Perception, 2010, volume 39, pages 931 – 941

doi:10.1068/p6633

Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation

Andrew D Engell¶, Alexander Todorov¶, James V Haxby
Department of Psychology and Center for the Study of Brain, Mind and Behavior, Princeton University, Princeton, NJ 08540, USA
Received 9 December 2009, in revised form 22 February 2010; published online 28 June 2010

Abstract. People rapidly and automatically evaluate faces along many social dimensions. Here, we focus on judgments of trustworthiness, which approximate basic valence evaluation of faces, and test whether these judgments are an overgeneralization of the perception of emotional expressions. We used a behavioral adaptation paradigm to investigate whether the previously noted perceptual similarities between trustworthiness and emotional expressions of anger and happiness extend to their underlying neural representations. We found that adapting to angry or happy facial expressions causes trustworthiness evaluations of subsequently rated neutral faces to increase or decrease, respectively. Further, we found no such modulation of trustworthiness evaluations after participants were adapted to fearful expressions, suggesting that this effect is specific to angry and happy expressions. We conclude that, in line with the overgeneralization hypothesis, a common neural system is engaged during the evaluation of facial trustworthiness and expressions of anger and happiness.

1 Introduction

Individuals form trait impressions along dimensions ranging from trustworthiness to sexual orientation with a single glance at an unfamiliar face (Rule and Ambady 2008; Willis and Todorov 2006). These fast evaluations are highly reliable, as judgments made after 50 – 100 ms of exposure to faces are strongly correlated with those made with no time constraints (eg Bar et al 2006; Todorov et al 2009). Although these trait impressions are formed in a minimal amount of time, they have been shown to have a significant real-world impact. Political election outcomes (Ballew and Todorov 2007) and sentencing in criminal trials (Blair et al 2004) can be predicted by evaluations of faces along trait dimensions. However, despite the extent of their influence and their efficient and reliable nature, there is little evidence that these face evaluations are valid. Some studies have found modest positive correlations between behavior and trait judgments from faces (Bond 1994), but others have failed to find such correlations (Hassin and Trope 2000; Zebrowitz et al 1996), and some have even found an inverse relationship (Zebrowitz et al 1998).

From an evolutionary point of view, it is puzzling why people reliably make seemingly invalid judgments. The overgeneralization hypothesis suggests a potential solution to this puzzle, positing that rapid trait impressions from facial appearance are due to overgeneralization of other, more veridical, evaluations (Knutson 1996; Montepare and Dobish 2003; Zebrowitz and Montepare 2008; Zebrowitz et al 2003, 2010). For instance, childlike traits are often attributed to adults who have facial features reminiscent of babies (eg round faces and large eyes—McArthur and Apatow 1983).

Here, consistent with the emotion overgeneralization hypothesis (Knutson 1996; Montepare and Dobish 2003; Said et al 2009; Zebrowitz et al, 2010), we test whether the neuronal populations supporting perception of emotional expressions signaling

¶ Please address correspondence to Andrew Engell, Department of Psychology, Yale University, New Haven, CT 06520, USA; e-mail: andrew.engell@yale.edu or Alexander Todorov, Department of Psychology, Princeton University, Princeton, NJ 08540, USA; e-mail: atodorov@princeton.edu

approach (ie angry and happy expressions) also support perception of trustworthiness in emotionally neutral faces. We focus on trustworthiness, because it reliably approximates the general valence evaluation of faces (Todorov 2008; Todorov and Engell 2008; Todorov et al 2008). Further, three lines of evidence suggest that trustworthiness judgments are associated with perceptions of expressions of anger and happiness. First, trustworthiness judgments of emotionally neutral faces are positively correlated with judgments of happiness and negatively correlated with judgments of anger (Todorov and Duchaine 2008). A similar pattern has been observed for judgments of affiliation, an attribute similar to trustworthiness (Montepare and Dobish 2003). Second, whereas dynamic changes in expressions from neutral to angry are perceived as more intense when accompanied by an identity change from a trustworthy to untrustworthy face, changes in expression from neutral to happy are perceived as more intense when accompanied by an identity change from an untrustworthy to trustworthy face (Oosterhof and Todorov 2009). In other words, angry and happy expressions are perceived as more intense when accompanied by congruent changes in structural features. Third, computer modeling of face trustworthiness suggests that these judgments are grounded in similarity to expressions of anger and happiness (Oosterhof and Todorov 2008; Todorov 2008).

In the current study we reasoned that, if trustworthiness evaluations are due to overgeneralization of the perception of angry and happy facial expressions, it should be possible to influence those evaluations by first adapting the neural populations which support the perception of those expressions. Specifically, we predicted that behavioral adaptation to angry faces should result in higher trustworthiness ratings of emotionally neutral faces, whereas adaptation to happy faces should result in lower trustworthiness ratings. Further, to the extent that these effects are specific to displays of anger and happiness, trustworthiness ratings should not be influenced by adaptation to fearful expressions.

To test this prediction, we used a behavioral adaptation paradigm. The central tenet of this paradigm is that extended exposure to a given stimulus creates a visual aftereffect such that the visual appearance of subsequently viewed stimuli is shifted away from the adapting stimulus along any shared dimensions. This idea can be clearly understood when one considers the motion aftereffect, or 'waterfall illusion'. Prolonged exposure to the downward motion of a waterfall results in subsequently viewed static stimuli to appear as if they are moving upward. One explanation for this is that, in the absence of moving stimuli, the direction-selective cells within area MT/V5, a motion sensitive region of the visual system, will randomly discharge. The random asynchronous firings for motion in various directions cancel each other out and the sum response of the entire neural population results in no perceived motion. While viewing the rush of falling water, cells that are specifically tuned to downward motion will fire vigorously, causing the sum discharge of the population to skew decidedly to downward motion. After prolonged exposure to a downward moving stimulus, neurons demonstrate decreased responsiveness (Kohn and Movshon 2003). This diminished response allows the baseline firing of neurons tuned to upward motion to have a greater impact on the sum output of MT in the absence of external stimulation. The resulting visual phenomenon is that static images appear to be moving upward.

We suggest that adaptation may be a useful tool for investigating overgeneralization effects in face evaluation (cf Buckingham et al 2006). High-level adaptation (ie adaptation effects that are not due to the low-level visual features of a stimulus) has proven useful for investigating multiple dimensions of the neural representation of faces, including gender, attractiveness, emotional expression, and identity (eg Fox and Barton 2007; Leopold et al 2001; Rhodes et al 2003; Webster et al 2004). For instance, Webster and colleagues showed that androgynous faces that had an equal probability of being categorized by participants as "male" or "female" were seen as distinctly "male" after extended exposure to female faces and as distinctly "female" after extended exposure to male faces.

To test our predictions, we asked participants to evaluate the trustworthiness of emotionally neutral faces both before and after perceptual adaptation to angry, happy, or fearful faces. We found that angry and happy, but not fearful, expressions shifted subsequent evaluations of trustworthiness. A follow-up experiment confirmed the efficacy of our computer-generated expressive faces as perceptual adapter stimuli. A third experiment was performed to investigate whether the effect was due to a shift in response bias rather than neural adaptation.

2 Experiment 1. Effects of adaptation on trustworthiness judgments of emotionally neutral faces

In the first experiment we tested the overgeneralization hypothesis by investigating whether adaptation to angry and happy expressions would influence subsequent evaluations of trustworthiness.

2.1 *Method*

2.1.1 *Participants.* Thirty-six Princeton University undergraduate students participated in the study for course credit.

2.1.2 *Stimuli.* 288 near-photorealistic faces with unique identities were created with the FaceGen software package (Singular Inversions, Vancouver, BC). Adapter faces were created by morphing 192 of these faces to display expressions of anger, fear, and happiness with the expression manipulation tools available in FaceGen. The remaining 96 emotionally neutral faces were used as the test faces (figure 1b). FaceGen supports

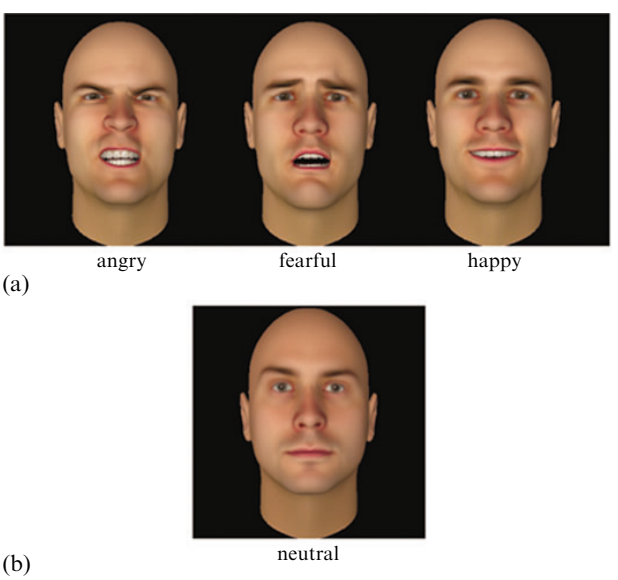

Figure 1. [In colour online, see http://dx.doi.org/10.1068/p6633] Example stimuli. (a) Angry, fearful, and happy adapter faces used in all three experiments. (b) Emotionally neutral test face used in experiments 1 and 3. (c) Emotionally ambiguous test faces used in experiment 2.

adding several emotional expressions to any face and the expression strengths can be set anywhere between 0% and 100%. We added expressions using FaceGen's Anger, SmileOpen (happy), and Fear expression controls. We used maximum emotion strength of 100% (figure 1a). Emotions of anger and happiness with intensity strength of 50% have been used by Oosterhof and Todorov (2008), and these expressions were clearly perceived as angry and happy, respectively. We also tested the ecological validity of these expressions in a separate study in which ten participants were shown ten faces of each expression in a random sequential order and asked to categorize the expression as "angry", "happy", or "fearful". Average categorization accuracy in this study was 98.7% and did not vary across emotions.

2.1.3 *Procedures.* Participants were first asked to rate the trustworthiness of each test face using a 9-point Likert scale. During the subsequent 'adaptation phase', participants were randomly assigned to the angry ($N = 13$), fearful ($N = 12$), or happy ($N = 12$) adapter conditions and passively viewed 30 emotionally expressive adapter faces from their respective condition. These faces were randomly drawn from the 192 possible identities. Participants were only adapted to the expression type to which they were randomly assigned, as adapter type was a between-subjects factor. Each adapter face

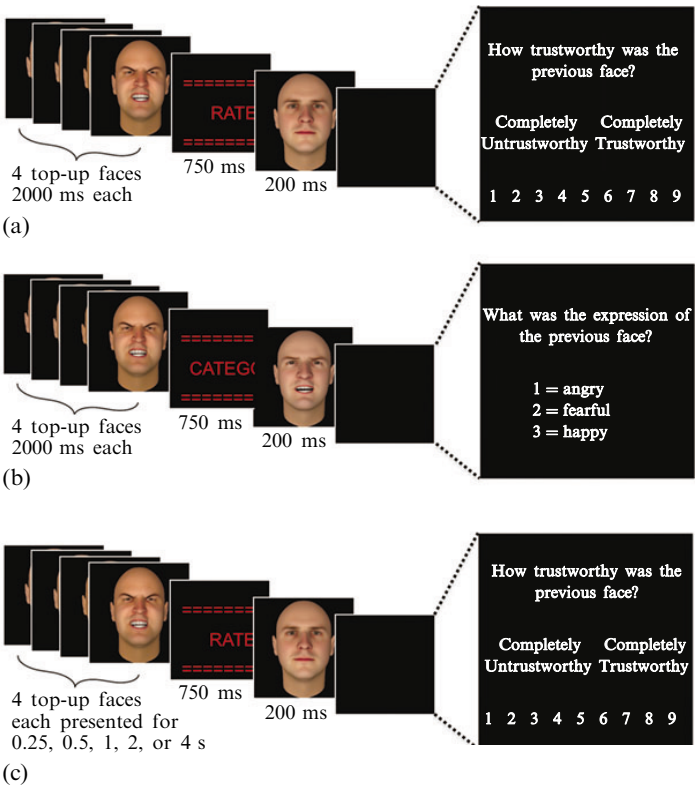

Figure 2. [In colour online.] Post-adaptation procedure. (a) Trustworthiness experiment. Prior to the onset of each test face, participants viewed four 'top-up' adapter faces (angry, fearful, or happy) for 2000 ms each. After the presentation of the top-up adapters a cue indicated that the participant should rate the trustworthiness of the emotionally neutral test faces on a 9-point Likert scale. (b) Stimulus-validation experiment. The procedure was identical to the validation experiment except that participants were asked to categorize the expression of the emotionally ambiguous face that followed the cue. (c) Response bias. The procedure was the same as in experiment 1 except that there was no long-term adaptation and the duration of 'top-up' adapters varied by trial such that duration randomly chosen to be 0.25, 0.5, 1, 2, or 4 s per face (there were always four top-up faces prior to presentation of the test face).

was preceded by a 200 ms blank display and remained on screen for 2000 ms. After the adaptation phase, participants were again asked to rate the trustworthiness of the test faces. Prior to the onset of each test face, participants viewed four 'top-up' adapter faces (angry, fearful, or happy) for 2000 ms each (see figure 2a). Importantly, all test faces were 80% of the size of the adapter faces in order to disrupt adaptation of low-level visual features. In an effort to achieve sufficient adaptation, participants rated 24 of the 96 test faces and then repeated the adaptation phase of the experiment. This cycle of adaptation and categorization was repeated four times in total. The difference between the average post-adaptation and pre-adaptation trustworthiness scores reflected the effect of adaptation to expression on trustworthiness evaluation.

2.2 *Results*

The data were analyzed with a mixed-model 2(test phase: pre or post) × 3(adapter type: angry, fearful, or happy) ANOVA with test phase a within-subjects factor and adapter type a between-subjects factor. There were no significant main effects of either test phase ($F_{1,34} = 0.45$) or adapter type ($F_{2,34} = 0.04$). There was, however, a significant interaction ($F_{2,34} = 20.54$, $p < 0.001$). To explicate this interaction, we analyzed the simple effects of test phase at each level of adapter type (figure 3). After adapting to angry faces, emotionally neutral faces were evaluated as more trustworthy ($t_{12} = 4.51$, $p = 0.001$), whereas after adapting to happy faces emotionally neutral faces were evaluated as less trustworthy ($t_{11} = 3.55$, $p = 0.005$). Adapting to fearful faces did not result in different mean evaluations of trustworthiness ($t_{11} < 1$). The change in trustworthiness evaluation in both the angry and happy adapter conditions was significantly larger than the change in the fearful adapter condition ($t_{23} = 4.05$, $p < 0.001$ and $t_{22} = 2.55$, $p = 0.018$, respectively).

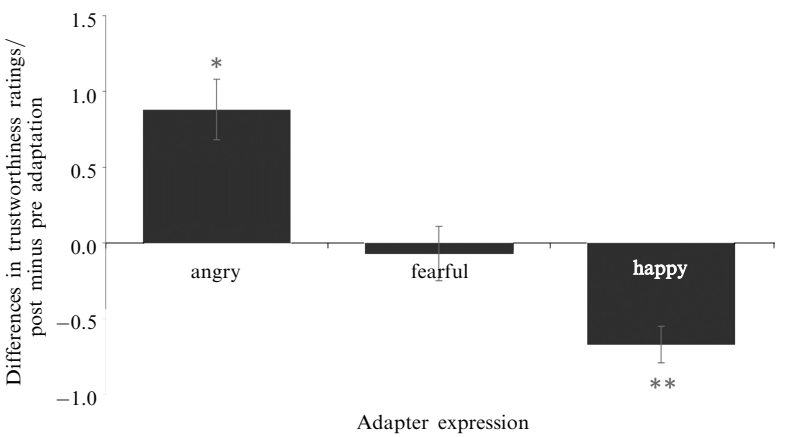

Figure 3. Effect of adaptation to angry, fearful, and happy faces on trustworthiness evaluation of emotionally neutral faces. Adaptation to angry faces caused a significant increase in trustworthiness evaluations, while adaptation to happy faces caused a significant decrease in trustworthiness evaluations. Adaptation to fearful faces had no effect on subsequent trustworthiness evaluations of neutral faces. Error bars show ±1 SEM. * $p = 0.001$, ** $p = 0.005$.

3 Experiment 2. Validating the effectiveness of adapter faces

In the second experiment we sought to confirm that the differential effects of the expression adapters in experiment 1 were not due to limitations of the stimuli and/or design. Particularly, we sought to confirm that the null effect of fearful adapters was not due to a general inability of these stimuli to evoke aftereffects. We therefore used the same adapter stimuli in a traditional expression adaptation paradigm in which participants categorized emotionally ambiguous faces before and after adaptation.

3.1 *Method*

3.1.1 *Participants.* Thirty-two Princeton University undergraduate students participated in the study for course credit.

3.1.2 *Stimuli.* The adapter faces were the same as those used in experiment 1 (see figure 1a). Test faces were emotionally ambiguous faces created by morphing between pairs of expressions (anger and fear, anger and happiness, fear and happiness) using a randomly selected subset of 90 of the 192 expressive faces (figure 1c). The morphed test faces showed 40%, 50%, or 60% of one emotion as compared to the other.

3.1.3 *Procedures.* The procedures for the validation experiment were the same as those in experiment 1 except that participants were asked to categorize the emotionally ambiguous test faces as "angry", "fearful", or "happy" before and after adaptation to emotionally expressive faces (figure 2b). This experiment used a forced-choice identification task as opposed to the Likert scale evaluation used in experiment 1, because, unlike trustworthiness, expressions represent discrete categories. This approach also more closely replicates the paradigms used in previous investigations of facial expression adaptation (eg Webster et al 2004). As in experiment 1, participants were randomly assigned to the angry-adapter ($N = 10$), fearful-adapter ($N = 9$), or happy-adapter ($N = 13$) condition.

3.2 *Results*

To measure the effect of adaptation, we first calculated for each test phase (pre- and post-adaptation) the proportion of trials in which an emotionally ambiguous face was categorized as each of the three expressions. The difference between the post-adaptation and pre-adaptation proportions for each adapter type reflected the effect of adaptation on expression perception. As shown in table 1, all three adapter conditions affected subsequent categorization of emotionally ambiguous faces, such that a face was less frequently categorized as the adapter condition and more often categorized as the expression reflecting the opposite end of the morph continuum. Participants in the angry-adapter condition were less likely to categorize faces as angry when viewing angry–happy ($t_9 = 6.29$, $p < 0.05$) or angry–fearful ($t_9 = 7.48$, $p < 0.05$) morphs, whereas they were more likely to categorize faces as happy ($t_9 = 7.19$, $p < 0.05$) and fearful ($t_9 = 7.86$, $p < 0.05$), respectively. Participants in the fearful-adapter condition were less likely to categorize faces as fearful when viewing fearful–angry ($t_8 = 5.24$, $p < 0.05$ or fearful–

Table 1. Experiment 2: Difference in proportions (mean, SEM p) of categorization of emotionally ambiguous faces after adaptation to expressions of anger, fear, and happiness, as compared to before adaptation. Positive and negative differences indicate that the emotionally ambiguous morph face was more or less often categorized as the given expression after adaptation.

Perceived as	Angry–Fearful	Angry–Happy	Happy–Fearful
Angry adapter: test face morph			
Angry	−0.31, 0.04, < 0.001	−0.40, 0.07, < 0.001	−0.02, 0.01, = 0.052
Fearful	0.32, 0.04, < 0.001	0.03, 0.04, = 0.522	0.01, 0.03, = 0.808
Happy	−0.01, 0.01, = 0.104	0.37, 0.07, < 0.001	0.01, 0.03, = 0.814
Happy adapter: test face morph			
Angry	−0.03, 0.04, = 0.523	0.10, 0.03, = 0.012	−0.01, 0.02, = 0.827
Fearful	0.03, 0.04, = 0.508	−0.03, 0.02, = 0.225	0.11, 0.03, = 0.007
Happy	0.0, 0.0, = 1	−0.07, 0.03, = 0.033	−0.11, 0.03, = 0.005
Fearful adapter: test face morph			
Angry	0.34, 0.08, = 0.002	0.09, 0.03, = 0.016	0.02, 0.02, = 0.239
Fearful	−0.38, 0.07, = 0.001	−0.03, 0.02, = 0.172	−0.37, 0.04, < 0.001
Happy	0.03, 0.03, = 0.347	−0.06, 0.01, = 0.002	0.35, 0.04, < 0.001

happy ($t_8 = 10.64$, $p < 0.05$) morphs, whereas they were more likely to categorize faces as happy ($t_8 = 8.74$, $p < 0.05$) and angry ($t_8 = 4.37$, $p < 0.05$), respectively. Participants in the happy-adapter condition were less likely to categorize faces as happy when viewing happy–angry ($t_{12} = 2.58$, $p < 0.05$) or happy–fearful ($t_{12} = 3.25$, $p < 0.05$) morphs, whereas they were more likely to categorize faces as angry ($t_{12} = 2.94$, $p < 0.05$), and fearful ($t_{12} = 2.91$, $p < 0.05$), respectively.

For angry and happy adapters, there was no effect on the emotionally ambiguous face that did not comprise the adapter expression—eg the happy–fearful morph in the angry-adapter condition (ps > 0.05). However, unexpectedly the fearful adapter did significantly affect classification of the angry–happy morph such that it was more often classified as angry after adaptation (ps < 0.05), although this effect was substantially smaller than the effects for categorization of ambiguous faces that comprised the fearful expression (angry–fearful and happy–fearful).

4 Experiment 3. Effect of adaptation time on cross-adaptation

In the third experiment we investigated whether the adaptation effects observed in experiment 1 were due to a response bias shift, which would suggest that the adaptation effects are conceptual rather than neural in nature. For example, if the adapter faces were implicitly used as a standard of comparison for the evaluated faces, one would predict similar effects to those seen in experiment 1. That is, relative to angry faces, emotionally neutral faces may seem more trustworthy. In contrast, relative to happy faces, emotionally neutral faces may seem less trustworthy. To rule out this alternative explanation, we manipulated the duration of adaptation prior to trustworthiness evaluation. Conceptual, or response-bias, effects should be relatively immune to changes in the duration of the adapting stimulus, whereas sensitivity to adapter duration is a hallmark of visual aftereffects (cf Leopold et al 2005). That is, increased adapter duration should result in increased changes in trustworthiness evaluation to the extent that this evaluation relies on the same mechanisms that are responsible for perception of angry and happy expressions.

4.1 *Method*

4.1.1 *Participants.* Forty-nine Princeton University undergraduate students participated in the study for course credit.

4.1.2 *Stimuli.* The stimuli for this experiment were the same as those used in experiment 1.

4.1.3 *Procedure.* The procedure for this experiment was similar to that of experiment 1 with the following exceptions. Participants were randomly assigned to either an angry-adapter ($N = 24$) or happy-adapter ($N = 25$) condition. There was no 'long' adaptation phase. Rather, four 'top-up' adapter faces (ie the adapter stimuli shown immediately prior to the test face) were presented for 250 ms, 500 ms, 1000 ms, 2000 ms, or 4000 ms each for total adaptation times of 1, 2, 4, 8, and 16 s, respectively (figure 2c). All 100 trials (20 trials of each adapter duration) were presented in a random order.

4.2 *Results*

A 3-way 2(adapter type: angry/happy) $\times$ 2(test phase: pre/post) $\times$ 5(adapter duration: 1/2/4/8/16 s) mixed ANOVA, in which adapter type was a between-subjects factor, revealed three significant interactions. A significant interaction between test phase and adapter type ($F_{1,47} = 17.09$, $p < 0.001$) was driven by an increase in trustworthiness evaluations after adaptation to angry faces and a decrease in trustworthiness evaluations after adaptation to happy faces. A significant interaction between adapter type and adapter duration ($F_{4,188} = 2.88$, $p < 0.024$) demonstrates that increasing the adapter duration results in higher trustworthiness evaluations after adaptation to angry faces

but lower trustworthiness evaluations after adaptation to happy faces. As a result of these two interactions, there was a significant 3-way interaction between adapter type, test phase, and adapter duration ($F_{4,188} = 3.00$, $p < 0.020$). That is, increasing the adaptation time resulted in larger differences between the pre- and post-test phases, but whether this difference was positive or negative was dependent on which adapter type the participant was shown (angry and happy, respectively).

We further unpacked the 3-way interaction by separately analyzing the data from the angry-adapter and happy-adapter conditions with 2(test phase: pre/post) × 5(adapter duration: 1/2/4/8/16 s) repeated-measures ANOVAs. The ANOVA for the angry-adapter condition revealed a significant interaction ($F_{4,92} = 2.99$, $p < 0.05$), such that the effect of adaptation increased as a function of adapter duration (figure 4, top). Trend analysis of this effect showed a significant linear component ($F_{1,23} = 9.08$, $p = 0.006$) and a non-significant quadratic component ($F_{1,23} = 2.24$, $p = 0.15$). The ANOVA for the happy-adapter condition revealed a main effect of test phase ($F_{1,24} = 18.08$, $p < 0.001$), such that trustworthiness evaluations were significantly lower in the post-adaptation test phase. The interaction was also significant ($F_{4,96} = 2.86$, $p < 0.05$), such that the effect of adaptation varied as a function of adapter duration (figure 4, bottom). Trend analysis of this effect showed a non-significant linear component ($F_{1,24} = 1.78$, $p = 0.20$) and a significant quadratic component ($F_{1,24} = 4.66$, $p = 0.011$).

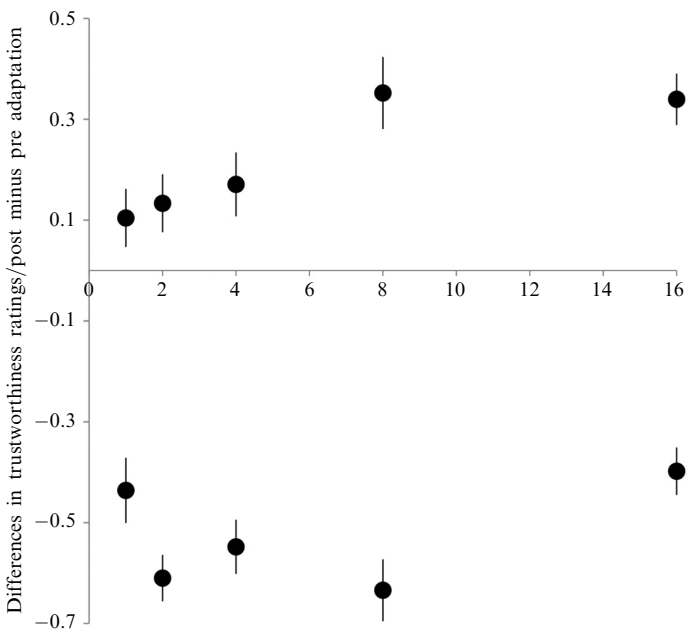

Figure 4. Top: effect of adapter duration on angry adapters. Bottom: effect of adapter duration on happy adapters. Error bars show ±1 SEM.

5 Discussion

The results of this study offer evidence in support of the emotion overgeneralization hypothesis for rapid trustworthiness evaluations of neutral faces. Namely that this evaluation relies on the same neural mechanisms that are involved in perception of emotional expressions. Adaptation of the brain's response to expressions of anger results in higher evaluations of trustworthiness, whereas adaptation to expressions of happiness results in lower evaluations of trustworthiness. This 'cross-adaptation' suggests that these facial dimensions (expression and trustworthiness) are, at minimum, subserved by partially overlapping neural populations. Moreover, the extent of the neural commonality

between expression and trustworthiness seems to be restricted to expressions of anger and happiness, as adaptation to a different expression with negative valence, fear, did not modulate trustworthiness evaluations. This is consistent with a previous study, which demonstrated that exaggerating emotionally neutral faces along the trustworthiness dimension resulted in changes in expressions of anger and happiness but not in disgust, sadness, fear, or surprise (Oosterhof and Todorov 2008).

In experiment 2, we addressed the possibility that the null effect of fear adaptation on trustworthiness evaluations was merely due to a more general failure of these stimuli to evoke aftereffects. We found that fearful adapters effectively shifted the participants' perception of emotionally ambiguous faces away from fear. That is, participants were significantly less likely to categorize angry – fearful or fearful – happy morphs as "fearful" after prolonged exposure to our fearful-face stimuli. Thus, the null effect of fearful adapters in the first experiment cannot be attributed to inadequacy of the stimuli. Interestingly, adaptation to fearful faces did increase the likelihood that participants would classify angry – happy morphs as 'angry', although this effect was weaker than the effect for classification of angry – fearful and happy – fearful morphs. It is not clear why we observed the former effect. One possibility is that the effect was partly driven by perceptual similarity. Recent research suggests that the emotion opposite to fear is disgust (Susskind et al 2008). For example, the anti-face of fearful expressions are disgust expressions. The latter are highly similar and easily confusable with angry expressions (Aviezer et al 2008). Hence, it is possible that this partial similarity resulted in increased classification of angry – happy morphs as 'angry' after adaptation to fearful expressions. Another possibility is that this 'adaptation' effect reflects mood induction. However, further investigation will be necessary to explicate this result.

An alternative explanation for our cross-adaptation results is that trustworthiness evaluations were influenced at a conceptual level, perhaps by shifting the response-bias of participants. If so, manipulating the duration of the adapter stimuli should have very little effect on the adaptation effect. On the other hand, neural adaptation is strongly modulated as a function of adapter duration (cf Leopold et al 2005). In experiment 3, we demonstrated that adapter duration significantly affected the strength of the aftereffect, suggesting at least a partially neural, rather than conceptual, basis of the aftereffect. The effects were particularly clear for the angry-adapter condition. The effect of adaptation increased as a function of the adapter duration. Interestingly, the effect of adapter duration for the happy-adapter was more variable than for the angry-adapter. Although there was a general trend toward stronger adaptation effects at longer happy-adapter durations, the longest duration (16 s) was similar to the shortest duration (1 s). This pattern of results suggests that perceptions of trustworthiness may be more closely related to similarity to expressions of anger than similarity to expressions of happiness. Future studies are needed to address this question.

The adaptation stimuli (emotionally expressive faces) were created with a software package whose algorithms for morphing neutral faces into emotionally expressive faces are not empirically validated. This raises the concern that perhaps the stimuli did not realistically reflect the emotional expressions that they were intended to portray. However, the results of experiment 2 should largely allay this concern as they show the same adaptation effects that would be expected from natural versions of the stimuli. For example, as with natural stimuli, adaptation to the computer-generated 'angry' faces resulted in participants perceiving faces morphed between angry and another expression as less angry. Moreover, in a separate study we demonstrated that participants were able to categorize the facial expressions with near perfect accuracy.

When exposed to a novel face, people extract a wealth of information within a very brief period of time. Much of this information is used to make valid categorizations (eg race or gender) or inferences of affective states (from emotional expressions)

and attentional focus (from eye-gaze direction). However, the information is often used to make dubious inferences regarding an individual's enduring traits. Our findings suggest that the perception of trustworthiness share a common neural substrate with the perceptions of anger and happiness. These findings are consistent with the idea that the perception of traces of anger or happiness in ostensibly neutral expressions leads to reliable, although not necessarily valid, evaluations of trustworthiness (Oosterhof and Todorov 2008). More generally, these results add evidence to the research that has implicated expression overgeneralization as one of the mechanisms responsible for face evaluation on social dimensions (eg Knutson 1996; Montepare and Dobish 2003; Oosterhof and Todorov 2008, 2009; Said et al 2009; Todorov 2008; Zebrowitz and Montepare 2008).

In a seminal paper, Gould and Lewontin (1979) criticized the widely held belief that all "traits" were the adaptive result of natural selection. To illustrate their point, they referred to the architectural notion of "spandrels", the space between adjoining arches. They argued that the creation of a given feature (the spandrel) could merely be the unintended byproducts of a deliberate decision (the adjoining arches). In this context, the current study can be interpreted to suggest that evaluative judgments of faces may be perceptual spandrels; a byproduct of evolutionary pressures exerted by the need to quickly perceive facial expressions.

References

Aviezer H, Ran H, Ryan J, Grady C, Susskind J M, Anderson A K, Moscovitch M, Schlomo B, 2008 "Angry, Disgusted or Afraid? Studies on the malleability of facial expression perception" *Psychological Science* **19** 724 – 732

Ballew C C, Todorov A, 2007 "Predicting political elections from rapid and unreflective face judgments" *Proceedings of the National Academy of Sciences of the USA* **104** 17948 – 17953

Bar M, Neta M, Linz H, 2006 "Very first impressions" *Emotion* **6** 269 – 278

Blair I V, Judd C M, Chapleau K M, 2004 "The influence of Afrocentric facial features in criminal sentencing" *Psychological Science* **15** 674 – 679

Bond C F Jr, 1994 "The kernel of truth in judgments of deceptiveness" *Basic and Applied Social Psychology* **15** 523 – 534

Buckingham G, DeBruine L M, Little A C, Welling L L M, Conway C A, Tiddeman B P, Jones B C, 2006 "Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness" *Evolution and Human Behavior* **27** 381 – 389

Fox C J, Barton J J S, 2007 "What is adapted in face adaptation? The neural representations of expression in the human visual system" *Brain Research* **1127** 80 – 89

Gould S J, Lewontin R C, 1979 "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme" *Proceedings of the Royal Society of London, Series B* **205** 581 – 598

Hassin R, Trope Y, 2000 "Facing faces: studies on the cognitive aspects of physiognomy" *Journal of Personality and Social Psychology* **78** 837 – 852

Knutson B, 1996 "Facial expressions of emotion influence interpersonal trait inferences" *Journal of Nonverbal Behavior* **20** 165 – 182

Kohn A, Movshon J A, 2003 "Neuronal adaptation to visual motion in area MT of the macaque" *Neuron* **39** 681 – 691

Leopold D A, O'Toole A J, Vetter T, Blanz V, 2001 "Prototype-referenced shape encoding revealed by high-level aftereffects" *Nature Neuroscience* **4** 89 – 94

Leopold D A, Rhodes G, Müller K M, Jeffery L, 2005 "The dynamics of visual adaptation to faces" *Proceedings of the Royal Society of London, Series B* **272** 897 – 904

McArthur L Z, Apatow K, 1983 "Impressions of babyfaced adults" *Social Cognition* **2** 315 – 342

Montepare J M, Dobish H, 2003 "The contribution of emotion perceptions and their overgeneralizations to trait impressions" *Journal of Nonverbal Behavior* **27** 237 – 254

Oosterhof N N, Todorov A, 2008 "The functional basis of face evaluation" *Proceedings of the National Academy of Sciences of the USA* **105** 11087 – 11092

Oosterhof N N, Todorov A, 2009 "Shared perceptual basis of emotional expressions and trustworthiness impressions from faces" *Emotion* **9** 128 – 133

Rhodes G, Jeffery L, Watson T L, Clifford C W, Nakayama K, 2003 "Fitting the mind to the world: face adaptation and attractiveness aftereffects" *Psychological Science* **14** 558 – 566

Rule N O, Ambady N, 2008 "Brief exposures: Male sexual orientation is accurately perceived at 50 ms" *Journal of Experimental Social Psychology* **44** 1100 – 1105

Said C P, Sebe N, Todorov A, 2009 "Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces" *Emotion* **9** 260 – 264

Susskind J, Lee D, Cusi A, Feinman R, Grabski W, Anderson A K, 2008 "Expressing fear enhances sensory acquisition" *Nature Neuroscience* **11** 843 – 850

Todorov A, 2008 "Evaluating faces on trustworthiness: an extension of systems for recognition of emotions signaling approach/avoidance behaviors" *Annals of the New York Academy of Sciences* **1124** 208 – 224

Todorov A, Duchaine B, 2008 "Reading trustworthiness in faces without recognizing faces" *Cognitive Neuropsychology* **25** 395 – 410

Todorov A, Engell A D, 2008 "The role of the amygdala in implicit evaluation of emotionally neutral faces" *Social Cognitive and Affective Neuroscience* **3** 303 – 312

Todorov A T, Pakrashi M P, Oosterhof N O, 2009 "Evaluating faces on trustworthiness after minimal time exposure" *Social Cognition* **27** 813 – 833

Todorov A, Said C P, Engell A D, Oosterhof N N, 2008 "Understanding evaluation of faces on social dimensions" *Trends in Cognitive Sciences* **12** 455 – 460

Webster M A, Kaping D, Mizokami Y, Duhamel P, 2004 "Adaptation to natural facial categories" *Nature* **428** 557 – 561

Willis J, Todorov A, 2006 "First impressions: making up your mind after a 100-ms exposure to a face" *Psychological Science* **17** 592 – 598

Zebrowitz L A, Andreoletti C, Collins M A, Lee S Y, Blumenthal J, 1998 "Bright, bad, babyfaced boys: appearance stereotypes do not always yield self-fulfilling prophecy effects" *Journal of Personality and Social Psychology* **75** 1300 – 1320

Zebrowitz L A, Fellous J M, Mignault A, Andreoletti C, 2003 "Trait impressions as overgeneralized responses to adaptively significant facial qualities: evidence from connectionist modeling" *Personality and Social Psychology* **7** 194 – 215

Zebrowitz L Z, Kikuchi M K, Fellous J M F, 2010 "Facial resemblance to emotions: Group differences, impression effects, and race stereotypes" *Journal of Personality and Social Pyschology* **98** 175 – 189

Zebrowitz L A, Montepare J M, 2008 "Social psychological face perception: why appearance matters" *Social and Personality Psychology Compass* **2** 1497 – 1517

Zebrowitz L A, Voinescu L, Collins M A, 1996 "'Wide-eyed' and 'Crooked faced': Determinants of perceived and real honesty across the life span" *Personality and Social Psychology Bulletin* **22** 1258 – 1269

© 2010 a Pion publication

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION
VOLUME 39 2010

www.perceptionweb.com

Conditions of use. This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.