Cognition 250 (2024) 105861

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

# Objectively quantifying subjective phenomena: Measuring the flashed face distortion effect

Yi Gao [a,*], Minzhi Wang [b], Dobromir Rahnev [a]

[a] School of Psychology, Georgia Institute of Technology, Atlanta, GA, United States of America
[b] School of Computer Science, Georgia Institute of Technology, Atlanta, GA, United States of America

ARTICLE INFO

Editor: J. Wagemans

*Keywords:*
Face distortion
Objective method
Temporal frequency
Eccentricity
Face inversion

ABSTRACT

Objectively quantifying subjective phenomena like visual illusions is challenging. We address this issue in the context of the Flashed Face Distortion Effect (FFDE), where faces presented in succession appear distorted and grotesque. We first show that the traditional method of quantifying FFDE – via subjective ratings of the level of distortion – is subject to substantial biases. Motivated by this finding, we develop an objective method for quantifying FFDE by introducing two design innovations. First, we create artificially distorted faces and ask subjects to discriminate between undistorted and objectively distorted faces. Second, we employ both an illusion condition, which includes a succession of 15 face flashes, and a control condition, which includes a single face flash and does not induce an illusion. Using these innovations, we quantify the strength of the face distortion illusion by comparing the response bias for identifying distorted faces between the illusion and control conditions. We find that our method successfully quantifies the face distortion, with subjects exhibiting a more liberal response bias in the illusion condition. Finally, we apply our new method to evaluate how the face distortion illusion is modulated by face eccentricity, face inversion, the temporal frequency of the face flashes, and presence of temporal gaps between consecutive faces. Our results demonstrate the utility of our objective method in quantifying the subjective illusion of face distortion. Critically, the method is general and can be applied to other phenomena that are inherently subjective.

## 1. Introduction

Quantifying subjective phenomena such as visual and auditory imagery presents a substantial challenge due to their inherently subjective nature. This challenge is particularly evident in conditions like aphantasia, where individuals are unable to form visual images (Dance, Ipser, & Simner, 2022; Zeman et al., 2020; Zeman, Dewar, & Della Sala, 2015). Aphantasia is often assessed using tools like the Vividness of Visual Imagery Questionnaire (VVIQ) (Marks, 1973, 1995). The VVIQ measures mental imagery vividness on a scale from 1 (absence of mental imagery) to 5 (highly vivid). However, these ratings are subjective and can vary significantly based on individual biases. For instance, one person's rating of 5 for highly vivid imagery might involve detailed features such as specific colors and shapes, whereas another might give the same rating to more generalized imagery.

Quantification challenges extend to other subjective phenomena, notably visual illusions. While simpler illusions like the Müller-Lyer Illusion can be quantified using behavioral measurements like comparative and adjustment tasks (Manning, Morgan, Allen, & Pellicano, 2017; Tudusciuc & Nieder, 2010) or neurophysiological measurements (Plewan, Weidner, Eickhoff, & Fink, 2012; Weidner, Boers, Mathiak, Dammers, & Fink, 2010), more complex visual illusions pose greater challenges. One such complex phenomenon is the Flashed Face Distortion Effect (FFDE) where eye-aligned faces presented in succession are perceived as distorted (Tangen, Murphy, & Thompson, 2011). The faces are typically described as deformed and grotesque, with faces being perceived as twisted, squished, or enlarged. Critically, there is no objective method of quantifying the strength of the face distortion. All the current studies on FFDE have adopted a rating method where subjects indicate the subjective level of distortion using a rating scale (Balas & Pearson, 2019; Bowden, Whitaker, & Dunn, 2019; Gao, 2021; Utz & Carbon, 2015; Wen & Kung, 2014).

Though the rating method is a convenient and efficient way to collect data and is widely used in surveys and many other psychological studies, it suffers from several drawbacks (Uher, 2018, 2019). Most importantly, raters may intuitively rate according to explicit features of the

* Corresponding author at: Georgia Institute of Technology, 654 Cherry Str. NW, Atlanta, GA 30332, United States of America.
*E-mail address:* yi.gao0525@outlook.com (Y. Gao).

https://doi.org/10.1016/j.cognition.2024.105861
Received 15 March 2024; Received in revised form 11 June 2024; Accepted 13 June 2024
Available online 17 June 2024

manipulations, such as size, duration, and location of the stimuli. For example, observers may rate more eccentric faces as more distorted or less distorted due to low detectability and pooling of details in the peripheral visual field (Greenwood, Bex, & Dakin, 2009; Rosenholtz, 2016). This kind of response strategy can save them time and effort by making it unnecessary to engage in effortful information processing and decision-making. The underlying mechanisms of what people base their ratings on are hidden, making subjective ratings a suboptimal method of quantifying face distortion.

To address these challenges, we introduce a novel method to objectively characterize the amount of distortion in FFDE. The method combines two design innovations. First, we created artificially distorted faces designed to mimic the illusory facial distortions, and then asked subjects to discriminate between undistorted and objectively distorted faces. Second, we employed an illusion and control conditions. The illusion condition included a succession of 15 face flashes and produced a strong illusory face distortion. The control condition included a single face flash and did not induce an illusion. We then quantified the magnitude of subjective face distortion as the difference in response criterion between the illusion and control conditions.

We first show that traditional subjective methods are corrupted by decisional bias that is unrelated to the true subjective experience. Further, we demonstrate the efficacy of our objective method in quantifying face distortions by analyzing the impact of various factors – namely, temporal frequency, eccentricity, temporal gap, and face inversion – on subjective face distortion. Overall, these results establish the robustness and validity of our proposed quantification technique.

## 2. Methods

### 2.1. Subjects

We analyzed data from three experiments ($N = 27$ for each). Experiments 1 and 2 were taken from a previous manuscript (Gao, Chen, & Rahnev, 2023). Experiment 3 was a new experiment and comprised 17 females and 10 males with a mean age of 20.04 (SD = 22.61, range = 18–31). No subjects were duplicated between the three experiments. In all experiments, the sample sizes were selected to achieve 80% statistical power for a medium effect size (0.5). All subjects had normal or corrected-to-normal vision and provided signed consent approved by the Institutional Review Board of the university. Subjects were recruited from the university and received compensation of 1 Student Research Participation (SONA) credit or $10 per hour.

### 2.2. Procedure

All experiments included an illusion condition (Fig. 1A) and a control condition (Fig. 1B). The illusion condition involved presenting 15 flashes of bilaterally presented faces known to induce a strong face distortion illusion known as the Flashed Face Distortion Effect (FFDE) (Balas & Pearson, 2019; Bowden et al., 2019; Tangen et al., 2011; Wen & Kung, 2014). Faces in this condition started to appear distorted and even grotesque after the initial few flashes. The control condition involved presenting a single flash of bilaterally presented faces, which does not induce a perceptual illusion of face distortion (Balas & Pearson, 2019; Tangen et al., 2011; Wen & Kung, 2014). Subjects performed an objective task in which they evaluated whether the last pair of faces presented exhibited any distortions. In half of the trials in each condition, one of the last two faces was artificially distorted (left panels of Fig. 1A,B). In the other half of the trials, both faces remained undistorted (right panels of Fig. 1A,B). In the illusion condition, the first 14 flashes included only undistorted faces.

Each trial began with a red dot presented at fixation for a random duration ranging from 800 to 1300 ms. Following this, either 15 flashes or a single flash of faces (height = 6 degrees, width = 6 degrees) were presented, depending on the experimental conditions. In Experiments 1 and 2, half of the trials featured flashes lasting 250 ms, while the other half had flashes lasting 1000 ms. Additionally, the location of the image varied, such that the edge closest to fixation was at 2-degree eccentricity in half the trials and at 8-degree eccentricity in the other half of the trials. Note that because the images were 6-degree wide, their centers were located 5 and 11 degrees away from fixation, respectively. Each unique combination of the number of flashes, duration, and eccentricity was replicated twice per block, with each block comprising 16 trials. After the face presentation, subjects encountered a single response screen featuring three questions. The first question prompted subjects to identify whether either of the last two faces was distorted. The second question required subjects to rate their confidence on a 4-point scale (not confident at all, somewhat confident, very confident, extremely confident). Note that we do not analyze the confidence ratings in the current paper. The third question asked subjects to provide a subjective rating of the level of distortion (not distorted, minor distortion, major distortion, extreme distortion). We adopted a 4-point scale for distortion ratings, aligning with the 4-point confidence scale, which is the most commonly used scale for confidence in perceptual studies (Rahnev et al., 2020). Utilizing a consistent scale for both distortion judgment and confidence rating minimizes cognitive load for participants, potentially enhancing the quality of the data collected. Subjects had unlimited time to respond, with each question displayed on the same response screen. Subsequent questions were revealed only after the participant responded to the previous one.

In Experiment 1, all faces were displayed in an upright orientation. Experiment 2 replicated Experiment 1 but with an inverted orientation for all faces. Both experiments featured a $2 \times 2$ factorial design, encompassing stimulus eccentricity (faces presented at 2- vs. 8-degree eccentricity) and temporal frequency of face presentation (1 Hz vs. 4 Hz) as factors, resulting in four conditions. Across 10 blocks in each experiment, totaling 160 trials, there were 40 trials per condition.

Experiment 3 involved upright face presentations lasting 1000 ms at an 8-degree eccentricity. In the illusion condition, continuous faces were interspersed with a gray screen and fixation point, with temporal gaps of 0, 100, 500, and 1000 ms (Fig. 2). In the control condition, a single flash of faces was presented. Each condition of the five conditions was repeated twice per block, and there were 10 blocks in total, resulting in 100 trials in total, with 20 trials per temporal gap.

### 2.3. Stimuli

We used the same stimuli from our previous work (Gao et al., 2023). We utilized thirty faces (15 female and 15 male faces) from the Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Ohman, 1998). We manually created distorted faces using five methods (see Fig. 1C for examples of undistorted and artificially distorted faces). The first four types of distortion were implemented using the nudging tool provided by PicMonkey Photo Editor and Graphic Design Maker (picmonkey.com). Each face was aligned on a $8 \times 8$ grid, with specific facial features manually adjusted to introduce various forms of distortion. These modifications entailed: (1) elongating the eyes, nose, and contour, (2) elongating the eyes and mouth, (3) squeezing the eyes and mouth, and (4) rotating the eyes and mouth. Additionally, a fifth method of distortion was achieved through MATLAB scripts (The MathWorks, Natick, MA), which algorithmically contracted the size of the eyes and mouth based on methods described by Webster and Maclin (1999) and Yamashita, Hardy, De Valois, and Webster (2005).

### 2.4. Analyses

In both the illusion and control conditions, half the trials included distorted faces and the other half included only undistorted faces. In each condition, we computed response bias (c) for each subject based on the responses to the first question in each trial (“Is either of the last two faces distorted?”). To do so, we coded objectively distorted faces as the

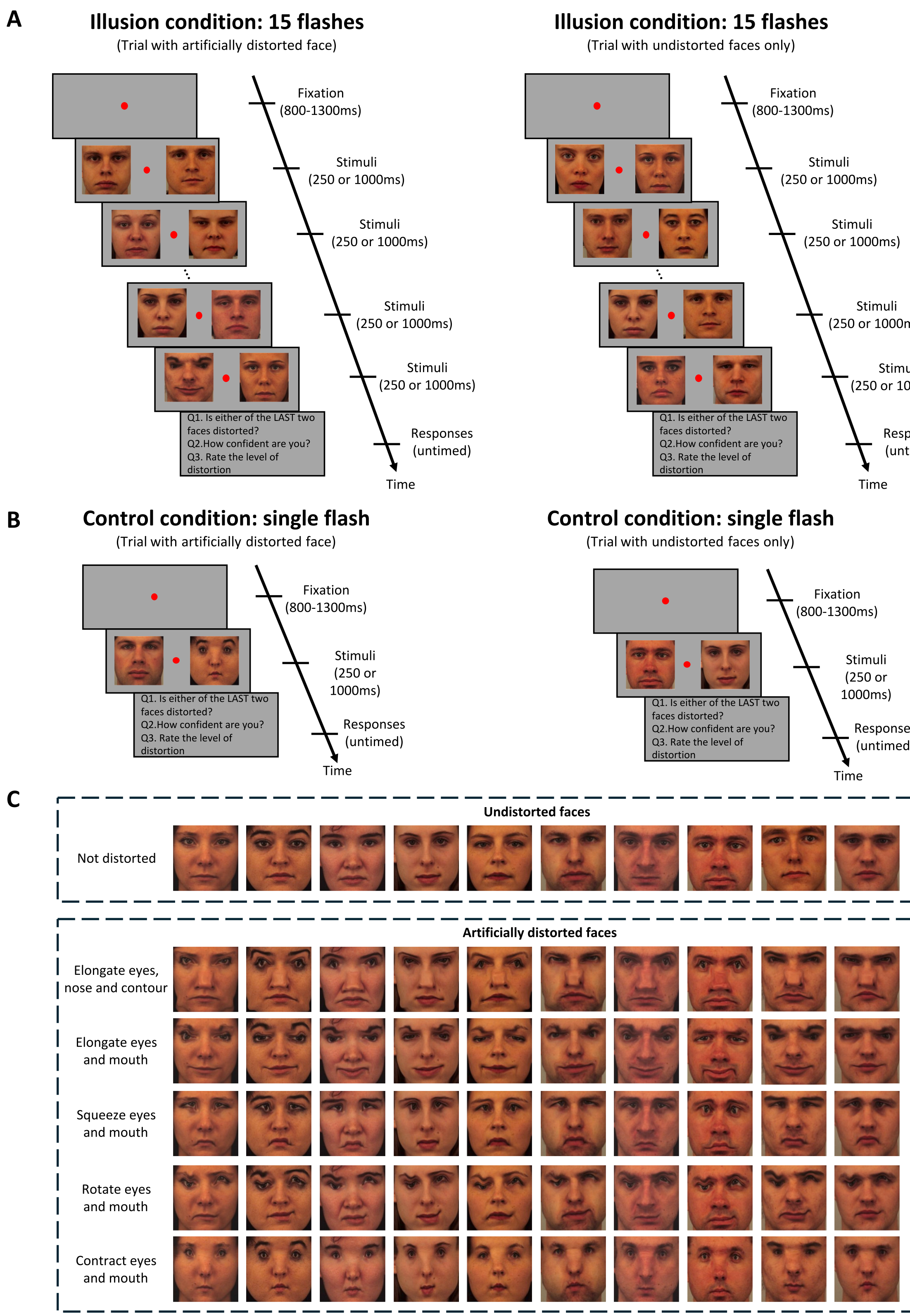


**Fig. 1.** Task and stimuli for Experiments 1 and 2. (A) Example trials from the illusion condition. The illusion condition comprised 15 flashes of faces inducing a strong visual illusion. Note that each example trial shows only four of the 15 flashes of faces. (B) Example trials from the control condition. The control condition involved a single flash of faces and did not induce illusionary distortion. The left panels of Figs. A and B display example trials that include artificially distorted faces, whereas the right panels illustrate example trials where no artificially distorted faces are present. In Experiments 1 and 2, the faces were presented for either 250 or 1000 ms (the same duration was used for all faces within a condition). Following each trial, subjects responded to three questions: Q1. Is either of the last two faces distorted? Q2. How confident are you? Q3. Rate the level of distortion. (C) Examples of undistorted faces and artificially distorted faces created by different methods.

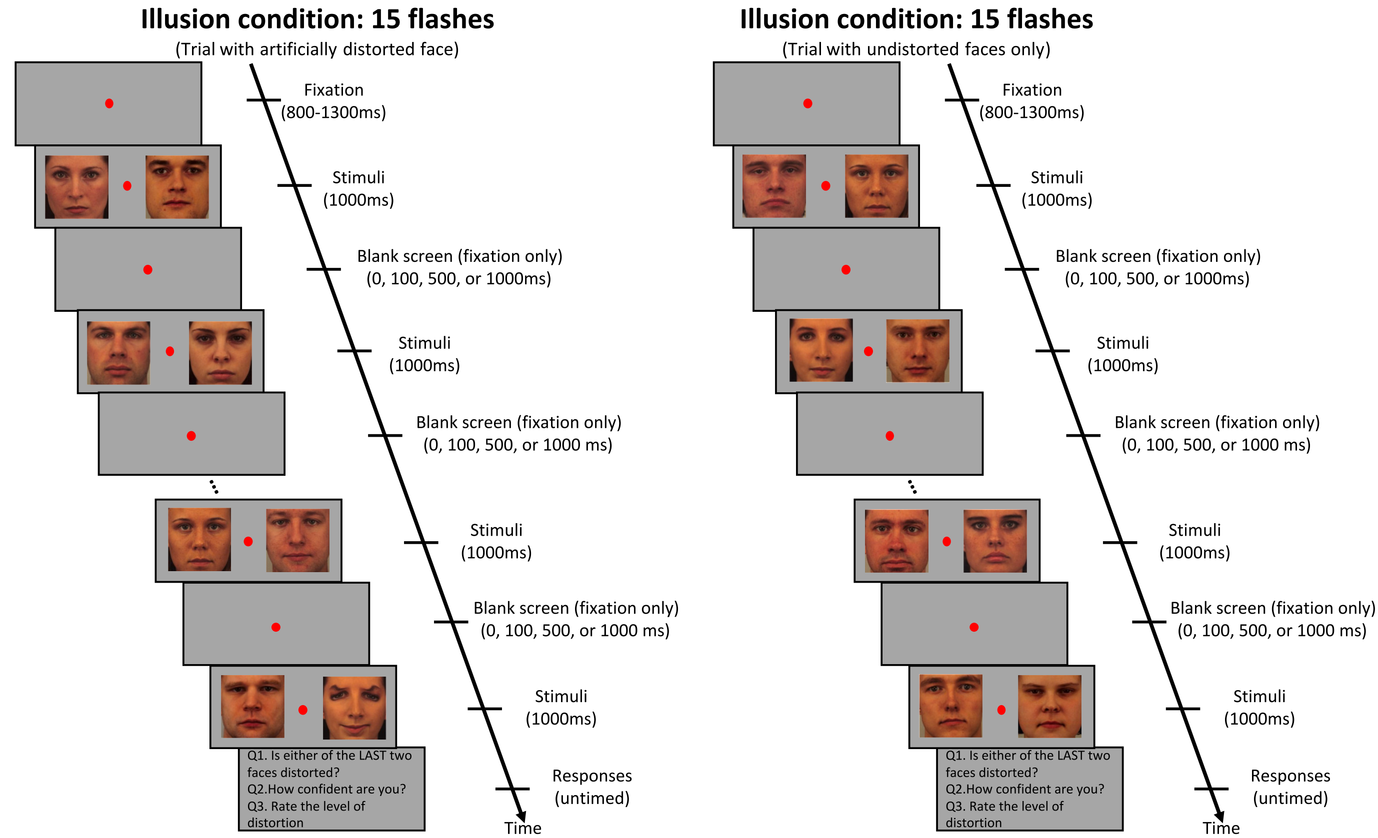


**Fig. 2.** Illusion condition design in Experiments 3. Example trials from the illusion condition in Experiment 3. The left panel displays an example trial that includes artificially distorted faces, whereas the right panel displays an example trial that does not include artificially distorted faces. Faces were always presented for 1000 ms. A blank screen with fixation point only was presented for 0, 100, 500, or 1000 ms (the same duration was used for all temporal gaps within a trial). Note that each example trial shows only four of the 15 flashes of faces. The control condition was the same as that of Experiments 1 and 2 (Fig. 1B) except that the flash lasted 1000 ms.

non-target and undistorted faces as the target. Consequently, the hit rate was calculated as the proportion of trials for which no artificial distortion was presented and subjects responded that there was no distortion. The false alarm rate was calculated as the proportion of trials for which an artificially distortion face was presented but subjects responded that there was no distortion. The response bias, c, was then computed using the formula from signal detection theory (Green & Swets, 1966):

$$c = -\frac{\varphi^{-1}(\text{hit rate}) + \varphi^{-1}(\text{false alarm rate})}{2}$$

where $\varphi^{-1}$ denotes the inverse of the cumulative standard normal distribution converting the hit rate and false alarm rate into Z scores. Note that a larger c value indicates a bias towards more frequently reporting "distorted," serving as an index for the presence of face distortion illusion.

We separately obtained subjective distortion ratings for each condition based on the responses to the third question on each trial ("Rate the level of distortion"). The responses to the second question ("How confident are you?") were analyzed in a separate manuscript for different research purposes (Gao et al., 2023) and are not examined here.

To examine whether distortion ratings are influenced by decisional biases, we conducted across-subject Pearson correlations between distortion ratings in the illusion and control conditions. To avoid contamination by the artificial distortions in the images, we limited these correlations to the ratings collected from the undistorted faces only. We performed these correlations for each experiment separately, as well as for the data combined across all three experiments. To explore how temporal frequency and the eccentricity of face presentation influence the face distortion illusion, we conducted two-tailed paired *t*-tests for response bias and distortion ratings for Experiments 1 and 2. These tests compared different eccentricities and temporal frequencies within both the illusion and control conditions, as well as the differences between the illusion and control conditions. For Experiment 3, we performed two-tailed paired t-tests on the response bias and distortion rating between the control condition and the illusion conditions with different temporal gaps. For all t-tests, we report Bayes factors computed using default priors (Krekelberg, 2022).

## 3. Results

We re-analyzed data from two previous experiments (Gao et al., 2023) and further collected data for a new experiment. Experiments 1 and 2 tested the effects of temporal frequency and eccentricity on FFDE, whereas Experiment 3 tested the effects of inserting temporal gaps between successive faces. Critically, in all three experiments, we collected traditional distortion ratings and also computed our novel objective metric of FFDE.

### *3.1. Subjective distortion ratings are subject to biases*

We first examined whether the subjective distortion ratings are subject to intrinsic bias. We reasoned that beyond the true level of distortion, subjective ratings may also reflect an overall propensity to give lower or higher distortion ratings. Further, such overall propensity for using low or high distortion ratings may vary from person to person. If this is indeed the case, one would expect to observe a positive correlation between the distortion ratings given to undistorted faces in the illusion condition (where true distortion exists) and undistorted faces in the control condition (which does not induce distortions). We checked for such correlation by first combining the data from all three experiments ($N = 81$) to increase power. We found a strong correlation between distortion ratings for undistorted faces in the illusion and control conditions ($r = 0.61$, $p = 1.9*10^{-9}$, $BF_{10} = 5.4*10^{6}$; Fig. 3). Further, the high correlation was present for each individual experiment (Experiment 1: r = 0.61, $p = .0007$, $BF_{10} = 43.33$; Experiment 2: $r = 0.65$, $p = .0003$, $BF_{10} = 3.86$; Experiment 3: r = 0.61, $p = .0008$, $BF_{10} = 39.27$). These findings demonstrate that subjective distortion ratings are inherently subject to intrinsic biases.

Previous work has used subjective distortion ratings to examine the effects of temporal frequency and eccentricity on FFDE, finding that both lower temporal frequency and larger eccentricity enhance the level of face distortion (Balas & Pearson, 2019). Here we examine whether these effects may be partially due to intrinsic biases too. We compared the effects of temporal frequency and eccentricity on the distortion ratings for undistorted faces in both the illusion condition (where one would expect differences) and the control condition (where one would not expect differences because there should be no distortion regardless of the stimulus manipulations) in Experiments 1 and 2.

Replicating the previous results (Balas & Pearson, 2019), we found that distortion ratings were significantly higher for more eccentric faces in the illusion condition (Experiment 1: t(26) = 4.30, $p = .0002$, Cohen's d = 0.83, $BF_{10} = 135.05$; Experiment 2: t(26) = 7.10, $p = .0000002$, Cohen's d = 1.37, $BF_{10} = 1.0*10^{5}$; Fig. 4A-B). However, the same difference in distortion ratings appeared in the control condition too (Experiment 1: t(26) = 4.56, $p = .0001$, Cohen's d = 0.88, $BF_{10} = 247.00$; Experiment 2: t(26) = 3.93, $p = .0006$, Cohen's d = 0.76, $BF_{10} = 56.53$). These results suggest that at least a part of the effect in the illusion condition stems from an intrinsic bias that is also observable in the control condition. The effect of eccentricity was significantly larger in the illusion than in the control condition in Experiment 2 (t(26) = 4.78, $p = .00006$, Cohen's d = 0.92, $BF_{10} = 424.34$), but this effect did not reach significance in Experiment 1 (t(26) = 1.59, $p = .13$, Cohen's d = 0.30, $BF_{01} = 1.64$). Our findings suggest that eccentricity modulates FFDE above and beyond the influence of intrinsic bias, but the true effect would be smaller if the intrinsic bias is accounted for.

We observed similar effects for temporal frequency too. As in previous work, we found that distortion ratings were significantly higher for slower temporal frequency in the illusion condition for both experiments (Experiment 1: t(26) = 5.94, $p = .000003$, Cohen's d = 1.14, $BF_{10} = 6.7*10^{3}$; Experiment 2: t(26) = 6.36, $p = .000001$, Cohen's d = 1.22, $BF_{10} = 1.8*10^{4}$; Fig. 4C-D). However, the same difference in distortion ratings again appeared in the control condition too (Experiment 1: t(26) = 2.48, $p = .02$, Cohen's d = 0.48, $BF_{10} = 2.58$; Experiment 2: t(26) = 5.17, $p = .00002$, Cohen's d = 1.00, $BF_{10} = 1.1*10^{3}$). As with eccentricity, these results suggest that at least a part of the effect in the illusion condition stems from an intrinsic bias that is also observable in the control condition. Nevertheless, the difference between the low and high temporal frequency stimuli was larger in the illusion than in the control condition (Experiment 1: t(26) = 4.39, $p = .0002$, Cohen's d = 0.84, $BF_{10} = 164.44$; Experiment 2: t(26) = 3.54, $p = .002$, Cohen's d = 0.68, $BF_{10} = 23.66$). As with eccentricity, our findings suggest that temporal

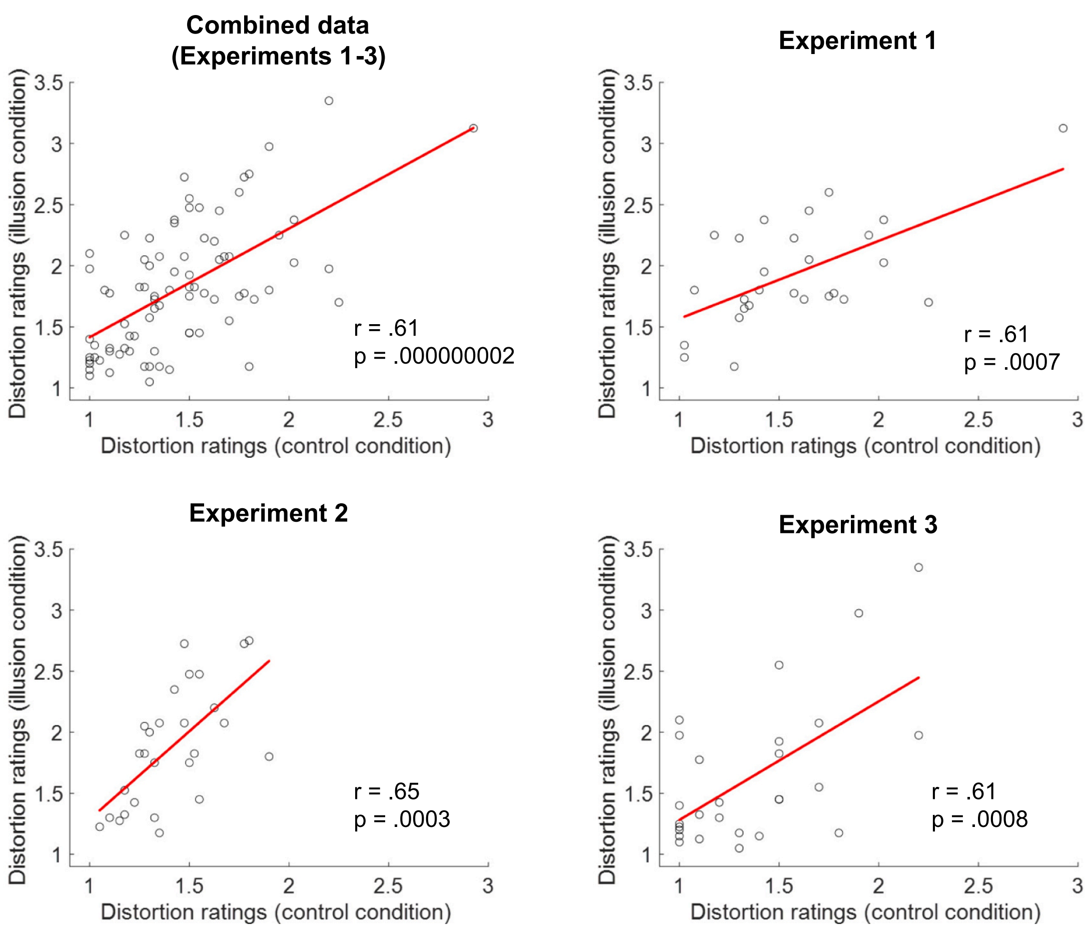


**Fig. 3.** Correlation of distortion ratings for undistorted faces in the illusion and control conditions. This figure illustrates the strong correlation between the perceived distortion ratings of undistorted faces in the illusion and control conditions across all three experiments, indicating a consistent pattern of subjective biases. The red line in each figure represents the linear regression model fitted to the data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

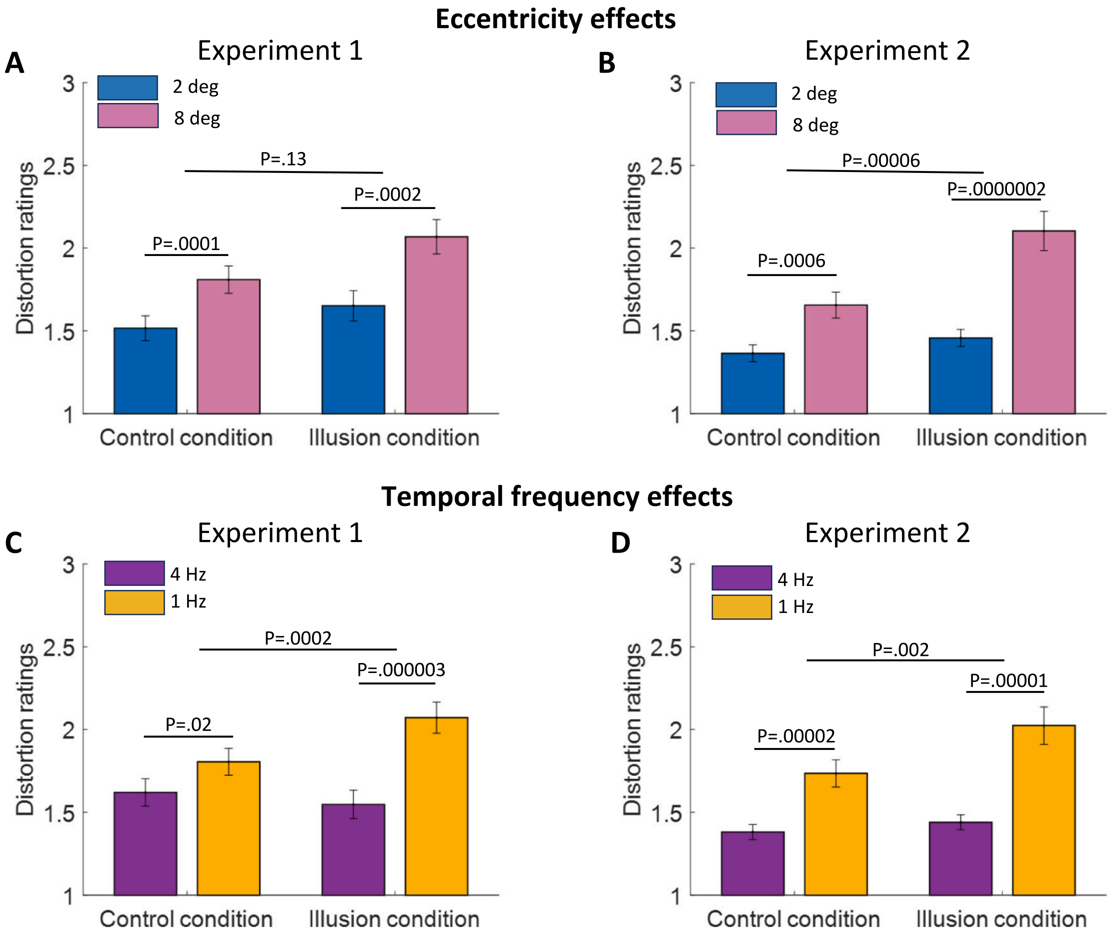


**Fig. 4.** Distortion ratings for undistorted faces in Experiments 1 and 2. (A,B) Effects of eccentricity on distortion ratings. Distortion ratings were higher for stimuli of larger eccentricity. However, this effect was present both in the illusion and the control conditions, suggesting that it was partly driven by intrinsic biases. Nevertheless, the effect of eccentricity was significantly larger in the illusion than in the control condition in Experiment 2, but this effect did not reach significance in Experiment 1. (C,D) Effects of temporal frequency on distortion ratings. Distortion ratings were higher for stimuli of lower temporal frequency. However, this effect was present both in the illusion and the control conditions, suggesting that it was partly driven by intrinsic biases. Nevertheless, the effect of temporal frequency was significantly larger in the illusion than in the control condition in both experiments. Error bars indicate SEM.

frequency does modulate FFDE above and beyond the influence of intrinsic bias. Nevertheless, the true effect would be smaller if the intrinsic bias is accounted for.

### 3.2. An objective method to quantify face distortion

Given the biases inherent in the standard subjective distortion ratings, here we introduce a novel, objective methodology for quantifying facial distortion. The goal of the method is to reduce the influence of subjective biases by giving subjects an objective task and explicitly accounting for any remaining biases by examining whether the effects in the illusion condition are larger than in the control condition. To validate our method, we apply it to the same data from Experiments 1 and 2 as above (this section) and to new data from Experiment 3 (next section).

We first checked whether our method produces similar response bias for different stimulus manipulations in the control condition. Ideally, one would see no significant differences given the fact that the control condition does not induce subjective distortion. Indeed, we found that response bias was matched for the two eccentricities in the control condition for both experiments (Experiment 1: t(26) = 0.26, $p$ = .80, Cohen's d = 0.05, $BF_{01}$ = 4.76; Experiment 2: t(26) = 0.50, $p$ = .62, Cohen's d = 0.10, $BF_{01}$ = 4.37, Fig. 5A-B). On the other hand, when examining temporal frequency, we found that response bias was matched in Experiment 1 (t(26) = 0.66, $p$ = .52, Cohen's d = 0.13, $BF_{01}$ = 4.03) but not in Experiment 2 (t(26) = 3.67, $p$ = .001, Cohen's d = 0.71, $BF_{10}$ = 31.19, Fig. 5C-D). In other words, asking subjects to complete an objective task seemed to reduce the intrinsic biases, though perhaps not eliminate them completely.

We also examined the effects of eccentricity and temporal frequency on FFDE using our new method by analyzing the data from the illusion condition. We found that face distortion was greater for more eccentric stimuli, though the effect was only significant for Experiment 2 (Experiment 1: t(26) = 1.62, $p$ = .12, Cohen's d = 0.65, $BF_{01}$ = 1.54; Experiment 2: t(26) = 2.94, $p$ = .007, Cohen's d = 0.57, $BF_{10}$ = 6.41, Fig. 5C-D). Face distortion was also greater for lower temporal frequency (Experiment 1: t(26) = 6.01, $p$ = .000002, Cohen's d = 1.16, $BF_{10} = 8.0*10^3$; Experiment 2: t(26) = 6.42, $p$ = .0000008, Cohen's d = 1.24, $BF_{10} = 2.1*10^4$), though only Experiment 1 showed a significant interaction (Experiment 1: t(26) = 3.40, $p$ = .002, Cohen's d = 0.65, $BF_{10}$ = 17.13; Experiment 2: t(26) = 1.73, $p$ = .10, Cohen's d = 0.33, $BF_{01}$ = 1.33). Overall, these results replicate previous conclusions (Balas & Pearson, 2019) that both eccentricity and temporal frequency modulate the strength of FFDE.

### 3.3. Our new method quantifies face distortion in the presence of temporal gap between successive faces

It has typically been assumed that FFDE only occurs when successive faces are presented without any gaps between them (Tangen et al., 2011). In Experiment 3, we directly tested this assumption and also

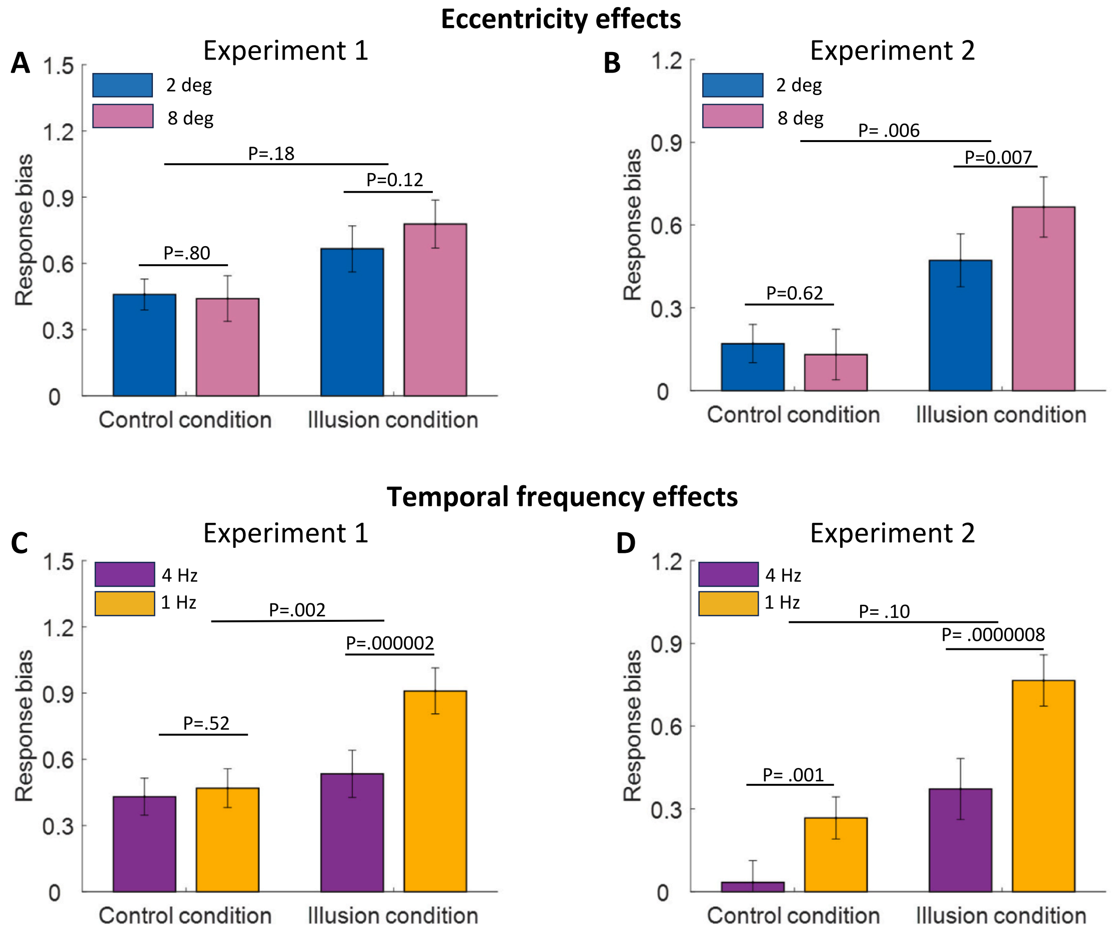


**Fig. 5.** Response bias in Experiments 1 and 2. (A,B) Effects of eccentricity on response bias. Response bias was matched in the control condition, suggesting that our method is robust to intrinsic biases. In the illusion condition, response bias was larger for more eccentric stimuli, though the effect (as well as the interaction between the control and illusion condition) was only significant for Experiment 2. (C,D) Effects of temporal frequency on response bias. In the control condition, response bias was matched in Experiment 1 but not in Experiment 2, suggesting that in the case of temporal frequency, our objective method mitigated intrinsic biases but without eliminating them completely. In the illusion condition, face distortion was greater for lower temporal frequency, though only Experiment 1 showed a significant interaction. Error bars indicate SEM.

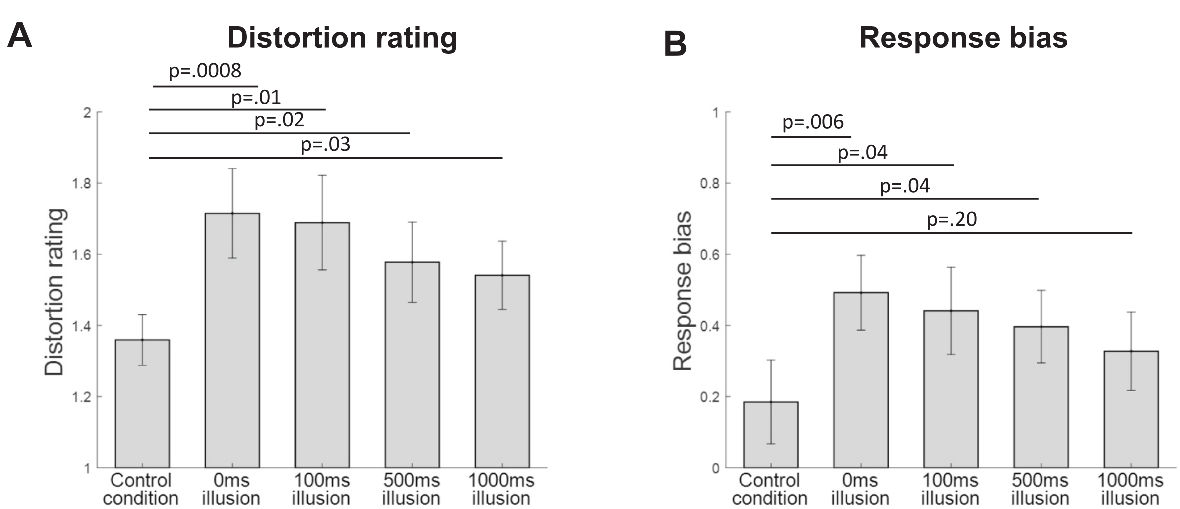


**Fig. 6.** Results for Experiment 3. (A) Results for subjective rating method applied to undistorted faces only. The distortion ratings were higher in all four illusion conditions compared to the control condition. (B) Results for our objective method. The response bias was higher in the 0-, 100-, and 500-ms illusion conditions compared to the control condition. *P*-values indicate the results of uncorrected *t*-tests. Error bars indicate SEM.

compared the sensitivity of the subjective ratings and our new method. We included four illusion conditions with varying temporal gaps between images: 0, 100, 500, and 1000 ms. In addition, we again included a control condition where a single pair of faces was presented.

We first examined the subjective ratings for undistorted faces only, and found significant differences between the control condition and all illusion conditions regardless of temporal gap (all four p's $< 0.03$, all $BF_{10} > 1.91$; Fig. 6A). After Bonferroni correction for four comparisons, the difference between the control condition and the illusion conditions with temporal gaps of 500 and 1000 ms were no longer significant (both p's $> 0.07$), while the difference between the control condition and illusion conditions with temporal gaps of 0 and 100 ms remained significant (both p's $< 0.04$).

Consistent with the subjective rating method, our new objective method found a clear difference between the control and the 0-ms, 100-, and 500-ms illusion conditions (0 ms: $t(26) = 2.99$, $p = .006$, Cohen's $d = 0.58$, $BF_{10} = 7.23$, 100 ms: $t(26) = 2.21$, $p = .04$, Cohen's $d = 0.42$, $BF_{10} = 1.60$, 500 ms: $t(26) = 2.14$, $p = .04$, Cohen's $d = 0.41$, $BF_{10} = 1.44$; Fig. 6B). After Bonferroni correction, only the difference between the control and the 0-ms condition remained significant ($p = .024$). The effects for 100 and 500 ms were modest and warrant replication in further studies. Face distortion was no longer significantly different from the control condition when the faces were presented with a relatively long temporal gap of 1000 ms ($t(26) = 1.32$, $p = .20$, Cohen's $d = 0.25$, $BF_{01} = 2.25$). Overall, these analyses suggest that our objective method successfully quantifies the effects of temporal gaps on face distortion.

## 4. Discussion

The goal of this study was to develop a method for objectively quantifying the strength of inherently subjective phenomenon of FFDE. We devised artificially distorted faces and asked subjects to distinguish between undistorted and objectively distorted faces. We implemented two conditions: an illusion condition, consisting of a sequence of 15 face flashes, and a control condition, involving a single face flash without inducing any illusion. Across three experiments, we varied factors such as eccentricity, temporal frequency, face orientation, and temporal gap between successive faces while measuring distortion ratings and response bias. Our analysis revealed that subjective distortion ratings are inherently susceptible to intrinsic biases, while our method exhibits reduced vulnerability to such biases.

Our results add to a long literature about the limitations of subjective ratings in not only behavioral sciences but also in diverse fields such as economics, health, sociology, and public management (Bollen & Paxton, 1998; Jahedi & Méndez, 2014; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Poulton, 1977). This literature has pointed out that subjective ratings often exhibit systematic measurement errors. Systematic biases can even sometimes lead to subjective ratings being uncorrelated or even negatively correlated with objective assessments (Cote & Buckley, 1987; Olken, 2009; Podsakoff et al., 2003).

In line with these previous findings, here we showed that the subjective rating method for face distortion is subject to several systematic biases. First, subjects may give low or high distortion ratings that are influenced not only by the true distortion strength but also by various decisional biases. In line with this possibility, we found that the distortion ratings in the illusion condition are strongly related to the distortion ratings in the control condition. Second, subjects may be inherently biased towards using low or high distortion ratings in different conditions, independent of the true level of distortion in each condition. In other words, when giving distortion ratings, they may also rely on extraneous information beyond the true distorted percept (Bollen & Paxton, 1998). In line with this possibility, we showed that the discrepancies across different eccentricities and temporal frequencies appear not only in the illusion condition, but also in the control condition (where no effect should exist). Together, these findings confirm that subjective ratings may suffer from multiple biases. Importantly, subjective ratings also reflect people's true experiences and the ratings' variability across subjects likely reflect the true differences in percepts. Our results do not challenge this notion but simply show that subjective ratings are not pure readout of subjective experience, and instead are subject to additional decisional biases (Rahnev & Denison, 2018).

The limitations of the subjective rating method underscore the necessity of objective measures in evaluating subjective phenomena. Objective tasks, such as detection or discrimination, can reduce or even eliminate the influence of personal biases. We refer to our method of quantifying FFDE as “objective” because subjects perform an objective task (identifying if distortion is physically present in the images), in contrast to “subjective” tasks where subjects rate their own experience. Nevertheless, we note that even objective tasks involve subjective responses, and thus our task is not objective in absolute terms.

Our study presents a promising approach for quantifying subjective phenomena such as visual illusions. Previous studies have applied objective detection and discrimination tasks to quantify subjective effects such as the Müller-Lyer illusion and visual aftereffects (Gao, Pieller, Webster, & Jiang, 2022; Lown, 1988; Webster, Kaping, Mizokami, & Duhamel, 2004). However, objective methods are relatively easier to design for these effects because these illusions can be manipulated in a graded fashion (e.g., by manipulating the true length of the lines in the Müller-Lyer illusion). On the other hand, objective methods are harder to devise for illusions that cannot be easily manipulated in a graded fashion, such as illusions where subjects see motion in stationary images (Fraser & Wilcox, 1979; Seno, Kitaoka, & Palmisano, 2013) or perceive a shape that does not exist as in the Kanizsa triangle (Kanizsa, 1976).

Our method depends on subjects performing a forced-choice task between two categories of images and the creation of objective images that mimic the subjective percept induced by the illusion of interest. However, creating objective images that perfectly align with subjective experiences is challenging, especially since the illusory percepts may vary based on many factors. In the case of FFDE, the illusory distortion likely changes based on the exact sequence of presentation as well as between observers. In our experiments, there was likely a significant, though not complete, overlap between the artificially induced distortions and those induced by the illusion. Our method requires a certain degree of overlap between these two types of distortions, but a complete overlap is not necessary. While every illusion is unique and will require a slightly different approach, our method can be generalized and applied to many other illusions that have not yet been studied using objective tasks.

One would expect that the closer the artificially distorted faces are to the illusory percepts people experience, subjects' ability to perform the objective task would diminish. In the extreme case, if the artificially distorted faces perfectly mimic the illusory percepts, subjects will be completely unable to distinguish between the undistorted and artificially distorted faces. We indeed found that subjects' ability to distinguish between undistorted and artificially distorted faces was lower in the illusion compared to the control condition (Supplementary Fig. 1). This sensitivity difference between the control and the illusion conditions could ostensibly serve as a metric of the strength of FFDE. However, the control condition may be inherently easier given that it features a single flash of faces, whereas the illusion condition is likely more difficult because the initial 14 flashes of faces potentially act as distractors. Therefore, quantifying FFDE by the difference in task sensitivity between the control and illusion condition would require additional control conditions to account for any task sensitivity difference that may exist between the single- and 15-flash conditions.

Applying our objective method, we found that FFDE increased with larger eccentricity and diminished with higher temporal frequencies of face presentation, consistent with previous findings that used the subjective rating method (Balas & Pearson, 2019). Beyond previous findings, we discovered that temporal gaps of 100 and 500 ms preserved the face distortion, whereas gaps of 1000 ms reduced it significantly. This finding illustrates the resilience of discontinuous presentation in eliciting face distortion, aligning with established findings on face aftereffects, where aftereffects significantly decrease with test durations of 1000 ms (Leopold, Rhodes, Mueller, & Jeffery, 2005; Rhodes, Jeffery, Clifford, & Leopold, 2007). Such results suggest that the FFDE may be indicative of a specific form of visual adaptation.

In summary, we develop an objective approach for quantifying subjective visual phenomena. The approach is robust to inherent biases unlike traditional subjective rating methods. Our method revealed new insights into the nature of face distortion and can be generalized to study a range of subjective phenomena.

## CRediT authorship contribution statement

**Yi Gao:** Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Minzhi Wang:** Data curation. **Dobromir Rahnev:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Formal analysis, Conceptualization.

## Declaration of competing interest

None.

## Data availability

I have shared the link to my data. The link is included in the manuscript.

## Acknowledgments

This work was supported by the National Institute of Health (award: R01MH119189) and the Office of Naval Research (award: N00014-20-1-2622). We thank Corey J. Breeland for his help with creating distorted faces.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.105861.

## References

Balas, B., & Pearson, H. (2019). The flashed face distortion effect does not depend on face-specific mechanisms. *Scientific Reports, 9*(1). https://doi.org/10.1038/s41598-018-37991-9
Bollen, K. A., & Paxton, P. (1998). Detection and determinants of Bias in subjective measures. *American Sociological Review, 63*(3), 465. https://doi.org/10.2307/2657559
Bowden, J., Whitaker, D., & Dunn, M. J. (2019). The role of peripheral vision in the flashed face distortion effect. *Perception, 48*(1), 93–101. https://doi.org/10.1177/0301006618817419
Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24*(3), 315. https://doi.org/10.2307/3151642
Dance, C. J., Ipser, A., & Simner, J. (2022). The prevalence of aphantasia (imagery weakness) in the general population. *Consciousness and Cognition, 97*, Article 103243. https://doi.org/10.1016/j.concog.2021.103243
Fraser, A., & Wilcox, K. J. (1979). Perception of illusory movement. *Nature, 281*(5732), 565–566. https://doi.org/10.1038/281565a0
Gao, Y. (2021). *Visual adaptation in central and peripheral vision* (Doctoral dissertation. Reno: University of Nevada.
Gao, Y., Chen, S., & Rahnev, D. (2023, October 7). Dynamic competition between sensory and decisional biases in perceptual decision making. *PsyArXiv*. https://doi.org/10.31234/osf.io/tbq97
Gao, Y., Pieller, J., Webster, M. A., & Jiang, F. (2022). Temporal dynamics of face adaptation. *Journal of Vision, 22*(11), 14. https://doi.org/10.1167/jov.22.11.14
Green, D. M., & Swets, J. A. (1966)*, 1. Signal detection theory and psychophysics* (pp. 1969–2012). John Wiley.
Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences, 106*(31), 13130–13135. https://doi.org/10.1073/pnas.0901352106
Jahedi, S., & Méndez, F. (2014). On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior & Organization, 98*, 97–114. https://doi.org/10.1016/j.jebo.2013.12.016
Kanizsa, G. (1976). Subjective Contours. *Scientific American, 234*(4), 48–52. https://doi.org/10.1038/scientificamerican0476-48
Krekelberg, B. (2022). *BayesFactor (2.3.0). Zenodo*. https://doi.org/10.5281/zenodo.7006300
Leopold, D. A., Rhodes, G., Mueller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Journal of Vision, 5*(8), 830. https://doi.org/10.1167/5.8.830
Lown, B. A. (1988). Quantification of the Müller-Lyer illusion using signal detection theory. *Perceptual and Motor Skills, 67*(1), 101–102. https://doi.org/10.2466/pms.1988.67.1.101
Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska directed emotional faces (KDEF). In *CD ROM from Department of Clinical Neuroscience*. Psychology section, Karolinska Institutet.
Manning, C., Morgan, M. J., Allen, C. T. W., & Pellicano, E. (2017). Susceptibility to Ebbinghaus and Müller-Lyer illusions in autistic children: A comparison of three different methods. *Molecular Autism, 8*(1), 16. https://doi.org/10.1186/s13229-017-0127-y
Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x
Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery, 19*(3–4), 153–167.
Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *Journal of Public Economics, 93*(7–8), 950–964. https://doi.org/10.1016/j.jpubeco.2009.03.001
Plewan, T., Weidner, R., Eickhoff, S. B., & Fink, G. R. (2012). Ventral and dorsal stream interactions during the perception of the Müller-Lyer illusion: Evidence derived from fMRI and dynamic causal modeling. *Journal of Cognitive Neuroscience, 24*(10), 2015–2029. https://doi.org/10.1162/jocn_a_00258
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879
Poulton, E. C. (1977). Quantitative subjective assessments are almost always biased, sometimes completely misleading*. *British Journal of Psychology, 68*(4), 409–425. https://doi.org/10.1111/j.2044-8295.1977.tb01607.x
Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences, 41*. https://doi.org/10.1017/S0140525X18000936
Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., … Zylberberg, A. (2020). The confidence database. *Nature Human Behaviour, 4*(3), 317–325. https://doi.org/10.1038/s41562-019-0813-1
Rhodes, G., Jeffery, L., Clifford, C. W. G., & Leopold, D. A. (2007). The timecourse of higher-level face aftereffects. *Vision Research, 47*(17), 2291–2296. https://doi.org/10.1016/j.visres.2007.05.012
Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science, 2*(1), 437–457. https://doi.org/10.1146/annurev-vision-082114-035733
Seno, T., Kitaoka, A., & Palmisano, S. (2013). Vection induced by illusory motion in a stationary image. *Perception, 42*(9), 1001–1005. https://doi.org/10.1068/p7511
Tangen, J. M., Murphy, S. C., & Thompson, M. B. (2011). Flashed face distortion effect: Grotesque faces from relative spaces. *Perception, 40*(5), 628–630. https://doi.org/10.1068/p6968
Tudusciuc, O., & Nieder, A. (2010). Comparison of length judgments and the Müller-Lyer illusion in monkeys and humans. *Experimental Brain Research, 207*(3–4), 221–231. https://doi.org/10.1007/s00221-010-2452-7
Uher, J. (2018). Quantitative data from rating scales: An epistemological and methodological enquiry. *Frontiers in Psychology, 9*(DEC). https://doi.org/10.3389/fpsyg.2018.02599
Uher, J. (2019). Data generation methods across the empirical sciences: Differences in the study phenomena's accessibility and the processes of data encoding. *Quality and Quantity, 53*(1), 221–246. https://doi.org/10.1007/s11135-018-0744-3
Utz, S., & Carbon, C.-C. (2015). Is the flashed face distortion effect expertise-based? - a systematic experimental investigation. *Journal of Vision, 15*(147).
Webster, M., & Maclin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review, 6*(4), 647–653. https://doi.org/10.3758/bf03212974
Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*(6982), 557–561. https://doi.org/10.1038/nature02420
Weidner, R., Boers, F., Mathiak, K., Dammers, J., & Fink, G. R. (2010). The temporal dynamics of the Müller-Lyer illusion. *Cerebral Cortex, 20*(7), 1586–1595. https://doi.org/10.1093/cercor/bhp217
Wen, T., & Kung, C. C. (2014). Using functional magnetic resonance imaging to explore the flashed face distortion effect. *Journal of Vision, 14*(12). https://doi.org/10.1167/14.12.29
Yamashita, J. A., Hardy, J. L., De Valois, K. K., & Webster, M. A. (2005). Stimulus selectivity of figural aftereffects for faces. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 420–437. https://doi.org/10.1037/0096-1523.31.3.420
Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery – Congenital aphantasia. *Cortex, 73*, 378–380. https://doi.org/10.1016/j.cortex.2015.05.019
Zeman, A., Milton, F., Della Sala, S., Dewar, M., Frayling, T., Gaddum, J., Hattersley, A., Heuerman-Williamson, B., Jones, K., MacKisack, M., & Winlove, C. (2020). Phantasia–The psychological significance of lifelong visual imagery vividness extremes. *Cortex, 130*, 426–440. https://doi.org/10.1016/j.cortex.2020.04.003