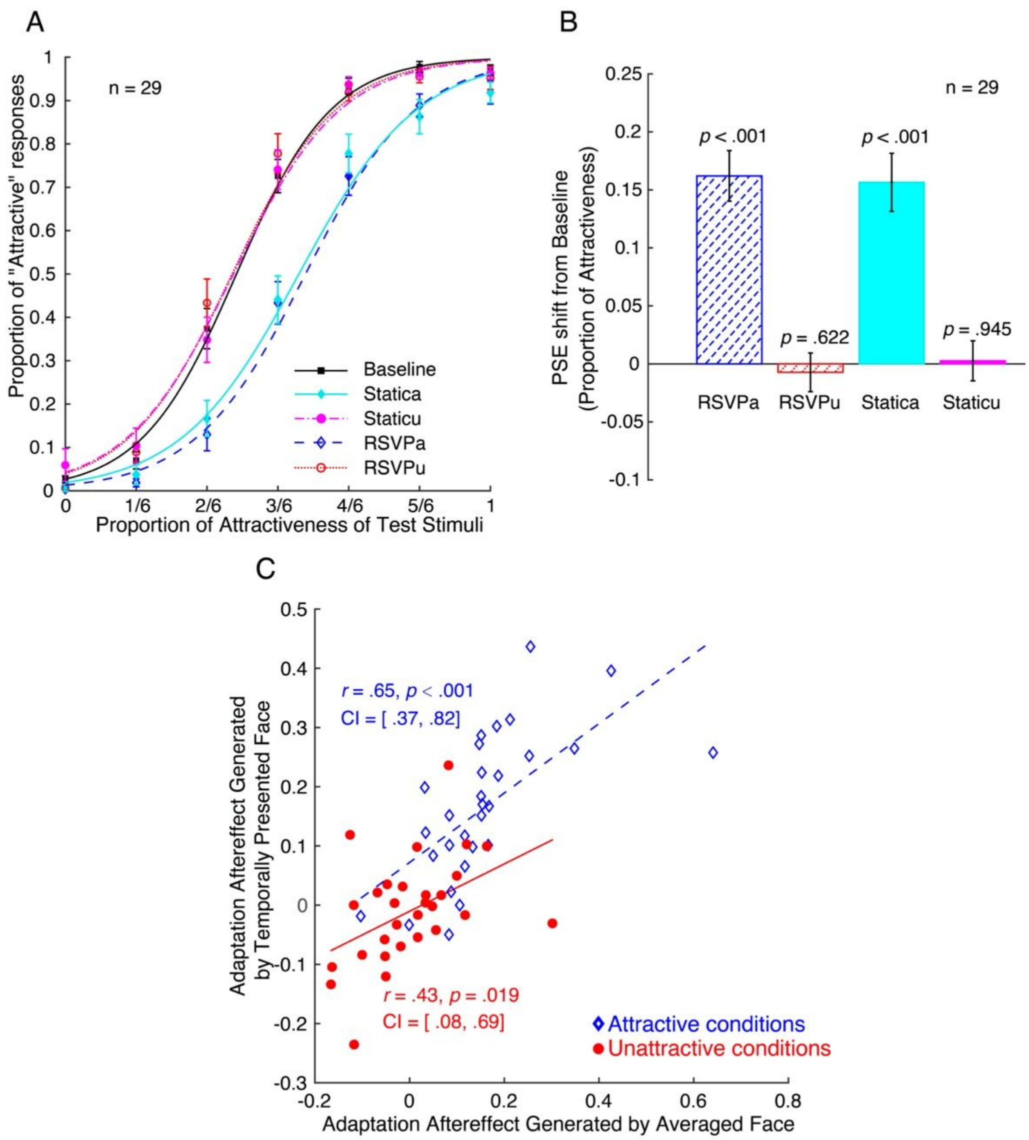


Figure 2. The RSVP and computer-generated morph average aftereffects (Experiment 1). (A) The psychometric functions of all participants averaged together. Error bars indicate the standard error of the mean. (B) Summary of all participants' results. For each condition, the adaptation aftereffect measured by PSE shift relative to baseline and the SEMs were plotted. The p-value shown for each condition in the figure was calculated using paired t-tests. Noticeably, a positive adaptation aftereffect measured by PSE shift indicates the target faces were perceived as less attractive than during baseline. The following figures adopt the same statistical analyses. (C) The relationship between the RSVP conditions and the paired morph average conditions. Each dot represents data from one participant: blue open diamond for the attractive conditions, and red filled circle for the unattractive conditions.

We found that adapting to an RSVP stream and its computer-generated morphed average face led to comparable, and correlated, facial attractiveness aftereffects. While these findings replicate prior work that showed similar effects for facial emotion (Ying & Xu, 2017), our results clarify what characteristics of a face are extracted in order to produce temporal ensemble perception. For example, the lack of differences between the morphed average faces and their RSVP groups suggest that morph, rather than gist, averaging occurs during temporal ensemble coding. If gist averaging had been occurring, then adapting to the unattractive face stream should have induced aftereffects where the viewer rated subsequently presented test faces as attractive more often than in the baseline. We did not observe this effect here with our unattractive RSVP group, instead, these faces produced no aftereffects, with aftereffects actually comparable to their morphed average counterpart. However, we do not think that this finding indicates that these faces were not processed at all during adaptation. We believe that the data simply fits with the hypothesis that the participants were morph averaging these faces together so that the group of unattractive faces were processed as more attractive (i.e., roughly equal to baseline levels) than what they were (i.e., unattractive). A similar lack of differences was found between the aftereffects produced by the attractive group and its morphed average face. To our knowledge, this is the first time that the morph averaging hypothesis of ensemble perception has been demonstrated as having empirical support over the gist hypothesis.

Experiment 2: Temporal ensemble coding is driven by the underlying mean attractiveness of the group

In Experiment 1, adapting to unattractive RSVP faces produced no significant adaptation aftereffects. We do not believe that this was due our participants not processing the unattractive faces. Instead, we posit that participants simply processed this face stream as more attractive than the gist attractiveness of the individual faces in the group (i.e., unattractive). If this is the case, then adding in a new mixed (‘MIX’) condition, comprised of attractive and unattractive faces, should induce aftereffects somewhere in between those observed for the attractive and unattractive conditions in Experiment 1. Moreover, the magnitudes of these aftereffects across all conditions should also be associated with the underlying mean attractiveness of the individual faces, thereby demonstrating that our visual system adapts to the RSVP of face streams in a linear fashion that is consistent with the principles of ensemble coding.

Experiment 2: Methods

Twenty new participants (10 Females; Mean Age: 22.84) participated in this experiment. We selected this sample size for two reasons: firstly, a power analysis based upon the effect size of Experiment 1 (η_p^2 = .65; using G*Power 3.1 software; Faul, Erdfelder, Buchner, & Lang, 2009), with α-value at .05, and power (1 – β) at .80 indicated that we needed at least 7 participants. However, considering the differences in experimental design, we chose to greatly expand this number to roughly triple that sample size.

We used the same adaptation procedure as in Experiment 1, except there were three adaptation conditions in addition to the baseline: RSVP of attractive faces (‘ATT’, four attractive faces), RSVP of mixed faces ('MIX', four attractive faces and four unattractive faces), and RSVP of unattractive faces ('UNA', four unattractive faces) at a reduced adaptation duration (1.88 s in Exp 2 vs. 3.764 s in Exp1). Note that in the 'MIX' condition the adapting RSVP streams were presented for the same duration as the 'ATT' and 'UNA' conditions (see Experiment 1, Methods section). Thus, in the 'MIX' condition, each adapting face was only presented 10 times during the adaptation phase, so that the adapting duration is equated across different conditions. Also, each test face in each block appeared 12 times in a random order. Additionally, after the main experiment, we asked the participants to rate the mean attractiveness of the RSVP sequences on a 7-point scale (1 for most unattractive and 7 for most attractive), with each stream presented 10 times. These RSVP sequences were randomly presented for the same duration (42.5 Hz; 80 frames × 23.5 ms; in total 1.88 s) as that during the adapting stage in the main experiment.

Since our data consisted of repeated measures from three observations (i.e., an observation from each of the unattractive, mixed, and attractive conditions) for each participant, we used the repeated measures correlation analysis (Bakdash & Marusich, 2017) to quantify the strength of the relationship between the attractiveness ratings of the faces and the adaptation aftereffects produced by those faces. It uses the analysis of covariance (ANCOVA) to 'statistically adjust for inter-participant variability', thus 'estimates the common regression slope' (generating the same slope), in other words, the association shared among individuals.

Experiment 2: Results and Discussion

The mean adaptation results from all participants are shown in Figure 6A. Similar to Experiment 1, the RSVP of the Attractive condition generated a significant rightward shift of the psychometric curve, while the RSVP of the Unattractive condition failed to produce a shift. Interestingly, the RSVP of the Mixed condition generated a smaller yet substantial rightward shift. Relative to baseline, significant aftereffects were generated by the RSVPs of attractive (Figure 3A, M = .22, SEM = .004; $t(19)$ = 5.85, p < .001) and mixed (M = .12, SEM = .002; $t(19)$ = 5.06, p < .001) but not the unattractive (M = .01, SEM = .02; $t(19)$ = .47, p = .64) faces. Bayesian analyses suggested that the lack of aftereffects in the unattractive condition was in favor of the null hypothesis (BF_{01} = 3.89); i.e., no adaptation aftereffect relative to baseline. Participants therefore rated the test faces as less attractive after adapting to the attractive and mixed RSVP streams (Figure 3C). Moreover, we replicated Experiment 1 in showing no aftereffects in the unattractive group, suggesting participants were not processing the RSVP stream as unattractive. An ANOVA yielded significant differences among all three adaptation conditions (with Greenhouse-Geisser correction, $F(1.55, 29.36)$ = 33.22, p < .001, η_p^2 = .64). Subsidiary Bonferroni corrected comparisons showed significant differences between the attractive and unattractive ($t(19)$ = 6.73, p < .001), attractive and mixed ($t(19)$ = 3.88, p = . 003), and mixed and unattractive ($t(19)$ = 5.86, p < .001) conditions.

As the ‘Mixed’ condition contains the adapting stimuli from the ‘Attractive’ and the ‘Unattractive’ conditions, it should in theory yield an aftereffect which is roughly equal to the mean of those of two conditions. We therefore compared the adaptation aftereffects of the ‘MIX’ condition with the average of the aftereffects from those two conditions. The paired samples t-test suggested that there was no significant difference between this pair ($t(19)$ = .28, p = . 78, BF_{01} = 4.15). Therefore, the ‘Mixed’ condition closely resembles the midpoint of the ‘Attractive’ and ‘Unattractive’ conditions. This indicates that the participants perceived the attractiveness of the adapting stream in a graded fashion consistent with ensemble coding.

An ANOVA on the participants' attractiveness ratings of the RSVP streams showed they were also significantly different from one another ($F(2, 38) = 112.55$, $p < .001$, $\eta_p^2 = .86$). Further comparisons indicated that participants judged the RSVP of the attractive faces (M = 4.89, SEM = .013) as the most attractive, followed by the RSVP of mixed faces (M = 3.98, SEM = .016), and the RSVP of unattractive faces (M = 2.65, SEM = .017) were judged as least attractive (all $ps < .001$). Further repeated measures correlation analyses (Bakdash & Marusich, 2017) revealed a significant positive correlation between the attractiveness ratings of the RSVP streams and the adaptation aftereffects (Figure 3D, $r = .71$, $p < .001$, 95% CI [0.50, 0.84]); indicating that the brain performs temporal ensemble statistics in a linear fashion from the underlying attractiveness of the stream.

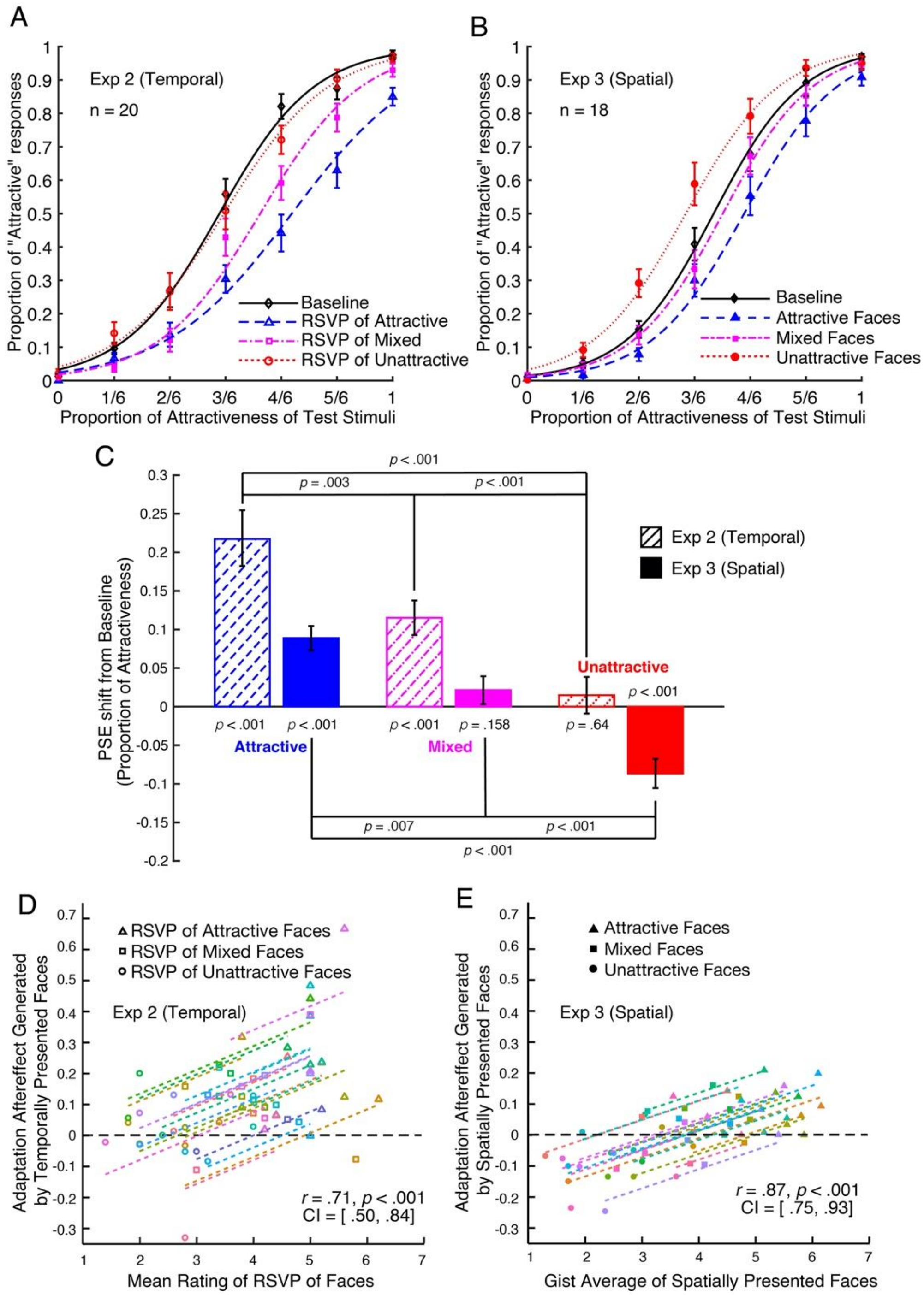


Figure 3. Adaptation aftereffects to temporally presented RSVPs (Experiment 2) and spatially presented faces (Experiment 3). (A) The psychometric functions of Experiment 2's participants averaged together. 'Error bars indicate the standard error of the mean. (B) The psychometric functions of Experiment 3's participants averaged together. (C) Combined summary of all participants' results from Experiment 2 and Experiment 3. The hatched bars indicate Temporal Presentation RSVP conditions (Experiment 2), and the solid bars represent Spatial Presentation conditions (Experiment 3). (D) The adaptation aftereffect as a function of the attractiveness rating of the RSVP of faces in Experiment 2. (E) The adaptation aftereffect as a function of the mean attractiveness rating of the adapting

faces in Experiment 3. In both (D) & (E), each color represents the data from one individual participant. The horizontal dashed black auxiliary line indicates no adaptation aftereffect.

In Experiment 2, we replicated the results from Experiment 1, but further illustrated the linear fashion in which the brain morph averages the attractiveness of a temporal stream of attractive, unattractive and mixed faces. These results therefore lend further support to our morph averaging hypothesis for temporally presented face groups. Interestingly, although the ensemble representation of the unattractive face RSVP stream was not processed as unattractive, as reflected by the lack of aftereffects, the direct ratings of these unattractive RSVPs did appear to be perceived as unattractive to some extent (*M* = 2.65 out of a 1-to-7 scale, see the above Results section for more details). Previous work has shown that adaptation aftereffects can yield insights into perceptual operations even in the absence of differences in direct ratings (Liu et al., 2014). Thus, adaptation and direct rating may reflect two distinct visual processes: perceptual vs. cognitive process.

Experiment 3: Spatial ensemble statistics represent the gist

Across Experiments 1 and 2 we have shown temporal ensemble perception extracts the morph average. However, is this also true for spatial ensemble coding when a group of faces is presented simultaneously? We previously showed that the adaptation aftereffects produced by spatially presented faces (i.e., a group presented onscreen at the same time) generated aftereffects in the direction that we would expect if the gist averaging hypothesis was true (Ying et al., 2019); i.e., the unattractive faces made subsequently presented faces appear more attractive, and adapting to a mix of unattractive and attractive faces produced no aftereffects relative to the baseline no adaptation condition. This result is at odds with the morph averaging that we have observed from our RSVP paradigms in Experiments 1 and 2. We therefore wanted to replicate this gist averaging in a spatial adaptation paradigm by using the same adapting faces from Experiment 2. By using identical adapting faces, we could directly compare the aftereffects derived from temporal and spatial ensemble coding. If the aftereffects between Experiment 2 and 3 are indistinguishable, then it would imply that a similar mechanism is at work both temporally and spatially; i.e., the faces are being morph averaged from their fiducial points and surface characteristics. However, if the aftereffects between the two experiments are different, then it would provide the first evidence that temporal and spatial ensemble statistics may reflect qualitatively distinct calculations. For example, if gist averaging occurs during spatial ensemble coding, then we would expect an overall negative shift for all of the adapting face conditions relative to those effects observed in Experiment 2: e.g., the unattractive group will now elicit negative aftereffects, the mixed group will be no different from baseline, and the attractive group will elicit smaller positive aftereffects than the attractive group in Experiment 2. We test these hypotheses in Experiment 3.

Experiment 3: Methods

Eighteen new participants (11 Females; Mean Age: 22.78) participated in this experiment; we had initially aimed for 20, but two dropped out during the experiment. Here we used the same adapting faces and blocks from Experiment 2, except the mixed condition only contained two attractive and two unattractive faces so that there were only four faces in the adapting group. During adaptation, the four adapting faces were presented around the central fixation cross (Figure 4), with the test face presented at the center of the screen. The center-center difference between each adaptor and the central fixation cross is around 3° (124.5 pixels). This spatial layout is similar to our recent study on ensemble coding of facial attractiveness (Ying et al., 2019). The trial sequence was otherwise similar to Experiments 1 and 2. After the experiment we asked the participants to rate the attractiveness of the eight individual adapting faces to compute an average from the ratings, thereby reflecting the gist average.

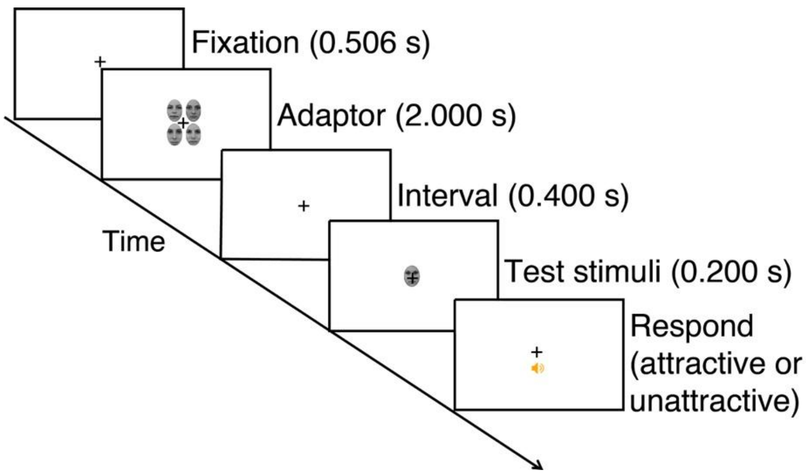


Figure 4. Example trial sequence from a spatial adaptation condition (the demonstrated faces are AF01NES, AF05NES, AF06NES, AF07NES and AF34NES from KDEF database). Participants fixated on the cross at all times. After 0.506 s, four adapting faces simultaneously appeared for 2 s. After a 0.4 s interval, the test face appeared on the screen for 0.2 s. Then a beep sound indicated participants should judge the attractiveness of the target face by pressing the ‘A’ button for attractive, or the ‘S’ button for unattractive. Experimental parameters for all conditions and experiments are detailed in the Methods section.

Experiment 3: Results and Discussion

The mean adaptation results from all participants are shown in the psychometric curves in Figure 3B. Unlike Experiments 1 and 2, the Unattractive condition generated a leftward shift away from baseline; this direction is what we would expect if our participants were adapting to the unattractive group as though they were unattractive (Ying et al., 2019). Such differences relative to Experiment 2 were also observed for the Mixed condition, which failed to generate any significant aftereffects. We statistically examined what aftereffects our spatial conditions produced relative to the baseline condition. Significant aftereffects were generated by both the attractive (Figure 3C, *M* = .093, *SEM* = .016; $t(17) = 5.91$, $p < .001$) and unattractive (*M* = -.083, *SEM* = .020; $t(17) = -4.20$, $p = .001$) groups. Test faces were rated as unattractive following adaptation to the attractive group, and conversely rated as attractive more frequently following the unattractive groups adaptation, all relative to baseline. By contrast, the mixed faces evoked no aftereffects (*M* = .028, *SEM* = .019; $t(17) = 1.48$, $p = .16$).

An ANOVA on the three adaptation conditions was significant ($F(2, 34) = 50.42$, $p < .001$, $\eta_p^2 = .75$). Bonferroni corrected comparisons showed that the attractive and unattractive ($t(17) = 8.69$, $p < .001$), attractive and mixed ($t(17) = 3.56$, $p = .007$), and mixed and unattractive ($t(17) = 7.93$, $p < .001$) conditions were all significantly different from one another. As in the case of Experiment 2, there was a significant positive repeated measures correlation ($r = .87$, $p < .001$, 95% CI [0.75, 0.93]; Figure 3E) between the mean attractiveness ratings of the groups of adapting faces and their aftereffects.

A side by side comparison between Experiment 2 and 3 (Figure 3C), shows qualitative differences between the aftereffects of our RSVP experiments and the spatial aftereffects here; note that these differences are apparent despite us using the same adapting faces between the experiments. To confirm these differences statistically, a mixed model ANOVA on the adaptation aftereffects was performed, with a between subject factor of Group (Experiment 2: Temporal vs. Experiment 3: Spatial) and a within subject factor of Attractiveness (unattractive vs. mixed vs. attractive). We found a significant main effect of Attractiveness (with Greenhouse-Geisser correction, $F(1.60, 57.46) = 73.30$, $p < .001$, $\eta_p^2 = .67$) due to differences between the adaptation aftereffects (i.e., attractive > mixed > unattractive, Figure 3A, all *ps* $< .001$). Similarly, there was also a significant main effect of Group ($F(1,36) = 12.19$, $p = .001$, $\eta_p^2 = .25$) due to the Experiment 2 Temporal group exhibiting more positive aftereffects in contrast to our current Spatial group (Exp 2 $M = .12$ vs. Exp 3 $M = .012$). Finally, the Group × Attractiveness was not significant (with Greenhouse-Geisser correction, $F(1.59, 57.36) = .80$, $p = .45$, $\eta_p^2 = .02$). These findings therefore indicate that while our participants were producing aftereffects that were comparably distinct between attractiveness conditions, the actual perceptual outcomes as reflected by adaptation aftereffects, appeared qualitatively different between Experiments 2 and 3.

To test whether the above differences in adaptation were also present in the direct ratings, we performed a mixed model ANOVA on the mean attractiveness ratings of the adapting faces with a between subjects factor of Group (Temporal Experiment 2 vs. Spatial Experiment 3) and a within subjects factor of Attractiveness (unattractive vs. mixed vs. attractive). There was a significant main effect of Attractiveness ($F(2, 72) = 302.74$, $p < .001$, $\eta_p^2 = .89$) due to the faces being rated significantly different from one another (i.e., attractive > mixed > unattractive, all *ps* $< .001$), but no main effect of Group ($F(1, 36) = .025$, $p = .88$, $\eta_p^2 = .001$; Bayesian analyses provided further support for the null hypothesis; $BF_{01} = 4.08$). There was, however, a significant interaction between the effects of Attractiveness and Group ($F(2, 72) = 3.64$, $p = .031$, $\eta_p^2 = .092$). Despite this interaction, there were no significant between group differences in the mean attractiveness ratings of the adapting faces for each of the attractiveness blocks (attractive $p = .11$, mixed $p = .82$, unattractive $p = .43$). Thus, presenting the adapting faces spatially or temporally (RSVP) did not change participants' ratings of the adapting faces' attractiveness. These results suggest that the qualitative differences in adaptation aftereffects derived from temporal and spatial ensemble coding are not due to differences in the perceptions of the adapting faces' attractiveness.

While there were some minor differences between the adaptation durations in Experiments 2 & 3, we do not believe that these differences affect our interpretations of the data. Research into the time course of face adaptation has revealed (e.g., facial identity: Rhodes, Jeffery, Clifford, & Leopold, 2007; facial expression: Burton, Jeffery, Bonner, & Rhodes, 2016) that adaptation aftereffects follow the classic time course pattern of 'logarithmic build-up' and 'exponential decay'. This means that the adaptation aftereffect can be altered *quantitatively* by some changes in time (like the adaptation duration), but not *qualitatively*. We recently found that facial expression adaptation aftereffect can be generated after as brief as 34 ms of adaptation (Sou & Xu, 2019). Thus, the qualitative differences in aftereffects from temporal and spatial ensemble coding here are likely to be maintained, even if the adaptation duration was matched across conditions. To confirm this fact though, we ran a new experiment.

Experiment 4: Spatial-Temporal ensemble statistics induce morph averaging

While the attractive and unattractive temporal face streams generated asymmetrical aftereffects in Experiment 2 (i.e., the attractive group generated aftereffects, but the unattractive faces did not), the spatial face groups generated symmetrical aftereffects in Experiment 3 (Figure 3C, attractive group generated aftereffects, as too did the unattractive group). While there are other minor differences between the procedures across Experiments 2 and 3, such as the locations of the RSVP versus the static spatial adaptor locations, we do not believe these are causing the qualitative differences we observe between temporal and spatial ensemble coding. Instead, we believe that these effects reflect the fact that temporal and spatial ensemble coding computations are distinctly different. However, to be certain of this belief, we decided to run Experiment 3 again, except this time, we added an RSVP manipulation to the adapting faces. This meant that we could directly compare ‘pure’ spatial ensemble coding (i.e., that derived from the static groups of faces in Experiment 3) versus temporal ensemble coding (i.e., that derived from the RSVP of faces presented at the same four locations as the static spatial groups).

Furthermore, we had participants directly rate the mean attractiveness of the groups of adapting faces in both the spatial and temporal conditions so that we could assess whether the direct rating and the adaptation measures of ensemble coding were similar across presentation methods.

Experiment 4: Methods

Twenty new participants (13 Females; Mean Age: 21.75) participated in this experiment. We matched the sample size of the current experiment with the previous two experiments. The general design was adapted from Experiments 2 and 3. The trial sequence was similar to that of Experiment 3. During adaptation, there were four RSVP face streams simultaneously presented surrounding the central fixation cross (Figure 5). The spatial locations of the four streams were identical to those in Experiment 3 (3° away from the fixation cross). Thus, we name this manipulation the Spatial-Temporal condition. Within each RSVP stream, the faces were presented at 42.5 Hz (the same as Experiments 1 and 2) for 1.98 s (84 faces in total, and each presented for 23.5 ms; the adaptation duration is almost identical to Experiment 3: 2 s). All of the faces presented within the Spatial-Temporal streams were the faces used in Experiment 3, with ‘ATT’, ‘MIX’, and ‘UNA’ conditions. These faces were presented in a pseudo-random order, so that within each frame, the four faces presented onscreen together were always of different identities.

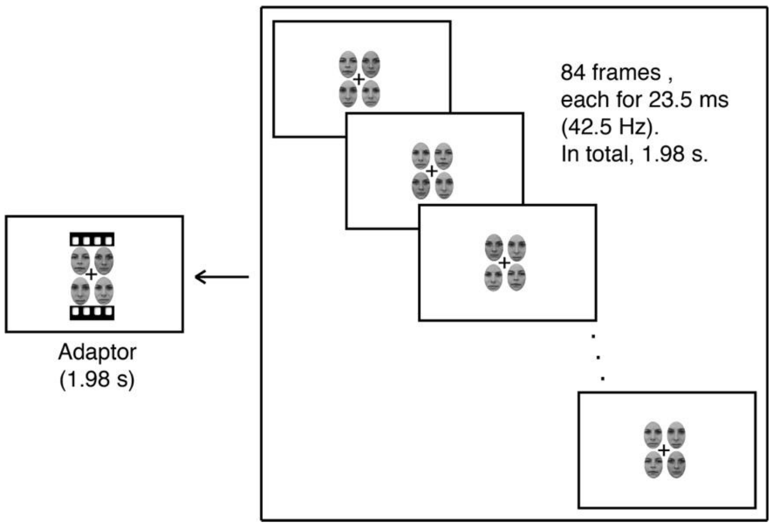


Figure 5. The Spatial-Temporal adaptor for Experiment 4 (the demonstrated faces are AF01NES, AF05NES, AF06NES, AF07NES and AF34NES from KDEF database). The adaptor is four simultaneous streams of RSVPs of faces (42.5 Hz, the same as Experiment 2), presented for 1.98 s in total. The spatial relationships of the four streams (3° away from the central fixation cross) were the same as that in Experiment 3. Thus, the Spatial-Temporal adaptor is a combination of the adaptation manipulations from Experiments 2 and 3.

In addition to our adaptation paradigm, we also measured ensemble perception of facial attractiveness via direct ratings. We asked our participants to rate the attractiveness of each adapting face, and these adapting faces as a group in the spatial-temporal configuration on a 7-point scale. Each group of faces was presented for 1 s. We chose 1 s for direct rating because it has been shown that this is sufficiently long for the participants to make judgments on attractiveness (e.g., Ying et al., 2019). Moreover, to clarify whether the computer-generated averaged face is indeed more attractive than the mean of its components, we also asked participants to rate the computer-generated averaged face of the attractive and unattractive groups. The order of the stimuli in direct rating tasks was randomized for each participant.

Experiment 4: Results and Discussion

The mean adaptation results from all participants are summarized in Figure 6A. After exposure to the attractive Spatial-Temporal faces (blue dotted line), there was a rightward shift in the psychometric curve relative to baseline, indicating that the ensemble representation of this group is attractive. A similar shift, albeit smaller in magnitude, is observed in the 'MIX' condition (magenta dash-dotted line). By contrast, the 'UNA' condition (red dotted line) failed to generate a significant shift from the baseline condition. This finding replicates our temporal ensemble coding results in Experiment 2, and appears qualitatively different from the aftereffects induced via spatial adaptation in Experiment 3.

Overall, significant aftereffects were generated by the Spatial-Temporal attractive (Figure 6B, $M = .16$, $SEM = .029$; $t(19) = 5.57$, $p < .001$) and mixed ($M = .068$, $SEM = .022$; $t(19) = 2.97$, $p = .008$) but not the unattractive ($M = -.013$, $SEM = .02$; $t(19) = -.85$, $p = .41$) faces. Bayesian analyses suggested that the lack of aftereffects in the unattractive Spatial-Temporal condition was in favor of the null hypothesis ($BF_{01} = 3.122$). Thus, the observed data indicates that there was indeed no adaptation aftereffect in the unattractive Spatial-Temporal condition. To compare the three adaptation conditions, we conducted an ANOVA and found significant differences among all three adaptation conditions (with Greenhouse-Geisser correction, $F(1.47,27.98) = 26.36$, $p < .001$, $\eta_p^2 = .58$). Subsidiary Bonferroni corrected comparisons showed significant differences between the attractive and unattractive ($t(19) = 5.85$, $p < .001$), attractive and mixed

($t(19) = 3.92$, $p = .003$), and mixed and unattractive ($t(19) = 4.82$, $p < .001$) conditions. These findings confirm our hypothesis that temporal ensemble coding induces morph averaging, whereas ensemble coding for spatially presented face groups (i.e., Experiment 3) results in gist averaging.

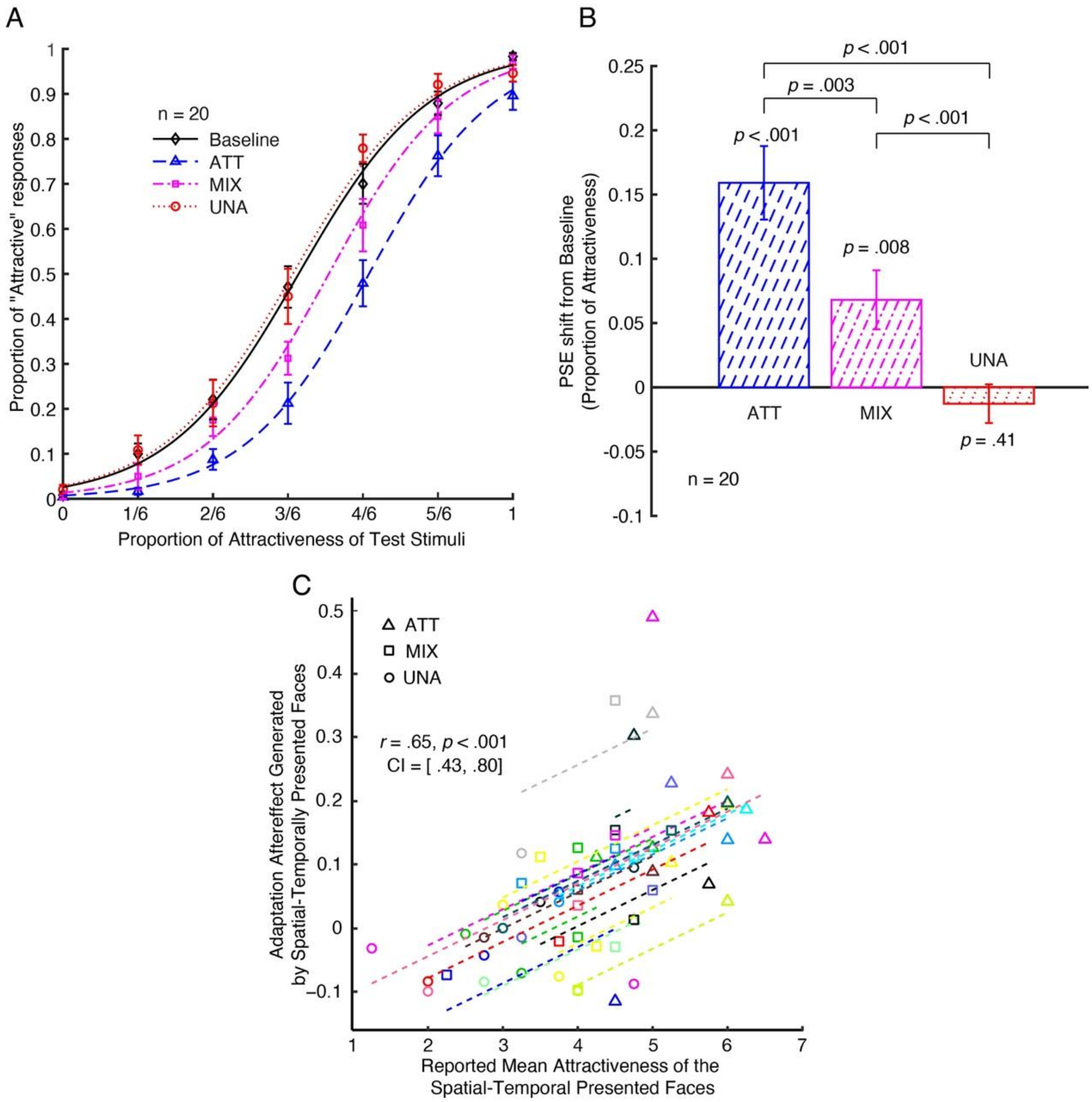


Figure 6. Spatial-Temporal adaptation aftereffects (Experiment 4). (A) The psychometric functions of all participants averaged together. Error bar indicates the SEM. (B) Summary of all 20 participants' results from Experiment 4. (C) The adaptation aftereffect as a function of the reported mean attractiveness of the adapting faces in Experiment 4. Each color represents the data from one individual participant.

To directly test whether the RSVP spatial manipulation we employed here was similar to the effects observed from the RSVP streams in Experiment 2, we ran an ANOVA on the aftereffects of Experiments 2 and 4, with Group (Exp 2, Exp 4) being the between subject factor, and Attractiveness (ATT, MIX, UNA) being the within subject factor. The results showed that there were no significant differences between these two experiments ($F(1, 38) = 2.20$, $p = .15$, $\eta_p^2 = .055$), nor any interaction between them and the attractiveness of the faces (with Greenhouse-Geisser correction, $F(1.51, 57.47) = .65$, $p = .49$, $\eta_p^2 = .017$). Instead, there was only a significant difference among the three attractiveness conditions (with Greenhouse-Geisser correction, $F(1.51, 57.47) = 59.51$, $p < .001$, $\eta_p^2 = .61$). Thus, the aftereffects induced by a single RSVP stream (Experiment 2) and multiple RSVP streams (Experiment 4) were comparable, and reflective of morph averaging.

To confirm that temporal and spatial ensemble coding reflect distinct perceptual outcomes, we compared the aftereffects between Experiments 3 and 4 using the same ANOVA. While we did not find any significant interaction between the experiments and the attractiveness of the faces (with Greenhouse-Geisser correction, $F(1.56, 55.97) = .58$, $p = .52$, $\eta_p^2 = .016$), there was a significant difference among three attractiveness conditions (with Greenhouse-Geisser correction, $F(1.56, 55.97) = 66.85$, $p < .001$, $\eta_p^2 = .65$). However, in addition, there was also a significant difference between the two experiments ($F(1, 36) = 6.07$, $p = .019$, $\eta_p^2 = .14$); the Spatial-Temporal aftereffects in Experiment 4 were more positive than those induced by the spatial group in Experiment 3. Taken together, the pattern of observed aftereffects in Spatial-Temporal adaptation is more similar to temporal ensemble coding, than to the static spatial ensemble coding we observed in Experiment 3. In other words, the Spatial-Temporal ensemble is largely driven by morph averaging of the faces from the temporal streams.

To examine the attractiveness ratings of the adapting faces directly, we conducted an ANOVA on the participants' attractiveness ratings of the Spatial-Temporal streams and found a significant difference among the three types of attractiveness adaptors ($F(2, 38) = 47.72$, $p < .001$, $\eta_p^2 = .72$). Further comparisons revealed that participants rated the Spatial-Temporal streams of the attractive faces ($M = 5.38$, $SEM = .015$) as the most attractive, followed by the Spatial-Temporal streams of mixed faces ($M = 4.16$, $SEM = .015$), with the Spatial-Temporal streams of unattractive faces ($M = 3.20$, $SEM = .020$) being rated as least attractive (all *ps* < .001). We further compared the direct ratings between Experiments 3 and 4 with a mixed-model ANOVA. There was a significant difference among the three attractiveness conditions, as expected (with Greenhouse-Geisser correction, $F(1.43,51.30) = 169.33$, $p < .001$, $\eta_p^2 = .83$). Importantly, there was also a significant difference between the two experiments ($F(1,36) = 4.42$, $p = .043$, $\eta_p^2 = .11$); the spatial-temporal streams (Exp 4) were rated as more attractive than the 'spatial group' (Exp 3). Thus, both rating and adaptation aftereffects data suggest that spatial (Exp 3) and spatial-temporal (Exp 4) ensemble coding are distinct from each other. There was no significant interaction between the experiments and the attractiveness of the faces (with Greenhouse-Geisser correction, $F(1.43,51.30) = 2.74$, $p = .09$, $\eta_p^2 = .071$).

Why were there similar ratings between Experiments 2 and 3, but different ratings between Experiments 3 and 4? We believe the reason was in the tasks in rating. In Experiment 2, 'mean attractiveness' was measured by participants rating the mean attractiveness of each RSVP stream; while in Experiment 3, the 'gist/mean attractive' was measured by the mean rating of individual adapting faces by another group of participants. By contrast, in Experiment 4, 'mean attractiveness' was measured by participants rating the mean attractiveness of Spatial-Temporal streams. Due to these differences, future experiments on the comparisons on ratings in the same task should be conducted.

Further repeated measures correlation analyses (Bakdash & Marusich, 2017) revealed a significant positive correlation between the attractiveness ratings of the Spatial-Temporal streams and the adaptation aftereffects (Figure 6C, r = .65, p < .001, 95% CI [0.42, 0.80]). This indicates that the observed attractiveness aftereffects were driven by the ensemble coding of the attractiveness of the adapting stimuli.

To test whether the computer-generated averaged face was more attractive than its components, we compared the mean ratings of individual attractive (M = 4.35, SEM = .18) and unattractive (M = 1.92, SEM = .19) faces with their computer-generated average faces (attractive: M = 5.65, SEM = .19; unattractive: M = 2.59, SEM = .17) conditions. We found that in both the attractive (t(19) = 7.46, p < . 001) and unattractive (t(19) = 5.58, p < . 001) conditions, the computer-generated average faces were more attractive than their components.

General Discussion

We investigated the perceptual calculations performed during ensemble statistics across four experiments. Experiment 1 showed that RSVP streams and their paired computer-generated morphed averages led to comparable, and correlated, facial attractiveness aftereffects. Experiment 2 replicated the findings from Experiment 1, thereby further supporting the morph average hypothesis; i.e., no aftereffects in the unattractive condition, such that the unattractive group was perceived as more attractive than the gist of the group (i.e., these faces are unattractive), and positive aftereffects in the mixed condition. Moreover, in Experiment 2 we found that aftereffects increased as a function of the underlying RSVP stream's attractiveness, suggesting that temporal ensemble perception occurs in a linear fashion. In contrast to the first two experiments, however, Experiment 3 showed that spatial ensemble statistics favored the gist averaging hypothesis; i.e., no aftereffects in the mixed condition, and negative aftereffects in the unattractive condition. Combining the manipulations in Experiments 2 (temporal) and 3 (spatial) together, Experiment 4 showed that ensemble coding of a Spatial-Temporal presentation of faces is formed by morph averaging, and not the gist. This confirms that the observed differences between Experiments 2 and 3 were not driven by the minor differences in presentation formats, but by distinct ensemble coding operations. Taking all four experiments together, it is clear that temporal and spatial ensemble statistics stem from qualitatively different extraction processes.

While a number of prior studies have examined spatial ensemble coding and temporal ensemble coding (Haberman et al., 2015; Haberman & Whitney, 2007, 2009; Whitney & Levi, 2011; Whitney & Yamanashi Leib, 2017; Wolfe et al., 2015; Ying & Xu, 2017; Ying et al., 2019), no study to our knowledge has compared the effects of both. Moreover, even if researchers had compared the averaging of facial traits other than attractiveness (e.g., emotion) across these two presentation formats, it would have been highly unlikely that they would have observed differences between temporal and spatial ensemble coding anyway. This is because adapting to facial emotion, via either a morph or gist averaging process, would result in the same outcome (as illustrated in Figure 7A with hypothetical data). Here, we took advantage of the fact that averaging faces together from their morphed properties makes them more attractive (DeBruine et al., 2007; Leder, Goller, Forster, Schlageter, & Paul, 2017; Perrett et al., 1994; Valentine et al., 2004; as illustrated in Figure 7B with hypothetical data). By doing so, we confirmed that there are qualitative differences between how ensemble coding mechanisms extract distinct information across spatial and temporal presentations.

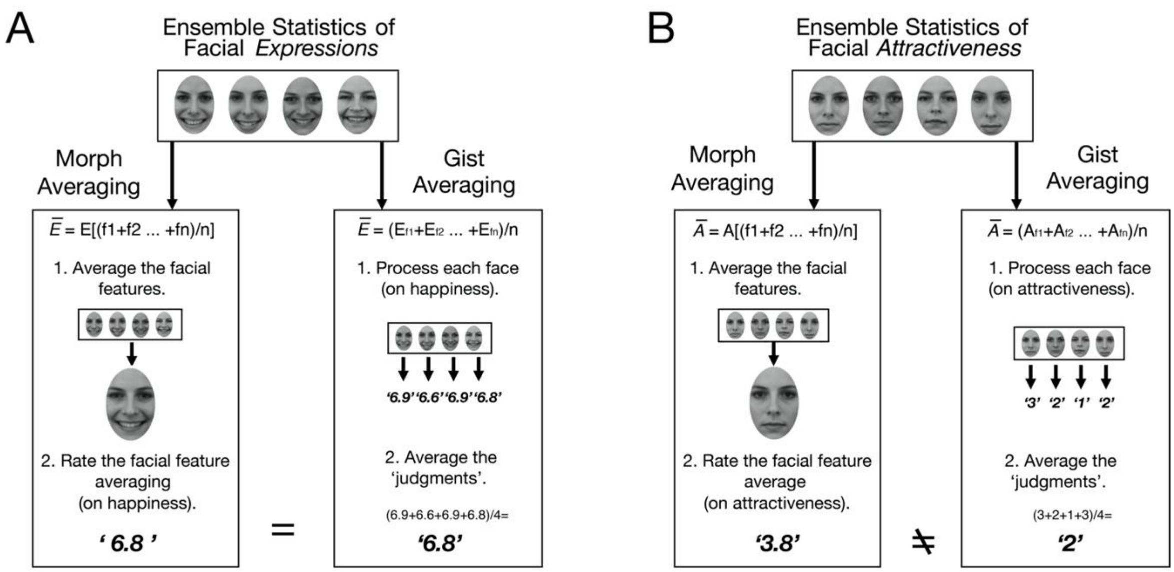


Figure 7. The 'morph averaging' and 'gist averaging' hypotheses (the demonstrated faces are AF01NES, AF05NES, AF06NES and AF07NES from KDEF database; the digits are from hypothetical data only for demonstration purposes). (A) Ensemble coding for facial expressions: the 'morph averaging' and 'gist averaging' hypotheses predict the same perceptual outcome for emotion; i.e., happy intensity rating of 6.8. (B) Ensemble coding of facial attractiveness: the 'morph averaged' face is more attractive than the mean attractiveness of its individual component faces, with the averaged face not equal to the mean 'judgments' (i.e., attractiveness rating of 3.8 versus 2).

We should explicitly clarify to readers that the null results found in Experiments 1, 2 and 4 (i.e., in the unattractiveness conditions) actually support our morph averaging hypothesis of temporal ensemble coding. These findings were not due to the unattractive faces not being unattractive enough to elicit negative aftereffects, nor are the lack of effects due to a lack of power. First, we used the very same stimuli in Experiments 2, 3, and 4, with the presentation methods being the largely the only difference among the 3 studies. The negative aftereffects generated by the unattractive condition in Experiment 3 shows that the unattractive group was processed by the participants as unattractive; i.e., the participants perceived the subsequently presented faces as attractive (which replicates the results from Ying et al., 2019). In other words, the very same faces generated asymmetrical aftereffects when being presented *temporally*, but generated symmetrical aftereffects when presented *spatially*.

This finding is at odds with the suggestion that the unattractive adapting faces in Experiments 1, 2, and 4 were simply insufficient in unattractiveness to elicit the aftereffects expected from an unattractive group. This point is further strengthened by the large effect size in the unattractive condition's aftereffects in Experiment 3. Moreover, by analyzing the data via Bayes Factors (Dienes, 2014; Rouder et al., 2009), we found evidence supporting the null hypothesis (i.e., the unattractive RSVP faces are equivalent to baseline and their computer-generated average face), thus countering any suggestion that the null effects across Experiments 1, 2, and 4 were a result of low statistical power. Simply put, the current data strongly favors the notion that the RSVP streams of unattractive faces are perceived as neither attractive nor unattractive relative to participants' baseline norms of attractiveness, and that this perceptual outcome was not due to these faces not being unattractive enough to elicit negative aftereffects. Instead, participants must have been averaging the unattractive RSVP stream in such a fashion that it made the faces be processed as more attractive than their underlying gist (i.e., unattractive). This was clarified by the fact that the aftereffects of the RSVP streams were equivalent to, and correlated with, their computer-generated averaged morph face counterparts.

The qualitative differences between the adaptation aftereffects produced by RSVP streams and spatial presentations of faces likely reveal the hierarchical nature of the human face perception system (Bartolomeo, Vuilleumier, & Behrmann, 2015; Behrmann & Plaut, 2013; Duchaine & Yovel, 2015; Eimer, 2000; Gobbini & Haxby, 2007; Haxby, Hoffman, & Gobbini, 2000, 2002; Haxby & Gobbini, 2012; Liu, Harris, & Kanwisher, 2002; Young & Bruce, 2011;

Zhao, Zhen, Liu, Song, & Liu, 2017). For example, extracting the morph averaging properties of a face arguably occurs at an earlier stage of encoding (Eimer, 2000; Gauthier et al., 2000; Grill-Spector, Knouf, & Kanwisher, 2004; Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yovel, 2006; Pitcher, Walsh, Yovel, & Duchaine, 2007) in comparison to when the brain can conceptually calculate the aspects of a face that make it unattractive (i.e., gist; O'Doherty et al., 2003). If we consider the visual features processing as perceptual, and the assessment of attractiveness as cognitive, we therefore provide the first direct evidence for distinct ensemble processing of temporal and spatial stimuli such that ensemble coding for temporal stimuli occurs at a perceptual level, whereas ensemble coding for spatial stimuli occurs at a cognitive level. This spatial process may be based on 'local support' such that "data coming from spatially local components of the image tend to use parallel computations, rather than global or serial methods" (e.g., Firestone & Scholl, 2016; Pylyshyn, 1999; Dawson & Pylyshyn, 1988; Marr & Poggio, 1979; Rosenfeld, Hummel, & Zucker, 1976). On the other hand, the refresh rate of the RSVP and spatial-temporal conditions in our experiments are really high. However, we are yet sure that whether the temporal (morph) averaging occurs before or after the attractiveness of the individual faces has been determined. We suspect that a new face norm is continuously being updated as each face is presented in the RSVP stream, and its information extracted. Only once this information has been extracted in the form of a new morphed face norm, can it then produce a conceptual appraisal (e.g., this group of unattractive faces' information morphs together to then be judged as moderately attractive) that drives subsequent adaptation aftereffects. We anticipate future neuroimaging and electrophysiological work will confirm these distinct neural stages responsible for driving ensemble statistics derived from temporal versus spatial averaging.

Conclusions

Researchers have long speculated as to the composition of the neural calculations performed during ensemble coding. We have shown for the first time that temporal ensemble statistics do not simply reflect the 'gist' of the attractiveness judgements attributed to a group of faces, but are instead extracted by morph averaging the group's fiducial points and surface characteristics together. By contrast, spatial ensemble coding appears reflective of a gist averaging process in which the group's general characteristics of attractiveness (e.g., this group is unattractive), can be maintained. This reveals two distinct levels of ensemble statistics that can occur for the same facial trait: the gist averaging we observed during static spatial ensemble coding, and the morph averaging for temporal ensemble coding. We anticipate that these results will help inform a broader theoretical framework to understand ensemble perception, but also enhance our knowledge of face processing and appraisal mechanisms.

Author Contributions

H. Ying, A. Choo, E. Burns and H. Xu developed the study concept and contributed to the study design. H. Ying and A. Choo performed testing and data collection. H. Ying performed the data analysis and interpretation under the supervision of H. Xu. H. Ying drafted the manuscript, and E. Burns and H. Xu provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgements

Supported by Nanyang Technological University Research Scholarship (HY), Undergraduate Research Experience on Campus (AC), College of Arts, Humanities and Social Sciences Incentive Scheme (HX), and Singapore Ministry of Education Academic Research Fund (AcRF) Tier 1 (HX). H. Ying is also supported by the Social Science Foundation of Jiangsu Province (17JYC006), the City & University strategy-Soochow University Leading Research Team in Humanities and Social Sciences. Parts of this research (data from Exp 1) were presented at the Annual Meeting of Visual Science Society (VSS), May 2017, St. Pete Beach, Florida. The research reported here forms part of H. Ying's Ph.D. thesis at Nanyang Technological University. All data have been made publicly available via the Open Science Framework (OSF) and can be accessed at https://osf.io/rgdja/.

References

1. Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn Sci, 15*(3), 122-131. doi:10.1016/j.tics.2011.01.003
2. Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological science, 12*(2), 157-162. doi:Doi 10.1111/1467-9280.00327
3. Bachy, R., & Zaidi, Q. (2014). Factors governing the speed of color adaptation in foveal versus peripheral vision. *JOSA A, 31*(4), A220–A225.
4. Bartolomeo, P., Vuilleumier, P., & Behrmann, M. (2015). The whole is greater than the sum of the parts: Distributed circuits in visual cognition. *Cortex, 72*, 1-4. doi:10.1016/j.cortex.2015.09.001
5. Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in psychology, 8*, 456. doi: 10.3389/fpsyg.2017.00456.
6. Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition (vol 17, pg 210, 2013). *Trends in Cognitive Sciences, 17*(7), 361-361. doi:10.1016/j.tics.2013.05.009
7. Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.
8. Burns, E. J., Martin, J., Chan, A. H. D., & Xu, H. (2017). Impaired processing of facial happiness, with or without awareness, in developmental prosopagnosia. *Neuropsychologia, 102*, 217-228. doi:10.1016/j.neuropsychologia.2017.06.020

9. Burton, N., Jeffery, L., Bonner, J., & Rhodes, G. (2016). The timecourse of expression aftereffects. *Journal of Vision, 16*(15), 1-1.
10. Chen, C., Chen, X., Gao, M., Yang, Q., & Yan, H. (2015). Contextual influence on the tilt after-effect in foveal and para-foveal vision. *Neuroscience Bulletin, 31*(3), 307–316.
11. Dawson, M. & Pylyshyn, Z. W. (1988) Natural constraints in apparent motion. In: *Computational processes in human vision: An interdisciplinary perspective*, ed. Z. W. Pylyshyn. Ablex Publishing.
12. Debruine, L., Jones, B. C., Unger, L., & Little, A. C. (2007). Dissociating averageness and attractiveness: attractive faces are not always average. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 11. doi:0.1037/0096-1523.33.6.1420.
13. Debruine, L., & Tiddeman, B. (2017). *WebMorph.*, Retrieved from http://webmorph.org/.
14. Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*. doi:ARTN 78110.3389/fpsyg.2014.00781
15. Duchaine, B., & Yovel, G. (2015). A Revised Neural Framework for Face Processing. *Annual Review of Vision Science, Vol 1, 1*, 393-416. doi:10.1146/annurev-vision-082114-035518
16. Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble Perception of Dynamic Emotional Groups. *Psychological Science, 28*, 193-203. doi:10.1177/0956797616678188
17. Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport*, *11*(10), 2319-2324.
18. Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.
19. Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences, 39,* e229.
20. Fox, C. J., & Barton, J. J. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain research, 1127*, 80-89.
21. Galton, F. (1878). Composite portraits. *Journal of the Anthropological Institute of Great Britain & Ireland, 8*, 132-144.
22. Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12*(5), 912-912.
23. Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32-41. doi:10.1016/j.neuropsychologia.2006.04.015
24. Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7*(5), 555-562. doi:10.1038/nn1224
25. Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *J Exp Psychol Gen, 144*(2), 432-446. doi:10.1037/xge0000053
26. Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of vision, 15*(4), 16-16. doi:10.1167/15.4.16
27. Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Curr Biol, 17*(17), R751-753. doi:10.1016/j.cub.2007.06.039

28. Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *J Exp Psychol Hum Percept Perform, 35*(3), 718-734. doi:10.1037/a0013899
29. Haberman, J., & Whitney, D. (2012). Ensemble Perception:summarizing the scene and broadening the limits of visual processing.
30. Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223-233. doi:Doi 10.1016/S1364-6613(00)01482-0
31. Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*(1), 59-67. doi:Doi 10.1016/S0006-3223(01)01330-0
32. Haxby, J. V., & Gobbini, M. I. Distributed Neural Systems for Face Perception, Oxford University Press (2011).
33. Hsu, S. M., & Young, A. (2004). Adaptation effects in facial expression recognition. *Visual Cognition*, *11*(7), 871-899.
34. Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.
35. Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B-Biological Sciences, 361*(1476), 2109-2128. doi:10.1098/rstb.2006.1934
36. Keysers, C., Xiao, D. K., Foldiak, P., & Perrett, D. I. (2001). The speed of sight. *Jouranl of Cogntive Neuroscience, 13*(1), 90-101.
37. Leder, H., Goller, J., Forster, M., Schlageter, L., & Paul, M. A. (2017). Face inversion increases attractiveness. *Acta psychologica*, *178*, 25-31.
38. Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience, 4*(1), 89-94. doi:10.1038/82947
39. Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat Neurosci, 5*(9), 910-916. doi:10.1038/nn909
40. Liu, P., Montaser-kouhsari, L., & Xu, H. (2014). Effects of face feature and contour crowding in facial expression adaptation. *Vision Research*, *105*, 189–198. https://doi.org/10.1016/j.visres.2014.10.014
41. Luo, C., Burns, E., & Xu, H. (2017). Association between autistic traits and emotion adaptation to partially occluded faces. *Vision research, 133*, 21-36.
42. Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, 91-630.
43. Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *204*(1156), 301-328.
44. McKeeff, T. J., Remus, D. A., & Tong, F. (2007). Temporal limitations in object processing across the human ventral visual pathway. *Journal of Neurophysiology, 98*(1), 382-393. doi:10.1152/jn.00568.2006
45. Morey, R. D., & Rouder, J. N. (2015). BayesFactor (Version 0.9.11-3)[Computer software].

46. O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia, 41*(2), 147-155. doi:10.1016/s0028-3932(02)00145-8
47. 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.
48. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442.
49. Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shape and judgements of female attractiveness. *Nature, 368*, 239-242. doi:10.1038/368239a0
50. Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17*(18), 1568-1573. doi:10.1016/j.cub.2007.07.063
51. Potter, M. C. (1976). Short-term conceptual memory for pictures. *J Exp Psychol Hum Learn, 2*(5), 509-22. doi: 10.1037/0278-7393.2.5.509
52. Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and brain sciences, 22*(3), 341-365.
53. R Core Team (2017). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.
54. Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Res, 46*(18), 2977-2987. doi:10.1016/j.visres.2006.03.002
55. Rhodes, G., Jeffery, L., Clifford, C. W., & Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. *Vision Research, 47*(17), 2291-2296.
56. Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W., & Nakayama, K. (2003). Fitting the mind to the world: face adaptation and attractiveness aftereffects. *Psychol Sci, 14*(6), 558-566.
57. Rouder, J.N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356-374.
58. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.
59. Rosenfeld, A., Hummel, R. A., & Zucker, S. W. (1976). Scene labeling by relaxation operations. *IEEE Transactions on Systems, Man, and Cybernetics*, (6), 420-433.
60. Sou, K. L., & Xu, H. (2019). Brief facial emotion aftereffect occurs earlier for angry than happy adaptation. *Vision research, 162*, 35-42
61. Tiddeman, B., Burt, D.M., & Perrett, D. (2001). Computer Graphics in Facial Perception Research, *IEEE Computer Graphics and Applications, 21*(5), 42-50.
62. Valentine, T., Darling, S., & Donnelly, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin & Review, 11*(3), 482-487.
63. Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*(6982), 557-561. doi:10.1038/nature02420
64. Webster, M. A., & MacLeod, D. I. (2011). Visual adaptation and face perception. *Philos Trans R Soc Lond B Biol Sci, 366*(1571), 1702-1725. doi:10.1098/rstb.2010.0360

65. Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cognitve Science, 15*(4), 160-168. doi:10.1016/j.tics.2011.02.005
66. Whitney, D., & Yamanashi Leib, A. (2017). Ensemble Perception. *Annu Rev Psychol*. doi:10.1146/annurev-psych-010416-044232
67. Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42*(3), 671-684. doi:10.3758/Brm.42.3.671
68. Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Jouranl of Vission, 15*(4), 11. doi:10.1167/15.4.11
69. Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the cortical hierarchy: Low-level curve adaptation affects high-level facial-expression judgments. *Journal of Neuroscience, 28*(13), 3374-3383.
70. Yap, W.J., Chan, E., & Christopoulos, G.I. (July 2016). *Nanyang Facial Emotional Expression [N-FEE] Database - Development and Validation*. Poster presented at the 23rd Congress of the International Association for Cross-Cultural Psychology, Nagoya, Japan.
71. Ying, H., & Xu, H. (2017). Adaptation reveals that facial expression averaging occurs during rapid serial presentation. *Journal of Vision, 17*(1), 15. doi:10.1167/17.1.15
72. Ying, H., Burns, E. J., Lin, X., & Xu, H. (2019). Ensemble statistics shape face adaptation and the cheerleader effect. *Journal of Experimental Psychology: General , 148(3), 421.*
73. Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959-974. doi:10.1111/j.2044-8295.2011.02045.x
74. Zhao, L., & Chubb, C. (2001). The size-tuning of the face-distortion after-effect. *Vision Research, 41*(23), 2979-2994. doi:Doi 10.1016/S0042-6989(01)00202-4
75. Zhang, G. L., Li, A. S., Miao, C. G., He, X., Zhang, M., & Zhang, Y. (2018). A consumer-grade LCD monitor for precise visual stimulation. *Behavior research methods, 50*(4), 1496-1502.
76. Zhao, Y., Zhen, Z., Liu, X., Song, Y., & Liu, J. (2017). The neural network for face recognition: Insights from an fMRI study on developmental prosopagnosia. *Neuroimage, 169*, 151-161. doi:10.1016/j.neuroimage.2017.12.023

Figure 1

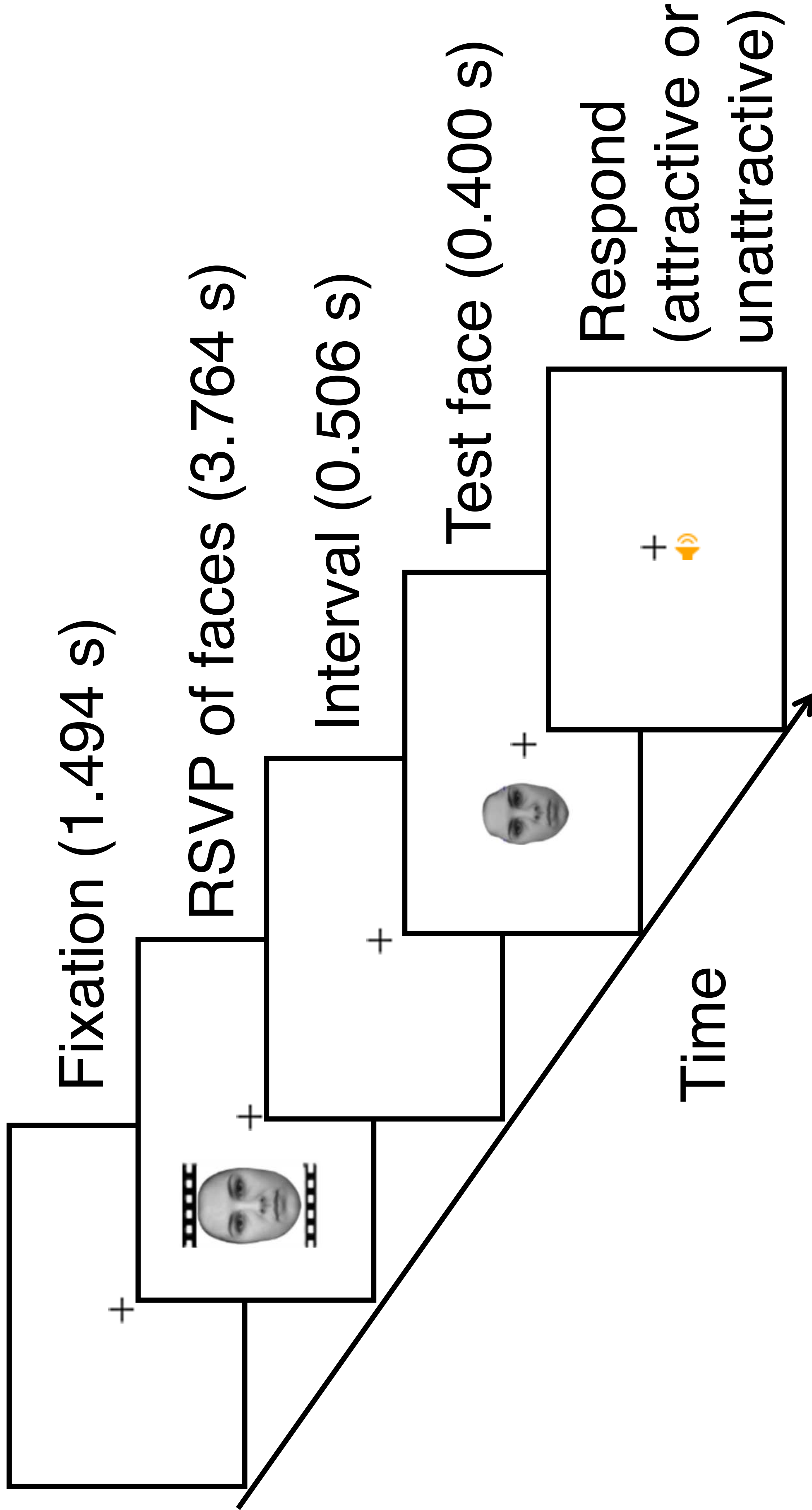

Fixation (1.494 s)
RSVP of faces (3.764 s)
Interval (0.506 s)
Test face (0.400 s)
Respond (attractive or unattractive)
Time

Figure 2

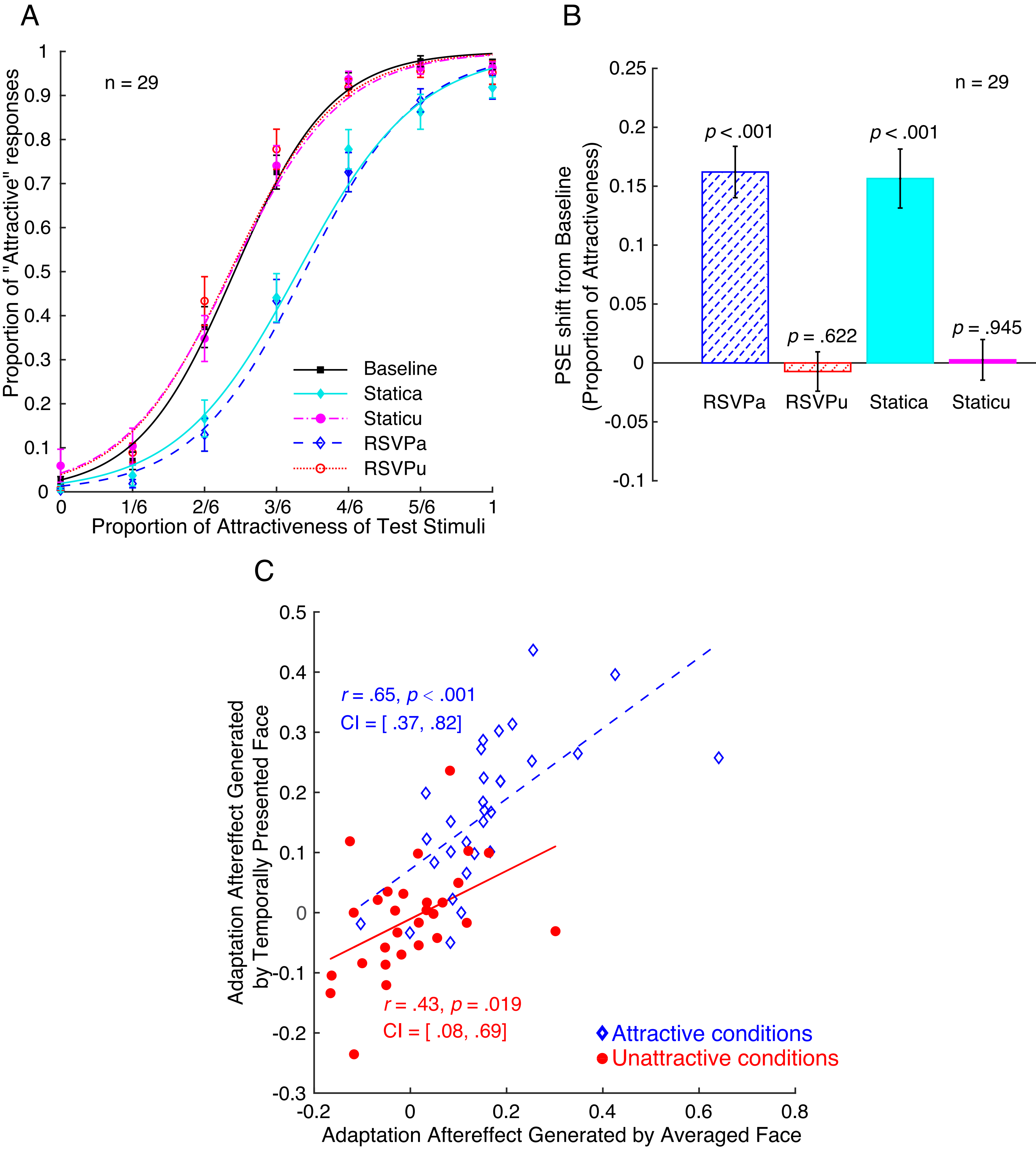

A
n = 29
Proportion of "Attractive" responses
Proportion of Attractiveness of Test Stimuli
0
1/6
2/6
3/6
4/6
5/6
1
Baseline
Statica
Staticu
RSVPa
RSVPu
B
n = 29
PSE shift from Baseline (Proportion of Attractiveness)
p < .001
p = .622
p < .001
p = .945
RSVPa
RSVPu
Statica
Staticu
C
Adaptation Aftereffect Generated by Temporally Presented Face
Adaptation Aftereffect Generated by Averaged Face
r = .65, p < .001
CI = [.37, .82]
r = .43, p = .019
CI = [.08, .69]
Attractive conditions
Unattractive conditions

Figure 3

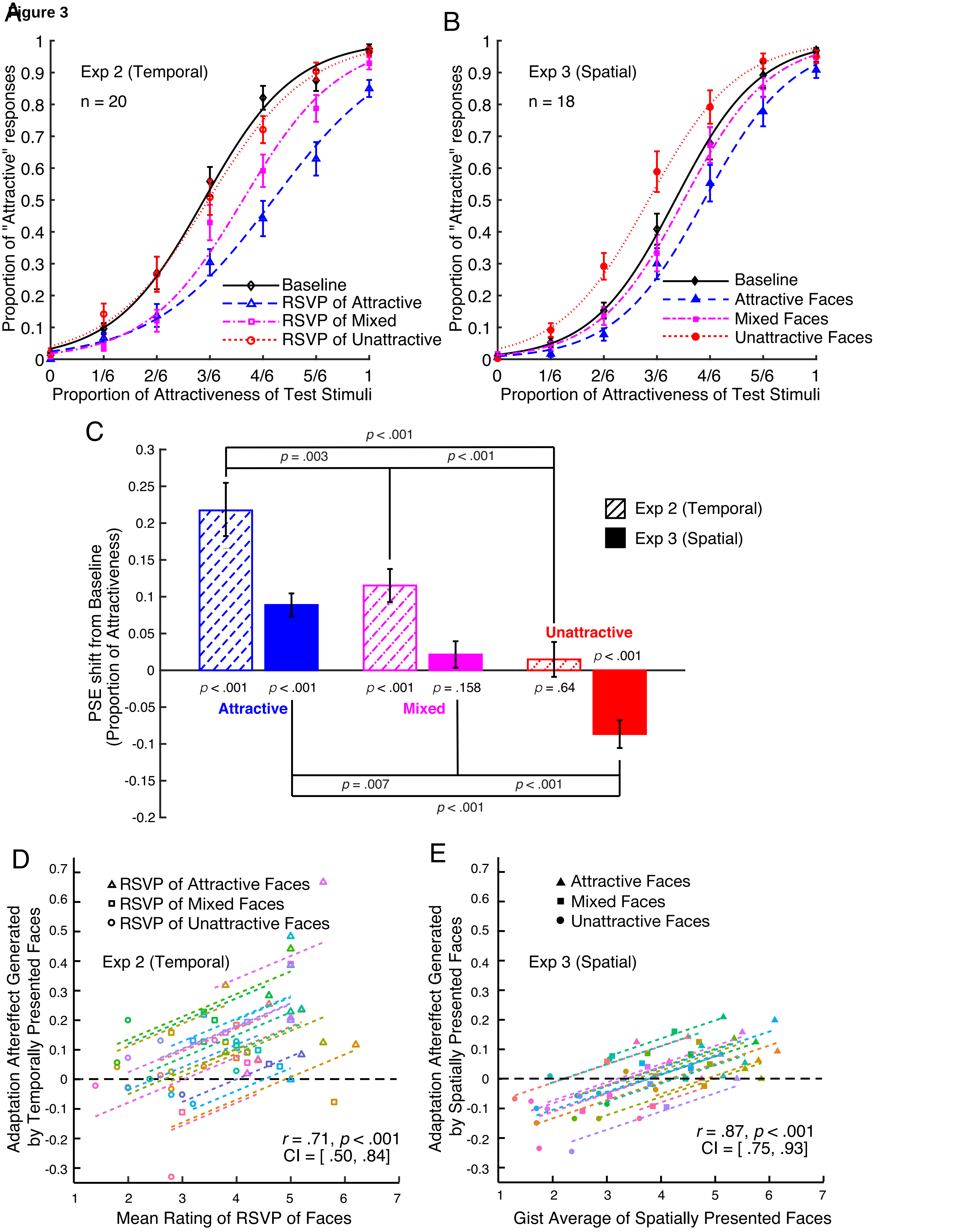

A
Exp 2 (Temporal)
n = 20
Proportion of "Attractive" responses
Baseline
RSVP of Attractive
RSVP of Mixed
RSVP of Unattractive
Proportion of Attractiveness of Test Stimuli
B
Exp 3 (Spatial)
n = 18
Proportion of "Attractive" responses
Baseline
Attractive Faces
Mixed Faces
Unattractive Faces
Proportion of Attractiveness of Test Stimuli
C
PSE shift from Baseline (Proportion of Attractiveness)
p < .001
p = .003
p < .001
Exp 2 (Temporal)
Exp 3 (Spatial)
Unattractive
p < .001
p < .001
p < .001
p < .001
p = .158
p = .64
Attractive
Mixed
p = .007
p < .001
p < .001
D
RSVP of Attractive Faces
RSVP of Mixed Faces
RSVP of Unattractive Faces
Exp 2 (Temporal)
Adaptation Aftereffect Generated by Temporally Presented Faces
r = .71, p < .001
CI = [.50, .84]
Mean Rating of RSVP of Faces
E
Attractive Faces
Mixed Faces
Unattractive Faces
Exp 3 (Spatial)
Adaptation Aftereffect Generated by Spatially Presented Faces
r = .87, p < .001
CI = [.75, .93]
Gist Average of Spatially Presented Faces

Figure 4

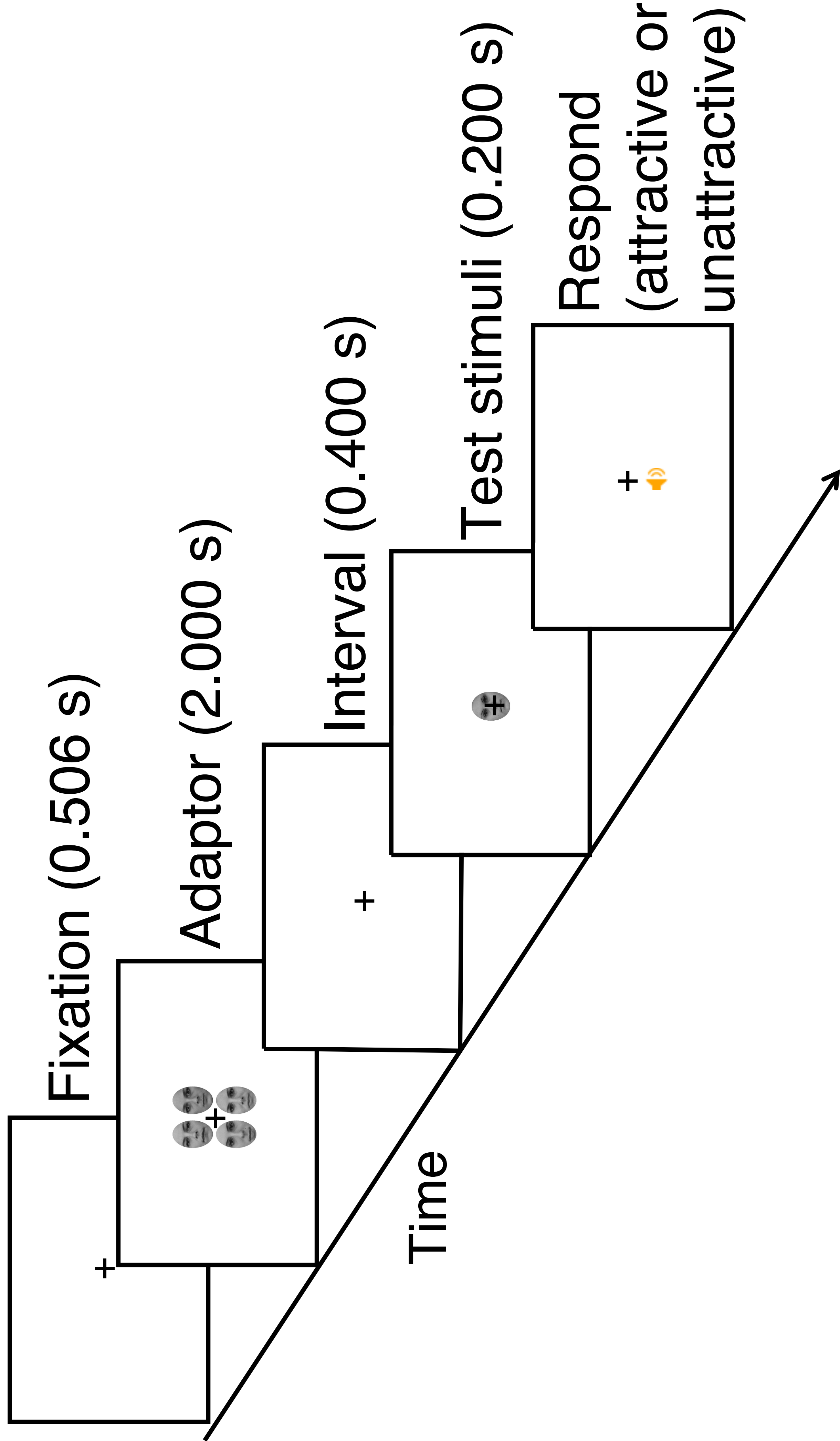

Fixation (0.506 s)
Adaptor (2.000 s)
Interval (0.400 s)
Test stimuli (0.200 s)
Respond (attractive or unattractive)
Time

Figure 5

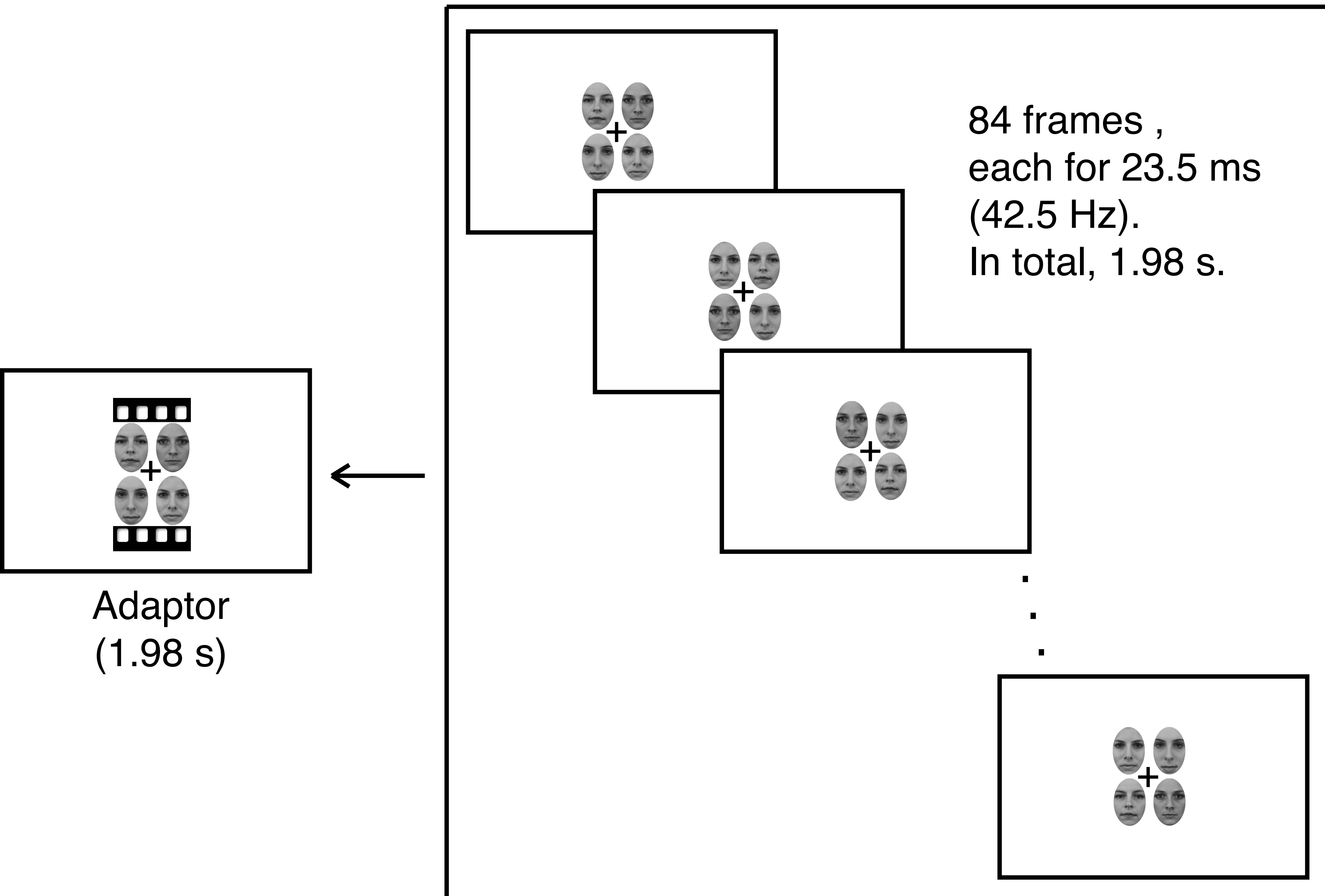
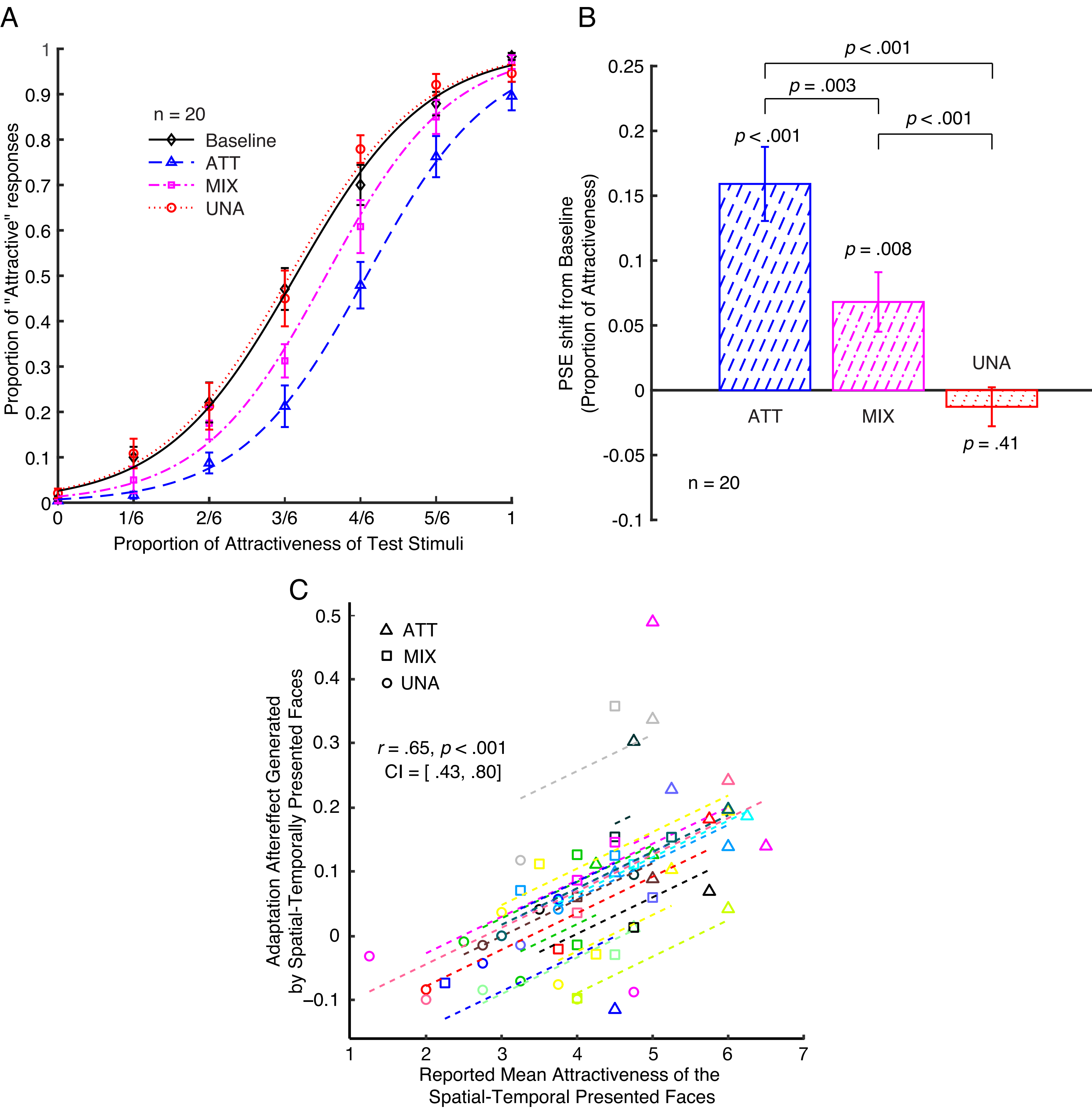
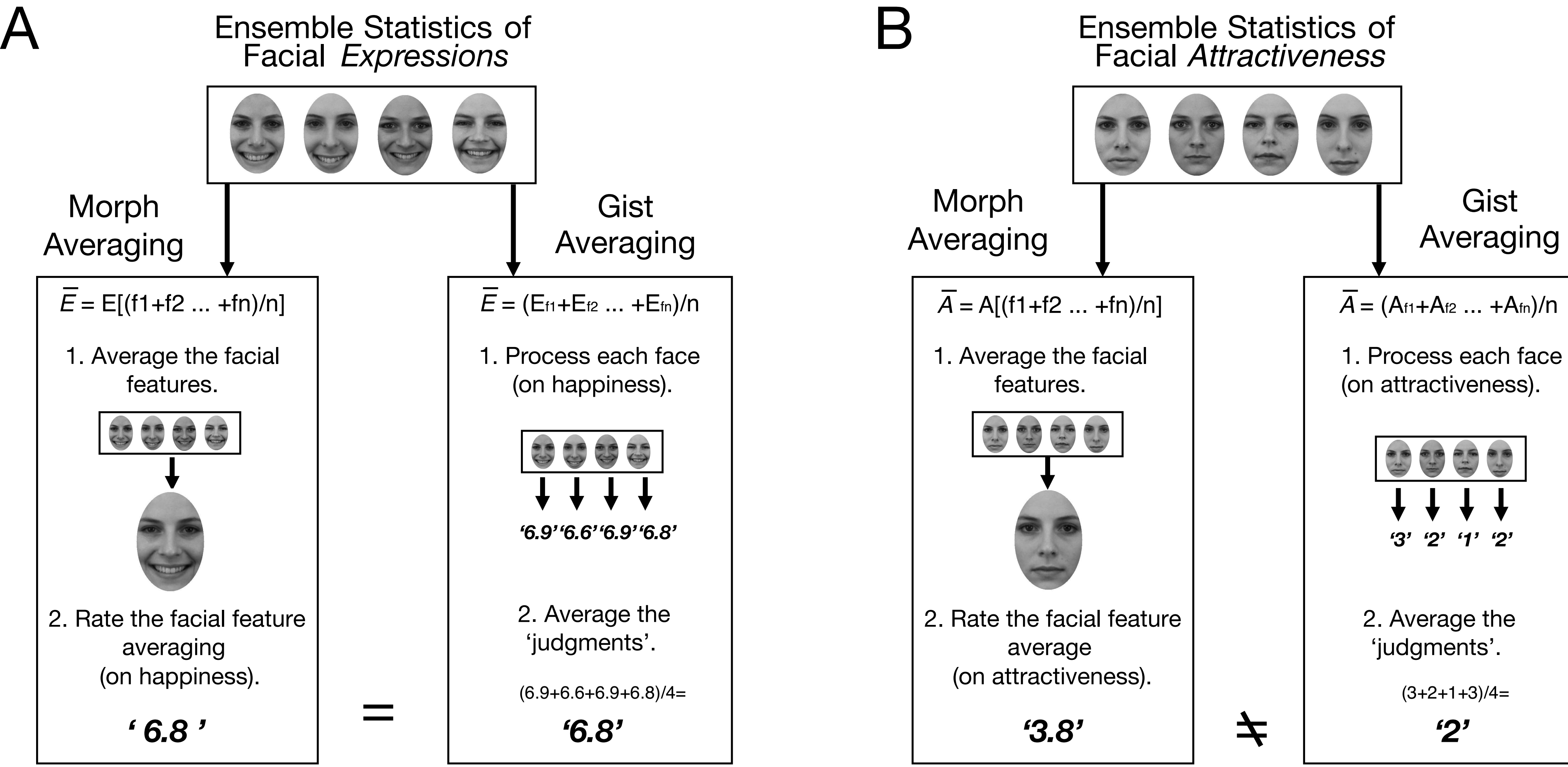


Figure 6



A
n = 20
Baseline
ATT
MIX
UNA
Proportion of "Attractive" responses
Proportion of Attractiveness of Test Stimuli
0
1/6
2/6
3/6
4/6
5/6
1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
B
p < .001
p = .003
p < .001
p < .001
p = .008
UNA
ATT
MIX
p = .41
n = 20
PSE shift from Baseline
(Proportion of Attractiveness)
0.25
0.2
0.15
0.1
0.05
-0.05
-0.1
C
ATT
MIX
UNA
r = .65, p < .001
CI = [.43, .80]
Adaptation Aftereffect Generated
by Spatial-Temporally Presented Faces
Reported Mean Attractiveness of the
Spatial-Temporal Presented Faces
0.5
0.4
0.3
0.2
0.1
−0.1
2
3
4
5
6
7

Figure 7


A
Ensemble Statistics of
Facial *Expressions*
Morph
Averaging
$\bar{E}$ = E[(f1+f2 ... +fn)/n]
1. Average the facial features.
2. Rate the facial feature averaging (on happiness).
' 6.8 '
=
Gist
Averaging
$\bar{E}$ = (E_{f1}+E_{f2} ... +E_{fn})/n
1. Process each face (on happiness).
'6.9' '6.6' '6.9' '6.8'
2. Average the 'judgments'.
(6.9+6.6+6.9+6.8)/4=
'6.8'
B
Ensemble Statistics of
Facial *Attractiveness*
Morph
Averaging
$\bar{A}$ = A[(f1+f2 ... +fn)/n]
1. Average the facial features.
2. Rate the facial feature average (on attractiveness).
'3.8'
≠
Gist
Averaging
$\bar{A}$ = (A_{f1}+A_{f2} ... +A_{fn})/n
1. Process each face (on attractiveness).
'3' '2' '1' '2'
2. Average the 'judgments'.
(3+2+1+3)/4=
'2'