Routledge
Taylor & Francis Group

Volume 31 – Issue 9 – October 2023
Visual Cognition

Visual Cognition

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pvis20

Negative aftereffects of face trait impressions are modulated by emotional expressions

Fiammetta Marini, Clare A. M. Sutherland, Linda Jeffery, Sarah D. Maisey & Mauro Manassi

To cite this article: Fiammetta Marini, Clare A. M. Sutherland, Linda Jeffery, Sarah D. Maisey & Mauro Manassi (01 Oct 2024): Negative aftereffects of face trait impressions are modulated by emotional expressions, Visual Cognition, DOI: 10.1080/13506285.2024.2407873

To link to this article: https://doi.org/10.1080/13506285.2024.2407873

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

View supplementary material

Published online: 01 Oct 2024.

Submit your article to this journal

Article views: 51

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=pvis20

VISUAL COGNITION
https://doi.org/10.1080/13506285.2024.2407873

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Negative aftereffects of face trait impressions are modulated by emotional expressions

Fiammetta Marini [a], Clare A. M. Sutherland [a,b], Linda Jeffery [b,c], Sarah D. Maisey [b] and Mauro Manassi [a]

[a]School of Psychology, University of Aberdeen, King's College, Aberdeen, UK; [b]School of Psychological Science, University of Western Australia, Crawley, Western Australia, Australia; [c]School of Population Health, Curtin University, Bentley, Australia

ABSTRACT

Facial trustworthiness impressions critically shape our everyday social interactions. While previous research has predominantly considered trustworthiness impressions to be stable over time, preliminary evidence has shown that they are affected by visual adaptation, such that long exposure to (un)trustworthy-looking faces biases the perception of following faces in the opposite trustworthiness direction. Here, by employing a visual adaptation task across two experiments, we sought further evidence that trustworthiness impressions are shaped by the temporal context. In Experiment 1, we investigated whether visual adaptation affect trustworthiness judgements and found evidence of robust negative face aftereffect. In Experiment 2, we focused our investigation on whether emotional expressions, key cues involved in trait impressions, influence trustworthiness and dominance impressions. We found that adaptation to anti-expressions, which were expected to bias subsequent neutral faces to resemble the original expression (happiness, anger, and fear), significantly modulated subsequent evaluations of trustworthiness and dominance. This result confirms the critical role of emotion perception in trait evaluations. Importantly, using anti-expressions minimised semantic adaptation, thus highlighting the perceptual nature of this aftereffect. Taken together, our findings confirm that temporal context shapes trustworthiness impressions, by showing that visual adaptation affects trust judgements, and that past emotional expressions influence following impressions of trustworthiness and dominance.

ARTICLE HISTORY
Received 1 May 2024
Accepted 16 September 2024

KEYWORDS
Visual adaptation; trustworthiness; face emotional expression; dominance; negative aftereffect

Faces are a rich source of information exploited to form sophisticated expectations about others' personality traits along many social dimensions, including trustworthiness (Todorov et al., 2015). Judgements of trustworthiness are one of the most important dimensions at the core of face evaluation (Lin et al., 2021; Oosterhof & Todorov, 2008; Sutherland et al., 2013), rapidly formed even after 33 ms of exposure to a new face (Todorov et al., 2009; Willis & Todorov, 2006). While there is little evidence that supports the accuracy of these impressions, people tend to broadly agree on how trustworthy a face looks (Foo et al., 2021; Korva et al., 2013). Importantly, these critical evaluations influence various socio-cognitive processes, including attention processes (Todorov et al., 2009; Willis & Todorov, 2006) and memory processes (Giraudier et al., 2022; Wendt et al., 2019; Weymar et al., 2019). As a consequence, trustworthiness impressions strongly influence people's social behaviour (Olivola et al., 2014) in a pervasive way in many aspects of society. For example, their consequences extend to legal contexts (Flowe, 2012; Wilson & Rule, 2015), employment settings (Linke et al., 2016; Rule & Ambady, 2008), and romantic relationships (Appel et al., 2023; South Palomares & Young, 2017; Valentine et al., 2020).

Although the majority of research has focused on the facial properties that make a face appear trustworthy (Sutherland & Young, 2022; Todorov et al., 2015), previous research has devoted considerably less attention to understanding the ways in which the spatio-temporal context we are immersed

CONTACT Fiammetta Marini f.marini.21@abdn.ac.uk School of Psychology, University of Aberdeen, King's College, Aberdeen, AB24 3FX, UK
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13506285.2024.2407873.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

everyday influences face perception of social traits. In general, perception of many face attributes, such as identity and emotional expressions, are dynamically influenced and distorted by the spatial (Carragher et al., 2021; Haberman & Whitney, 2009; Marini et al., 2023) and temporal context in which faces are embedded. In the temporal domain, research has shown that face perception is modulated by visual adaptation for a variety of face attributes, including identity (Afraz & Cavanagh, 2008; Leopold et al., 2001; Rhodes et al., 2015), gender (Afraz & Cavanagh, 2009; Webster et al., 2004) and emotional expression (Burton et al., 2016; Fox & Barton, 2007; Skinner & Benton, 2010; Webster & Macleod, 2011). Adaptation occurs when long exposure to a stimulus results in a negative aftereffect that biases the perception of a subsequent stimulus away from the characteristics of the previously seen one (Burton et al., 2016; Clifford & Rhodes, 2005; Rhodes et al., 2007; Webster & Macleod, 2011). A classic example of visual adaptation is the motion aftereffect demonstrated by the "waterfall illusion", where prolonged exposure to downward motion biases the perception of subsequently viewed static stimuli to appear as if moving upward (Anstis et al., 1998). Visual adaptation was proposed to play a significant role in contributing to efficient coding and discrimination of stimuli (Anstis et al., 1998; Bosten et al., 2022a; Rhodes & Leopold, 2011; Webster & Macleod, 2011). After long exposure to a stimulus, our visual system adapts to the prevailing inputs in the visual environment and flexibly "recalibrates" with a shift in perception, increasing sensitivity to changes in the scene (Kohn, 2007).

Since trustworthiness impressions rely on visual cues associated with identity, gender, and emotional expressions (Oosterhof & Todorov, 2009; Sutherland et al., 2015), which are affected by visual adaptation (Burton et al., 2016; Rhodes & Jeffery, 2006; Swe et al., 2019; Yang et al., 2011), it is reasonable to expect that visual adaptation would, in turn, affect trustworthiness impressions. However, surprisingly only three studies to date have investigated visual adaptation to facial trustworthiness (Engell et al., 2010; Keefe et al., 2013; Wincenciak et al., 2013). Keefe et al. (2013) examined the impact of adaptation to different levels of facial trustworthiness on sensitivity to trustworthiness discrimination and found that adaptation to a face improved observers' sensitivity only when assessing subsequent test faces with a similar trustworthiness level. However, this study did not directly test for trustworthiness negative aftereffect. Wincenciak et al. (2013) did directly test for negative trustworthiness aftereffects by studying the influence of prior exposure to trustworthy and untrustworthy faces on the perception of subsequent test faces, and found evidence consistent with classic visual adaptation for trustworthiness. Intriguingly, gender effects were found, where female but not male observers exhibited a trustworthiness aftereffect. However, issues with the levels of trustworthiness of the face stimuli displayed (as the authors acknowledged), may have resulted in a weaker aftereffect, which eventually could have contributed to the observed gender differences. Although these two studies provide the first tentative evidence that observers can show adaptation effects for trustworthiness impressions, the mixed results and methodological limitations present make trustworthiness negative aftereffect still not as clearly defined and well-replicated as other face attributes.

Importantly, faces in everyday life are often preceded by many other facial cues that could influence trustworthiness impressions, such as emotional expressions, one of the key underlying cues of trustworthiness impressions (Oosterhof & Todorov, 2008). In this respect, the emotion overgeneralization hypothesis argues that people infer personality traits from subtle resemblance to emotional expressions in the structural aspects of individual faces (Montepare & Dobish, 2003; Olivola et al., 2014; Zebrowitz, 1997). According to this hypothesis, faces with a neutral expression that subtly resemble a positive emotional expression, such as with high surprised-looking eyebrows and an upward shaped mouth, are perceived as trustworthy, while those resembling negative emotions, for example with lower eyebrows, are seen as untrustworthy (Flowe, 2012; Oosterhof & Todorov, 2008; Todorov et al., 2015). Accordingly, Engell et al. (2010) showed that visual adaptation to facial emotion expressions can bias the perception of facial trustworthiness, suggesting shared common mechanisms engaged during the perception of happy and angry expressions and trustworthiness. However, observers in the Engell et al. (2010) study adapted to easily recognizable emotional expression, which carry substantial semantic content, and this could have

potentially induced a semantic aftereffect. Indeed, when the adapting stimuli have a semantic connection to the test stimuli, aftereffects may also arise from the semantic meaning of the adapting stimuli (Hills et al., 2010; Javadi et al., 2012; Ghuman et al., 2010). For example, according to Javadi et al. (2012), adapting to images associated with femininity, such as lipstick and earrings, can bias the perception of gender ambiguous faces to appear as masculine, showing that negative aftereffect can be driven by a semantic representation, not just by a perceptual one.

Although these three studies constitute important first steps in the investigation of the influence of visual adaptation on trustworthiness impressions, they provide limited and somewhat inconclusive evidence. Here, through two studies with complementary paradigms we considered again the ways in which the temporal context influences face trustworthiness impressions. In Experiment 1, we investigated whether observers adapt to trustworthiness cues and exhibit trustworthiness negative aftereffect through an experimental paradigm typical of visual adaptation studies (Burton et al., 2016; Leopold et al., 2001; Webster & MacLin, 1999), in order to replicate Wincenciak et al. (2013)'s findings (and also overcome some of their experimental limitations).

In Experiment 2, we investigated the importance of subtle visual cues to emotion as contributors to impressions of trustworthiness and gather further evidence of how impressions of facial trustworthiness might be influenced by preceding faces in the temporal context. Specifically, we tested whether emotional expressions in the past might bias the perception of trustworthiness in subsequently seen faces. Experiment 2 intended to, first, replicate Engell et al. (2010)'s findings on a common mechanism of emotion perception and trustworthiness impressions, and, second, to minimise the contribution of semantic adaptation by using anti-expressions (i.e., stimuli that elicit an expression aftereffect in the direction of the original expression; Ghuman et al., 2010; Javadi et al., 2012). Third, Experiment 2 aimed to expand the investigation to dominance judgements as a comparison, given that they represent another important face judgement potentially sharing similar emotional perceptual underlying cues (Witham et al., 2021) and they are considered orthogonal to trustworthiness impressions (Oosterhof & Todorov, 2008), thus different results after adaptation to anti-expressions would indicate a difference in the specific emotions involved in each face impression.

Experiment 1: Negative aftereffect in face trustworthiness perception

Experiment 1 investigates whether visual adaptation influences trustworthiness impressions. For this purpose, we assessed whether adaptation to (un)trustworthy-looking computer-generated faces biases the perception of trustworthiness in subsequent neutral faces. Importantly, this experiment intended to replicate Wincenciak et al. (2013)'s findings, while also addressing limitations of previous studies with three main experimental improvements. First, we used FaceGen computer-generated faces (Todorov et al., 2013a), which are meticulously designed to vary primarily in trustworthiness appearance, thereby increasing confidence that any effects reflect adaptation of visual representations of trustworthiness. Second, to better quantify the absolute adaptation effect, we included a "no adaptation" control condition which was not expected to influence subsequent face appearance, for a baseline comparison. Third, we inserted an attention check task to encourage consistent attention to the adapting faces, given that attention can modulate the strength of face aftereffects (Rhodes et al., 2011a). Though trustworthiness impressions may be rapidly and automatically formed, suggesting attention may not be critical to forming trust impressions (Lischke et al., 2017; Swe et al., 2022; Willis & Todorov, 2006) it is nevertheless possible that adaptation effects could be influenced by attention. We predicted that if visual adaptation occurs for facial trustworthiness impressions, long exposure to an untrustworthy face would make the subsequent test face appear more trustworthy. Conversely, adaptation to a trustworthy looking face would result in the subsequent test face being perceived as less trustworthy.

Methods

Participants

A total of fifty-three participants were recruited via multiple recruitment channels, including the SONA recruitment system, social media advertisements, and word-of-mouth. Undergraduate students

received course credits for their participation. The study was approved by the Ethics Committee of the School of Psychology of the University of Aberdeen (UK) and participants gave informed consent at the beginning of the study.

The final sample included fifty-two participants (34 females, 15 males, 3 others; M = 24.3 years, s.d. = 5.2 years). As exclusion criteria, we removed from the analysis participants that had an error rate greater than 30% in an attentional control task (Asterisk task; see task description in Procedure section), suggesting a lack of attention on the adaptor (one participant removed). Additionally, we planned to exclude from the analysis any participants who reported problems or distractions during the experiment or misunderstanding of the instructions (asked at the end of the experiment), but none reported any issue. An a-priori power run on G*Power analysis guided the determination of the sample size, indicating that 52 participants would yield adequate statistical power (80%) to detect an effect size of .40 at a standard alpha error probability of .05 in a one sample t-test. This power calculation was based on the effect size observed in previous studies that utilised a similar paradigm to investigate facial perception, including trustworthiness impressions (Burton et al., 2016: d = .40; Wincenciak et al., 2013: $\eta_p^2 = .13$). The study was pre-registered on https://osf.io/2xb8k/.

Apparatus and stimuli

The experiment was programmed and run using PsychoPy (https://psychopy.org/, Peirce et al., 2019), on a 21" Sony TRINITRON CPD-G500 monitor. Stimuli were viewed from 50 cm with the head placed in a chin rest.

The adaptor and test stimuli consisted of computer-generated emotionally neutral faces that varied in their trustworthiness levels. These face images were obtained from the FaceGen computer-generated images dataset, and they were specifically designed to have both trustworthy and untrustworthy appearance (Todorov et al., 2013a). From this dataset, we selected three facial identities, each of which had a trustworthy and untrustworthy looking version. We controlled for low-level features by using the SHINE toolbox (Willenbockel et al., 2010) in MATLAB R2017b (The MathWorks, USA). Specifically, this process involved converting the images to grayscale and adjusting their luminance and contrast to match the average values of all images in the dataset. Given that previous research suggested that gender interacts with trustworthiness (Sutherland et al., 2015), only male-looking faces were selected. For the same reason, the trustworthy-looking versions of each identity were set at +1 standard deviation on the trustworthiness dimension, while the untrustworthy-looking morphed faces were set at −3 standard deviations on the trustworthiness dimension. This asymmetric approach was implemented to prevent the faces from becoming increasingly androgynous and feminine as trustworthiness levels increased, as a very trustworthy appearance can make a male face appear more feminine (Marini et al., 2023; Oliveira et al., 2020a).

A face morph continuum of five grayscale images was created between the initial trustworthy and untrustworthy versions of each identity by using Psychomorph software (Sutherland et al., 2017; Tiddeman et al., 2001). The continuum was composed of six different trustworthiness strengths (0%, 25%, 50%, 75%, 100%), ranging from the least trustworthy (0% trustworthy, labelled as 1) and the most trustworthy (100% trustworthy, labelled as 5) morphed face (Figure 1A). All stimuli were presented on a grey background.

The adapting stimuli, which were the untrustworthy or trustworthy extreme faces of the morph continuum, were displayed at size of approximately 12.68° in height and 9.7° in width. The test face stimuli, which were the trustworthy neutral faces in the middle of the morph continuum (50% morph), were presented in the same location but at 80% of the size of the adaptors to minimize the influence of retinotopic adaptation between adaptors and test faces.

Procedure

On each trial, an adaptor face was displayed for 8000 ms. In one-third of the trials (see Figure 1B for the trial sequence), the adaptor face was very untrustworthy looking (morph 0% trustworthy), while in one-third of the trials the adaptor face was very trustworthy looking (morph 100% trustworthy). The identity of the face was randomly selected between the three possible identities and remained the same between

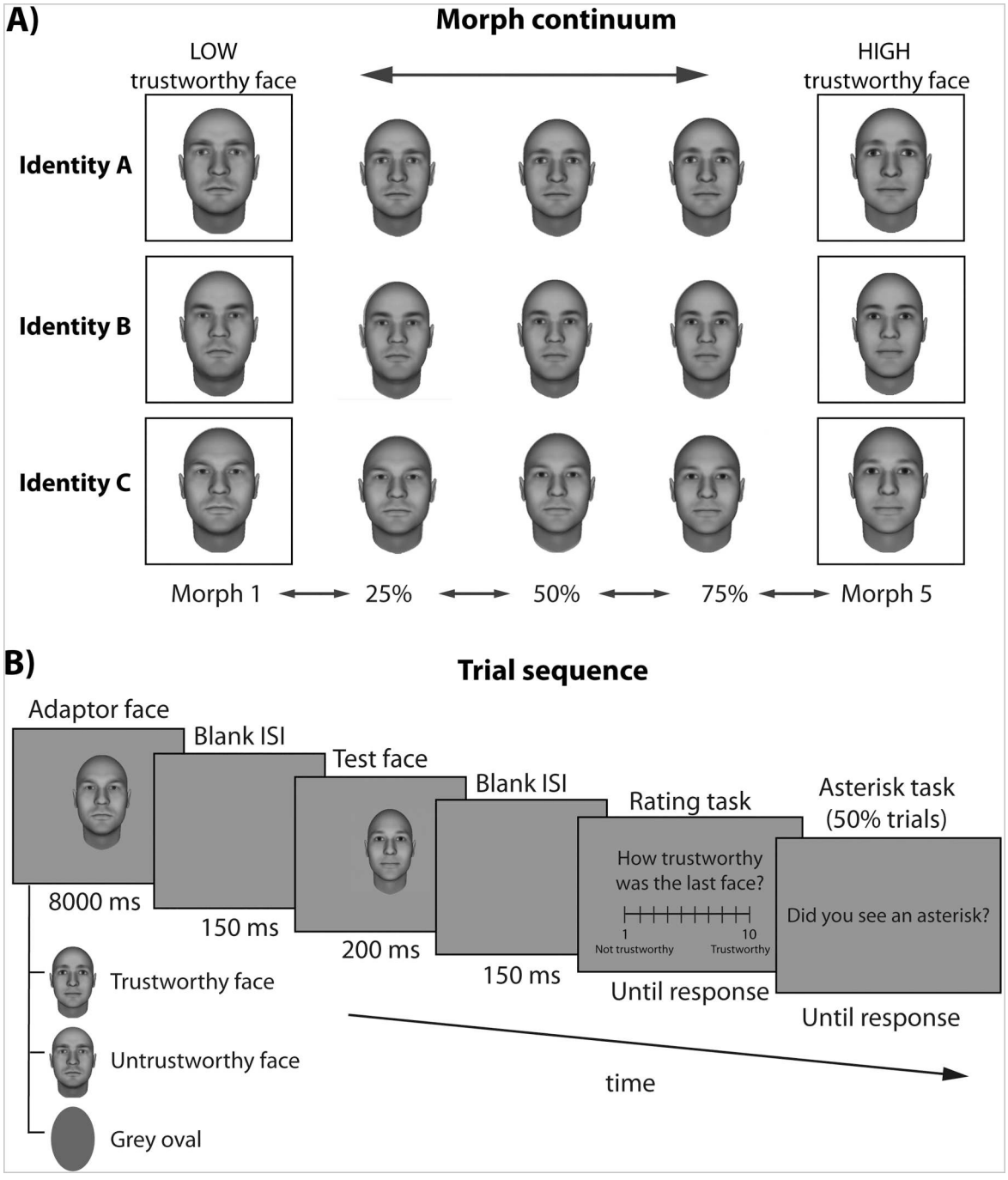

Figure 1. (A) Stimuli. Three pairs of computer-generated FaceGen images with extreme trustworthiness levels (Todorov et al., 2013a) were used to create a morph continuum of 5 faces ranging from the least trustworthy morphed face (labelled as 1) and the most trustworthy morphed face (labelled as 5). (B) Trial sequence of Experiment 1. On each trial participants saw an adaptor face (8000 ms) followed by a 150 ms ISI and a test face (200 ms). Following an 150 ms ISI, participants were asked to indicate on a rating scale from 1 to 10 the level of trustworthiness of the test face and whether they saw or not an asterisk during the adaptor face exposure.

adaptor and test face. For the remaining one-third of the trials, a grey oval replaced the face on the screen. This served as a control "no adaptation" condition since adapting to the oval was not expected to influence subsequent face perception.

Following a 150 ms inter-stimulus-interval (ISI), a test face was presented for 200 ms. In two-thirds of the trials, the test face was neutral in trustworthiness appearance, displaying neither untrustworthiness nor trustworthiness (morph 50% trustworthy). In the

remaining one-third of the trials, the test face was a slightly untrustworthy (morph 25% trustworthy) or slightly trustworthy (morph 75% trustworthy) face. We included these trials to maintain participant motivation, as continuously rating the same three test faces may have led to reduced attention, but they were not included in the analysis. After a 150 ms ISI following the test face presentation, participants were instructed to rate the perceived trustworthiness of the test face by clicking on a Likert scale ranging from 1 (not trustworthy at all) to 10 (extremely trustworthy).

The durations of both the adaptor (8000 ms) and test face stimuli (200 ms) were selected based on previous research (Burton et al., 2016; Witham et al., 2021) which showed these timings produced robust face aftereffects. Trials were presented in random order to prevent adaptation to any specific adaptor condition. To ensure that participants were attentive to the adapting stimuli throughout the entire exposure duration, we also included an attention check. Specifically, in 50% of the trials, an asterisk appeared for 150 ms on the adaptor face at 1850 ms, 3850 ms, or 5850 ms into the adaptation period. Next, at the end of each trial participants were asked to indicate whether they saw the asterisk by clicking the left (yes) or right (no) arrow key. Participants whose error rate exceeded 30% were excluded from the analysis as it indicated a lack of attention to the adapting stimuli. Three blocks of 18 trials were presented, resulting in 54 trials in total per observer. Among these, thirty-six trials had the average morph face as the test face, with twelve trials allocated per adaptor condition (untrustworthy, trustworthy, grey-oval).

Before starting the experiment, participants were familiarised with the stimuli and the task. First, they completed 10 practice trials in which they saw a randomly selected face from the continuum for 400 ms and were instructed to indicate the level of trustworthiness of the face on a 1–10 rating scale. Second, they completed a training block of 10 example trials. The experiment took approximately 30 minutes to complete. At the end of the experiment, participants were required to indicate whether their data should be excluded for any reason (e.g., experienced problems, misunderstanding of the task instructions) by typing their response (none reported any issues).

Results and discussion

Adaptation impact scores

To quantify the difference in perception of trustworthiness following adaptation to untrustworthy or trustworthy faces versus adaptation to a grey oval, we calculated individual adaptation impact scores. These scores were obtained by subtracting the average trustworthiness ratings in the grey-oval adaptation condition from the average trustworthiness ratings in the other two adaptor conditions (untrustworthy, trustworthy) respectively, for each participant (see Supplementary Materials for the raw average scores for each adaptation condition). First, to see the general adaptation impact on participants' ratings, we flipped the sign of the adaptation impact scores in the "trustworthy adaptor" condition in order to ignore the direction of the effect, and then we averaged together the adaptation impact scores for both the trustworthy and untrustworthy adaptor conditions (Figure 2B). Positive values indicate a negative repulsive bias away from the previously seen trustworthiness level (i.e., negative aftereffect), zero value indicates no bias, and negative values indicate a positive attractive bias toward the past (i.e., positive serial dependence; Manassi et al., 2023; Manassi & Whitney, 2022; Marini et al., 2024). A stronger aftereffect is indicated by a greater deviation from zero. As predicted, almost all participants showed a negative aftereffect (Mean: $1 \pm$ s.e: 0.1). Moreover, a one-sample t-tests against zero showed that the trend was significantly different from zero ($t(51) = 14.03$, $p < .001$, $d = 1.94$). Second, we investigated the difference in the impact of adaptation between the trustworthy and untrustworthy adaptation conditions. For this analysis, the adaptation impact scores for each adaptor condition were calculated (Figure 2A). Here positive values indicate higher trustworthiness ratings for test faces following adaptation to the adaptor face compared to the grey-oval, while negative values indicated lower trustworthiness ratings compared to the grey-oval. On average, participants showed a negative adaptation impact score in the trust adaptor face condition ($-0.5 \pm$ s.e: 0.1), whereas a positive adaptation impact score was found in the untrust adaptor face condition (Mean: $1.4 \pm$ s.e: 0.1). These results suggest that the test face was perceived as more untrustworthy after adaptation to trustworthy looking

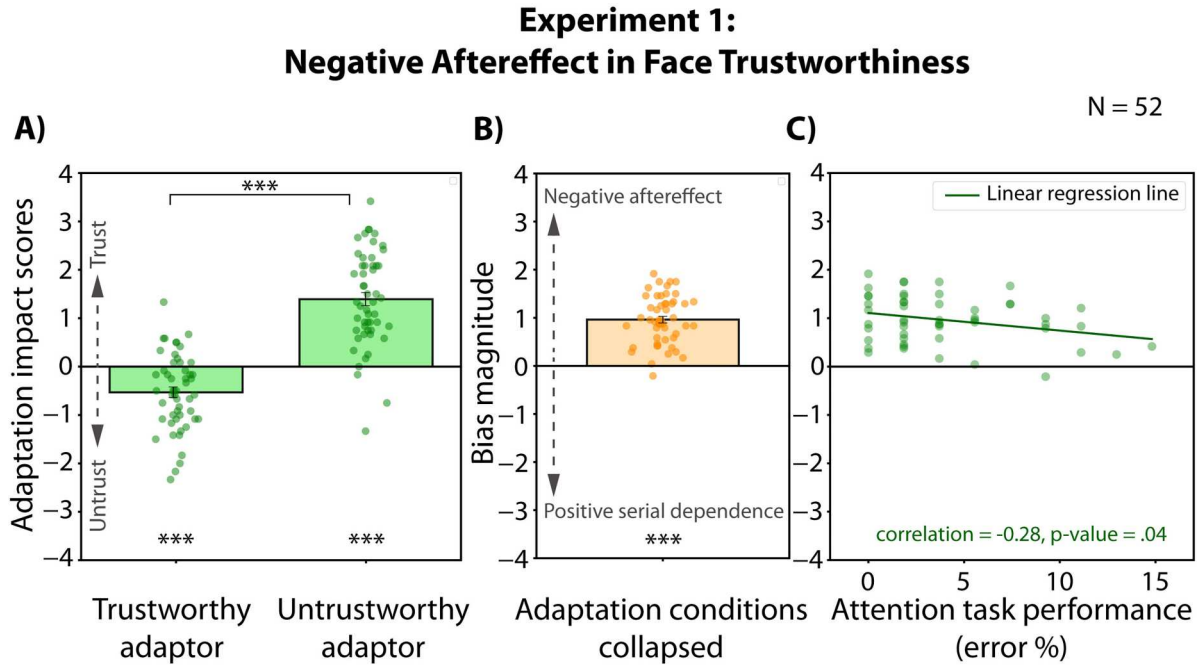

Figure 2. Experiment 1 results. (A) For each participant, the average trustworthiness ratings for the control (grey oval) adaptor condition were subtracted from the average trustworthiness ratings for the other two adaptor conditions (untrustworthy, trustworthy). The group average aftereffect trustworthiness score is plotted on the y-axis. The adaptation conditions are plotted on the x-axis. Each dot represents one participant. Error bars represent the standard error of the mean. Below each bar is indicated the aftereffect impact scores significance level of difference from zero ($p < .05$ showed as *, $p < .001$ showed as ***, $p < .01$ showed as **) for each adaptation condition. Moreover, over the bars the comparison in strength of aftereffect across the different adaptation conditions is illustrated. (B) This bar graph shows for each participant the two adaptor conditions aftereffect impact scores collapsed together, indicating the negative aftereffect bias magnitude. Error bars represent the standard error of the mean. The aftereffect impact scores were significantly different from zero ($p < .001$, ***). (C) The collapsed aftereffect impact scores were correlated with the percentage of errors in the Asterisk attention task. The group average collapsed is plotted on the y-axis, whereas the performance in the attention task is plotted on the x-axis. A linear regression line is plotted in green on the data.

adaptors and more trustworthy after adaptation to untrustworthy looking adaptors.

To establish whether there was a significant aftereffect for each adaptor condition separately, we ran two one-sample t-tests against zero. Both the trustworthy adaptation condition ($t(51) = -4.93$, $p < .001$, $d = 0.33$) and the untrustworthy adaptation condition ($t(51) = 10.24$, $p < .001$, $d = -0.68$) adaptation impact scores were significantly different from zero, indicating that the adaptors biased trustworthiness ratings in the predicted directions (i.e., away from the adaptors).

Finally, to test whether the absolute strength of the adaptation effect was different across the two adaptation conditions (ignoring the direction of the effect), a pairwise t-test was conducted by taking into consideration the collapsed adaptation impact score across both untrustworthy and trustworthy adaptor conditions displayed in Figure 2B. This test showed a significantly stronger aftereffect after adaptation to untrustworthy compared to trustworthy faces ($t(51) = -4.24$, $p < .001$). We will return to this finding in the General Discussion.

Attention check analysis

For each observer, we calculated the percentage of error rate for the Asterisk task via the number of incorrect responses as an attention check (Figure 2C). All the participants had an error rate lower than 30% in the Asterisk task, except for one participant, which was removed from the analysis. This suggests participants were consistently attending to the adapting faces. However, though overall performance was good, there was some variation among participants on this task. Given this variation and evidence that greater attention leads to stronger aftereffects for adaptation to other face attributes, we investigated the correlation between each participant's adaptation impact score and their error rate in the Asterisk task.

For each participant, we averaged the absolute adaptation impact scores across both untrustworthy and trustworthy adaptor conditions. We found a significant negative correlation ($r = -0.28$, $p = .04$) between the collapsed adaptation impact score and the error percentage in the Asterisk task, suggesting a potential link between attention and negative aftereffect strength (see General Discussion).

In Experiment 1, our results showed that long exposure to untrustworthy or trustworthy looking faces biased the perception of trustworthiness in subsequent faces, consistent with a visual adaptation pattern. Therefore, our study successfully replicated the findings of Wincenciak et al. (2013) and addressed limitations observed in prior research by using a different set of stimuli and a different paradigm. By using computer generated FaceGen images created to vary in their level of trustworthiness we highly controlled trustworthiness cues in the face stimuli. Moreover, by adding a "no adaptation" control condition as a baseline comparison to better quantify the absolute adaptation effect, and an attention control task to measure participant's attention to the stimuli during the task we improved the experimental paradigm. Interestingly, we found that participants who had a better performance in the attention check task had also a higher adaptation impact score, suggesting an effect of attention on the adaptation strength (Clifford & Rhodes, 2005; Rhodes et al., 2011a). Taken together, our results not only replicated Wincenciak et al. (2013)'s findings, but due to methodological enhancements, strengthened existing evidence for visual adaptation to trustworthiness.

Experiment 2: The role of emotional expressions in subsequent trustworthiness and dominance perception

Experiment 1 found that visual adaptation affects trustworthiness impressions by using controlled faces with neutral emotional expressions. Building upon Experiment 1, Experiment 2 investigated how emotional expressions in the visual history influence subsequent trustworthiness impressions by using more realistic faces not controlled for gender appearance. Using a visual adaptation paradigm (Burton et al., 2016), we investigated whether trustworthiness judgements are driven by subtle facial cues resembling emotion, in line with the emotional overgeneralization hypothesis (Montepare & Dobish, 2003; Olivola et al., 2014; Zebrowitz, 1997). Experiment 2 aimed to extend the findings of Engell et al. (2010) on the influence of visual adaptation to emotional expressions on trustworthiness perceptions while minimising semantic adaptation by using anti-expressions as adaptors (faces with opposite visual characteristics of the original corresponding emotion, which bias neutral faces toward the original expression; Juricevic & Webster, 2012; Skinner & Benton, 2010). Importantly, anti-expressions are not clearly identifiable expressions, thus they are less likely to elicit semantic adaptation than the readily identifiable emotional expressions used by Engell et al. (2010). Therefore, an aftereffect for trustworthiness following adaptation to anti-expressions would suggest that a common mechanism underlies the visual representation of emotional expressions and evaluations of trustworthiness, as predicted by the emotion overgeneralisation hypothesis. Conversely, if no significant aftereffect is observed following adaptation to anti-expressions, this result would suggest that Engell et al. (2010) findings could be better explained by a semantic influence resulting from adapting to recognized emotional expressions.

Three different anti-expressions (anti-happy, anti-angry, anti-fearful) were used as adaptors to obtain clear evidence that very subtle cues to emotion contribute to trait impressions in specific ways, depending on the emotion observers adapted to. Specifically, anti-happy and anti-angry expressions were chosen because evaluations of trustworthiness are argued to arise from an overgeneralisation of facial cues resembling happy and angry expressions (Oosterhof & Todorov, 2008; Oosterhof & Todorov, 2009). Thus, we predicted that a test face showing an ambiguous expression (an average expression, see Methods) would be perceived as more trustworthy after adaptation to anti-happy faces and less trustworthy after adaptation to anti-anger faces. Conversely, anti-fear expression was chosen because trustworthiness impressions are not believed to be influenced by fearful emotional expressions. Regarding this, Engell et al., 2010 found that adapting to fear expressions resulted in no modulation of trustworthiness judgements in the following face. Thus, we predicted adaptation to an anti-fear face would make the subsequent face appear more fearful with no anticipated influence on trustworthiness ratings.

Moreover, we also extended our investigation to evaluations of facial dominance, which is a theoretically important facial judgement argued to account for most of the variation in facial impressions along with trustworthiness in models of first impressions (Oosterhof & Todorov, 2008), and is considered to be orthogonal to trustworthiness (Oosterhof & Todorov, 2008). Therefore, a different impact of anti-expressions on dominance ratings would indicate that specific emotional expressions are involved in each face trait impression. Moreover, testing dominance is informative to understand whether the effect of emotion adaptation extends to other face traits or whether it is unique for trustworthiness. Specifically, we predicted that an average expression would be perceived as more dominant after adaptation to anti-happy and anti-angry faces and less dominant after adaptation to anti-fear faces. Indeed, evaluations of facial dominance had been positively related to the perception of angry and happy expressions, and negatively to fearful expressions in neutral faces (Montepare & Dobish, 2003; Oosterhof & Todorov, 2008; Said et al., 2009; Witham et al., 2021).

Methods

Participants

A total of sixty-nine participants took part in this study at the School of Psychological Science at the University of Western Australia, recruited via the SONA system, social media advertisements, and word-of-mouth. Undergraduate students exchanged their participation for course credit. The study was approved by the Human Research Ethics Office of The University of Western Australia and participants gave informed consent at the beginning of the study.

The final sample included 68 participants (48 females, M = 22.4 years, s.d. = 5.9 years). The exclusion criteria were the same of Experiment 1. Here, we excluded from the analysis one participant with an error rate in the attention control task (asterisk task) higher than 30%. Only white participants were invited to participate to minimise an "own-race advantage" for processing stimuli derived from white faces (Meissner & Brigham, 2001; Skinner & Benton, 2010; restricting participants was not deemed necessary in Experiment 1 because the stimuli were computer-generated). Power calculations were based on the emotion aftereffect found in Burton et al. 2016, (d = 0.40). A larger sample size was recruited in Experiment 2 compared to Experiment 1 as the anticipated effect size was slightly smaller (d = 0.35), thus this approach was conservative. The power analysis indicated that 68 participants would yield adequate statistical power (80%) to detect an effect size of .35 at a standard alpha error probability of .05 in a one sample t-test.

Unfortunately, due to an oversight, this experiment was not pre-registered. Nevertheless, we pre-specified analyses and did not analyse data until the data collection was complete.

Apparatus and stimuli

The experiment was run using SuperLab Pro Version 4, on a 21.5inch iMac.

Anti-expression adaptor faces

In this experiment the adaptor faces were anti-expressions, which are faces characterised by visual attributes that are the opposite of their corresponding target expression (Figure 3A). Anti-expressions are created by morphing a typical expression (e.g., happy) along a linear path through a central point (i.e., the average expression) to a point equidistant in the opposite direction (i.e., anti-happy) (Juricevic & Webster, 2012; Skinner & Benton, 2010). When used as adaptors, they bias the perception of the following face toward the original expression (Juricevic & Webster, 2012; Skinner & Benton, 2010). For instance, adapting to "anti-anger" biases perception towards anger, making a subsequent neutral expression look angrier. Importantly, these aftereffects are based on the visual representation of the stimuli, as anti-expressions are not readily semantically identifiable as one of Ekman's (1973) six basic expressions (i.e., happy, angry, sad, fearful, disgust, or surprise). These anti-expressions were adapted by Burton et al. (2016) from Skinner and Benton's (2010) anti-expressions. In this experiment, anti-happy, anti-anger, and anti-fear expressions were used. These stimuli were originally created by first averaging with Psychomorph (Sutherland et al., 2017; Tiddeman et al., 2001) 50 face images of 25 females and 25 males individuals posing for each Ekman's (1973) six basic expression, resulting in a

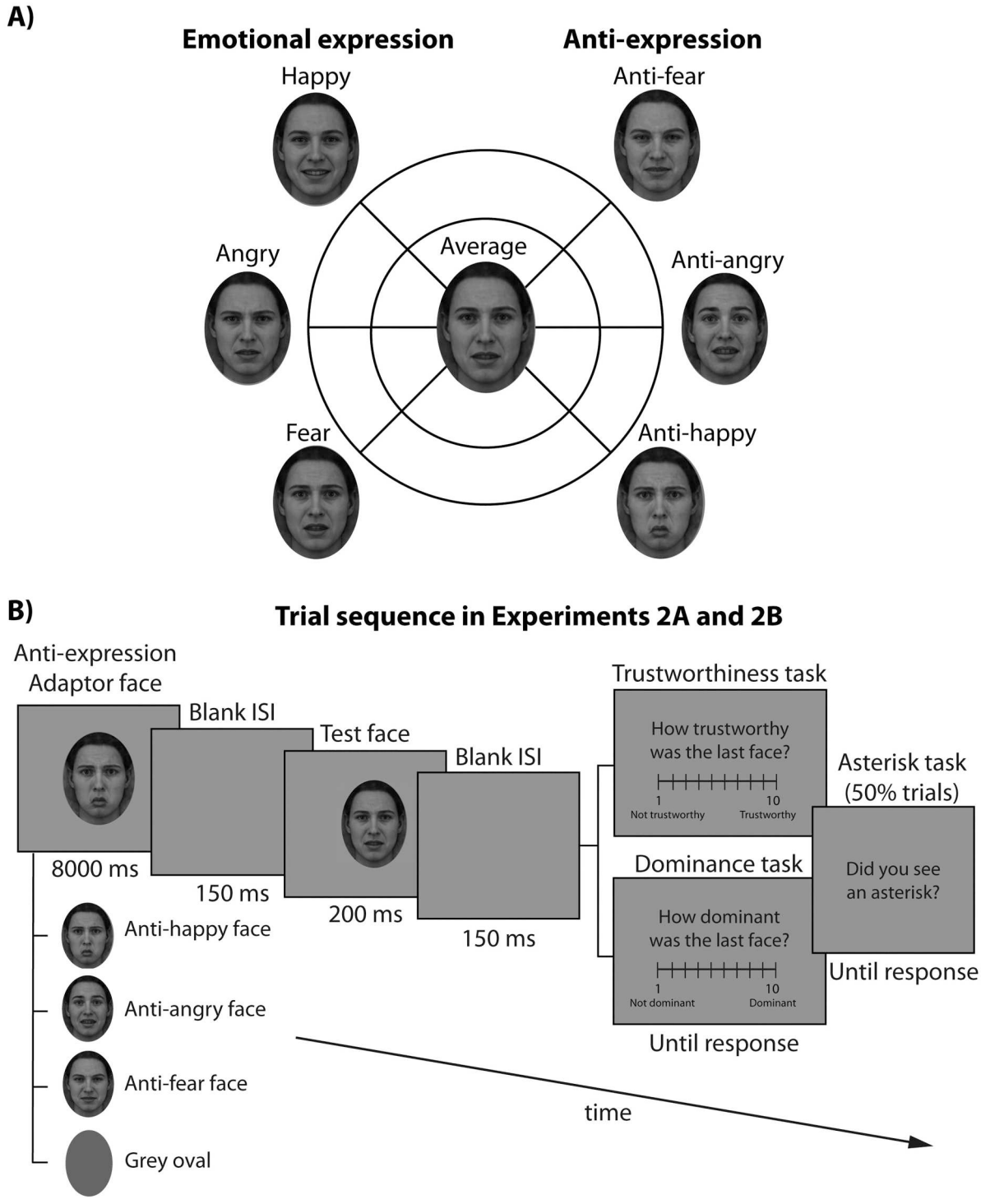

Figure 3. (A) The stimuli morph continuum ranged from the original facial expressions of happiness, anger and fear, an average face emotionally neutral, and their corresponding anti-expressions. In this experiment, anti-expressions were used as adaptor faces, whereas the average face was used as test face. (B) Trial sequence of Experiment 2 measuring trustworthiness and dominance. On each trial an adaptor face was displayed for 8000 ms. After 150 ms ISI, a test face appeared on the screen for 200 ms. Following an ISI of 150 ms, participants were asked to rate the trustworthiness/dominance of the test face on a scale from 1 to 10. Subsequently, participants indicated whether they observed an asterisk during the adaptor face exposure.

gender-neutral and identity-neutral facial expression prototype. Second, each emotion prototype was morphed linearly through the average expression to a point equidistant in the opposite direction. Stimuli were exported at −100% strength so that their relative distance from the emotionally neutral average

face (0%) was equal to the relative distance from the original expression (100%).

All anti-expression stimuli were masked with a grey oval to exclude non-facial cues. As a "no adaptation" control condition, the same mask was applied to a uniform grey oval with identical dimensions and with an average colour derived from all three anti-expressions. Adaptation to the grey oval served as a comparison for trait judgements of the average expression test face, as it was not expected to influence perception of the test faces. The adapting stimuli subtended a visual angle of approximately 12.68° high by 9.7° wide.

Test faces

The test face used was Burton et al. (2016) version of Skinner and Benton's (2010) "average expression" face. To create this average expression face, seven basic expressions (happiness, anger, fear, disgust, sadness, surprise and neutral) taken from 50 individual faces (25 female and 25 male) were morphed together to create seven separate average expressions, which captured the key characteristics of each expression while minimising any individual peculiarities. Next, these seven average expressions were merged to produce a single average expression face. This resulting gender, identity, and expression ambiguous face is known to be readily affected by adaptation (Burton et al., 2015, p. 2016; Skinner & Benton, 2010). The average expression was masked with a grey oval to exclude non-facial cues and used as the test face on two-thirds of the adaptation task trials. The remaining one-third of the trials used weak happy, angry, and fearful expressions as test face. These weak expressions were generated by taking a 100% emotional expression (e.g., happiness) and morphing it 50% towards the average expression face (0%).

Procedure

In Experiment 2, participants completed two adaptation tasks, rating test face trustworthiness and face dominance. The procedure in Experiment 2 was identical to that in Experiment 1, except that participants were exposed to one of the anti-expression adaptors (anti-happy, anti-anger, anti-fear, grey-oval) and were asked to rate either the trustworthiness or dominance of the test face, depending on the task. The trials sequence for both trustworthiness and dominance rating task are illustrated in Figure 3B.

The adaptation task included 48 trials, grouped into four blocks of 12, with allocated breaks in between each block. Of these 48 trials, 32 involved trait (trustworthiness/dominance) ratings of the average expression test face, with eight trials allocated to each adaptor condition (anti-happy, anti-anger, anti-fear, grey-oval). The remaining 16 trials utilised a weak (happy, angry, or fearful) expression as the test face, and ratings of these trials were not analysed.

Finally, to check for the possibility of gender after-effects (Pond et al., 2013a) participants were asked to indicate the gender of the test face and the three anti-adaptors ("male", "female" or "unsure"). In contrast to Experiment 1, where faces were highly controlled computer-generated stimuli and we were able to control the gender appearance of face stimuli, here the composite faces utilised were created by morphing real faces. Even if an equal number of male and female faces were averaged in order to obtain gender-neutral prototypes, it is not guaranteed that they were perceived as gender neutral. Given that previous research suggested that gender interacts with trustworthiness (Sutherland et al., 2015), it is important to understand how the ambiguous test face and the three anti-expressions were categorised as appearing male, female or ambiguous (unsure).

Before starting the experiment, participants familiarised with the stimuli and the task in a manner slightly different from Experiment 1. During the practice trials, participants rated the trustworthiness or dominance of the average expression face and three expressions (happy, angry, fear) presented at six different strengths (10%, 20%, 30%, 40%, 50%, 60%) on a scale of 1 to 10 (not-trustworthy – trustworthy / not-dominant – dominant). These stimuli were presented for 400 milliseconds in phase one and 200 milliseconds in phase two. Each phase comprised 19 trials, with one trial per expression at each strength, presented in a pseudorandom order to prevent consecutive presentations of the same expression.

All participants completed the practice and adaptation tasks for trustworthiness first, before completing both tasks for dominance. We carried out testing in this order because the trustworthiness task was our primary measure of interest, and the dominance task was secondary. Overall, the

experiment required approximately 45 minutes to complete.

Results and discussion

Effects of emotion adaptation on trustworthiness

Trustworthiness adaptation impact scores

To measure the influence of anti-expressions on trustworthiness ratings, we computed individual adaptation impact scores. To this purpose, the average trustworthiness ratings for the grey-oval adaptor condition were subtracted from the other three anti-expression adaptor conditions for each participant, resulting in three adaptation impact scores calculated for each participant (Figure 4A). The raw average scores for each adaptation condition are reported in the Supplementary Materials. A positive score indicated a higher trustworthy rating after adaptation to the anti-expression, whereas a negative score indicates a lower trustworthy rating after adaptation to the anti-expression. A stronger aftereffect is indicated by a greater deviation from zero. Consistent with Engell et al. (2010), adapting to anti-happy increased trustworthiness ratings (0.8 ± s.e: 0.1), and adapting to anti-anger decreased trustworthiness (−0.6 ± s.e: 0.1). Surprisingly, adapting to anti-fear resulted in a negative trustworthiness aftereffect (−0.5 ± s.e: 0.1). One-sample t-test against zero showed that the anti-happy ($t(67) = 7.53$, $p < .001$, $d = -0.58$), anti-angry ($t(67) = -7.10$, $p < .001$, $d = .54$), and anti-fear ($t(67) = -3.95$, $p < .001$, $d = .30$) adaptation condition were significantly different from zero, indicating an influence of the adaptor on the average test face ratings. Moreover, to test whether the absolute strength of the adaptation effect was different across the three adaptation conditions (ignoring the direction of the effect), a pairwise t-test was conducted by taking into consideration the adaptation impact score across adaptor conditions. We applied a Bonferroni correction for multiple comparisons ($\alpha = .008$). The t-test results showed that adaptation to anti-happy faces led to significantly higher adaptation impact score compared to the anti-fear condition ($t(67) = -1.93$, $p = .05$), whereas no significant difference emerged between adaptation impact scores after exposure to anti-happy and anti-angry ($t(67) = -1.63$, $p < .11$) or anti-angry and anti-fear ($t(67) = -0.79$, $p = .42$) adaptor faces.

Overall, these findings suggested that adapting to anti-happy faces increased the perceived trustworthiness in the test face, whereas adapting to anti-angry and anti-fear faces decreased the perceived level of trustworthiness in the test face. Contrary to the prediction of no effect, adapting to anti-fear had a smaller but still significant effect on trustworthiness, leading to decreased trustworthiness ratings; we will return to this point in the General Discussion.

Effects of emotion adaptation on dominance

Dominance adaptation impact scores

To measure the difference in dominance ratings following adaptation to an anti-expression versus adaptation to the grey-oval, we measured the individual adaptation impact scores, with the same procedure as in the trustworthiness rating task. The raw average scores for each adaptation condition are included in the Supplementary Materials. For each participant, three adaptation impact scores were calculated (Figure 4B). A positive score indicates a higher dominance rating after adaptation to the anti-expression and a negative score indicates a lower dominance rating after adaptation to the anti-expression. A stronger aftereffect is indicated by a greater deviation from zero.

Consistent with predictions, adapting to anti-happy (0.3 ± s.e: 0.1) and anti-anger (1.4 ± s.e: 0.1) increased perception of dominance, and adapting to anti-fear (−1.1 ± s.e: 0.1) decreased perception of dominance. These observations were confirmed by a series of one-sample t-tests which showed that the anti-happy ($t(67) = 3.25$, $p < .001$, $d = -0.24$), anti-angry ($t(67) = 13.27$, $p < .001$, $d = -1.01$), and anti-fear ($t(67) = -9.50$, $p < .001$, $d = .72$) adaptation condition were significantly different from zero, indicating an influence of the adaptor on the average test face ratings. Moreover, to test whether the absolute strength of the adaptation effect (ignoring the direction of the effect) was different across the three adaptation conditions we ran paired sample t-tests. We applied a Bonferroni correction for multiple comparisons ($\alpha = .008$). We found that adaptation to anti-angry faces led to significantly higher adaptation impact scores compared to adaptation to anti-happy faces ($t(67) = 9.25$, $p < .001$) and anti-fear faces ($t(67) = -2.20$, $p = .03$), and that adaptation to anti-fear faces resulted in significantly higher

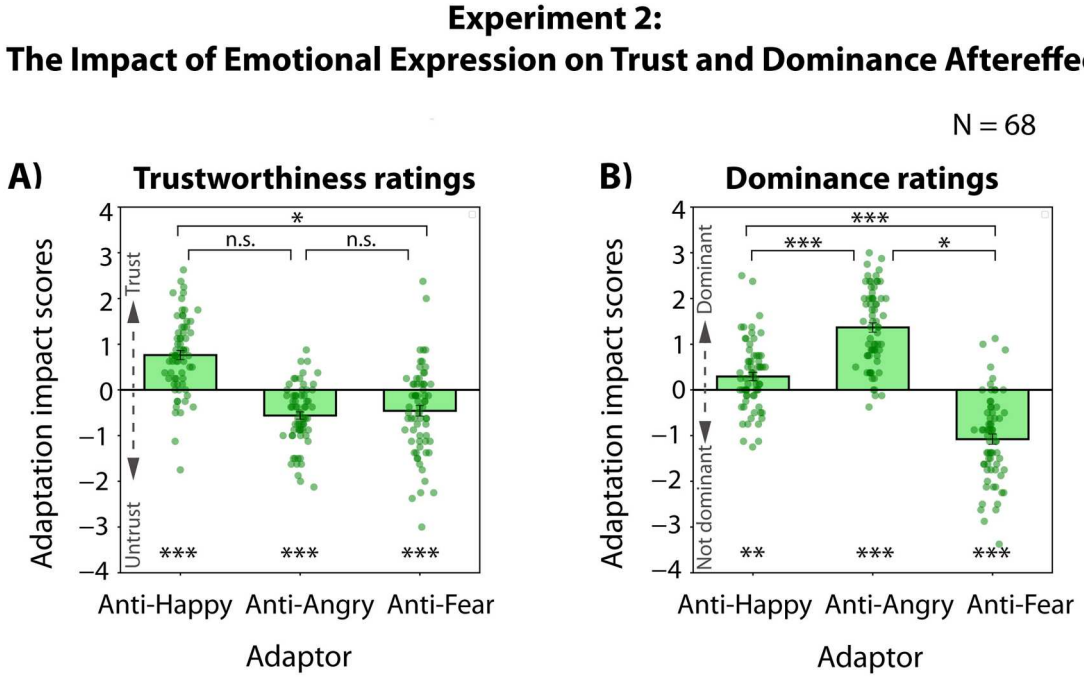

Figure 4. Experiment 2 Results. Trustworthiness (A) and Dominance (B) Adaptation impact scores. For each participant, the average trustworthiness/dominance ratings for the grey-oval adaptor condition were subtracted from the average trustworthiness/dominance ratings for the other three anti-expression adaptor conditions (anti-happy, anti-angry, anti-fear). The group average aftereffect trustworthiness/dominance impact score is plotted on the y-axis. The adaptation conditions are plotted on the x-axis. Each dot represents the score of one participant. Error bars represent the standard error of the mean. Below each bar is indicated the aftereffect impact scores significance level of difference from zero ($p < .05$ showed as *, $p < .001$ showed as ***, $p < .01$ showed as **) for each adaptation condition. Moreover, over the bars the comparisons in strength of aftereffect across the different adaptation conditions are illustrated.

adaptation impact scores compared to anti-happy condition ($t(67) = 4.81$, $p < .001$).

Overall, these findings suggest that adapting to anti-angry and anti-happy faces increased dominance in the test face, whereas adapting to anti-fear decreased dominance.

Attention check analysis

To assess participants' attention to the adaptor in the trustworthiness and dominance rating task, we calculated the error rate percentage for the Asterisk task in both tasks. All participants had an error rate lower than 30% in the Asterisk task, except one participant who was excluded from the analysis. Additionally, we examined the correlation between each participant's adaptation impact score and their error rate in the Asterisk task, considering the adaptation impact scores for all anti-expression adaptor conditions (anti-happy, anti-angry, anti-fear) in the trustworthiness and dominance judgement rating task.

First, we calculated the correlation between all the anti-expression adaptor conditions collapsed together (anti-happy, anti-angry, anti-fear, ignoring the direction of the effect by flipping the aftereffect scores for the anti-expressions with a negative average, multiplying each value for −1) and the error percentage in the Asterisk task for both the trustworthiness and dominance judgement rating tasks. We did not find any significant correlations for the collapsed scores in either the trustworthiness rating task ($r = -0.17$, $p = .15$) or the dominance rating task ($r = -0.17$, $p = .14$). However, given that adaption to anti-fear expressions was not predicted to influence the test face ratings for trustworthiness judgements we next collapsed just the values of anti-happy and anti-angry adaptation conditions (ignoring the direction of the effect), given these two conditions were predicted to impact both trustworthiness and dominance ratings. We calculated the correlation between the average collapsed adaptation impact scores and the error percentage in the Asterisk task, and we found no significant correlation for the collapsed scores neither in the trustworthiness rating task ($r = -0.18$, $p = .13$) and in the dominance

rating task (r = −0.23, p = .06). Next, we also compared the single adaptation impact scores after adaptation to each anti-expression. In the trustworthiness rating task, we found a small significant correlation (r = −0.26, p = .02) between the adaptation impact scores for the anti-angry adaptation and the error percentage in the Asterisk task. However, the correlation was not significant for the scores in the anti-happy (r = −0.04, p = .74) and anti-fear (r = −0.15, p = .19) adaptation condition. In the dominance rating task, we found no significant correlation between the adaptation impact scores for the anti-happy (r = −0.17, p = .15), anti-angry (r = −0.19, p = .11) and anti-fear (r = −0.21, p = .07) adaptation condition. Therefore, contrary to Experiment 1 results, here we did not find a correlation between attention and aftereffect strength.

Gender ratings of test faces and anti-expressions

Finally, we considered whether gender aftereffects (Pond et al., 2013a) could have contributed to the aftereffect in this experiment, given that previous research showed that gender influences trait impressions (Sutherland et al., 2015). Indeed, even if the face prototypes were created by averaging an equal number of female and male faces, it could be possible that they were not perceived as gender neutral. The frequency at which the ambiguous test face and the three anti-expressions were classified as male, female, or ambiguous (unsure) is reported in Table 1. Each participant classified the gender of the faces as the concluding step of the experiment.

Participants were more likely to judge the test face as female (X^2 (2) = 114.75, $p < .001$), and a similar trend can be found for the gender perception of the anti-anger (X^2 (2) = 32.52, $p < .001$), anti-happy (X^2 (2) = 16.39, $p < 0.001$) and anti-fear (X^2 (2) = 13.54, $p < 0.05$) anti-expression faces. Based on this analysis, there is a possibility that gender aftereffect also occurred in this experiment, so that adaptation to a feminine looking face could have biased the perception of the test face as more masculine (Pond et al., 2013a). This effect could have additionally influenced the perception of trustworthiness and dominance of the test face, given that these social judgements are influenced by gender (Oh et al., 2020; Oliveira et al., 2020b; Sutherland et al., 2015). Specifically, the test face could have appeared as more masculine, thus being perceived as less trustworthy and more dominant. Consequently, this could have reduced the aftereffect in the trustworthiness rating experiment and enhanced the aftereffect in the dominance rating experiment. We will return to this finding in the General Discussion.

Table 1. Experiment 2 frequency of gender choice for test and anti-expressions.

	Female	Male	Unsure
Test face	57 (83.8%)	6 (8.8%)	5 (7.4%)
Anti-anger	41 (59.4%)	15 (21.7%)	12 (17.39%)
Anti-happy	35 (51.5%)	20 (29.4%)	13 (19.1%)
Anti-fear	30 (44.1%)	27 (39.7%)	11 (16.2%)

In Experiment 2, we investigated whether adaptation to anti-expressions influenced the facial trait perception of the subsequent face, and we found that long exposure to anti-happy, anti-angry and anti-fear expressions biased the perception of trustworthiness and dominance of the following seen faces in ways consistent with the expression aftereffect we expected to induce. Overall, these findings suggest that a common mechanism underlies the visual representation of emotional expressions and evaluations of trustworthiness and dominance, as predicted by the emotion overgeneralisation hypothesis. Moreover, these results provided further support of Engell et al.'s, (2010) findings, by helping to rule out a semantic influence resulting from adapting to recognized emotional expressions and thus confirming a perceptual contribution of emotional cues seen in faces in forming trait impressions.

General discussion

The present studies investigated the temporal dynamics of face trait perception with an adaptation paradigm (Burton et al., 2016; Witham et al., 2021). Experiment 1 provided compelling evidence of negative aftereffects in impressions of trustworthiness, by showing that adaptation to (un)trustworthy-looking faces biased observers' perception of the trustworthiness of a subsequently presented neutral face in the opposite direction. Experiment 2 showed that evaluations of trustworthiness and dominance are influenced by the visual representation of emotional expressions in faces. These findings align with previous studies that reported visual adaptation for trait impressions (Engell et al., 2010; Keefe et al., 2013; Wincenciak et al., 2013) and are in line with

the emotion overgeneralisation hypothesis (Engell et al., 2010; Montepare & Dobish, 2003; Olivola et al., 2014; Oosterhof & Todorov, 2008; Zebrowitz, 1997).

Our findings are also informative about the cues that contribute to trustworthiness impressions. Emotional expression played an important role in trustworthiness aftereffects, consistent with previous research (Engell et al., 2010; Keefe et al., 2013; Wincenciak et al., 2013). Intriguingly, in Experiment 2 we found that adapting to different anti-expressions had a different influence on the perception of trustworthiness and dominance. With regard to trustworthiness impressions, we found that after adaptation to an anti-happy face, the test face was rated as more trustworthy, presumably because the test face appeared happier whereas after adaptation to an anti-angry face the test face was perceived as less trustworthy, due to looking angrier. However, unexpected results were observed after adaptation to anti-fear faces, where the average test face was rated as less trustworthy (Figure 4A). This finding is surprising as we did not predict that the perception of fear would influence trustworthiness. Moreover, our result differs from Engell et al. (2010) findings that adaptation to fearful faces did not affect trustworthiness impressions. One reason for this difference is that in our study we can be reasonably confident that adapting to anti-fear made the test face resemble a fear expression (Burton et al., 2016; Skinner & Benton, 2010; Witham et al., 2021). However, when adapting to fear, as in Engel's study, perception of the test face is biased away from fear, but it is less clear what it is biased toward. Consequently, it is more difficult to be certain what emotional expression is perceived in the test face after adaptation to a fearful face. In general, our finding aligns with a previous study suggesting that a negative valence associated with fear might explain lower trustworthiness ratings (Montepare & Dobish, 2003). Indeed, Montepare and Dobish (2003) found that faces with more fearful expressions were also perceived as low in affiliation. Affiliation and trustworthiness are partially related traits as they both reflect a positive valence, which suggest a person's approachability. Regarding this, the facial emotional valence has been suggested to be a significant mediator for impressions of trustworthiness (Tsankova et al., 2015). Thus, our findings reinforce the idea that trustworthiness evaluations are not specifically dependent on the perception of happiness and anger as hypothesised by Engell et al. (2010), but might rely on the valence of the face.

In the dominance impression domain, we found that the average test face was evaluated as more dominant after adapting to anti-happy and anti-anger expressions and less dominant after adapting to anti-fear expressions, consistent with our hypothesis. These results are in accordance with Witham et al. (2021) findings, who also showed how faces biased to look angry and happy following adaptation were rated higher in dominance and faces biased to look fearful after adaptation appeared less dominant. Previous research suggests that the display of anger functions as a cue to enhance strength and that the display of fear functions as a cue to signal weakness (Montepare & Dobish, 2003; Sell et al., 2014; Witham et al., 2021). Witham et al. (2021) found that adaptation to anti-anger expressions made a subsequent neutral face appear more dominant and stronger, supporting the idea that the expression of anger enhances the strength facial cues of a face, consequently increasing the perceived dominance (Toscano et al., 2016).

Importantly, using anti-expressions in Experiment 2 allowed us to distinguish between the perceptual or semantic nature of the mechanism underlying trustworthiness aftereffects. Our findings showed that the effect of emotion perception on these facial evaluations was predominately based on the visual representation of expression, rather than a semantic representation. This finding is consistent with Engell et al.'s (2010) finding that the magnitude of the aftereffect increased with increasing adaptor duration, which they interpreted as evidence for a perceptual locus of the aftereffect. Together these results further support the emotion overgeneralisation hypothesis suggesting that cues of emotional expression drive face trustworthiness impressions by signalling whether a person displaying such emotion should be approached or avoided (Adams & Kleck, 2003, p. 2005; Albohn et al., 2022), and strengthen the evidence that perception of trustworthiness in faces is based on a shared perceptual representation of emotional cues. Understanding whether trait judgements are driven by perceptual information related to visual representation of facial expressions or semantic representation of expressions

is relevant, as interventions to reduce the influence of inaccurate facial evaluations would differ if overgeneralisation were predominantly based on a visual or semantic representation. Additionally, the aftereffects in both experiments are unlikely to reflect purely low-level retinotopic adaptation because we displayed the adaptor and test faces at different sizes, so they may reflect adaptation of higher level face-selective representations.

Another cue that could have potentially contributed to the trustworthiness aftereffect in Experiment 2 was gender perception. For this reason, in Experiment 2 we additionally tested the perceived gender of face stimuli given that the composite faces utilised were created by morphing real faces, thus it was not guaranteed that they would have been perceived as gender neutral. We found that the adaptor faces were perceived as somewhat feminine, and thus gender aftereffects could have also occurred, biasing perception of the test face to appear more masculine (Pond et al., 2013a). Given that the perceived gender has an impact in trait impressions (Sutherland et al., 2015), a more masculine-looking test face could have made the trustworthiness aftereffect slightly less strong (as masculine faces are generally perceived as less trustworthy), and the dominance aftereffect slightly stronger (as masculine faces are perceived as more dominant). However, the whole aftereffect is unlikely to be explained by just gender aftereffect as the faces displayed were not strongly feminine or masculine. Even in the trustworthiness rating in Experiment 2, where test faces could have been perceived as more masculine and thus more untrustworthy, observers showed a strong bias in perceiving test faces as more trustworthy after adaptation to anti-happy faces. Additionally, in Experiment 1 the computer-generated face stimuli were designed to appear male-looking to avoid gender interaction with trustworthiness, and thus one could speculate that the opposite gender aftereffect could have occurred, resulting in the test faces appearing as more feminine. However, in Experiment 1 we still found a significant aftereffect after adaptation to both trustworthy and untrustworthy face adaptor condition, suggesting that the trustworthiness aftereffect was strong enough to overcome a potential gender aftereffect. In this view, across two experiments we found that the trustworthiness aftereffect generalises regardless of the perceived gender of the test face. Moreover, Wincenciak et al. (2013) found that aftereffects were of comparable strength regardless of the gender combinations of adaptors and test faces. These results further suggest that the perceived gender of our face adaptors is unlikely to have been critical in determining the pattern of aftereffects observed. Moreover, in the study by Wincenciak et al. (2013) gender effects in trustworthiness aftereffect were found, where female but not male observers exhibited a trustworthiness aftereffect. For this reason, we ran a post-hoc analysis focused on the relationship between observer's gender and trait impression aftereffect (results reported in Supplementary Materials). Notably, this analysis was not pre-registered, and unfortunately our sample was not gender balanced as investigating gender differences was not the scope of our study. Overall, our findings in both experiments indicated that males also exhibit a trustworthiness aftereffect, even if weaker compared to females. Therefore, while these results diverge from the gender effect observed by Wincenciak et al. (2013), they are not entirely contradictory with prior research, as males showed a less strong aftereffect. Future research could investigate potential interactions between perceived face gender and observer's gender and trustworthiness negative aftereffect.

Additionally, our findings suggest that attention influences trustworthiness aftereffects. We found a modest correlation between participants' attention check asterisk task and the strength of the aftereffect in Experiment 1, but not in Experiment 2. This result could suggest a possible involvement of attention in enhancing the strength of the aftereffect in Experiment 1, in line with previous work that have shown that attention enhances adaptation for simple stimulus attributes, such as motion (Von Grünau et al., 1998), oriented gratings (Spivey & Spirn, 2000), and face adaptation (Boynton, 2004; Rhodes et al., 2011). For example, Rhodes et al. (2011) manipulated attention to adapting faces, and found that identity aftereffect increased in the attention condition compared to the passive viewing condition. They proposed that attention could increase relevant neural activity related to face processing, and eventually lead to increased adaptation (Clifford & Rhodes, 2005; Rhodes et al., 2011a). On the other hand, in Experiment 2, we did not replicate this effect, in line with other work which proposed that attention is

not strongly involved in modulating negative aftereffects (Knapen et al., 2010; Suzuki & Grabowecky, 2003; Van Boxtel et al., 2010). However, the modest and non-significant associations we observed were all in the predicted direction, so it is important to highlight that our attention manipulation was primarily designed to encourage participants to maintain attention to the adapting faces and may not have been sufficiently sensitive to individual variation in attention to reliably predict the strength of the adaptation effects. Nor did we have sufficient power to detect small associations. Therefore, we are extremely cautious in drawing definitive conclusions on the results obtained in this attention analysis. An interesting future line of research could directly investigate the role of attention in trustworthiness visual adaptation.

Our results raise the possibility that a negativity bias for untrustworthy faces could influence trustworthiness aftereffects. In Experiment 1 we found a stronger aftereffect after adaptation to untrustworthy- compared to trustworthy-looking faces (Figure 2A). This asymmetry was not found in previous studies that investigated trustworthiness negative aftereffect (Keefe et al., 2013; Wincenciak et al., 2013). This result might reflect a general negativity bias which makes untrustworthy faces more salient and relevant, as their detection could potentially minimise the risk of harm in threatening situations, and eventually more impactful during adaptation. Accordingly, previous work has found enhanced processing for untrustworthy faces (Lischke et al., 2018) and better memorisation of untrustworthy compared to trustworthy faces (Giraudier et al., 2022; Meconi et al., 2014; Rule et al., 2012; Weymar et al., 2019). However, it might be also the case that an asymmetry in the face stimuli used in Experiment 1 influenced the aftereffects. The trustworthy-looking versions of each identity were made to be +1 standard deviation on the trustworthiness dimension and the untrustworthy faces set to −3 standard deviations on the trustworthiness dimension, as faces became particularly feminine with higher levels of trustworthiness (Oliveira et al., 2020b; Sutherland et al., 2015). This asymmetric approach might have resulted in untrustworthy looking faces producing stronger aftereffects, because they were more extreme adaptors (McKone et al., 2014; Pond et al., 2013b; Susilo et al., 2010). Therefore, we suggest caution in drawing conclusions regarding a negativity bias from the asymmetry in adaptation impact scores in Experiment 1.

Our findings cast a light on how prior visual experience influences trustworthiness impressions, suggesting that these fundamental social judgements are not fixed over time, but are instead dynamically influenced by the faces we saw before. A large literature has focused on the morphological facial characteristics that determine trustworthiness impressions (Sutherland & Young, 2022; Todorov et al., 2015) and their social consequences (Olivola et al., 2014) by using isolated faces and considering trust judgements as stable over time. However, in everyday life, faces are not often perceived in isolation, but they are often embedded in a spatial and temporal context, where we perceive a series of faces one after the other in the visual environment. In this light, our findings represent a further step towards a more ecological approach to the study of trustworthiness judgements. In the spatial context domain, previous research has shed light on the role of situational context in which we perceive a face, by expanding the investigation on how we form trust impressions of groups (Chwe & Freeman, 2023; Marini et al., 2023). For example, Marini et al. (2023) showed that when we encounter a crowd of people the visual system is able to extract the average trustworthiness of the group we are facing – a phenomenon called ensemble perception (Haberman & Whitney, 2009). Similarly, Carragher et al. (2021) found that when individuals perceived in a group are rated as more trustworthy looking compared to when seen alone, in line with the cheerleader effect (Walker & Vul, 2014; Ying et al., 2019). Taken together, these studies along with the current and previous work with trustworthiness adaptation suggest that these critical social judgements are shaped not only by the temporal but also by the spatial context in which faces are embedded. With regard to the present studies, future research directions could focus on exploring the practical implications of the temporal context influence in various real-life contexts. For example, in scenarios such as police lineups, where a sequence of suspects' faces is sequentially displayed for identification purposes, it could be relevant to investigate whether the perceived trustworthiness of a suspect's face may be shaped by exposure to a preceding face, with detrimental effects on eyewitness identification. Similarly, in the

context of dating apps, where users scan through streams of successive face photos in search of a mate (preferably trustworthy; Appel et al., 2023), it is important to understand whether the current trustworthiness judgment from a face could be biased by past faces seen, as this could eventually influence mate choice.

Taken together, these findings highlight emotion perception as one of the underlying mechanisms of trustworthiness and dominance, and thus contribute to understanding why we formulate facial trait impressions in the first place despite the lack of evidence that they accurately reflect underlying personality characteristics (Todorov et al., 2015; Wilson & Rule, 2017). Indeed, trustworthiness inferences are not considered accurate or reliable enough to be useful for making social decisions (Foo et al., 2021; Korva et al., 2013; Todorov et al., 2015; Wilson & Rule, 2015), to the point that a new line of research started to even wonder how to mitigate facial trustworthiness biases in decision-making processes (Jaeger et al., 2020). In this view, trustworthiness impressions can be viewed as "visual heuristics" exploited by the visual system to rapidly form sophisticated – yet often inaccurate – expectations about others' behaviour and intentions (Siddique et al., 2022).

Future research directions and limitations

The neural mechanisms underlying the adaptation of visual perceptions related to trustworthiness impressions remain to be fully understood. In general, visual adaptation has been proposed to have functional advantages and to contribute to efficient coding and discrimination of stimuli by calibrating our visual system to the prevailing inputs present in our visual environment (Barlow, 1990; Bosten et al., 2022b; Clifford & Rhodes, 2005; Mather et al., 2008; Rhodes & Leopold, 2011; Webster & Macleod, 2011). Visual adaptation of faces has been suggested to occur within the context of the norm-based model (Clifford & Rhodes, 2005; Rhodes & Leopold, 2011; Valentine, 1991), which proposes that faces are represented in our brain in relation to an average or prototype face at the centre of a computational face-space, or in the context of the exemplar-based model (Lewis, 2004; Ross et al., 2014), in which the average at the centre of face-space has no special status. Evidence suggests that face attributes of emotion, gender and identity may be norm-based coded (Burton et al., 2015; Jeffery et al., 2018; Pond et al., 2013; Rhodes & Jeffery, 2006; Rhodes et al., 2007). Given that these face attributes also influence trait impressions it is plausible that trustworthiness and dominance may likewise be represented by a norm-based system, with trustworthiness and dominance opponent-coded by neural channels tuned to high and low values on dimensions that underly these traits.

Adaptation to faces induces a change in sensitivity to the current stimulus (Clifford & Rhodes, 2005; Webster, 2015; Webster & Macleod, 2011; Webster & MacLin, 1999) which is argued to result from a decrease in the firing rate of cells that encode facial features (Kohn, 2007). The observed trustworthiness aftereffects could be explained by a temporary reduction in sensitivity of the cell population representing trustworthy and untrustworthy faces after exposure to such faces. Previous work found neural markers of trustworthiness perception in faces in the electrophysiological cortical responses to trustworthy or untrustworthy looking faces with the FPVS (Fast Periodic Visual Stimulation) paradigm, which does not require explicit judgements from observers (Siddique et al., 2023; Swe et al., 2020, 2022; Verosky et al., 2020). However, trustworthiness impressions are high-level complex social judgements that depends on a variety of bottom-up facial features, including eyebrow high and mouth curvature (Oosterhof & Todorov, 2008). As noted above, they are also connected to other facial judgements such as gender and attractiveness (Oosterhof & Todorov, 2008, p. 2009; Sutherland et al., 2015), and as further confirmed in the present study, emotional expression. Therefore, visual adaptation to trustworthiness is likely to influence multiple neuron populations coding various facial features. Future research could clarify where trustworthiness adaptation occurs at the neural level.

We used highly-controlled face stimuli that boost confidence that our adaptation effects reflect the traits we sought to manipulate. The complexity of face trait judgements in terms of visual cues (Vernon et al., 2014) and their intercorrelation with other social judgements (Keles et al., 2021; Oh et al., 2022; Siddique et al., 2022) makes it challenging, and probably impossible, to observe real faces that vary only in the level of trustworthiness or

dominance. That is why we used computer-generated faces and controlled for gender in Experiment 1, and used morphed faces and anti-expressions in Experiment 2, that also afforded good control. However, these stimuli lack ecological validity and may not fully reflect the way natural faces are neurally represented (Lischke et al., 2017). It will be important for future research to test the generalisability of our findings using more naturalistic face images that have greater ecological validity (Lischke et al., 2017; Sutherland & Young, 2022; Todorov et al., 2013b).

Further studies could also explore whether adaptation to trustworthiness occurs when faces are perceived in more naturalistic contexts, where both visual and auditory situational information might convey relevant information for trustworthiness judgements. In this view, it would be interesting to investigate whether visual adaptation to trustworthiness is affected by the context. For example, prior research has shown that the visual background surrounding faces influences trustworthiness evaluations (for a review, see Brambilla et al., 2024). Moreover, faces embedded in threatening auditory contexts are perceived as more untrustworthy than those embedded in non-threatening auditory contexts (Brambilla et al., 2021), suggesting cross-modal integration of facial and auditory information. Given that recent work showed that perceivers form first impressions of personality characteristics from individuals' voices (Mileva & Lavan, 2023), and considering also previous evidence of auditory adaptation (Pérez-González & Malmierca, 2014; Zäske et al., 2010), future research could investigate whether visual adaptation to trustworthiness occurs in a more naturalistic situation where we meet new individuals and are exposed to both their face and voice. Specifically, investigating how adaptation to trustworthiness occurs when faces are paired with trustworthy or untrustworthy voices could help determine whether visual adaptation to trustworthiness is affected by higher order multi-modal integration of various sources of information from the same individual, and how they integrate in a unitary impression of trustworthiness of that person.

Conclusion

Previous literature focused on trustworthiness impressions by considering the morphological determinants and the social consequences of these important judgements, but considerably less attention has been dedicated to understanding the ways in which the temporal context influences face perception of social traits. Our findings provide further robust evidence that visual adaptation influences trustworthiness impressions by replicating findings from previous work while also overcoming experimental limitations of prior research. Moreover, our results offer additional strong support on the interplay between facial trait impressions and emotional expressions, by showing that emotional expressions in the visual history influence subsequent impressions of trustworthiness and dominance.

Acknowledgements

We thank Alexander Todorov for the use of FaceGen images utilised in Experiment 1, and Andrew Skinner and Christopher Benton for the use of the original test and anti-expression face images shown in Experiment 2. We also thank Yong Foo and Cody Witham for helpful discussions about the design of Experiment 2. Finally, we thank R. Chakravarthi for his helpful advice and comments on an earlier version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Credit author statement

Fiammetta Marini: Conceptualization, Data curation, Software, Formal analysis, Visualization, Writing – original draft. **Clare A.M. Sutherland**: Conceptualization, Supervision, Methodology, Writing – review & editing. **Linda Jeffery**: Conceptualization, Data curation, Methodology, Writing – review & editing. **Sarah D. Maisey:** Conceptualization, Methodology, Data collection, Writing – first draft of Experiment 2. **Mauro Manassi**: Conceptualization, Methodology, Supervision, Writing – review & editing.

ORCID

Fiammetta Marini http://orcid.org/0009-0006-5860-7712
Clare A. M. Sutherland http://orcid.org/0000-0003-0443-3412
Linda Jeffery http://orcid.org/0000-0002-3980-5864
Sarah D. Maisey http://orcid.org/0000-0003-1733-6232
Mauro Manassi http://orcid.org/0000-0003-4210-7570

References

Adams, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, *14*(6), 644–647. doi:10.1046/j.0956-7976.2003.psci_1479.x

Afraz, S. R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research*, *48*(1), 42–54. doi:10.1016/j.visres.2007.10.028

Afraz, A., & Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. *Journal of Vision*, *9*(10), 10–10. doi:10.1167/9.10.10

Albohn, D. N., Brandenburg, J. C., Kveraga, K., & Adams, R. B. (2022). The shared signal hypothesis: Facial and bodily expressions of emotion mutually inform one another. *Attention, Perception, and Psychophysics*, *84*(7), 2271–2280. doi:10.3758/s13414-022-02548-6

Anstis, S., Verstraten, F. A. J., & Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences*, *2*(3), 111–117. doi:10.1016/S1364-6613(98)01142-5

Appel, M., Hutmacher, F., Politt, T., & Stein, J. P. (2023). Swipe right? Using beauty filters in male Tinder profiles reduces women's evaluations of trustworthiness but increases physical attractiveness and dating intention. *Computers in Human Behavior*, *201*, 107871. doi:10.1016/j.chb.2023.107871

Barlow, H. B. (1990). A theory about the functional role and synaptic mechanism of visual after-effects. *Vision : Coding and Efficiency*, 363–375. https://cir.nii.ac.jp/crid/1573387449871915648

Bosten, J. M., Coen-Cagli, R., Franklin, A., Solomon, S. G., & Webster, M. A. (2022). Calibrating vision: Concepts and questions. *Vision Research*, *201*, 108131. doi:10.1016/j.visres.2022.108131

Boynton, G. M. (2004). Adaptation and attentional selection. *Nature Neuroscience*, *7*(1), 8–10. doi:10.1038/nn0104-8

Brambilla, M., Masi, M., Mattavelli, S., & Biella, M. (2021). Faces and sounds becoming one: Cross-modal integration of facial and auditory cues in judging trustworthiness. *Social Cognition*, *39*(3), 315–327. doi:10.1521/soco.2021.39.3.315

Brambilla, M., Mattavelli, S., & Masi, M. (2024). Face–context integration and trustworthiness evaluation. *European Review of Social Psychology*, *35*(2), 378–423.

Burton, N., Jeffery, L., Bonner, J., & Rhodes, G. (2016). The timecourse of expression aftereffects. *Journal of Vision*, *16*(15), 1–1. doi:10.1167/16.15.1.

Burton, N., Jeffery, L., Calder, A. J., & Rhodes, G. (2015). How is facial expression coded? *Journal of Vision*, *15*(1), 1–1. doi:10.1167/15.1.1

Carragher, D. J., Thomas, N. A., & Nicholls, M. E. R. (2021). The dissociable influence of social context on judgements of facial attractiveness and trustworthiness. *British Journal of Psychology*, *112*(4), 902–933. doi:10.1111/bjop.12501

Chwe, J., & Freeman, J. B. (2023). Trustworthiness of crowds is gleaned in half a second. *Social Psychological and Personality Science*, *15*(3), 351–359.

Clifford, C. W. G., & Rhodes, G. (2005). *Fitting the Mind to the World: Adaptation and After-Effects in High-Level Vision*. https://philpapers.org/rec/CLIFTM

Ekman, P. (1973). Cross-cultural studies of facial expression. *Darwin and Facial Expression : A Century of Research in Review*. https://cir.nii.ac.jp/crid/1572261550529940736

Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception*, *39*(7), 931–941. doi:10.1068/p6633

Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PLoS One*, *7*(6), e37253. doi:10.1371/journal.pone.0037253

Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2021). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin*, *48*(11), 1580–1596.

Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research*, *1127*(1), 80–89. doi:10.1016/j.brainres.2006.09.104

Ghuman, A., Mcdaniel, J. R., & Martin, A. (2010). Report face adaptation without a face. *Current Biology*, *20*(1), 32–36. doi:10.1016/j.cub.2009.10.077

Giraudier, M., Ventura-Bort, C., Wendt, J., Lischke, A., & Weymar, M. (2022). Memory advantage for untrustworthy faces: Replication across lab- and web-based studies. *PLoS One*, *17*(2), 1–11. doi:10.1371/journal.pone.0264034

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. doi:10.1037/a0013899

Hills, P. J., Elward, R. L., & Lewis, M. B. (2010). Cross-modal face identity aftereffects and their relation to priming. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 876–891. doi:10.1037/a0018731

Jaeger, B., Todorov, A. T., Evans, A. M., & van Beest, I. (2020). Can we reduce facial biases? Persistent effects of facial trustworthiness on sentencing decisions. *Journal of Experimental Social Psychology*, *90*, 296–303. doi:10.1016/j.jesp.2020.104004

Javadi, A. H., Wee, N., Jeremy, J., & Barton, S. (2012). Cross-category adaptation: Objects produce gender adaptation in the perception of faces. *PLoS One*, *7*(9), e46079. doi:10.1371/journal.pone.0046079

Jeffery, L., Burton, N., Pond, S., Clifford, C. W., & Rhodes, G. (2018). Beyond opponent coding of facial identity: Evidence for an additional channel tuned to the average face. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 243.

Juricevic, I., & Webster, M. A. (2012). Selectivity of face aftereffects for expressions and anti-expressions. *Frontiers in Psychology*, *3*(JAN), 17105.

Keefe, B. D., Dzhelyova, M. P., Perrett, D. I., & Barraclough, N. E. (2013). Adaptation improves face trustworthiness

discrimination. *Frontiers in Psychology*, *4*, 1–7. doi:10.3389/fpsyg.2013.00358

Keles, U., Lin, C., & Adolphs, R. (2021). A cautionary note on predicting social judgments from faces with deep neural networks. *Affective Science*, *2*(4), 438–454. doi:10.1007/s42761-021-00075-5

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2010). The reference frame of the tilt aftereffect. *Journal of Vision*, *10*(1), 8–8. doi:10.1167/10.1.8

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. doi:10.1152/jn.00086.2007

Korva, N., Porter, S., O'Connor, B. P., Shaw, J., & ten Brinke, L. (2013). Dangerous decisions: Influence of juror attitudes and defendant appearance on legal decision-making. *Psychiatry, Psychology and Law*, *20*(3), 384–398. doi:10.1080/13218719.2012.692931

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

Lewis, M. B. (2004). Face-space-R: Towards a unified account of face recognition. *Visual Cognition*, *11*(1), 29–69. doi:10.1080/13506280344000194

Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. *Nature Communications*, *12*(1), 1–15. doi:10.1038/s41467-020-20314-w

Linke, L., Saribay, S. A., & Kleisner, K. (2016). Perceived trustworthiness is associated with position in a corporate hierarchy. *Personality and Individual Differences*, *99*, 22–27. doi:10.1016/j.paid.2016.04.076

Lischke, A., Junge, M., Hamm, A. O., & Weymar, M. (2017). Enhanced processing of untrustworthiness in natural faces with neutral expressions.

Lischke, A., Junge, M., Hamm, A. O., & Weymar, M. (2018). Enhanced processing of untrustworthiness in natural faces with neutral expressions. *Emotion*, *18*(2), 181–189. doi:10.1037/emo0000318

Manassi, M., Murai, Y., & Whitney, D. (2023). Serial dependence in visual perception: A meta-analysis and review. *Journal of Vision*, *23*(8), 18–18. doi:10.1167/jov.23.8.18

Manassi, M., & Whitney, D. (2022). Illusion of visual stability through active perceptual serial dependence. *Science Advances*, *8*(2), 2480. doi:10.1126/sciadv.abk2480

Marini, F., Manassi, M., & Ramon, M. (2024). Super recognizers: Increased sensitivity or reduced biases? Insights from serial dependence. *Journal of Vision*, *24*(7), 13–13. doi:10.1167/jov.24.7.13

Marini, F., Sutherland, C. A. M., Ostrovska, B., & Manassi, M. (2023). Three's a crowd: Fast ensemble perception of first impressions of trustworthiness. *Cognition*, *239*, 105540. doi:10.1016/j.cognition.2023.105540

Mather, G., Pavan, A., Campana, G., & Casco, C. (2008). The motion aftereffect reloaded. *Trends in Cognitive Sciences*, *12*(12), 481–487. doi:10.1016/j.tics.2008.09.002

McKone, E., Jeffery, L., Boeing, A., Clifford, C. W. G., & Rhodes, G. (2014). Face identity aftereffects increase monotonically with adaptor extremity over, but not beyond, the range of natural faces. *Vision Research*, *98*, 1–13. doi:10.1016/j.visres.2014.01.007

Meconi, F., Luria, R., & Sessa, P. (2014). Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. *Social Cognitive and Affective Neuroscience*, *9*(12), 1872–1879. doi:10.1093/scan/nst189

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. doi:10.1037/1076-8971.7.1.3

Mileva, M., & Lavan, N. (2023). Trait impressions from voices are formed rapidly within 400 ms of exposure. *Journal of Experimental Psychology: General*, *152*(6), 1539–1550. doi:10.1037/xge0001325

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, *27*(4), 237–254. doi:10.1023/A:1027332800296

Oh, D. W., Dotsch, R., Porter, J., & Todorov, A. (2020). Gender biases in impressions from faces: Empirical studies and computational models. *Journal of Experimental Psychology: General*, *149*(2), 323–342.

Oh, D. W., Martin, J. D., & Freeman, J. B. (2022). Personality across world regions predicts variability in the structure of face impressions. *Psychological Science*, *33*(8), 1240–1256. doi:10.1177/09567976211072814

Oliveira, M., Garcia-Marques, T., Garcia-Marques, L., & Dotsch, R. (2020). Good to bad or bad to bad? What is the relationship between valence and the trait content of the big two? *European Journal of Social Psychology*, *50*(2), 463–483. doi:10.1002/ejsp.2618

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 566–570. doi:10.1016/j.tics.2014.09.007

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. doi:10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128–133. doi:10.1037/a0014520

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. doi:10.3758/s13428-018-01193-y

Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: An overview. *Frontiers in Integrative Neuroscience*, *8*(FEB), 19.

Pond, S., Kloth, N., McKone, E., Jeffery, L., Irons, J., & Rhodes, G. (2013). Aftereffects support opponent coding of face gender. *Journal of Vision*, *13*(14), 16–16. doi:10.1167/13.14.16

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, *46*(18), 2977–2987. doi:10.1016/j.visres.2006.03.002

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. doi:10.1016/j.visres.2007.05.012

Rhodes, G., Jeffery, L., Evangelista, E., Ewing, L., Peters, M., & Taylor, L. (2011). Enhanced attention amplifies face adaptation. *Vision Research*, *51*(16), 1811–1819. doi:10.1016/j.visres.2011.06.008

Rhodes, G., & Leopold, D. A. (2011). *Adaptive norm-based coding of face identity*. Oxford Handbook of Face Perception.

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., Ewing, L., Calder, A. J., & Palermo, R. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition*, *142*, 123–137. doi:10.1016/j.cognition.2015.05.012

Ross, D. A., Deroche, M., & Palmeri, T. J. (2014). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin and Review*, *21*(1), 47–70. doi:10.3758/s13423-013-0449-5

Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits: Short report. *Psychological Science*, *19*(2), 109–111. doi:10.1111/j.1467-9280.2008.02054.x

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition*, *125*(2), 207–218. doi:10.1016/j.cognition.2012.06.017

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260–264. doi:10.1037/a0014681

Sell, A., Cosmides, L., & Tooby, J. (2014). The human anger face evolved to enhance cues of strength. *Evolution and Human Behavior*, *35*(5), 425–429. doi:10.1016/j.evolhumbehav.2014.05.008

Siddique, S., Sutherland, C. A. M., Jeffery, L., Swe, D., Gwinn, O. S., & Palermo, R. (2023). Children show neural sensitivity to facial trustworthiness as measured by fast periodic visual stimulation. *Neuropsychologia*, *180*, 108488. doi:10.1016/j.neuropsychologia.2023.108488

Siddique, S., Sutherland, C. A. M., Palermo, R., Foo, Y. Z., Swe, D. C., & Jeffery, L. (2022). Development of face-based trustworthiness impressions in childhood: A systematic review and metaanalysis. *Cognitive Development*, *61*, 101131. doi:10.1016/j.cogdev.2021.101131

Skinner, A. L., & Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychological Science*, *21*(9), 1248–1253. doi:10.1177/0956797610380702

South Palomares, J. K., & Young, A. W. (2017). Facial first impressions of partner preference traits: Trustworthiness, status, and attractiveness. *Social Psychological and Personality Science*, *9*(8), 990–1000. doi:10.1177/1948550617732388

Spivey, M. J., & Spirn, M. J. (2000). Selective visual attention modulates the direct tilt aftereffect. *Perception and Psychophysics*, *62*(8), 1525–1533. doi:10.3758/BF03212153

Susilo, T., McKone, E., & Edwards, M. (2010). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research*, *50*(3), 300–314. doi:10.1016/j.visres.2009.11.016

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. doi:10.1016/j.cognition.2012.12.001

Sutherland, C. A. M., Rhodes, G., & Young, A. W. (2017). Facial image manipulation: A tool for investigating social perception. *Social Psychological and Personality Science*, *8*(5), 538–551. doi:10.1177/1948550617697176

Sutherland, C. A. M., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology*, *113*(4), 1056–1078.

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, *106*(2), 186–208. doi:10.1111/bjop.12085

Suzuki, S., & Grabowecky, M. (2003). Attention during adaptation weakens negative afterimages. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(4), 793–807. doi:10.1037/0096-1523.29.4.793

Swe, D. C., Burton, N. S., & Rhodes, G. (2019). Are expression aftereffects fully explained by tilt adaptation? *Journal of Vision*, *19*(14), 21–21. doi:10.1167/19.14.21

Swe, D. C., Palermo, R., Gwinn, O. S., Bell, J., Nakanishi, A., Collova, J., & Sutherland, C. A. M. (2022). Trustworthiness perception is mandatory: Task instructions do not modulate fast periodic visual stimulation trustworthiness responses. *Journal of Vision*, *22*(11), 17–17. doi:10.1167/jov.22.11.17

Swe, D. C., Palermo, R., Gwinn, O. S., Rhodes, G., Neumann, M., Payart, S., & Sutherland, C. A. M. (2020). An objective and reliable electrophysiological marker for implicit trustworthiness perception. *Social Cognitive and Affective Neuroscience*, *15*(3), 337–346. doi:10.1093/scan/nsaa043

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*(5), 42–50. doi:10.1109/38.946630

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724–738. doi:10.1037/a0032335

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. doi:10.1146/annurev-psych-113011-143831

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. doi:10.1521/soco.2009.27.6.813

Toscano, H., Schubert, T. W., Dotsch, R., Falvello, V., & Todorov, A. (2016). Physical strength as a cue to dominance: A data-driven approach. *Personality and Social Psychology Bulletin*, *42*(12), 1603–1616. doi:10.1177/0146167216666266

Tsankova, E., Krumhuber, E., Aubrey, A., Kappas, A., Möllering, G., Marshall, D., & Rosin, P. (2015). The multi-modal nature of trustworthiness perception. In *Proceedings of the 1st joint conference on facial analysis, animation, and auditory-visual speech processing (FAAVSP)* (pp. 147–152). ISCA. http://www.isca-speech.org/archive/avsp15/av15_147.html

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*(2), 161–204. doi:10.1080/14640749108400966

Valentine, K. A., Li, N. P., Meltzer, A. L., & Tsai, M. H. (2020). Mate preferences for warmth-trustworthiness predict romantic attraction in the early stages of mate selection and satisfaction in ongoing relationships. *Personality & Social Psychology Bulletin*, *46*(2), 298–311. doi:10.1177/0146167219855048

Van Boxtel, J. J. A., Tsuchiya, N., & Koch, C. (2010). Opposing effects of attention and consciousness on afterimages. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8883–8888. doi:10.1073/pnas.0913292107

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(32), 3353–3361.

Verosky, S. C., Zoner, K. A., Marble, C. W., Sammon, M. M., & Babarinsa, C. O. (2020). Implicit responses to face trustworthiness measured with fast periodic visual stimulation. *Journal of Vision*, *20*(7), 29–29. doi:10.1167/jov.20.7.29

Von Grünau, M. W., Bertone, A., & Pakneshan, P. (1998). Attentional selection of motion states. *Spatial Vision*, *11*(4), 329–347. doi:10.1163/156856898X00068

Walker, D., & Vul, E. (2014). Hierarchical encoding makes individuals in a group seem more attractive. *Psychological Science*, *25*(1), 230–235. doi:10.1177/0956797613497969

Webster, M. A. (2015). Visual adaptation. *Annual Review of Vision Science*, *1*(1), 547–567. doi:10.1146/annurev-vision-082114-035509

Webster, M. A., Kaping, D., Mizukami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557–561. doi:10.1038/nature02420

Webster, M. A., & Macleod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1702–1725. doi:10.1098/rstb.2010.0360

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin and Review*, *6*(4), 647–653. doi:10.3758/BF03212974

Wendt, J., Weymar, M., Junge, M., Hamm, A. O., & Lischke, A. (2019). Heartfelt memories: Cardiac vagal tone correlates with increased memory for untrustworthy faces. *Emotion (Washington, D.C.)*, *19*(1), 178–182. doi:10.1037/emo0000396

Weymar, M., Ventura-Bort, C., Wendt, J., & Lischke, A. (2019). Behavioral and neural evidence of enhanced long-term memory for untrustworthy faces. *Scientific Reports*, *9*(1), 1–8. doi:10.1038/s41598-019-55705-7

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

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. doi:10.1111/j.1467-9280.2006.01750.x

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331. doi:10.1177/0956797615590992

Wilson, J. P., & Rule, N. O. (2017). Advances in understanding the detectability of trustworthiness from the face: Toward a taxonomy of a multifaceted construct. *Current Directions in Psychological Science*, *26*(4), 396–400. doi:10.1177/0963721416686211

Wincenciak, J., Dzhelyova, M., Perrett, D. I., & Barraclough, N. E. (2013). Adaptation to facial trustworthiness is different in female and male observers. *Vision Research*, *87*, 30–34. doi:10.1016/j.visres.2013.05.007

Witham, C., Foo, Y. Z., Jeffery, L., Burton, N. S., & Rhodes, G. (2021). Anger and fearful expressions influence perceptions of physical strength: Testing the signalling functions of emotional facial expressions with a visual aftereffects paradigm. *Evolution and Human Behavior*, *42*(6), 547–555. doi:10.1016/j.evolhumbehav.2021.05.005

Yang, H., Shen, J., Chen, J., & Fang, F. (2011). Face adaptation improves gender discrimination. *Vision Research*, *51*(1), 105–110. doi:10.1016/j.visres.2010.10.006

Ying, H., Burns, E., Lin, X., & Xu, H. (2019). Ensemble statistics shape face adaptation and the cheerleader effect. *Journal of Experimental Psychology: General*, *148*(3), 421–436. doi:10.1037/xge0000564

Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, *268*(1–2), 38–45. doi:10.1016/j.heares.2010.04.011

Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?*. Boulder, CO: West View Press.