Published in final edited form as:
J Pers Soc Psychol. 2018 March ; 114(3): 380–396. doi:10.1037/pspa0000108.

Facial Expression Predictions as Drivers of Social Perception

Lorena Chanes[1,2,3,4,*], **Jolie Baumann Wormwood**[1], **Nicole Betz**[1], and **Lisa Feldman Barrett**[1,2,*]

[1]Northeastern University, Department of Psychology, Boston, MA

[2]Massachusetts General Hospital, Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA

[3]Department of Clinical and Health Psychology, Universitat Autònoma de Barcelona, Catalunya, Spain

[4]Serra Húnter Fellow

Abstract

Emerging perspectives in neuroscience indicate that the brain functions predictively, constantly anticipating sensory input based on past experience. According to these perspectives, prediction signals impact perception, guiding and constraining experience. In a series of six behavioral experiments, we show that predictions about facial expressions *drive* social perception, deeply influencing how others are evaluated: individuals are judged as more likable and trustworthy when their facial expressions are anticipated, even in the absence of any conscious changes in felt affect. Moreover, the effect of predictions on social judgments extends to both real-world situations where such judgments have particularly high consequence (i.e., evaluating presidential candidates for an upcoming election), as well as to more basic perceptual processes that may underlie judgment (i.e., facilitated visual processing of expected expressions). The implications of these findings, including relevance for cross-cultural interactions, social stereotypes and mental illness, are discussed.

Keywords

Predictions; emotion; facial expressions; social perception; evaluative judgments

Predictions as Drivers of Social Perception

Emerging perspectives indicate that the brain functions predictively, as it constantly anticipates sensory input based on past experience (see, e.g., Barrett & Simmons, 2015; Chanes & Barrett, 2016; Clark, 2013; Friston, 2005; Hohwy, 2013). Predictions are

[*]Correspondence concerning this article should be addressed to Lorena Chanes, Department of Clinical and Health Psychology, Universitat Autònoma de Barcelona, Carrer de la Fortuna – Edifici B – B5B/053, Campus de la UAB – 08193 Bellaterra, Cerdanyola del Vallès, Barcelona, Spain, or to Lisa F. Barrett, Department of Psychology, 125 Nightingale Hall, Northeastern University, Boston MA 02115, USA. lorena.chanes@uab.cat or l.barrett@neu.edu.

Author Contributions
LC and LFB developed the study concept. LC, LFB and JBW contributed to study design. LC, JBW and NB performed data collection and analyses. All authors wrote the manuscript and approved final version for submission.

continuously issued and compared with incoming sensory inputs, which are in turn used to update the predictions. Predictions are thought to occur at multiple time scales and levels of specificity, from very specialized perceptual levels related to real-time sampling of the environment, to more abstract (i.e., general) and stable levels based on our core internal model of the world (e.g., Kiebel, Daunizeau, & Friston, 2008). Research within this framework has primarily come from the neurosciences, where predictions are assessed as top-down neural signals (see, e.g., Gilbert & Li, 2013 for a review on visual processing). According to these neuroscience perspectives, predictions importantly impact perception, guiding and constraining how we experience the world. However, in the behavioral domain, the posited role of predictions has largely been assessed only indirectly (although see Pinto, Gaal, Lange, Lamme, & Seth, 2015) under various labels (for recent discussions in terms of predictive coding see, e.g., Otten, Seth, & Pinto, 2017; Panichello, Cheung, & Bar, 2013). In this paper, we explicitly test the hypothesis that predictions significantly shape high-level social perception with important consequences for everyday life, in particular, judgments of others' likability and trustworthiness. In so doing, we suggest that predictive coding accounts of perception offer a unique explanatory lens that can contribute to unifying a wide variety of social perception effects within a common framework by identifying predictive signals as a potential shared mechanism.

Evidence of Predictions in Social Perception Across Different Research Domains

Research on stereotypes offers a prominent example of how predictions are often indirectly studied in social psychology. Stereotypes can be thought of as predictions (Barrett, 2017; Otten et al., 2017) as they represent implicitly-held expectations about shared properties of category members, and recent theoretical work has proposed that they constrain initial perceptions rather than being downstream products of social perception as long presumed (Freeman & Johnson, 2016). Moreover, research has shown that violating stereotypes can carry negative consequences for the violators. For example, women who violate gender stereotypes at work, e.g., women who succeed in stereotypically 'male' professions, are less liked and perceived as more interpersonally hostile than their male counterparts (Heilman, Wallen, Fuchs, & Tamkins, 2004), with potential career-affecting outcomes (Heilman, Wallen, Fuchs, & Tamkins, 2004; Heilman, 2001). Similarly, Mendes and colleagues (2007) found that participants rated individuals whose socioeconomic status violated racial/ethnic stereotypes (e.g., a white individual with low socioeconomic status, a Latino/a individual with high socioeconomic status) less favorably than individuals whose socioeconomic status matched those stereotypes. Although these findings are consistent with our hypothesis that predictions can influence high-level social perception, stereotypes are stable predictions that are deeply rooted in society; they frequently involve historical and political dimensions. Moreover, stereotype predictions are often assumed by the researchers rather than evaluated at the individual level. Thus, isolating the role of flexible, dynamic predictions on social perception requires novel or modified paradigms that can assess a dynamic range of predictions at the individual level.

Research examining emotion congruence in verbal and nonverbal communication has also indirectly approached the role of predictions on social perception. Findings in this domain have suggested a preference for individuals who present prediction-consistent information in social interaction. For example, leaders whose affect is congruent with the emotional content of their message (i.e., sad nonverbal displays when conveying saddening information) are rated more positively than leaders whose affect is incongruent with their message (Newcombe & Ashkanasy, 2002). Similarly, school-age children prefer to request information from consistent speakers (e.g., speakers providing a negative statement with negative affect) than from inconsistent or unfamiliar speakers (Gillis & Nilsen, 2016), and they rely on emotional congruence to determine whether a speaker is lying or telling the truth (Rotenberg, Simourd, & Moore, 1989). In these kinds of emotion congruence studies, however, two cues with equal or opposite valence are typically simultaneously displayed by a single target person (e.g., target person’s words and tone of voice). This could be perceived as a flagrant inconsistency of the target person’s internal state, making it difficult to assess the impact of the perceiver’s own dynamic expectations.

Recent behavioral work has also investigated the effects of nonverbal predictive cues (e.g., gaze direction) on social perception. This work has demonstrated that individuals are more liked and/or trusted when they display nonverbal cues that are predictive of task-relevant events (e.g., target location) for a perceiver (Bayliss, Griffiths, & Tipper, 2009; Bayliss & Tipper, 2006; Heerey & Velani, 2010). However, in these studies, individuals were more liked and trusted as the predictive cues they displayed had positive consequences for task performance but not necessarily otherwise, and thus the role of predictability was confounded with potential benefits for the perceiver. Thus, again, the effect of predictions *per se* was not explored.

Finally, a variety of research has investigated the impact of context on emotion perception (for discussions, see Barrett, Mesquita, & Gendron, 2011, and Aviezer, Hassin, Bentin, & Trope, 2008), which is also relevant for social perception. Contextual cues set up predictions that importantly impact how we perceive emotion on other individuals’ faces (for a review, see Barrett, Mesquita, & Gendron, 2011). For example, participants rely on situational information, including presented scenarios (Carroll & Russell, 1996) and visual scenes (Righart & de Gelder, 2008), when judging emotional responses from facial configurations. Body configurations are also used as disambiguating contexts, creating variable interpretations of identical facial configurations depending on simultaneous contextual information from the body (Aviezer, Trope, & Todorov, 2012; Aviezer et al., 2008). These findings do not directly examine the influence of emotion predictions on social judgments, but do demonstrate that context guides emotion predictions, and can encourage participants to issue specific dynamic predictions about emotional displays. Thus, we can leverage emotional context in our experimental designs to explicitly explore the impact of predictions about emotional displays on social perception, which is the goal of the current research.

Although previous research provides some evidence that predictions play a role in social perception, these studies have typically not been discussed in terms of predictive coding, nor have they been designed to specifically assess the role of predictions as the underlying mechanism driving differences in social perception. Thus, the impact of flexible, dynamic, individualized predictions on social perception has remained largely unexplored and its study requires novel experimental designs.

The Present Studies

In this paper, we explicitly assessed how predictions of facial expressions embedded in emotional contexts impact individuals' evaluation of others. We directly examined whether facial expression predictions *drive* social perception, deeply influencing how individuals experience social input. We hypothesized that individuals would be evaluated as more likable and trustworthy when their facial expressions were anticipated (Experiments 1, 2, 3, and 4), an effect that we predicted would occur across and within emotion categories (e.g., for fear, happiness, and sadness) as well as across affective categories (e.g., pleasant and unpleasant, high and low arousal emotion categories). We also hypothesized that the effect of predictions would extend beyond flagrant violations of stereotypical expressions to more nuanced and individualized predictions concerning what one expects to see another person look like in a given emotional context (Experiments 3 and 4). We also explored whether the effect of predictions on social perception would extend beyond well-controlled lab environments to real-world situations where such judgments have particularly high consequence, such as evaluating presidential candidates for an upcoming election (Experiment 4). Finally, we hypothesized that changes in consciously felt affect would not underlie these effects (i.e., that they would not be attributable to affective misattribution mechanisms, Clore, Gasper, & Garvin, 2001; Experiment 5). Instead, we hypothesized that predictions would drive social perception directly, operating as a kind of processing fluency (Winkielman, Schwarz, Fazendeiro, & Reber, 2003), such that the perceptual processing of predicted facial expressions would be facilitated, leading to more positive ratings (Experiment 6). To the extent that our hypotheses are supported, the present studies will demonstrate that predictive coding theories offer a unique explanatory lens to integrate seemingly disparate social perception effects as driven by explicit or implicit predictions at various levels of specificity. Those levels would span from specific visual predictions for what one should expect to see on another's face in the very next moment, to very abstract, general, predictions about what another person may do over a much longer time course, such as over the next several minutes, days, weeks, or years.

Experiments 1 and 2: Stereotypical Facial Expressions and Social Perception

In a first set of experiments, we explored whether an individual's perception of another person is shaped by whether that person's facial expression matches or mismatches predicted stereotypical expressions. We hypothesized that individuals would be experienced as more likable and trustworthy when their facial expressions matched the perceiver's predictions compared to when they did not. To test this, we implemented a novel task design wherein, on each trial, participants read an emotionally evocative story (scenario) about a target person and were asked to imagine how the person would look in that scenario, leading them to generate specific perceptual predictions. Next, participants saw the person's facial expression. On each trial, the facial expression either matched the prediction evoked by the scenario (e.g., a smiling face after a happy scenario) or did not match it (e.g., a smiling face after a sad scenario). Thereafter, participants rated the likability (Experiment 1) or trustworthiness (Experiment 2) of the target person. These experiments represent a critical first step to look for evidence that facial expression predictions can impact social perception. Accordingly, our design focused on highly stereotypical facial expressions for different emotion categories and flagrant mismatches from the evoked predictions elicited by the scenarios, allowing us to first attempt to detect effects on social perception under conditions with robust and apparent prediction violations. If our hypotheses are supported within this paradigm, then we can begin to examine how facial expression predictions may shape the perception of social others in a more nuanced and individualized basis.

Method

Participants—Two groups of subjects participated in the two experiments. For each experiment, participants were 35[1] young adults recruited from Northeastern University (Mean Age±SD: 19±2 y.o., 22 female for Experiment 1; and Mean Age±SD: 19±1 y.o., 19 female for Experiment 2). All participants reported normal or corrected-to-normal visual acuity, were native English speakers and received course credit for their participation. The target sample size of *n*=35 was chosen because samples of 30–40 are thought to provide enough power to detect a medium to large behavioral effect (see Wilson VanVoorhis & Morgan, 2007).

Materials and Procedure—Participants completed 3 practice trials and 45 experimental trials of a social perception task on a computer. Instructions and stimuli were presented using E-Prime 2 running on a Dell Optiplex 745 and a 17-inch Samsung LCD flat-screen monitor (1280×1024). A diagram of the structure for each trial of the task is given in Figure 1a. Each trial started with a fixation screen (7 s), followed by a photograph of a neutral face of a target person (5 s). A short written story (scenario) about the target person, designed to be emotionally evocative, was then displayed for 20 s. The scenario was meant to evoke one of three emotions: fear, sadness, or happiness. Participants were asked to imagine how the target person would look in that scenario while reading. Then a second photograph of the target person was displayed, either portraying a neutral facial expression or a stereotypical facial expression for one of the three emotions (e.g., a pout depicting sadness), for 5 s. The face could "match" the evoked emotion (e.g., fear scenario followed by a stereotypical fear face; 21 trials), be neutral (e.g., fear scenario followed by a neutral face; 12 trials) or "not match" the evoked emotion (e.g., fear scenario followed by a stereotypical sad or happy face; 12 trials). In these experiments, trials with neutral faces were included only to generate variability (i.e., to avoid having only stereotypical facial expressions). They were not included in the analyses, which contrasted performance on trials with the same category of faces (stereotypical happy, sad, fear) when they were predicted (matched) and not predicted (non-matched).

[1] Informed consent was collected for 38 participants in Experiment 1 but 3 were excluded from analyses due to a detected error in the task program (n=1) and noncompliance with the inclusion criterion of being native English speakers (n=2). Informed consent was collected for 36 participants in Experiment 2 but 1 was excluded from analyses due to noncompliance with the inclusion criterion of being native English speaker.

Participants were then asked to quickly perform one rating (up to 2 s). In Experiment 1, participants rated how likable the target person was. In Experiment 2, they rated how trustworthy the target person was. Ratings were performed on a scale from 1 to 4 (1=unlikable/4=very likable; 1=untrustworthy/4=very trustworthy).

We used photographs from the Interdisciplinary Affective Science Laboratory Face Set (www.affective-science.org). We used a different target person (identity) for each of the 48 trials (3 practice trials: 2 female, 1 male; 45 experimental trials: 28 female, 17 male). For each stereotypical facial expression (sad, fear, happy, neutral) and identity, two versions were available: one with mouth closed and one with mouth open. Half of the initial neutral faces corresponded to each version (i.e., mouth open or closed). If the second face was displaying a neutral expression, then the other version (i.e., mouth open or closed) was used to avoid repetition of the same image. If the second face was displaying a fear, happy or sad expression, then the open or closed mouth version was used randomly. Normed ratings of intensity, attractiveness, and stereotypicality (i.e., accuracy in identifying the emotion category) for the mouth-closed face set stimuli for the emotion categories used in this experiment are provided in the supplemental materials (Table S1). See Figure S1 in the supplemental materials for sample face stimuli.

Written stories (scenarios) were taken from a set of emotion scenarios developed and pilot-tested in a prior set of experiments (Wilson-Mendenhall, Barrett, & Barsalou, 2013). The scenarios used sampled from the four quadrants of the affective circumplex. Positive valence was represented by happy scenarios, whereas negative valence was represented by sad and fear scenarios. All emotion categories (happiness, fear, sadness) included high and low arousal scenarios. As described by Wilson-Mendenhall and colleagues (2013), an independent set of participants rated the scenarios to verify that they elicited the intended variation in subjectively experienced valence and arousal, and participants reported that it was relatively easy to immerse themselves in these scenarios (details can be found in the supplemental materials of Wilson-Mendenhall, Barrett, & Barsalou, 2013). For the purpose of the present experiment, we removed the last sentence of each scenario, which explicitly mentioned the specific emotion category. See Table S2 in the supplemental materials for sample scenarios.

Results

Experiment 1: Perceived Likability—A 2 (face match: matched, non-matched) by 3 (face emotion category: sad, fear, happy) repeated-measures analysis of variance (ANOVA) on likability judgments supported our hypothesis; target persons were rated as more likable when displaying predicted (matched) facial expressions (M=2.94, SE=.07) than when displaying non-matched facial expressions (M=2.19, SE=.07), F(1, 34)=75.97, p<.001, η_p^2=.691. We also observed a main effect of face emotion category, F(2, 68)=12.00, p<.001, η_p^2=.261. Bonferroni comparisons revealed that target persons displaying stereotypical happy faces were rated as more likable (M=2.81, SE=.07) than those displaying stereotypical sad (M=2.46, SE=.08), p=.003, or fear faces (M=2.42, SE=.08), p=.001 (sad vs. fear faces: p=1.00).

This analysis also revealed a significant interaction between face match and face emotion category on ratings of likability, $F(2, 68)=55.65$, $p<.001$, $\eta_p^2 =.621$. To explore this interaction, we conducted a series of repeated-measures ANOVAs, one for each face emotion category, with face match as the within-subjects factor. As anticipated, for all three emotion categories, individuals were rated as more likable when displaying predicted facial expressions (matched) than when displaying non-matched facial expressions (fear faces: $F(1, 34)=22.53$, $p<.001$, $\eta_p^2 =.399$; sad faces: $F(1, 34)=5.70$, $p=0.023$, $\eta_p^2 =.144$; happy faces: $F(1, 34)=126.52$, p<.001, $\eta_p^2 =.788$)[2]. See Figure 1b.

Experiment 2: Perceived Trustworthiness—A 2 (face match: matched, non-matched) by 3 (face emotion category: fearful, sad, happy) repeated-measures ANOVA on trustworthiness judgments supported our hypothesis; target persons were rated as more trustworthy when displaying predicted (matched) facial expressions (M=3.04, SE=.07) than when displaying non-matched facial expressions (M=1.92, SE=.07), $F(1, 34)=152.89$, $p<.001$, $\eta_p^2 =.818$. No main effect of face emotion category was observed ($F<1$).

This analysis also revealed a significant interaction between face match and face emotion category on trustworthiness ratings, $F(2, 68)=72.69$, $p<.001$, $\eta_p^2 =.681$. To explore this interaction, we conducted a series of repeated-measures ANOVAs, one for each face emotion category, with face match as the within-subjects factor. As anticipated, for all three emotion categories, individuals were rated as more trustworthy when displaying predicted facial expressions (matched) than when displaying non-matched facial expressions (fear faces: $F(1, 34)=36.04$, $p<.001$, $\eta_p^2 =.515$; sad faces: $F(1, 34)=70.63$, $p<0.001$, $\eta_p^2 =.675$; happy faces: $F(1, 34)=222.56$, $p<.001$, $\eta_p^2 =.867$)[3]. See Figure 1c[4].

Discussion

In Experiments 1 and 2, we observed initial evidence that perceivers evaluate individuals more positively when provided with social information (in these experiments, facial expressions) that matches their predictions. Specifically, we found that participants rated individuals as more likable and trustworthy when those individuals exhibited expected facial expressions. Of import, the impact of predictions on social perception held across facial emotion categories (i.e., even individuals displaying stereotypical sad expressions were liked and trusted more if their expression was expected).

These experiments provide initial evidence that predictions influence social perception, but critically, we did not directly assess participants' predictions in these experiments. Instead, we assumed the displayed facial expression roughly matched or mismatched the perceiver's predictions based on the presentation of highly stereotypical facial expressions and normed

[2] Additional analysis to fully unpack this interaction can be found on supplementary materials

[3] Additional analysis to fully unpack this interaction can be found on supplementary materials

[4] Some of the non-match trials involved cross-valence violations (e.g., a negatively-valenced pouting or startled face followed by a positively-valenced happy scenario) while other non-match trials involved within-valence violations (e.g., a negatively-valenced pouting face following a negatively-valenced fear scenario). These experiments were not designed nor powered to compare the impact of these various types of non-match trials. However, means and standard errors are provided in the supplemental materials in Figure S2 for evaluative ratings in Experiments 1 and 2 broken down by the various non-match trial types for each facial emotion category. The pattern of results is consistent with an interpretation that cross-valence violations were perceived as more blatant or pronounced than within-valence mismatches.

emotion scenarios. Thus, we were not able to rule out the possibility that the effect observed was driven by blatant violations of stereotypes in non-match trials, where the target person might be perceived as displaying a highly inappropriate response (e.g., smiling in response to a tragedy or displaying a fearful face in response to meeting a good friend for coffee). Such blatantly inappropriate responses may have led participants to judge target persons in those trials as bizarre or maladaptive. Indeed, many mental illnesses are characterized by symptoms involving contextually-inappropriate affect displays (e.g., Liddle, 1987). Thus, blatant mismatches may have triggered judgments regarding that individual's emotional or mental stability, with subsequent consequences regarding judgments of their character.

Thus, in Experiment 3, we attempted to replicate our observations while additionally asking participants to explicitly report the degree to which each facial expression matched their predictions on each trial. This approach allows for a more direct measure of the effects of predictability of facial expressions on social judgements and, critically, allows us to examine the effects of predictability within only those trials in which the target person presented an "appropriate" facial expression, with no blatant violations of stereotypes or conventionality (i.e., within match trials).

Experiment 3: Predictions and Social Perception

In Experiment 3, we explicitly assessed whether individual facial expression predictions underlie participants' social perception (in particular, likability ratings) by asking participants to additionally rate how similar each target person looked to what they expected (predictability). These trial-by-trial predictability ratings provide us with a measure of the degree of subjective prediction fulfillment, as participants were instructed to imagine how the target person would look in each scenario as they read. Moreover, the inclusion of a trial-by-trial measure of predictability allowed us to examine whether social perception was driven by more subtle shifts in individual facial expression predictions, as opposed to blatant violations of stereotypicality (e.g., a smiling face in what should be terrifying scenario). That is, we were able to examine whether a target individual was perceived as more likable when his or her displayed facial expression more closely matched a specific perceiver's predictions, irrespective of whether that stereotypical expression matched the emotion scenario on that trial. We examined this, not only across all experimental trials where we expected robust differences in predictability to emerge (i.e., across match trials and non-match trials), but also across trials where we expected predictability to vary less and reflect more ideographic differences in facial expression predictions (e.g., within match trials only, where predictability should be higher on average, or within non-match trials only where predictability should be lower on average). Analyzing only match trials, where the target person displays an "appropriate" facial expression, allows us to overcome an important potential confound of Experiments 1 and 2: that target identities displaying blatant norm/ stereotype violations (e.g., cross-valence mismatches) may have been judged as callous, maladaptive, or even mentally disabled, with consequent impact on evaluative judgments of likability and trustworthiness.

We predicted that individuals would be judged as more likable on trials where their facial expression more closely matched the perceiver's predictions, even when controlling for blatant stereotypical face matching and mismatching, and that predictability would mediate the impact of stereotypical match and mismatch on social perception. Importantly, we also predicted that the effect of predictions on social perception would emerge when considering only those trials in which the individuals presented an "appropriate" facial expression (match trials). To the extent that we find that predictability of facial expressions still predicts social judgment ratings in the match trials, we will be able to rule out the possible confound that our effect is driven by judgments of blatant (e.g., cross-valence) facial expression and scenario mismatches.

Method

Participants—Participants were 35[5] young adults recruited from Northeastern University (Mean Age±SD: 20±2 y.o., 25 female). All participants reported normal or corrected-to-normal visual acuity, were native English speakers and received course credit for their participation. The target sample size of n=35 was based on the results of Experiments 1 and 2.

Materials and Procedure—For Experiment 3, materials and procedure were identical to Experiments 1 and 2 except that participants first rated how similar the facial expression looked to what they expected (predictability rating), then how likable the target person was (likability rating) on scales from 1 to 4 (1=not at all similar/4=very similar; 1=unlikable/4=very likable).

Results

We utilized hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002), which allowed us to avoid aggregation across trials and model variability in trial-by-trial performance nested within each participant. We utilized a continuous sampling model with random effects, and a restricted maximum likelihood method of estimation for model parameters. All Level-1 (trial-level) variables were group centered (i.e., centered around each participant's mean; Enders & Tofighi, 2007).

Mediational Analyses—Consistent with findings from Experiments 1 and 2, our HLM analysis revealed that individuals were rated as significantly more likable when displaying matched facial expressions (M=2.97, SE=.08) than when displaying non-matched facial expressions (M=2.30, SE=.06), t(34)=8.62, p<.001. As expected, this analysis also showed that matched facial expressions were rated as significantly more predicted (M=3.10, SE=.08) than non-matched facial expressions (M=1.68, SE=.05), t(34)=17.62, p<.001. Crucially, analyses also revealed that the relationship between face match condition (matched, non-matched) and likability ratings was significantly mediated by ratings of predictability. Predictability ratings accounted for 97.5% of the relationship between face match condition and ratings of likability and remained a significant predictor of likability ratings in this model (B=.47, SE=.04, t(34)=10.95, p<.001). A Sobel test confirmed significant mediation (Z=9.30, p<.001). When controlling for the effect of predictability, match condition was no

[5] Informed consent was collected for 36 participants in Experiment 3 but 1 was excluded from analyses due to noncompliance with the inclusion criterion of being native English speaker.

longer a significant predictor of likability (*B*=.02, SE=.07; *t*(34)=0.25, *p* =.804), suggesting the relationship was fully mediated by ratings of predictability.

Predictability Predicts Likability Ratings—In addition, to move beyond broad condition-based differences, we utilized HLM to examine whether within-subject differences in ratings of predictability significantly predicted within-subject differences in perceptions of likability, ignoring match condition and including neutral face trials. This analysis revealed that individuals that portrayed expressions that were rated as more predicted were also rated as more likable (*B*=.46, SE=.04; *t*(34)=11.45, *p*<.001). We also examined this relationship separately within matched, non-matched, and neutral facial expression conditions. This analysis revealed that, within-subjects, ratings of predictability predicted ratings of likability when looking within just the match trials (*B*=.50, SE=.05; *t*(34)=10.77, *p*<.001), within just the non-match trials (*B*=.41, SE=.06; *t*(34)=6.77, *p*<.001), and within only the neutral face trials (*B*=.42, SE=.05; *t*(34)=8.50, *p*<.001). Moreover, the strength of the effect did not differ across conditions (matched vs. non-matched: *B*=.08, SE=.07; *t*(34)=1.28, *p*=.201; matched vs. neutral: *B*=.08, SE=.05; *t*(34)=1.57, *p*=.127; non-matched vs. neutral: *B*=.01, SE=.05; *t*(34)=0.12, *p*=.905). See Figure 2a[6].

Discussion

In Experiment 3, we replicated and extended the findings of Experiments 1 and 2, including a trial-by-trial measure of facial expression predictions that allowed us to assess predictions directly instead of making assumptions about a perceiver's predictions based on stereotypicality. Using HLM, we replicated the results of Experiment 1, finding that an individual is judged as more likable when displaying a facial expression that matches the emotion category of the preceding scenario than when his or her facial expression does not match it. Moreover, we extended these results by conducting a mediational analysis that demonstrated that participants' individual predictions about facial expressions underlie this effect.

Importantly, the trial-by-trial measure of predictability additionally allowed us to evaluate whether the effect of predictions was driven primarily by trials in which there was a violation of stereotype or conventionality (non-match trials) that could have led participants to judge the individual as bizarre or maladaptive. We demonstrated that the effect of predictions on evaluative judgments was still observed when considering only the match trials where an "appropriate" facial expression was always displayed (i.e., an expression congruent with the emotion evoked by the scenario: smiling face after a happy scenario, pouting face after a sad scenario, startled face after a fear scenario). Similarly, as expected, the effect of predictions on perception was also observed when examining only trials with stereotypical facial expressions that did not match the emotion category evoked by the scenario (non-match trials), and also when examining only trials with neutral facial expressions (where predictability ratings may be expected to fall between the match and non-match trials on average). This observation reveals that the impact of predictability on

[6] Additional exploratory analyses comparing cross-valence and within-valence violations can be found in the supplemental materials.

social perception goes beyond extreme violations of stereotypical faces, holding when deviations from expectations are more nuanced and individualized.

Experiment 4: Generalizability

In Experiment 4, we sought to replicate the results of Experiment 2 and extend our findings from Experiment 3 to trustworthiness by explicitly investigating whether individual facial expression predictions also underlie perceptions of trustworthiness. As in Experiment 3, we used predictability ratings on a trial-by-trial basis as a measure of the degree of subjective prediction fulfillment, and examined the relationship between prediction fulfillment and judgments of trustworthiness. We predicted that individuals would be judged as more trustworthy when their facial expressions fulfilled the perceiver's predictions, and that predictability ratings would mediate the impact of stereotypical matching on judgments of trustworthiness (as it did for judgments of likability in Experiment 3). Similar to Experiment 3, we also performed additional analyses separately for match trials, in which only 'appropriate', stereotype-congruent, facial expressions were presented, to rule out the possibility that the impact of predictions on social perception is driven by non-match trials in which a blatant stereotype/norm violation occurred.

Critically, in Experiment 4, we also assessed whether perceived predictability impacted social judgments in a real-world situation of high consequence: the 2016 United States presidential election. To do so, we asked participants to rate the perceived predictability, likability and trustworthiness of the candidates of the two major political parties (Hillary Clinton and Donald Trump). In the context of a political campaign, basic social judgments of a candidate (i.e., likability or trustworthiness) are critical factors that shape voting behavior and could sway the results of an election (Kinder, 1983; Miller & Miller, 1976; Rosenberg & Sedlak, 1972). Consistent with our findings in Experiments 1–3, we hypothesized that candidates' perceived predictability would be positively related to perceptions of the candidates' likability and trustworthiness.

Furthermore, both presidential candidates consistently violated various stereotypes throughout the electoral campaigns. Hillary Clinton was an unconventional presidential candidate because of her gender; she was the only female presidential candidate in American history to be nominated by a major political party. Donald Trump was an unconventional presidential candidate in that he lacked a background in politics and consistently violated expectations of the decorum with which a presidential candidate is expected to behave (e.g., by repeatedly using non-normatively negative language; for a review, see Conway, Repke, & Houck, 2017). Thus, we also explored whether the degree to which participants' expected stereotypical facial expressions (measured through our paradigm) predicted social judgments of the candidates or voting behavior. That is, we wanted to explore whether individuals whose predictions are strongly tied to stereotype consistency/violation within our task are also particularly sensitive to stereotype consistency/violation in a real-world context. We used ratings of predictability in our task to assess the degree to which participants' expected stereotypical expressions: when a perceiver experienced non-matched faces as less similar to their expectations (e.g., a pouting face after a happy scenario was given a lower rating of predictability), this indicates that their facial expression predictions were fairly stereotypical (e.g., the participant was likely expecting a smiling face after a happy scenario); by contrast, when a perceiver experienced non-matched faces as more similar to their expectations, this indicates that their facial predictions were less stereotypical. We utilized these metrics to examine whether the stereotypicality of participants' predictions related to their perceptions of the candidates. We hypothesized that individuals who made more stereotypical facial expression predictions would like and trust the candidates less and would even be less likely to vote for them, given that the candidates routinely violated various stereotypes/norms about presidential candidates throughout their campaigns.

Finally, we tested the generalizability of our effect by recruiting Experiment 4 participants from Amazon Mechanical Turk, which provides an older and more diverse population than typical university-based samples (Ross, Zaldivar, Irani, & Tomlinson, 2010). To the extent that our hypotheses for Experiment 4 are supported, it suggests that our findings may generalize to how people make real-world social judgments with widespread consequences.

Method

Participants—Participants were 90 young adult United States citizens recruited through Amazon Mechanical Turk (Mean Age±SD: 31±5 y.o., 38 female, gender missing for one participant who did not report it). Sample size was based on Experiment 3. It was adjusted (increased) in order to account for the decreased number of trials of the task and anticipated increases in response attrition for participants completing the task online compared to in the controlled lab environment. Participants received $10 or $5 for their participation. All participants had HIT approval ratings of at least 95%. To protect against negligent participation, only individuals who responded to at least 2/3 of the experimental trials within the task were remunerated for their participation and included in the analyses[7]. Sixteen participants identified themselves as Republicans, 49 identified themselves as Democrats, and 25 did not identify themselves as Republicans or Democrats. The majority of participants (82 out of 90) were registered voters.

Materials and Procedure—For Experiment 4, instructions and stimuli were presented using Qualtrics online research platform. Participants performed the same task as in Experiment 3 but with the following changes: (i) the total number of trials was 39 (3 practice trials: 2 female, 1 male; 36 experimental trials: 18 female, 18 male); (ii) only sad, fear and happy faces were presented after the scenario (no neutral faces); (iii) only the mouth closed version of the faces was used; and (iv) participants firstly rated how similar the target person looked to what they expected (predictability rating) and then how trustworthy the target person was (trustworthiness rating) on scales from 1 to 4 (1=not at all similar/4=very similar; 1=untrustworthy/4=very trustworthy). Eight trials (in total) were excluded from analyses due to the presentation of a repeated target person caused by a detected error in the task program. Additionally, at the end of the task, participants responded to the following questions regarding their political preferences and their evaluations of the presidential candidates for the then upcoming 2016 presidential election in the United States: (1) Do you consider yourself a Republican? (yes/no); (2) Do you consider yourself a Democrat? (yes/

[7] Informed consent was collected for 113 participants but only 90 met this criterion.

no); (3) Are you a registered voter? (yes/no); (4) and (5) How likely are you to vote for Donald Trump [Hillary Clinton]? (1=unlikely, 4=very likely); (6) and (7) How predictable is Donald Trump [Hillary Clinton]? (1=unpredictable, 4=very predictable); (8) and (9) How likable is Donald Trump [Hillary Clinton]? (1=unlikable, 4=very likable); (10) and (11) How trustworthy is Donald Trump [Hillary Clinton]? (1=untrustworthy, 4=very trustworthy). The order of the questions regarding the two presidential candidates was counterbalanced across participants. Means, standard deviations, and inferential statistics comparing ratings of Trump and Clinton in our sample can be found in Table S3 in the supplemental materials.

Results

Mediational Analyses—Data were again analyzed using HLM. Consistent with findings from Experiments 1–3, our HLM analysis revealed that individuals exhibiting matched facial expressions were rated as significantly more trustworthy (M=3.13, SE=.06) than those exhibiting non-matched facial expressions (M=2.23, SE=.05), t(89)=15.56, p<.001. As expected, this analysis also revealed that individuals exhibiting matched facial expressions were rated as significantly more predicted (M=3.13, SE=.06) than those exhibiting non-matched facial expressions (M=1.68, SE=.06), t(89)=25.95, p<.001. Crucially, analyses also revealed that the relationship between face match condition and trustworthiness was significantly mediated by ratings of predictability. Predictability ratings explained 85.7% of the relationship between face match condition and ratings of trustworthiness and remained a significant predictor of trustworthiness ratings in this model (B=.52, SE=.03, t(89)=16.62, p<.001). A Sobel test confirmed significant mediation (Z=13.99, p<.001). However, facial expression match condition also remained a significant predictor of trustworthiness ratings in this model (B=.13, SE=.03, t(89)=4.24, p<.001), suggesting that ratings of predictability only partially mediated this relationship.

Predictability Predicts Trustworthiness Ratings—In addition, we again utilized HLM to examine whether within-person differences in ratings of predictability significantly predicted within-person differences in perceptions of trustworthiness, ignoring match condition. This analysis revealed that individuals that portrayed expressions that were rated as more predicted by a perceiver were also rated as more trustworthy (B=.56, SE=.03; t(89)=18.12, p<.001). As in Experiment 3, we also found that this effect reflected more nuanced changes in predictability and was not simply driven by larger changes in predictability across face match conditions (matched and non-matched). That is, ratings of predictability predicted ratings of trustworthiness within-subjects when looking at just the matched trials (B=.48, SE=.03; t(89)=14.84, p<.001), and when looking at just the non-matched trials (B=.59, SE=.03; t(89)=15.61, p<.001), though the effect was significantly stronger within the non-matched than matched trials (B=−.11, SE=.03; t(89)=3.75, p<.001). See Figure 2b[8].

Judgments of Presidential Candidates—To assess whether predictability was a determinant for perceptions of likability and trustworthiness in situations where such social perceptions have high real-world consequence, we regressed the likability and

[8] Additional exploratory analyses comparing cross-valence and within-valence violations can be found in the supplemental materials.

trustworthiness ratings of each presidential candidate (Clinton and Trump) onto ratings of their predictability. As hypothesized, perceptions that the candidate was more predictable predicted greater perceived likability of the candidate ($\beta_{Clinton}$=.17, *F*(1,89)=2.73, *p*=.102; β_{Trump}=.29, *F*(1,89)=8.06, *p*=.006) and greater perceived trustworthiness of the candidate ($\beta_{Clinton}$=.23, *F*(1,89)=4.98, *p*=.028; β_{Trump}=.23, *F*(1,89)=4.86, *p*=.030). However, the relationship between predictability and likability failed to reach traditional levels of significance for Clinton.

Stereotypicality of Predictions and Real-World Judgments—Further, we examined whether reactions to stereotypical prediction violations within our paradigm were able to predict reactions to individuals violating stereotypes in real-world situations of high consequence. Specifically, we assessed whether holding more stereotypical facial expression predictions (assessed as lower predictability ratings for non-matched faces) predicted perceptions of trustworthiness and likability for the 2016 US presidential candidates, both of whom regularly violated behavioral and/or stereotype-based norms.

As predicted, a series of linear regressions revealed that lower ratings of predictability for non-matched faces within our task (i.e., holding more stereotypical facial expression predictions) significantly predicted perceiving Clinton as less likable (β=.37, *F*(1,89)=13.64, *p*<.001) and less trustworthy (β=.30, *F*(1,89)=8.48, *p*=.005), and even predicted a decreased likelihood of voting for her (β=.25, *F*(1,89)=5.67, *p*=.019). For Trump, holding more stereotypical facial expression predictions (i.e. lower predictability ratings of non-matched faces within our task) significantly predicted perceiving Trump as less trustworthy (β=.23, *F*(1,89)=4.85, *p*=.030), but they failed to predict ratings of his perceived likability (β=.06, *F*(1,89)=0.27, *p*=.602) or the likelihood of voting for him (β=.07, *F*(1,89)=0.43, *p*=.513).

Discussion

In Experiment 4, we extended our findings from Experiment 3 to trustworthiness and showed that facial expression predictions influence social judgments for a more diverse group of participants, demonstrating the generalizability of the effect. Critically, this study also demonstrated that the perceived predictability of individuals in the real world, the 2016 US presidential candidates, influenced judgments of their likability and trustworthiness—two important social judgments that can impact voting decisions (Kinder, 1983; Miller & Miller, 1976; Rosenberg & Sedlak, 1972). In line with our hypotheses, for both Trump and Clinton, the candidates were judged as more likable and more trustworthy when they were thought to be more predictable. However, the relationship between predictability and likability was only marginally significant for Clinton.

Both Clinton and Trump consistently violated stereotypes about presidential candidates throughout the presidential campaign (e.g., gender stereotypes in the case of Clinton and conventionality stereotypes in the case of Trump). Thus, in Experiment 4 we also examined whether holding more stereotypical facial expression predictions within our experimental task predicted social perceptions of these two candidates, who regularly violated behavioral norms and expectations throughout their presidential campaigns. As expected, we found that participants who held more stereotypical facial expression predictions perceived Clinton and

Trump as less trustworthy than those who held less stereotypical facial expression predictions. Holding stereotypical facial expression predictions was also associated with perceiving Clinton (but not Trump) as less likable and being less likely to vote for her. A possible explanation for these partially different patterns may relate to whether the norm/stereotype violations of the two candidates were, or were perceived to be, similarly expectancy-violating and equally relevant to the social context. That is, although both candidates did violate normative expectancies for presidential candidates, we did not assess the degree to which their violations were perceived as equivalent by our sample. Of note, Clinton was rated as significantly more predictable than Trump by our sample (see Table S3 in the supplemental materials). Thus, it is possible that a violation of gender stereotypes (i.e., a woman running for United States president) has less impact on how predictable a person is judged to be than constant violations of behavioral and social norms. At the same time, a violation of gender stereotypes might be considered a more serious norm violation in the context of a presidential election than a violation of conventionality, such that holding more stereotypical expectations would negatively impact Clinton more than Trump. Future research should examine the extent to which such effects are moderated by awareness of stereotype violations and perceptions of the severity or relevance of such violations. In addition, our sample consisted of more individuals identifying themselves as Democrats than as Republicans (or as not affiliated with either party), which may have contributed to Clinton being rated more positively than Trump overall in the present study. Overall, however, our findings demonstrate that our task has real-world implications, predicting perceptions about presidential candidates and even real-world decisions of great import (i.e., voting behavior).

Thus far, we have demonstrated that the extent to which social information blatantly confirms or violates predictions influences social judgments (Experiments 1 and 2), that this effect holds for more nuanced prediction violations that do not necessarily involve blatant norm violations, that this effect is mediated by explicit predictions of the perceivers (Experiments 3 and 4), and that this effect extends to real-world social evaluative judgments of high import (Experiment 4). However, we have yet to examine potential mechanisms for this effect. In the remaining studies, we examine the role of changes in felt affect and processing fluency as two potential causal explanations for our findings.

Experiment 5: Predictions and Reported Affect

Findings from Experiment 4 demonstrate that expectations about facial expressions are important drivers of social perception that extend beyond laboratory-based measures to influence real-world social judgments of high consequence. However, the mechanism by which predictions influence social perception remains unexplored. In the present experiment, we examine whether changes in felt (conscious) affect offer a viable explanation for our previous findings.

According to affect-as-information theory (Clore et al., 2001; Clore & Huntsinger, 2007; Schwarz & Clore, 1983), individuals utilize their feelings as a source of information when making decisions or social judgments, particularly when they are unsure of the source of their feelings. Indeed, previous research suggests that incidental affect (i.e., affect unrelated to the decision at hand) can influence how individuals perceive and respond to social others, including judgments about whether they pose a threat (Baumann & Desteno, 2010; Wormwood, Lynn, Barrett, & Quigley, 2016), judgments about whether to cooperate with them or offer assistance (Bartlett, Condon, Cruz, Baumann, & Desteno, 2012; Isen & Levin, 1972), and even judgments about whether they should be admitted into medical school (Redelmeier & Baxter, 2009). Most relevant to the current investigation, previous research has also demonstrated that incidental affect can influence the perceived trustworthiness and likability of others (Anderson, Siegel, White, & Barrett, 2012). Thus, it is possible in our experiments that a perceiver may experience positive affect when his or her expectations about a facial expression are fulfilled, and that they, in turn, misattribute those positive feelings as a reaction to the individual being perceived, leading them to evaluate the target person more positively (i.e., as more likable and trustworthy). If so, predictions may drive social perception when perceivers misattribute the positive or negative affective feelings that result from a met or violated prediction, respectively, as their affective reaction to the individual being perceived.

In order to test this possibility, we conducted an experiment using a paradigm nearly identical to that of Experiments 1 and 2, but instead of asking participants to make judgments about the individuals being displayed, participants were asked to self-report their own affective feelings on each trial (i.e., their felt valence and arousal). To the extent that changes in conscious affective feelings causally underlie our findings, we would expect to see more positive affect reported on matched trials than on non-matched trials, regardless of the specific emotion category being evoked by the scenario or the specific face emotion category being displayed. Conversely, if changes in felt affect are associated with the emotion categories evoked by the scenarios (e.g., greater self-reported positive affect following happy scenarios than fear and sad scenarios) or with the emotion categories of the stereotypical facial expressions presented (e.g., greater self-reported positive affect following stereotypical happy expressions than stereotypical fear and sad expressions), this pattern of findings would suggest that affective misattribution is not a viable mechanism underlying the influence of predictions on social perception.

Method

Participants—Participants were 37[9] young adults recruited from Northeastern University (Mean Age±SD: 19±1 y.o., 21 female). All participants reported normal or corrected-to-normal visual acuity, were native English speakers and received course credit for their participation. The target sample size of *n*=37 was based on the results of Experiments 1 and 2.

Materials and Procedure—Materials and procedure were identical to Experiments 1 and 2 except that, instead of rating the likability or trustworthiness of the target person, participants rated how pleasant they felt (valence rating) followed by how activated they felt (arousal rating) on 4-point scales (1=unpleasant/4=very pleasant; 1=deactivated/4=very activated).

[9] Informed consent was collected for 38 participants in Experiment 5 but 1 was excluded from analyses due to noncompliance with the inclusion criterion of being native English speaker.

Results

Valence Ratings—A 2 (face match: matched, non-matched) by 3 (face emotion category: sad, fearful, happy) repeated-measures ANOVA on valence ratings revealed a significant main effect for face emotion category, *F*(2, 72)=93.40, *p*<.001, η_p^2 =.722. Bonferroni tests demonstrated that participants reported significantly more positive affect on trials with happy faces (*M*=2.58, SE=.04) than on trials with either sad faces (*M*=1.87, SE=.05), *p*<. 001, or fear faces (*M*=1.96, SE=.06), *p*<.001 (sad v. fear: *p*=.077). Consistent with an affective-misattribution account, this analysis additionally revealed that participants reported more positive affective feelings after matched faces (*M*=2.21, SE=.03) than non-matched faces (*M*=2.07, SE=.06), *F*(1, 36)=5.98, *p*=.020, η_p^2 =.142.

Inconsistent with an affective-misattribution account, however, this analysis also revealed a significant interaction between face match and face emotion category on these ratings, *F*(2, 72)=233.57, *p*<.001, η_p^2 =.866. To examine this interaction, we conducted a series of repeated-measures ANOVAs, one for each face emotion category, with face match as the within-subject factor. For trials with stereotypical happy expressions, participants reported significantly more positive affect on matched trials than on non-matched trials, *F*(1, 36)=403.93, *p*<.001, η_p^2 =918. However, the opposite was true for trials with stereotypical fear and sad expressions (*F*(1, 36)=42.31, *p*<.001, η_p^2 =.540 and *F*(1, 36)=79.63, *p*<.001, η_p^2 =689, respectively), where participants reported significantly more positive affect on non-matched trials than on matched trials. See Figure 3a.

Arousal Ratings—A 2 (face match: matched, non-matched) by 3 (face emotion category: sad, fearful, happy) repeated-measures ANOVA failed to reveal a significant main effect of face emotion category on self-reported arousal, *F*(2, 72)=1.99, *p*=.145, η_p^2 =.052. Inconsistent with an affective-misattribution account, this analysis also failed to reveal a significant main effect of face match on self-reported arousal, *F*<1. Also inconsistent with an affective misattribution account, the interaction between face match and face emotion category reached significance, *F*(2,72)=4.09, *p*=.021, η_p^2 =.102. To examine this interaction, we conducted a series of repeated-measures ANOVAs, one for each face emotion category, with face match condition as the within-subjects factor. This analysis revealed that there were no differences in self-reported arousal across matched and non-matched trials for either happy or fear faces (*F*(1, 36)=1.40, *p*=.245, η_p^2 =.037 and *F*(1, 36)=1.74, p=.195, η_p^2 =.046, respectively). However, participants reported significantly lower arousal on matched trials (M=2.46, SE=.09) than on non-matched trials (M=2.69, SE=.11) for sad faces, *F*(1, 36)=4.28, *p*=.046, η_p^2 =.106. See Figure 3b.

Felt Affect By Emotion Scenario—In light of these results, we hypothesized that felt affect was driven by scenario emotion category rather than by predictions. In order to test this, we conducted two repeated-measures ANOVAs, one for ratings of valence and one for ratings of arousal, with emotion scenario condition as the within-subject factor. This analysis revealed a significant main effect of emotion scenario on ratings of valence, *F*(2, 72)=247.75, *p*<.001, η_p^2 =.873. Post-hoc Bonferroni comparisons revealed that participants reported significantly higher (more positive) valence on trials with happy scenarios (*M*=3.27, SE=.07) than sad scenarios (*M*=1.55, SE=.05), *p*<.001, or fear scenarios (*M*=1.62, SD=.06), *p*<.001. There were no differences in reported valence between trials with sad and fear scenarios, *p*=.207. See Figure 3c. This analysis also revealed a significant effect of emotion scenario condition on ratings of arousal, *F*(2, 72)=8.24, *p*=.001, $\eta_p{}^2$ =.186. Post-hoc Bonferroni comparisons revealed that participants reported significantly lower arousal ratings on trials with sad scenarios (*M*=2.42, SE=.07) than happy scenarios (*M*=2.62, SE=.09), *p*=.045, or fear scenarios (*M*=2.72, SE=.08), *p*<.001. There were no differences in reported arousal between trials with happy and fear scenarios, *p*=.590. See Figure 3d.

Discussion

The pattern of results observed suggests that changes in conscious affect are unlikely to underlie the effect of predictions on social perception. Across the four previous studies, we found that predicted faces were evaluated as more likable and trustworthy than non-predicted faces, regardless of the emotion category. In order for affective misattribution to account for these findings, Experiment 5 would need to have revealed that participants felt more affectively positive when presented with predicted faces than unpredicted faces across all face emotion categories. Instead, however, the present experiment revealed that changes in felt affect differed across face emotion categories and scenario emotion categories as opposed to across face match and non-match conditions. That is, while participants did experience changes in both felt pleasantness and activation across different conditions in Experiment 5, the participants’ affect changed in response to the affective value of scenarios and facial expressions rather than in response to the predictability of the facial expressions (i.e., whether the facial expression matched the emotion category evoked by the preceding scenario or not). The pattern of results appears to be best explained by a series of main effects that directly follow from existing literature on emotion and emotion perception (e.g., Russell & Pratt, 1980): 1) participants reported more positive affect following happy scenarios than sad or fear scenarios (e.g., Wilson-Mendenhall et al., 2013); 2) participants reported more positive affect following happy faces than pouting (sad) or startled (fear) faces (e.g., Wild, Erb, & Bartels, 2001); and (3) participants reported lower arousal following sad scenarios than happy or fear scenarios (e.g., Wilson-Mendenhall et al., 2013). Thus, our findings suggest that affective misattribution is unlikely to be the mechanism underlying the impact of predictions on social perception, as changes in felt affect accompanying prediction fulfillment and violation are not consistent across emotion categories.

Experiment 6: Processing of Predicted Faces

Another possible underlying mechanism for the impact of predictions on social perception, and one that should be consistent across emotion categories, is that prediction leads to a form of processing fluency (Winkielman et al., 2003) whereby the processing of predicted stimuli may be facilitated. According to the literature on processing fluency, ease of processing is associated with more positive evaluations (Winkielman et al., 2003). Thus, within this view, and consistent with our findings in Experiments 1–4, if the processing of predicted facial expressions is facilitated, then individuals displaying predicted expressions should be perceived more positively than those displaying non-predicted expressions. From a predictive coding perspective, we would expect facilitated processing of predicted stimuli, as the representation of expected sensory input has been shown to be highly efficient (Jehee, Rothkopf, Beck, & Ballard, 2006; Kok, Jehee, & de Lange, 2012).

A first step towards testing this explanation involves exploring whether the processing of predicted faces is indeed facilitated. In Experiment 6, we tested whether predicted (matched) faces exhibit privileged processing using a modified version of the paradigm from Experiments 1–5, in which the final facial expression was initially suppressed from conscious awareness using Continuous Flash Suppression (CFS; Tsuchiya & Koch, 2005). Instead of making person judgments on each trial, participants in Experiment 6 reported when they could first see each face as the contrast of the face image was slowly raised over the course of the trial, eventually breaking the suppression effect achieved through CFS. We hypothesized that expected facial expressions would be processed more efficiently than non-predicted facial expressions, and that this would be true across emotion categories. Thus, we predicted that participants would report seeing facial expressions earlier on trials where the stereotypical expression matched the preceding scenario’s emotion category than on trials where it did not.

Method

Participants—Participants were 42[10] (24 female) young adults recruited from Northeastern University and the surrounding Boston community through fliers and Craigslist.com advertisements (Mean Age±SD: 21±5 y.o.; age missing for 4 undergraduate participants from Northeastern University). Desired sample size was estimated based on previous work using a similar binocular suppression technique (Anderson, Siegel, Bliss-Moreau, & Barrett, 2011; Study 2). Participants received course credit or $10 for their participation. All participants reported normal or corrected-to-normal visual acuity. Participants wearing glasses were excluded from the analyses given that they interfere with the proper function of the mirror stereoscope, a device used to visualize stimuli in this experiment. Some of the participants (*n*=25) performed additional, unrelated tasks as part of a different study.

Materials and procedure—Instructions and stimuli were presented using Matlab R2011a running on a Dell Optiplex 980 and a 17-inch Dell LCD flat-screen monitor (resolution of 1280×1024). Participants sat with their head placed on chin and forehead rests and viewed stimuli displayed on the screen through a mirror stereoscope at a distance of approximately 47 cm. The stereoscope allows for the simultaneous presentation of a different stimulus to each eye. Stimuli were presented in a gray scale surrounded by a white frame to facilitate fusion. The task consisted of 3 practice trials followed by 45 experimental trials. Participants were requested to remain still during each trial with their forehead and chin on the rests. Prior to the task, eye dominance was determined for each participant using the Dolman method (Dolman, 1919; Fink, 1938), as research has shown that suppression is more effective under CFS when the image to be suppressed is presented to the non-dominant eye.

[10] Informed consent was collected for 51 participants but 9 were excluded from analyses because they wore glasses (n=8) or the stereoscope could not be calibrated and no data was collected (n=1).

As in the experiments reported above, on each trial, a photograph of a neutral face of a target person was displayed for 5 s to both eyes (Figure 4a). A scenario was then displayed for 20 s also to both eyes. Participants were asked to imagine the facial expression of the target person while reading. After a brief fixation screen (0.5 s), the dominant eye was presented with a series of high contrast Mondrian patterns. The patterns alternated at a rate of 10 Hz for a maximum duration of 10 s. These patterns decreased in contrast linearly, updated every 10 ms, over the first 5 s from full contrast to a final contrast of $\log_{10}$ contrast = −1. At the same time, the non-dominant eye was presented with an initially low-contrast face of the target person, either portraying a neutral facial expression or a stereotypical facial expression for one of the three emotions categories (e.g., a pout depicting sadness). The contrast of the suppressed face ramped up linearly, updated every 10 ms, over the first 1 s of the trial, from a very low initial contrast ($\log_{10}$ contrast = −3) to an ending contrast of $\log_{10}$ contrast = −0.5. The face was presented in one of the four corners of the image and, as in Experiments 1–3, could "match" the evoked emotion (matched faces; 21 trials), be neutral (12 trials) or "not match" the evoked emotion (non-matched faces; 12 trials). As in Experiments 1 and 2, we focused our analyses on comparing responses to the same facial expressions when they either matched or did not match the preceding emotion scenario (i.e., neutral faces were not included in the analyses).

When using CFS, participants typically experience seeing initially only the Mondrian patterns presented to the dominant eye, and then the face presented to the non-dominant eye becomes visible once it is sufficiently high contrast (and Mondrian patterns are sufficiently low contrast) to break the suppression effect. Participants were instructed to press the spacebar as soon as they saw the face (within 10 s) and reaction time (RT) was collected as the outcome measure. Participants were then requested to report in which of the four corners of the image the face was located (unlimited time). The number of errors was expected to be low; this question served as a control to ensure that participants performed the task correctly and did actually see the faces. There was an inter-trial interval of 3s.

We used the same set of photographs as for Experiments 1–5. In addition, we used a modified version of these photographs, manipulated with Adobe Photoshop, for faces presented during CFS (see Figure S1). A full-contrast black and white image of each face was cropped into an oval shape so that only the facial features remained and the hair, ears, and neck were all removed. This cropped image was placed on a neutral gray background and the edges of the oval were blended with the background. Finally, each image was cropped into a 113×113 pixel square such that the eyes, eyebrows, nose, mouth, and chin all remained in the square (see Figure S1 for images of sample stimuli). Twelve different identities were included (6 female, 6 male), each of which was used in 3 or 4 trials to yield the total 45 experimental trials. We used the same set of scenarios as in the experiments above.

Results

As expected, there were very few trials where participants incorrectly reported the location of the target face (Mean Accuracy±SD: 0.99±0.02). A 2 (face match: matched, non-matched) by 3 (face emotion category: sad, fear, happy) repeated-measures ANOVA revealed facilitated processing for expected facial expressions. Face match condition significantly impacted reaction time (RT), with matched faces yielding faster RTs (*M*=2.74 s, SE=.15 s) than non-matched faces (*M*=2.87 s, SE=.16 s), *F*(1, 41)=5.37, *p*=.026, η_p^2 =. 116. See Figure 4b.

This effect did not differ across emotion category (i.e., the interaction was not statistically significant; *F*(2, 82)=2.47, *p*=0.091, η_p^2 =.057). However, the analysis also revealed a significant main effect of face emotion category on RTs, *F*(2, 82)=7.78, *p*<.001, η_p^2 =.159. Bonferroni comparisons revealed that fear faces yielded faster RTs (*M*=2.64, SE=.13) than either sad (*M*=2.94, SE=.17), *p*=.001, or happy faces (*M*=2.84, SE=.17), *p*=.025, which did not differ, *p*=.675. This is in agreement with prior work reporting preferential processing for stereotypical fear faces (Yang, Zald, & Blake, 2007), most likely because faces stereotypically portray fear with widened eyes showing a lot of sclera, creating a higher contrast in fear facial expressions than other facial expressions displaying less sclera (see, e.g., Hedger, Adams, & Garner, 2015).

Discussion

Experiment 6 shows that the processing of predicted faces is facilitated compared to the processing of non-predicted faces. Given that the literature on processing fluency suggests that ease of processing is associated with more positive evaluations (Winkielman et al., 2003), these findings suggest that processing fluency could be a potential underlying mechanism for the observed effect of facial expression predictions on social perception. These findings are also consistent with the neuroscience perspectives on predictive coding that underlie this research, which have reported highly efficient low-level processing for predicted stimuli (Jehee et al., 2006; Kok et al., 2012).

It is worth noting, however, that it remains unclear the exact level at which the facilitated processing observed here occurs; there is continued debate concerning whether CFS paradigms can be used to isolate non-conscious or pre-conscious processing of suppressed stimuli or whether they merely capture differences in conscious processing of stimuli after the suppressed stimulus has already reached awareness (see, e.g., Stein & Sterzer, 2014; Yang, Brascamp, Kang, & Blake, 2014). Thus, while our results clearly indicate that predicted faces are processed more efficiently than non-predicted faces, the present paradigm is unable to address whether this facilitation extends to unconscious processing.

General Discussion

Taken together, data across these six experiments provide evidence that facial expression predictions strongly contribute to our experience of other people. We demonstrated that individuals are evaluated as more likable and more trustworthy when displaying predicted facial expressions than non-predicted facial expressions (Experiments 1–4). Moreover, the effects of facial expression predictions on social perception proved broad, emerging across the three emotion categories explored; even for instances of negative emotions, participants liked a person better for expressing negatively when it was expected (Experiments 1–4). Importantly, the observed effects also extended beyond a laboratory setting to ratings of a real-world situation, namely, the 2016 United States presidential elections: presidential candidates were evaluated as more likable and trustworthy if they were also perceived as more predictable (Experiment 4). Interestingly, we also found evidence suggesting that sensitivity to stereotypical prediction violations may represent a stable individual difference: participants who exhibited more stereotypical facial expression predictions within our experimental task also liked and trusted a presidential candidate who violated gender norms (Clinton) less, and were even less likely to vote for her (Experiment 4). Finally, in a series of two experiments, we demonstrated that processing fluency, rather than affective misattribution, appears to be a better candidate for the mechanism underlying the effect of facial expression predictions on social perception (Experiments 5 and 6); predicted faces were processed more efficiently, which could in turn lead to more positive evaluations.

Our experimental design allowed us to examine the role of predictions in social perception directly, moving beyond the existing literature that has typically studied the role of predictions only indirectly in a number of ways. First, we were able to assess the power of complex, multimodal predictions, beyond perceptual priming (e.g., mere exposure; see Zajonc, 1968), as predictions were generated by participants imagining the target individual’s facial expression, not by presenting images of potential facial expressions to shape predictions perceptually. Our findings also go beyond emotion congruence effects (e.g., Mehrabian & Wiener, 1967), where the source of predictions and prediction violations are simultaneously presented by the target individual. In our paradigm, predictions were evoked, not by the target person in the moment, but by the emotion scenario, and predictions were established prior to the presentation of the target person’s facial expression. The findings also demonstrate the effect of predictions on social perception above and beyond stereotype violation effects (e.g., Heilman, Wallen, Fuchs, & Tamkins, 2004; Heilman, 2001), as, importantly, the effect of individual predictions on social perception was observed both across and within trials where the facial expressions were stereotypical matches or mismatches to the emotion evoked by the preceding scenario. Participants liked and trusted a target individual more when the displayed facial expression more closely matched their own predictions, even when examining only trials where all displayed facial expressions would be considered a stereotypical match to the predictions evoked by the preceding emotion scenario (i.e., smiling face after a happy scenario).

The present experiments leave a number of intriguing questions open for future exploration. First, we here focused on emotion categories that varied across valence and arousal levels, but that might all be considered to have an affiliative social quality. Future research could assess whether the effects described here are also observed for other emotion categories that are stereotypically non-affiliative, such as anger or disgust. Based on the strong across-emotion effects observed here, we hypothesize that facial expression predictions would also drive social perception for stereotypically non-affiliative emotion displays, such that a scowling face would still be evaluated more positively when displayed following a matching (i.e., angering) scenario, like being stuck in traffic. Still, this remains an empirical question to be explored.

Second, while our findings from Experiments 3 and 4 demonstrate that the role of predictions in social perception are not limited to instances involving blatant mismatches (e.g., cross-valence mismatches, like smiling in a horrifying situation), future research should further examine social perception under these contexts involving extreme prediction violations. In particular, future work could assess whether inferences about the person's mental health or intellectual capabilities moderate the impact of predictions on evaluative judgments. Conversely, there is also evidence that norm violators may actually be preferred in specific situations in which unpredictable, inappropriate behavior, may be beneficial for the task at hand (Kiesler, 1973). For example, it is hypothesized that many voters favored Donald Trump's language choices *because* they violated conventionality and political-correctness norms (Conway et al., 2017). Situational demands could be manipulated in future studies in order to examine the factors that might lead perceivers to form more favorable impressions of prediction-violating behavior, as well as whether such situations extend to emotion expression displays.

Future research should also further examine the underlying mechanisms of the effects observed. Of specific interest is whether processing fluency mediates the impact of predictions on social perception. Although a critical underlying condition was established here—predicted facial expressions did indeed exhibit evidence of facilitated processing—future work should more directly examine whether prediction-facilitated processing of facial expressions leads to more positive social perceptions of the individuals displaying those expressions. Given the existing literature on the effect of other kinds of processing fluency on social evaluative judgments (Winkielman et al., 2003), we would anticipate this relationship to hold for the processing and evaluation of facial expressions as well. However, in future studies, researchers could examine whether the effect of predictions on social perception is eliminated under conditions where processing ease is manipulated or held constant.

In addition, future studies could further characterize the perceptual impact of predictions. Emotion context is known to importantly influence perceptual emotion categorization, with individuals using situational information over facial information to judge emotion (Carroll & Russell, 1996). An emotion category is a heterogeneous population of instances that are situated (i.e., tailored by the environment) (Barrett, 2017). It is well known that people do more than widen their eyes in fear – they can grimace in fear, squint in fear, cry in fear, and even laugh in the face of fear. In the context of the present research, it is possible that non-matched facial expressions are categorized more often as displaying the emotion evoked by the scenario than the same facial expressions would be outside that context (i.e., with no scenario or a matched scenario). Future research could explore whether facial expressions on non-match trials are categorized more often as matching the emotion category of the scenario rather than that of the face when predictability ratings are high as compared to low, thus reflecting that predictions were able to 'switch' emotion categorization and diminish potential prediction violation. Relatedly, future studies could address whether facial expressions are perceived as more emotionally intense when they are predicted (e.g., match trials/high predictability ratings) than when they are not.

Finally, the current research studied facial expression predictions as a paradigmatic case, allowing us to explore the role of socially-relevant predictions on social perception. However, predictive coding perspectives posit that predictions broadly impact experience, perception, and action at multiple levels (e.g., Barrett & Simmons, 2015; Chanes & Barrett, 2016; Clark, 2013; Friston, 2005; Hohwy, 2013). Thus, we expect the effect of predictions on social perception to extend to a broad variety of socially-relevant phenomena. Future research should examine the role of other sources of predictions on social perception beyond emotion scenarios, including verbal and nonverbal cues (e.g., body language, immediate environment, a perceiver's own emotions, etc.), as well as other targets for those predictions (other than facial expressions). Future research should also examine the role of predictions in other forms of socially-relevant perception and action; for example, whether predictable individuals are more likely to be recipients of pro-social behavior, like helping or cooperation, and whether prediction-violating individuals are more likely to be recipients of harm or ostracism.

Conclusion

The present findings have significant consequences for everyday life. They demonstrate the importance of predictions—particularly of facial expressions—in the emotion and social domain, including situations of great consequence, such as evaluating the candidates in an upcoming presidential election. Predictions drive how we perceive others. Being able to predict someone well makes us like him/her more, which suggests that predictions may be a mediator of empathy. The deep impact of facial expression predictions on social perception (from lower-level perceptual processing to abstract higher-level evaluative judgments) may contribute to a mechanistic understanding of why we dislike stereotype violators, e.g., why we distrust a political leader who shows inappropriate affect following a terrorist attack (e.g., Bucy, 2000) or why women may be less liked in positions of power or other positions where they violate expectations based on gender stereotypes (Heilman, Wallen, Fuchs, & Tamkins, 2004; Heilman, 2001). Examining predictions as a more general mechanism in social perception may be also relevant for cross-cultural contexts, providing a possible explanation for why people may have a more difficult time liking or trusting one another across cultural boundaries where emotion norms or concepts are not necessarily shared (for reviews of evidence that emotion concepts vary across geographically separate cultures, see Lillard, 1998; Mesquita et al., 1992; Russell, 1991). Finally, the present findings also provide insights into the possible mechanisms for the disruptions in emotion perception that are part of every mental disorder (see, e.g., Bourke et al., 2010; Domes et al., 2009; Leitman et al., 2010) as disruptions of predictions of emotion cues, including facial expressions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants to from the U.S. Army Research Institute for the Behavioral and Social Sciences (W911NF-16-1-0191 to Karen Quigley and J.B.W), the National Science Foundation (BCS-1422327 to Karen Quigley and J.B.W; CMMI 1638234 to Mario Sznaier and L.F.B.), the National Cancer Institute (U01 CA193632 to L.F.B.), the National Institute on Aging (R01 AG030311 to L.F.B. and Brad Dickerson), and a Fyssen Foundation postdoctoral fellowship and Alicia Koplowitz Foundation short-term fellowship to L.C. The views, opinions, and/or findings contained in this paper are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents.

References

Anderson E, Siegel EH, Bliss-Moreau E, Barrett LF. The visual impact of gossip. Science (New York, NY). 2011; 332(6036):1446–8. https://doi.org/10.1126/science.1201574.

Anderson E, Siegel E, White D, Barrett LF. Out of Sight but Not Out of Mind: Unseen Affective Faces Influence Evaluations and Social Impressions. Emotion. 2012; 12(6):1210–1221. https://doi.org/10.1037/a0027514. [PubMed: 22506501]

Aviezer, H., Hassin, R., Bentin, S., Trope, Y. Putting facial expressions back in context; First Impressions. 2008. p. 255-286.Retrieved from http://books.google.com/books?hl=en&lr=&id=poHGCvweVFsC&oi=fnd&pg=PA255&dq=Putting+Facial+Expressions+Back+in+Context&ots=PeDq2XxpwX&sig=uZIoNDThCiPce8uGuYHDmJY3FI

Aviezer H, Hassin RR, Ryan J, Grady C, Susskind J, Anderson A, Bentin S. Angry, disgusted, or afraid? Studies on the malleability of emotion perception: Research article. Psychological Science. 2008; 19(7):724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x. [PubMed: 18727789]

Aviezer H, Trope Y, Todorov A. Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. Science. 2012; 338(6111):1225–1229. https://doi.org/10.1126/science.1224313. [PubMed: 23197536]

Barrett, LF. How emotions are made. Houghton Mifflin; Harcourt: 2017.

Barrett LF, Mesquita B, Gendron M. Context in Emotion Perception. Current Directions in Psychological Science. 2011; 20(5):286–290. https://doi.org/10.1177/0963721411422522.

Barrett LF, Simmons WK. Interoceptive predictions in the brain. Nature Reviews Neuroscience. 2015 May.16 2014. https://doi.org/10.1038/nrn3950.

Bartlett MY, Condon P, Cruz J, Baumann J, Desteno D. Gratitude: Prompting behaviours that build relationships. Cognition & Emotion. 2012; 26(1):2–13. https://doi.org/10.1080/02699931.2011.561297. [PubMed: 21500044]

Baumann J, Desteno D. Emotion guided threat detection: Expecting guns where there are none. web of Personality and Social Psychology. 2010; 99(4):595–610. https://doi.org/10.1037/a0020665.

Bayliss AP, Griffiths D, Tipper SP. Predictive gaze cues affect face evaluations: The effect of facial emotion. European web of Cognitive Psychology. 2009; 21(7):1072–1084. https://doi.org/10.1080/09541440802553490.

Bayliss AP, Tipper SP. Predictive gaze cues and personality judgments: Should eye trust you? Psychological Science. 2006; 17(6):514–520. https://doi.org/10.1111/j.1467-9280.2006.01737.x. [PubMed: 16771802]

Bourke C, Douglas K, Porter R. Processing of facial emotion expression in major depression: a review. The Australian and New Zealand web of Psychiatry. 2010; 44(8):681–96. https://doi.org/10.3109/00048674.2010.496359.

Bucy EP. Emotional and Evaluative Consequences of Inappropriate Leader Displays. Communication Research. 2000; 27(2):194–226. https://doi.org/10.1177/009365000027002004.

Carroll JM, Russell JA. Do facial expressions signal specific emotions? Judging emotion from the face in context. web of Personality and Social Psychology. 1996; 70(2):205–218. https://doi.org/10.1037/0022-3514.70.2.205.

Chanes L, Barrett LF. Redefining the Role of Limbic Areas in Cortical Processing. Trends in Cognitive Sciences. 2016; 20(2):96–106. https://doi.org/10.1016/j.tics.2015.11.005. [PubMed: 26704857]

Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 2013; 36(3):181–204. https://doi.org/10.1017/S0140525X12000477. [PubMed: 23663408]

Clore GL, Gasper K, Garvin E. Affect as information. Handbook of Affect and Social Cognition. 2001:121–144.

Clore, GL., Huntsinger, JR. How emotions inform judgment and regulate thought. Trends in Cognitive Sciences. 2007. https://doi.org/10.1016/j.tics.2007.08.005

Conway LG, Repke MA, Houck SC. Donald Trump as a cultural revolt against perceived communication restriction: Priming political correctness norms causes more Trump support. web of Social and Political Psychology. 2017; 5(1):244–259. https://doi.org/10.5964/jspp.v5i1.732.

Dolman P. The Maddox rod screen test. Transactions of the American Ophtalmological Society. 1919; 17:235–249.

Domes G, Schulze L, Herpertz SC. EMOTION RECOGNITION IN BORDERLINE PERSONALITY DISORDER—A REVIEW OF THE LITERATURE. web of Personality Disorders. 2009; 23(1):6–19. https://doi.org/10.1521/pedi.2009.23.1.6.

Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. Psychol Methods. 2007; 12(2):121–138. https://doi.org/10.1037/1082-989X.12.2.121. [PubMed: 17563168]

Fink WH. The dominant eye: its clinical significance. Archives of Ophtalmology. 1938; 19(4):555–582.

Freeman, JB., Johnson, KL. More Than Meets the Eye: Split-Second Social Perception. Trends in Cognitive Sciences. 2016. https://doi.org/10.1016/j.tics.2016.03.003

Friston K. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci. 2005; 360(1456):815–836. https://doi.org/10.1098/rstb.2005.1622. [PubMed: 15937014]

Gilbert CD, Li W. Top-down influences on visual processing. Nature Reviews Neuroscience. 2013; 14(5):350–63. https://doi.org/10.1038/nrn3476. [PubMed: 23595013]

Gillis RL, Nilsen ES. Consistency between verbal and non-verbal affective cues: a clue to speaker credibility. Cognition and Emotion. 2016 Feb.9931:1–12. https://doi.org/10.1080/02699931.2016.1147422.

Hedger N, Adams WJ, Garner M. Fearful faces have a sensory advantage in the competition for awareness. web of Experimental Psychology: Human Perception and Performance. 2015; 41(6):1748–1757. https://doi.org/10.1037/xhp0000127.

Heerey EA, Velani H. Implicit learning of social predictions. web of Experimental Social Psychology. 2010; 46(3):577–581. https://doi.org/10.1016/j.jesp.2010.01.003.

Heilman ME, Wallen AS, Fuchs D, Tamkins MM. Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks. web of Applied Psychology. 2004; 89(3):416–427. https://doi.org/10.1037/0021-9010.89.3.416.

Heilman ME. Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder. web of Social Issues. 2001; 57(4):657–674. https://doi.org/10.1111/0022-4537.00234.

Hohwy, J. The predictive mind. New York: Oxford University Press; 2013.

Isen AM, Levin PF. Effect of feeling good on helping: cookies and kindness. Journal of Personality and Social Psychology. 1972; 21(3):384–388. [PubMed: 5060754]

Jehee JFM, Rothkopf C, Beck JM, Ballard DH. Learning receptive fields using predictive feedback. web of Physiology Paris. 2006; 100(1–3):125–132. https://doi.org/10.1016/j.jphysparis.2006.09.011.

Kiebel SJ, Daunizeau J, Friston KJ. A hierarchy of time-scales and the brain. PLoS Computational Biology. 2008; 4(11) https://doi.org/10.1371/web.pcbi.1000209.

Kiesler SB. Preference for predictability or unpredictability as a mediator of reactions to norm violations. Journal of Personality & Social Psychology. 1973; 27(3):354–359.

Kinder DR. Presidential Traits Report to the NES Board of Overseers. 1983

Kok P, Jehee JFM, de Lange FP. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron. 2012; 75(2):265–270. https://doi.org/10.1016/j.neuron.2012.04.034. [PubMed: 22841311]

Leitman DI, Laukka P, Juslin PN, Saccente E, Butler P, Javitt DC. Getting the cue: Sensory contributions to auditory emotion recognition impairments in schizophrenia. Schizophrenia Bulletin. 2010; 36(3):545–556. https://doi.org/10.1093/schbul/sbn115. [PubMed: 18791077]

Liddle PF. The symptoms of chronic schizophrenia. A re-examination of the positive-negative dichotomy. British web of Psychiatry. 1987 Aug.151:145–151. https://doi.org/10.1192/bjp.151.2.145.

Lillard A. Ethnopsychologies: Cultural Variations in Theories of Mind. Psychological Bulletin. 1998; 123(1):3–32. https://doi.org/10.1037/0033-2909.123.1.3. [PubMed: 9461850]

Mehrabian A. When feelings are communicated inconsistently. Journal of Experimental Research in Personality. 1970; 4(3):198–212.

Mehrabian A, Wiener M. Decoding of inconsistent communications. web of Personality and Social Psychology. 1967; 6(1):109–114. https://doi.org/10.1037/h0024532.

Mendes WB, Blascovich J, Hunter SB, Lickel B, Jost JT. Threatened by the unexpected: physiological responses during social interactions with expectancy-violating partners. web of Personality and Social Psychology. 2007; 92(4):698–716. https://doi.org/10.1037/0022-3514.92.4.698.

Mesquita B, Frijda NH, Ellsworth P, Fischer A, Van Goozen S, Van 't Hoff S, Scherer K. Cultural Variations in Emotions: A Review. Psychological Bulletin. 1992; 112(2):179–204. https://doi.org/10.1037/0033-2909.112.2.179. [PubMed: 1454891]

Miller AH, Miller WE. Ideology in the 1972 election: Myth or reality-A rejoinder. American Political Science Review. 1976; 70:832–849.

Newcombe MJ, Ashkanasy NM. The role of affect and affective congruence in perceptions of leaders: An experimental study. Leadership Quarterly. 2002; 13(5):601–614. https://doi.org/10.1016/S1048-9843(02)00146-7.

Otten M, Seth AK, Pinto Y. A social Bayesian brain: How social knowledge can shape visual perception. Brain and Cognition. 2017; 112:69–77. https://doi.org/10.1016/j.bandc.2016.05.002. [PubMed: 27221986]

Panichello, MF., Cheung, OS., Bar, M. Predictive feedback and conscious visual experience. Frontiers in Psychology. 2013. https://doi.org/10.3389/fpsyg.2012.00620

Pinto Y, Van Gaal S, De Lange FP, Lamme VaF, Seth AK. Expectations accelerate entry of visual stimuli into awareness. J Vision. 2015; 15(8):1–15. https://doi.org/10.1167/15.8.13.doi.

Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods. Advanced quantitative techniques in the social sciences 1. 2002; 2

Redelmeier DA, Baxter SD. Research: Rainy weather and medical school admission interviews. CMAJ. 2009; 181(12):933. https://doi.org/10.1503/cmaj.091546. [PubMed: 19969588]

Righart R, de Gelder B. Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. Social Cognitive and Affective Neuroscience. 2008; 3(3):270–278. https://doi.org/10.1093/scan/nsn021. [PubMed: 19015119]

Rosenberg S, Sedlak A. Structural Representations of Implicit Personality Theory. Advances in Experimental Social Psychology. 1972; 6(C):235–297. https://doi.org/10.1016/S0065-2601(08)60029-5.

Ross, J., Zaldivar, A., Irani, L., Tomlinson, B. Who are the Turkers? Worker Demographics in Amazon Mechanical Turk; Chi Ea 2010. 2010 Jul. 2016 p. 2863-2872.https://doi.org/10.1145/1753846.1753873

Rotenberg KJ, Simourd L, Moore D. Children's Use of a Verbal-Nonverbal Consistency Principle to Infer Truth and Lying. Child Development. 1989; 69(2):309–322.

Russell JA. Culture and the categorization of emotions. Psychological Bulletin. 1991; 110(3):426–450. https://doi.org/10.1037/0033-2909.110.3.426. [PubMed: 1758918]

Russell JA, Pratt G. A description of the affective quality attributed to environments. web of Personality and Social Psychology. 1980; 38(2):311–322. https://doi.org/10.1037/0022-3514.38.2.311.

Schwarz N, Clore GL. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. web of Personality and Social Psychology. 1983; 45(3):513–523. https://doi.org/10.1037/0022-3514.45.3.513.

Stein T, Sterzer P. Unconscious processing under interocular suppression: getting the right measure. Frontiers in Psychology. 2014; 5:387. [PubMed: 24834061]

Tsuchiya N, Koch C. Continuous flash suppression reduces negative afterimages. Nature Neuroscience. 2005; 8(8):1096–101. https://doi.org/10.1038/nn1500. [PubMed: 15995700]

Wild B, Erb M, Bartels M. Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: Quality, quantity, time course and gender differences. Psychiatry Research. 2001; 102(2):109–124. https://doi.org/10.1016/S0165-1781(01)00225-6. [PubMed: 11408051]

Wilson-Mendenhall CD, Barrett LF, Barsalou LW. Neural evidence that human emotions share core affective properties. Psychological Science. 2013; 24(6):947–56. https://doi.org/10.1177/0956797612464242. [PubMed: 23603916]

Wilson VanVoorhis CR, Morgan BL. Understanding power and rules of thumb for determining sample sizes. Tutorials in Quantitative Methods for Psychology. 2007; 3:43–50. https://doi.org/10.20982/tqmp.03.2.p043.

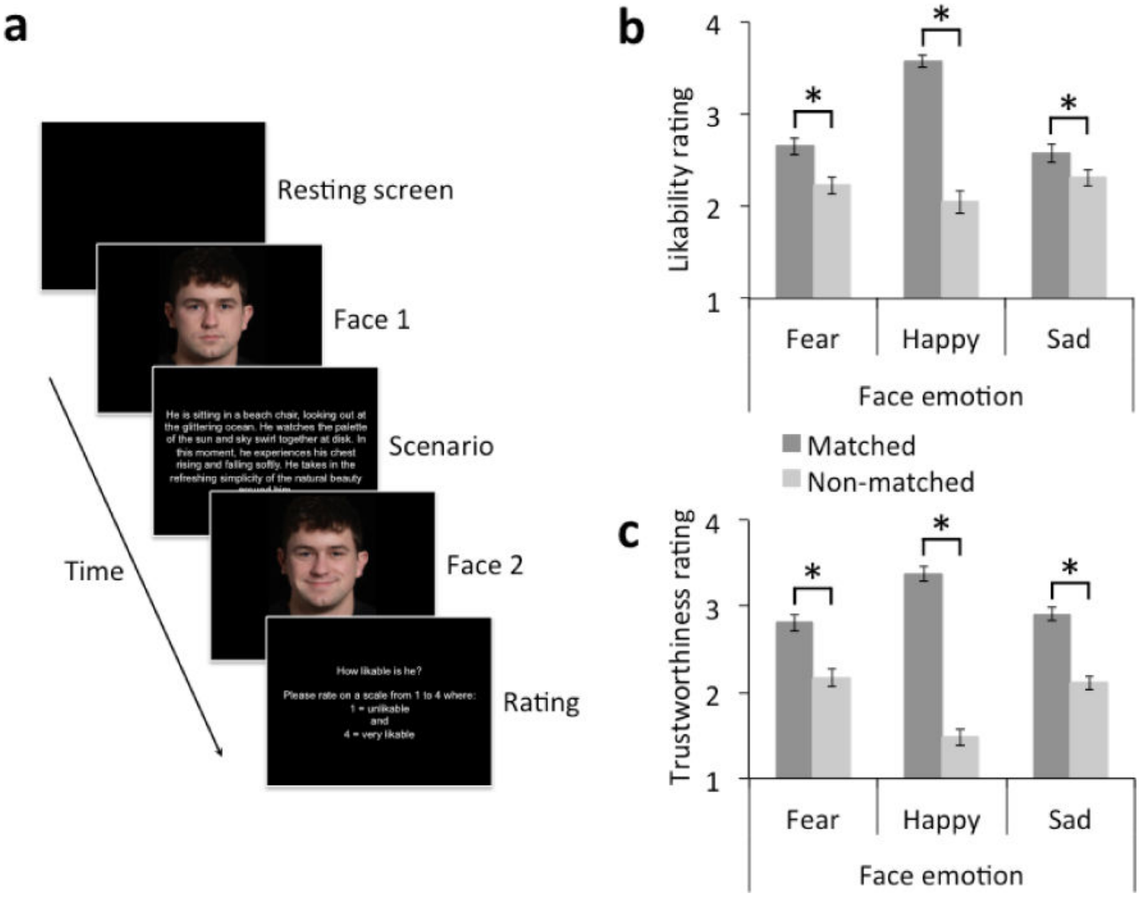
Winkielman, P., Schwarz, N., Fazendeiro, Ta, Reber, R. The hedonic marking of processing fluency: Implications for evaluative judgment; The Psychology of Evaluation: Affective Processes in Cognition and Emotion. 2003. p. 189-217.https://doi.org/http://dx.doi.org/10.4324/9781410606853

Wormwood JB, Lynn SK, Barrett LF, Quigley KS. Threat perception after the Boston Marathon bombings: The effects of personal relevance and conceptual framing. Cognition and Emotion. 2016; 30(3):539–549. [PubMed: 25707419]

Yang E, Brascamp J, Kang MS, Blake R. On the use of continuous flash suppression for the study of visual processing outside of awareness. Frontiers in Psychology. 2014; 5(724)

Yang E, Zald DH, Blake R. Fearful expressions gain preferential access to awareness during continuous flash suppression. Emotion (Washington, DC). 2007; 7(4):882–6. https://doi.org/10.1037/1528-3542.7.4.882.

Zajonc RB. The attitudinal effects of mere exposure. web of Personality & Social Psychology. 1968; 9:1–27. https://doi.org/10.1037/h0025848.



He is sitting in a beach chair, looking out at the glittering ocean. He watches the palette of the sun and sky swirl together at disk. In this moment, he experiences his chest rising and falling softly. He takes in the refreshing simplicity of the natural beauty around him.

How likable is he?

Please rate on a scale from 1 to 4 where:
1 = unlikable
and
4 = very likable

Fig. 1. Schematic representation of a trial and results in Experiments 1 and 2

Note: (a) Each trial started with the presentation of a fixation screen (7 s) followed by a photograph of a target person displaying a neutral expression (Face 1; 5 s) and then a short story (Scenario; 20 s). Then, a new photograph of the target person was presented, this time portraying a facial expression that could match the scenario emotion, be neutral, or not match the scenario emotion (Face 2; 5 s). Participants were asked to rate how likable (Experiment 1) or trustworthy (Experiment 2) the target person was. (b) Individuals exhibiting predicted facial expressions (matching the emotion evoked by the scenario) were rated as more likable than those exhibiting unpredicted ones (non-matching) across the three emotion categories explored (Experiment 1). (c) Similar results were observed for trustworthiness ratings (Experiment 2). Asterisks indicate *p*<.05.

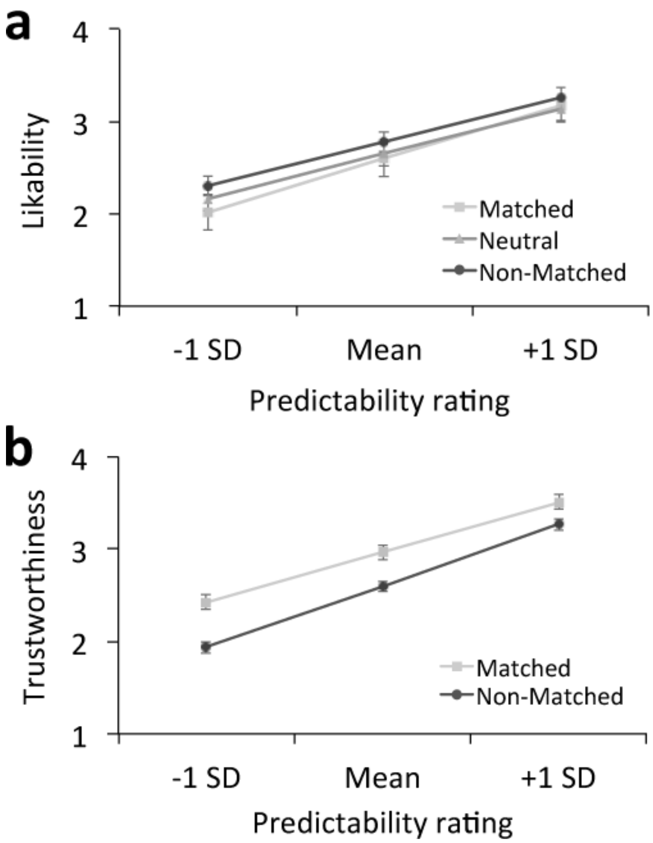


Fig. 2. Models for within-subjects predictability ratings predicting social perceptions in Experiments 3 and 4

Note: (a) HLM model for within-subject likability ratings predicted by within-subject predictability ratings, face match condition (match, non-match, neutral), and their interactions. Predictability ratings significantly predict perceived likability, and this effect does not differ across face match conditions (match, non-match, neutral). Consistent with the mediational analyses, this model shows that there is no longer a significant difference in average perceived likability between match and non-match trials when controlling for predictability ratings. (b) HLM model for within-subject trustworthiness ratings predicted by within-subject predictability ratings, face match condition (match, non-match), and their interaction. Predictability ratings significantly predict perceived trustworthiness, and this effect does not differ across face match conditions (match, non-match). Consistent with the mediational analyses, which found evidence of significant but only partial mediation, this model shows that there is still a significant difference in average perceived trustworthiness between match and non-match trials even when controlling for predictability ratings.

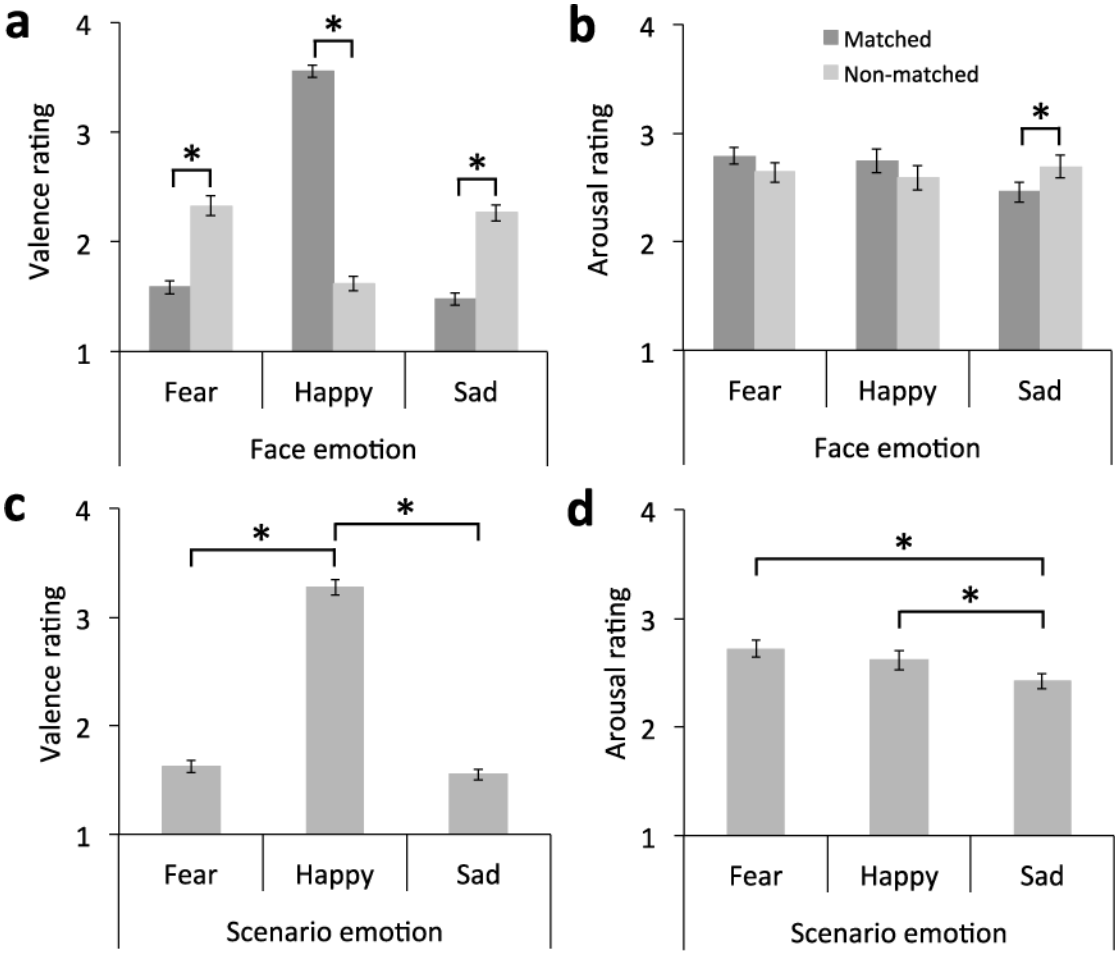


Figure 3. Results in Experiment 5

Note: (a) Trials with predicted facial expressions (matching the emotion evoked by the scenario) did not yield higher (more positive) valence ratings, which is inconsistent with affective misattribution as a viable mechanism underlying the effect of predictions on social perception. (b) Similarly, arousal ratings were not driven by prediction. (c) Valence ratings were higher (more positive) for happy than fear or sad scenarios. (d) Arousal ratings were lower for sad than happy and fear scenarios. Asterisks indicate *p*<.05.

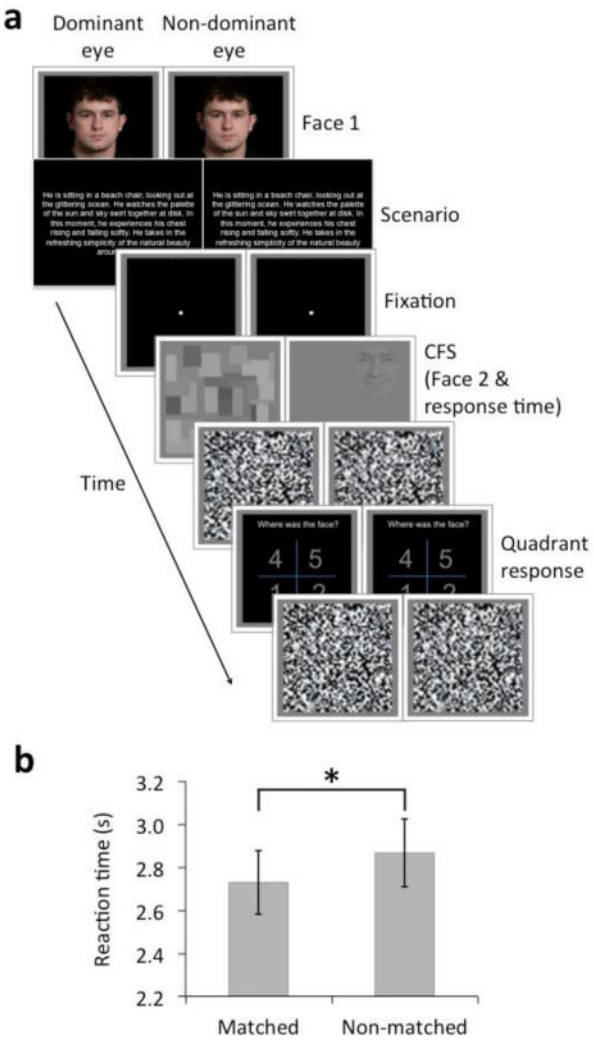

a
Dominant eye
Non-dominant eye
Face 1
He is sitting in a beach chair, looking out at the glittering ocean. He watches the palette of the sun and sky swirl together at disk. In this moment, he experiences his chest rising and falling softly. He takes in the refreshing simplicity of the natural beauty
Scenario
Fixation
CFS (Face 2 & response time)
Time
Where was the face?
4 5
Quadrant response
b
3.2
3.0
2.8
2.6
2.4
2.2
Reaction time (s)
*
Matched
Non-matched

Fig. 4. Schematic representation of a trial and results in Experiment 6

Note: (a) Each trial started with the presentation (to both eyes) of a photograph of a target person displaying a neutral face (Face 1; 5 s) followed by a short story (Scenario; 20 s). After a brief fixation screen, rapidly changing Mondrian patterns of decreasing contrast were presented to the dominant eye, while an image of the target person displaying a facial expression (that either matched or did not match the emotion scenario or was neutral) of increasing contrast was presented to the non-dominant eye (Face 2; max. time: 10 s). Participants were asked to press the spacebar as soon as they perceived the face and then to report in which quadrant the face was presented. (b) Faces portraying predicted facial expressions (matching the emotion evoked by the scenario) were processed faster than faces portraying unpredicted facial expressions (not matching), as revealed by faster reaction times. Asterisk indicates *p*<.05.