Behavioral and Neural Adaptation in Approach Behavior

Shuo Wang[1], Virginia Falvello[2], Jenny Porter[2], Christopher P. Said[2], and Alexander Todorov[2]

Abstract

■ People often make approachability decisions based on perceived facial trustworthiness. However, it remains unclear how people learn trustworthiness from a population of faces and whether this learning influences their approachability decisions. Here we investigated the neural underpinning of approach behavior and tested two important hypotheses: whether the amygdala adapts to different trustworthiness ranges and whether the amygdala is modulated by task instructions and evaluative goals. We showed that participants adapted to the stimulus range of perceived trustworthiness when making approach decisions and that these decisions were further modulated by the social context. The right amygdala showed both linear response and quadratic response to trustworthiness level, as observed in prior studies. Notably, the amygdala's response to trustworthiness was not modulated by stimulus range or social context, a possible neural dynamic adaptation. Together, our data have revealed a robust behavioral adaptation to different trustworthiness ranges as well as a neural substrate underlying approach behavior based on perceived facial trustworthiness. ■

INTRODUCTION

People often form judgments of others based purely on facial appearance, and these judgments predict a range of outcomes such as elections and sentencing decisions (Todorov, Mandisodza, Goren, & Hall, 2005; Blair, Judd, & Chapleau, 2004; see Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015, for a review). Impressions and judgments of unfamiliar people can be formed after a very brief exposure to faces as short as 100 msec (Todorov, Pakrashi, & Oosterhof, 2009; Bar, Neta, & Linz, 2006; Willis & Todorov, 2006). Such efficient face processing can be attributed to a dedicated neural system in humans including the amygdala (Mende-Siedlecki, Said, & Todorov, 2013; Adolphs, 2010; Haxby, Hoffman, & Gobbini, 2000; Kling & Brothers, 1992).

The human amygdala has long been associated with recognizing faces and facial emotions, as evidenced by lesion, functional neuroimaging, and human electrophysiological studies. Patients who lack a functional amygdala can have a selective impairment in recognizing fearful faces (Adolphs et al., 1999; Broks et al., 1998; Calder, 1996), and BOLD-fMRI shows activation within the amygdala for fearful faces (Whalen et al., 2004; Phillips et al., 1998; Morris et al., 1996). Single neurons in the human amygdala not only encode facial emotions in general (Fried, MacDonald, & Wilson, 1997) but also the subjective judgment (rather than stimulus properties) of facial emotions (Wang et al., 2014). Notably, although the large majority of work has focused on fearful faces (Adolphs, 2008), the amygdala responds to all facial expressions to some extent (Fitzgerald, Angstadt, Jelsone, Nathan, & Phan, 2006), and the amygdala's BOLD response to facial emotions largely depends on the task, even using the same faces (Kim et al., 2010).

The amygdala also plays a key role in processing complex facial attributes like perceived trustworthiness. Human patients with complete bilateral amygdala damage judge unfamiliar individuals to be more approachable and more trustworthy than do control participants (Adolphs, Tranel, & Damasio, 1998), and this is the case even when all internal facial features are occluded, indicating a default approach bias (Harrison, Hurlemann, & Adolphs, 2015). Functional neuroimaging studies using faces with varying levels of trustworthiness have shown both linear and nonlinear responses in the amygdala (Todorov, Baron, & Oosterhof, 2008), despite faces not being subjectively perceived (Freeman, Stolier, Ingbretsen, & Hehman, 2014). Furthermore, both low-frequency and high-frequency images provide sufficient information for the amygdala to differentiate faces on perceived trustworthiness (Said, Baron, & Todorov, 2009). Notably, the amygdala shows task-independent increased activity in response to faces judged untrustworthy when explicitly judging face trustworthiness and when judging on a dimension not related to trustworthiness, suggesting that the amygdala is automatically engaged in both explicit and implicit evaluation of trustworthiness (Todorov, Said, Oosterhof, & Engell, 2011; Winston, Strange, O'Doherty, & Dolan, 2002).

[1]West Virginia University, [2]Princeton University

Journal of Cognitive Neuroscience 30:6, pp. 885–897
doi:10.1162/jocn_a_01243

In this study, we test two important hypotheses of the role of the amygdala in encoding facial trustworthiness. First, previous studies show divergent amygdala responses to faces varying in perceived trustworthiness—some studies show linear response to stimulus levels, whereas others show nonlinear U-shaped response (i.e., quadratic response) to stimulus levels; some even show both responses in the same study (Freeman et al., 2014; Todorov et al., 2008; similarly for emotions (Wang et al., 2017). One possibility is that the amygdala encodes both responses. But it is also likely that, due to different stimuli used across studies and thus different ranges of the stimuli, the quadratic response can be explained by a combination of linear responses (Mende-Siedlecki, Said, et al., 2013), that is, two linear responses from two smaller, nonoverlapping ranges of the stimuli can collectively form a U-shaped response when the stimuli cover the whole range. Here, we test whether the amygdala's response varies as a function of the stimulus range. Furthermore, the amygdala has been shown to process stimulus relevance and evaluative goals and that it can be dynamically modulated by motivations of the perceiver (Cunningham & Brosch, 2012; Cunningham, Van Bavel, & Johnsen, 2008). The second hypothesis we test in this study is whether the amygdala is modulated by top–down instructions or evaluative goals. Both questions are important, because they not only reconcile previous divergent findings but also provide critical insights of the functional role that the amygdala plays in coding facial trustworthiness.

To test these hypotheses, we used faces from a computational model that parametrically manipulates the perceived trustworthiness of faces (Said, Dotsch, & Todorov, 2010; Oosterhof & Todorov, 2008). Specifically, to test the first hypothesis, that is, to understand whether the quadratic response in the amygdala only arises in specific ranges or whether the amygdala can adjust its response to any range of faces, we employed faces from different trustworthiness ranges: besides a full range of trustworthiness (−3 to +3 standard deviations [*SD*] away from the average face), we also used two smaller composing ranges: one negative (−3 to 0 *SD*) and one positive (0 to +3 *SD*). Notably, an identical stimulus can appear in different ranges but at different extremes (e.g., the average face [0 *SD*] is the most trustworthy face in the range of −3 to 0 *SD*, but the least trustworthy face in the range of 0 to +3 *SD*), so that the neural response can be independent of the stimulus. To test the second hypothesis, that is, context dependency of the amygdala's response, we employed two social contexts with the identical stimuli, which were framed as "approaching people for help" versus "looking out for people to avoid" in a dangerous environment. Together, we had six conditions (3 ranges × 2 contexts). It is worth noting that, to eliminate any order or carryover effects, we employed a between-subject design and assigned each participant to only one of the six conditions (see Methods and Discussion).

METHODS

Participants

Fifty-eight healthy, right-handed participants (25 women, mean age ± *SD*, 21.9 ± 4.28 years) with normal or corrected-to-normal vision and no history of neurological disorders participated in the experiment. All participants provided written informed consent according to protocols approved by the institutional review board of Princeton University. Five participants were excluded from analysis because their choice behavior did not monotonically increase as a function of trustworthiness level.

Stimuli and Task

Faces with different levels of trustworthiness were created by FaceGen Modeller (facegen.com/) and a computational model described in previous studies (Said et al., 2010; Oosterhof & Todorov, 2008). We had 3 bottom–up conditions (ranges of the trustworthiness of faces) × 2 top–down conditions (social contexts given by verbal instructions), resulting in six conditions in total, and each participant was randomly assigned to one of the six conditions. The trustworthiness of faces ranged from (1) −3 to +3 *SD* away from the average face, (2) −3 to 0 *SD*, and (3) 0 to +3 *SD* in eight equally spaced levels (Figure 1B).

We employed a between-subject design to avoid any order or carryover effects. In particular, different ranges of stimuli would interact and affect adaptation if they were presented consecutively. Although this reduced the number of participants per condition, thus reducing statistical power, we recruited a relatively large number of participants (53 in total), and we pooled participants of the same social context to study the effect of stimulus range (Figure 3B, F; each range had around 18 participants) and of the same stimulus range to study the effect of social context (Figure 3C, G; each context had around 27 participants) to further increase statistical power.

Before entering the scanner, participants were given one of two sets of verbal instructions describing a social context. The first set of instructions stated, "Imagine that you are in a dangerous neighborhood, and you are looking for people to approach for help." The second set of instructions stated, "Imagine that you are in a dangerous neighborhood, and you are looking out for people to avoid."

Inside the scanner, all participants, regardless of the verbal instructions given at the beginning of the experiment, performed a forced-choice task in which they decided whether to approach or avoid each face as it was shown on the screen. Participants were instructed that the faces might appear to be very similar but that there were subtle differences among them and that the participants should try to make a mixture of balanced approach and avoid responses. We used a rapid event-related

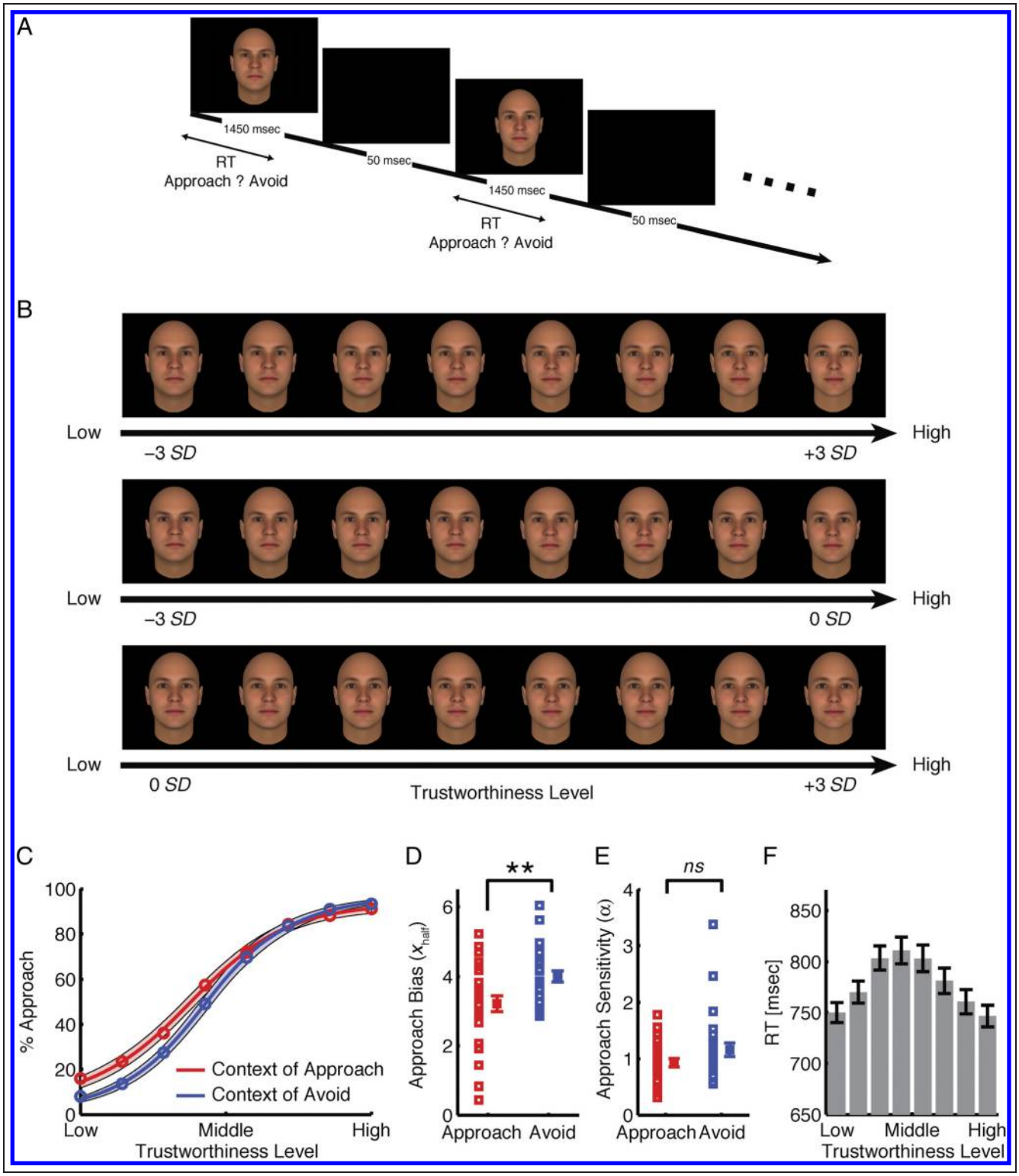

Figure 1. Task, stimuli, and behavioral results. (A) Task. Faces were presented in succession (1450 msec stimulus duration and 50 msec intertrial interval), and participants were asked to make "approach" or "avoid" response to each face. Faces are not shown to scale. (B) Faces with levels of trustworthiness ranging from −3 to +3 *SD* (top), −3 to 0 *SD* (middle), and 0 to +3 *SD* (bottom). (C) Group average of psychometric curves. The psychometric curves show the proportion of trials that participants chose to approach the face as a function of trustworthiness levels. Data were collapsed for the three ranges and averaged at each trustworthiness level (not at the actual *SD*). Shaded area denotes ±*SEM* across participants. (D) Approach bias x_{half} (inflection point of the logistic function). (E) Approach sensitivity α (steepness of the psychometric curve). Individual values are shown on the left, and average values are shown on the right. Error bars denote 1 *SEM* across participants. Asterisks indicate significant difference using unpaired two-tailed *t* test. $^{**}p < .01$. *ns* = not significant ($p > .05$). (F) RTs as a function of trustworthiness level.

design (Aguirre, 2007)—each face was presented for 1450 msec and followed by a 50-msec intertrial interval of black screen. Participants were instructed to respond as quickly as possible, and no feedback message was displayed.

The faces were presented with a Type 1–Index 1 sequence (Said et al., 2010; Aguirre, 2007). This type of sequence ensures that every face was preceded by every other face an equal number of times. Seventy-two rest trials, each 3000 msec long, were distributed throughout the sequence and included in the order counterbalancing. The experiment was divided into eight "runs," and each run had 146 face trials. The number of trials corresponding to each trustworthiness level, as well as the number of rests, was equally balanced among Runs 1–2, Runs 3–4, Runs 5–6, and Runs 7–8. We excluded the first four trials from each run to allow time for the MR signal to reach steady-state equilibrium and for the participants to orient to the task. This procedure also reinstated the adaptation effect for the first valid trial of every run and minimized effects caused by the discontinuity between runs. Participants practiced several trials before entering the scanner. Stimuli were presented using E-Prime software.

Behavioral Analysis

We fitted a logistic function to the behavioral data to obtain smooth psychometric curves:

$$P(x) = \frac{P_{\text{inf}}}{1 + e^{-\alpha(x - x_{\text{half}})}}$$

where P is the percentage of trials with approach response, x is the trustworthiness level, P_{inf} is the value when x approaches infinity (the curve's maximum value), x_{half} is the symmetric inflection point (the curve's midpoint; approach bias), and α is the steepness of the curve (approach sensitivity). P_{inf}, x_{half}, and α were fitted from the observed data (P and x). Flatter curves (smaller α) suggest that participants were less sensitive to the change in trustworthiness levels because they made similar approach/avoid choices given different trustworthiness levels and vice versa for steeper curves (larger α). We derived these parameters for each participant by pooling trials from all runs.

fMRI: Imaging Acquisition

MRI scanning was conducted at Princeton University on a 3-T Allegra MRI scanner (Siemens, Germany). Whole-brain data were acquired with echo-planar T2*-weighted imaging (EPI), sensitive to BOLD signal contrast (32 oblique axial slices, voxel size = 3 × 3 × 4 mm, repetition time = 2000 msec, echo time = 30 msec, flip angle = 80°). T1 weighted structural images were acquired at a resolution of 1 × 1 × 1 mm (repetition time = 2500 msec, echo time = 4.38 msec, flip angle = 8°).

fMRI: Face Localizer Task

We employed an established face localizer task (Said et al., 2010) to localize areas activated by faces. Two runs of the face localizer task were included at the end of the procedure. There were in total 32 alternating blocks of faces and chairs and 10 interleaved rest blocks, and each block was 14 sec long. One run had the fixed sequence of RFCFCRCFCFRFCFCRCFCFR (F = face block, C = chair block, R = rest block), and the other run had the fixed sequence of RCFCFRFCFCRCFCFRFCFCF. The order of the two runs was counterbalanced between participants.

The face localizer task revealed reliable differences in BOLD signal between faces and chairs in bilateral amygdala (peak: Montreal Neurological Institute [MNI] coordinate: $x = -21, y = -9, z = -12, Z = 6.44$, 25 voxels; $x = 18, y = -6, z = -16, Z = 7.34$, 36 voxels; $p < .001$ uncorrected, small volume corrected (SVC) based on anatomical amygdala ROIs), consistent with previous findings (Mende-Siedlecki, Verosky, Turk-Browne, & Todorov, 2013). We defined the functional ROI in the amygdala as all voxels in the amygdala identified by the localizer task (face > chairs; 25 voxels for left amygdala and 36 voxels for right amygdala). Other cortical regions known to be involved in face processing (family-wise error $p < .05$, whole brain), such as the fusiform face area, inferior frontal gyrus, STS/superior temporal gyrus, and dorsal medial pFC were also more activated for faces compared with chairs, consistent with the face-processing network described in previous studies (Mende-Siedlecki, Said, et al., 2013). The chairs minus faces contrast showed stronger activations in visual cortices, the ventral ACC, and posterior cingulate cortex.

fMRI: Imaging Analysis

Neuroimaging data were preprocessed and analyzed using SPM8 (www.fil.ion.ucl.ac.uk/spm/). The first three volumes were discarded to allow the MR signal to reach steady-state equilibrium. The EPI images were sinc interpolated in time for correction of slice-timing differences and realigned to the mean scan by rigid body transformations to correct for head movements. Utilizing linear and nonlinear transformations and smoothing with a Gaussian kernel of FWHM 6 mm, EPI and structural images were coregistered to the T1 MNI 152 template (MNI, International Consortium for Brain Mapping). Global changes were removed by high-pass temporal filtering with a cut-off of 128 sec to remove low-frequency drifts in signal.

In the localizer task, we used a block design and modeled BOLD responses using a general linear model (GLM), with the two regressors for face and chair conditions modeled as boxcar functions convolved with a 2-gamma hemodynamic response function (HRF).

In the main task, we used an event-related design. In the GLM design matrix, for every participant, we estimated a GLM with autoregressive order 1 (AR(1)) and

the following regressors: (1) linear effects of the trustworthiness level; (2) quadratic effects of the trustworthiness level; (3) linear carryover effects of the trustworthiness level, to account for variance caused by adaptation to the previous trial; (4) quadratic carryover effects of the trustworthiness level; and (5) a "face onset" regressor, indicating whenever a face was present. We also repeated the analysis by adding the z-normalized RT (for each participant) as one additional modulator and orthogonalized it to earlier modulators using the default SPM orthogonalization function. We derived similar results when adding RT as a modulator. Notably, we also found qualitatively the same results using a conventional orthogonal quadratic regressor, which confirmed the nonlinear response as observed in previous studies (Mende-Siedlecki, Said, et al., 2013; Said et al., 2010).

For all GLM analyses, six head motion regressors based on the SPM's realignment estimation routine were added to the model (aligned to the first slice of each scan). Multiple linear regression was then run to generate parameter estimates for each regressor for every voxel. The contrast (difference in beta values) images of the first-level analysis were entered into one-sample t tests for the second-level group analysis. For the localizer task, we used small volume correction defined by a priori ROIs of the structural amygdala (Tzourio-Mazoyer et al., 2002). For the main task, we used a functional ROI defined by the parts of the bilateral amygdala identified in the localizer task.

Lastly, to further confirm the nonlinear response in the amygdala, we also performed image analysis using AFNI (Cox, 1996) for individual trustworthiness levels as in a previous study (see Said et al., 2010, for details). We performed group analysis on the regression coefficients extracted from the polynomial regression. The coefficients were then used to determine whether the quadratic effects were significant and were compared between conditions.

RESULTS

Behavior: Adaptation to Face Range

We asked participants to make an "approach" or "avoid" two-alternative forced choice to each face (Figure 1A) while they imagined that they were in a dangerous neighborhood. Faces had levels of trustworthiness, ranging from (1) −3 to +3 *SD*, (2) −3 to 0 *SD*, and (3) 0 to +3 *SD* (Figure 1B), and each participant was given one of the two social contexts through verbal instructions before entering the scanner (see Methods). For each participant, we quantified behavior as the proportion of trials choosing approach as a function of trustworthiness level (Figure 1C). For each trustworthiness range and social context, we found a monotonically increasing relationship between the likelihood of approaching a face and the trustworthiness level of the face (Figures 1C and 2A), showing that participants could well track the gradual change in trustworthiness. Importantly, for all three ranges of trustworthiness, participants had similar psychometric curves, which rose from the least trustworthy face to the most trustworthy face in that particular range regardless of the absolute trustworthiness level in the face, suggesting that participants adapted to the trustworthiness range and found out the least and most trustworthy face dynamically. It is worth noting that the most trustworthy face in the range of −3 to 0 *SD* was identical to the least trustworthy face in the range of 0 to 3 *SD* (Figure 1B). However, participants chose to approach this face in 91.1% of the times in the range of −3 to 0 *SD* but only 17.0% of the times in the range of 0 to 3 *SD* (Figure 2A; two-tailed t test: $p = 2.16 \times 10^{-19}$; all t tests in this study were appropriate for unequal variance). Similar results were found for other identical faces in different ranges (Figure 2A). Moreover, even when analyzing the trials from the first run only, we found a similar behavioral adaptation (Supplementary Figure S1; supplemental material can be found at doi.org/10.6084/figshare.5852601.v1.m9), suggesting that participants could rapidly adapt to stimulus ranges. Together, our data suggest that, when sampling and evaluating faces along the trustworthiness dimension, participants can extract the ranges rapidly and adjust approach behavior dynamically.

Behavior: Modulation by Social Context

We further quantified each psychometric curve using two metrics derived from the logistic function: (i) approach bias x_{half}—the midpoint of the curve (in units of trustworthiness levels) at which participants were equally likely to approach or avoid a face, and (ii) approach sensitivity α—the steepness of the psychometric curve. Based on these two metrics, we found that, in the social context of "approaching people," participants had smaller x_{half} compared with the social context of "avoiding people" (Figure 1D; context of approach: 3.21 ± 1.19 (mean ± *SD*), context of avoid: 4.00 ± 0.80; two-tailed t test across participants: $t(51) = 2.81$, $p = .0070$, effect size in Hedges' g (standardized mean difference): $g = 0.76$, permutation test with 1000 runs, $p = .010$), confirming that participants chose to approach people more in the social context of "approaching people for help." This was primarily driven by the trustworthiness range of −3 to 0 *SD* and the trustworthiness range of 0 to +3 *SD* (Figure 2B; see figure legend for statistics); a two-way ANOVA of Range × Context confirmed the results (main effect of Range: $F(2, 47) = 2.95$, $p = .062$, $\eta^2 = .086$, main effect of Context: $F(1, 47) = 9.28$, $p = .0038$, $\eta^2 = .13$, interaction: $F(2, 47) = 3.20$, $p = .050$, $\eta^2 = .093$). In terms of approach sensitivity, α was similar for both social contexts (Figure 1E; context of approach: 0.93 ± 0.39, context of avoid: 1.16 ± 0.62; $t(51) = 1.64$, $p = .11$, $g = 0.44$, permutation $p = .11$), suggesting that participants had similar sensitivity to trustworthiness regardless of social contexts. This was the case for all separate

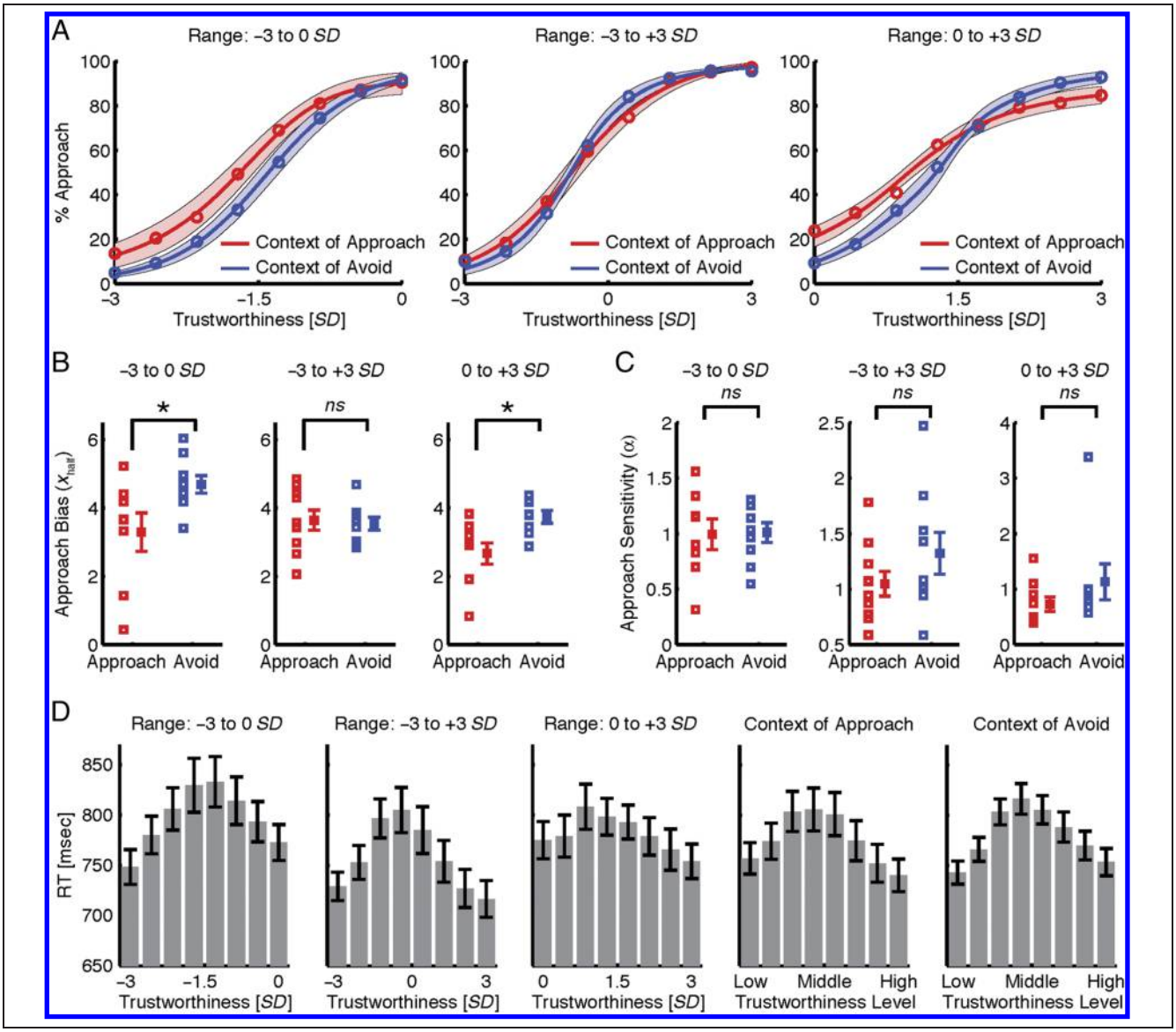

Figure 2. Behavioral results separately for each range of trustworthiness. (A) Group average of psychometric curves. The psychometric curves show the proportion of trials that participants chose to approach the face as a function of trustworthiness levels. Data were analyzed separately for each range of trustworthiness. Shaded area denotes ±*SEM* across participants. (B) Approach bias x_{half}. x_{half} differed between "approach" and "avoid" contexts for the trustworthiness range −3 to 0 *SD* (context of approach: 3.29 ± 1.59 [mean ± *SD*], context of avoid: 4.69 ± 0.78; two-tailed unpaired *t* test: $t(15) = 2.34$, $p = .033$, $g = 1.08$, permutation $p = .036$), and trustworthiness range 0 to +3 *SD* (context of approach: 2.67 ± 0.94, context of avoid: 3.74 ± 0.51; $t(15) = 2.87$, $p = .012$, $g = 1.33$, permutation $p = .008$), but not trustworthiness range −3 to +3 *SD* (context of approach: 3.64 ± 0.92, context of avoid: 3.54 ± 0.57; $t(17) = 0.29$, $p = .78$, $g = 0.13$, permutation $p = .79$). (C) Approach sensitivity α. α did not differ between "approach" and "avoid" contexts for all trustworthiness ranges (−3 to 0 *SD*: context of approach: 0.99 ± 0.39, context of avoid: 1.01 ± 0.27, $t(15) = 0.11$, $p = .92$, $g = 0.049$, permutation $p = .95$; −3 to +3 *SD*: context of approach: 1.05 ± 0.36, context of avoid: 1.32 ± 0.57, $t(17) = 1.28$, $p = .22$, $g = 0.56$, permutation $p = .24$; 0 to +3 *SD*: context of approach: 0.73 ± 0.40, context of avoid: 1.13 ± 0.92, $t(15) = 1.20$, $p = .25$, $g = 0.56$, permutation $p = .23$). Individual values are shown on the left, and average values are shown on the right. Error bars denote 1 *SEM* across participants. Asterisks indicate significant difference using unpaired two-tailed *t* test. *$p < .05$. *ns* = not significant ($p > .05$). (D) RTs for each range (one-way repeated-measures ANOVA of Trustworthiness level; −3 to 0 *SD*: $F(7, 112) = 11.1$, $p = 1.37 \times 10^{-10}$, $\eta^2 = .092$; −3 to +3 *SD*: $F(7, 126) = 25.3$, $p = 2.49 \times 10^{-21}$, $\eta^2 = .13$; 0 to +3 *SD*: $F(7, 112) = 7.02$, $p = 6.06 \times 10^{-7}$, $\eta^2 = .044$) and social context (context of approach: $F(7, 182) = 13.7$, $p = 3.82 \times 10^{-14}$, $\eta^2 = .056$; context of avoid: $F(7, 175) = 21.4$, $p = 1.05 \times 10^{-20}$, $\eta^2 = .12$) varied as a function of trustworthiness levels.

analyses within each trustworthiness range (Figure 2C; all *p*s > .05), and no difference was found between trustworthiness ranges (main effect of Range: $F(2, 47) = 1.18$, $p = .32$, $\eta^2 = .044$, main effect of Context: $F(1, 47) = 2.67$, $p = .11$, $\eta^2 = .050$, interaction: $F(2, 47) = 0.62$, $p = .54$, $\eta^2 = .023$).

Behavior: RT Further Confirmed Adaptation to Face Range

The RT for the approach/avoid decision can be considered as an implicit measure of confidence. Indeed, we found that RT for the approach/avoid judgment was

faster for faces at the extremes compared with faces in the middle of the continuum (Figure 1F; one-way repeated-measures ANOVA of Trustworthiness level: $F(7, 364) = 31.5$, $p = 4.72 \times 10^{-34}$, $\eta^2 = .073$). Notably, RT varied as a function of Trustworthiness level (inverted U-shape) similarly for each trustworthiness range and each social context (Figure 2D), again suggesting that participants adapted to each trustworthiness range. Consistent with this interpretation, even though the smaller physical difference in stimuli between adjacent trustworthiness levels made the faces more similar in the ranges of −3 to 0 *SD* and 0 to 3 *SD* compared with the range of −3 to 3 *SD*, the mean RT did not differ across ranges (one-way ANOVA of Range: $F(2, 50) = 1.11$, $p = .34$, $\eta^2 = .042$), nor at individual identical faces between ranges (all ps $> .05$). Similarly, RT did not differ between social contexts ($F(1, 51) = 0.043$, $p = .84$, $\eta^2 = 8.40 \times 10^{-4}$).

Notably, when we controlled for multiple comparisons using the Bonferroni correction, the above results still held.

Taken together, we found that participants dynamically adapted to the ranges of trustworthiness and used the statistically sampled face range to guide approach behavior. The approach behavior was further modulated by social context.

Both Linear and Quadratic Responses Were Observed in the Amygdala

We next analyzed the neuroimaging data (Figure 3). Each participant performed a separate face localizer task to identify a functional ROI within the amygdala sensitive to faces (see Methods).

We first collapsed all ranges and social contexts and compared the response of voxels within the functional ROI as a function of trustworthiness level. We found that increasing amygdala activation correlated positively with the trustworthiness level, and the activation was specifically within the right amygdala (Figure 3A; peak: MNI coordinate: $x = 27, y = -9, z = -12$, $Z = 3.69$, 5 voxels, $p < .01$, SVC; see Supplementary Figure S2 and Table S1 for whole-brain results), showing that the amygdala tracked the trustworthiness level. No activation was found for decreasing trustworthiness level in the amygdala.

Next, we investigated whether there was also a quadratic response in the amygdala. This time we still found a significant quadratic response primarily in the right amygdala (Figure 3E; peak: $x = 24, y = -3, z = -24$, $Z = 3.31$, 15 voxels, $p < .01$, SVC; see Supplementary Figure S2 and Table S1 for other areas)—the amygdala had the highest

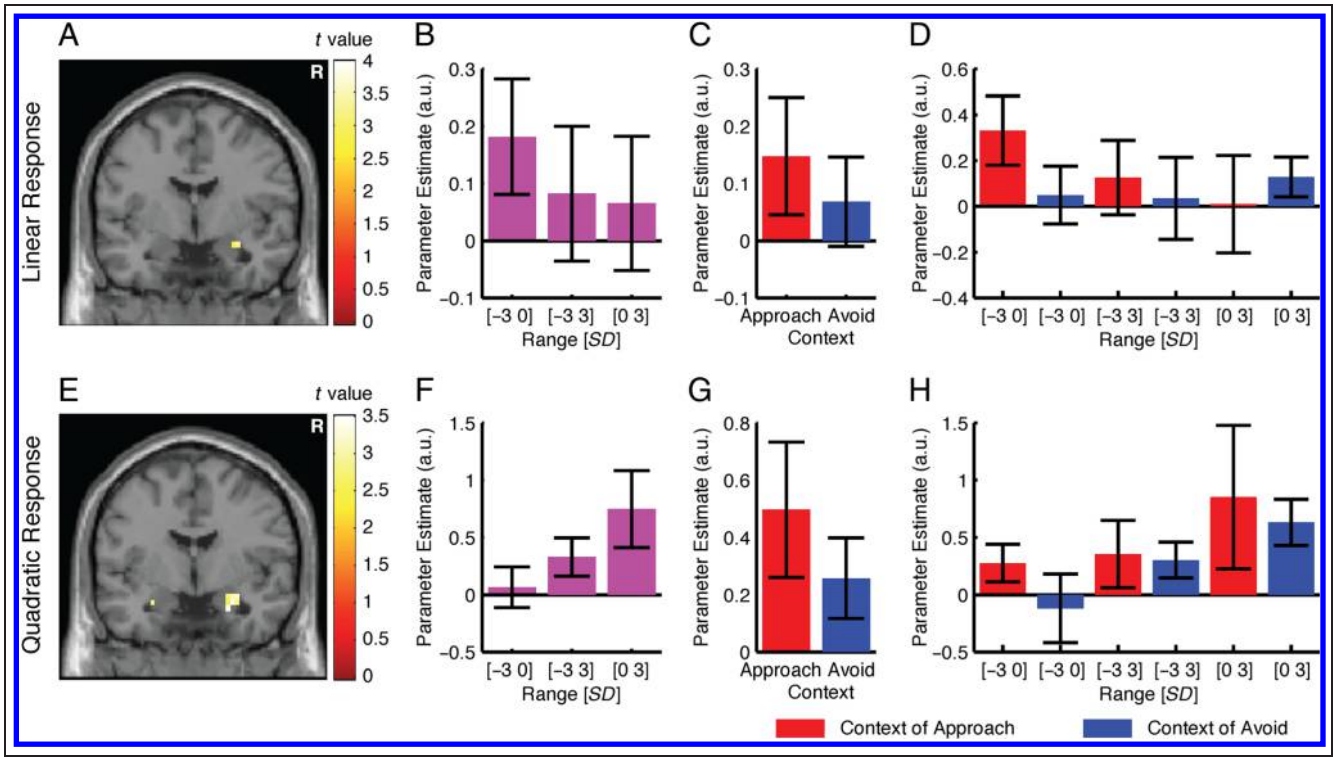

Figure 3. fMRI results. (A–D) Linear response to trustworthiness level. (E–H) Quadratic response to trustworthiness level. (A) Increasing linear response to trustworthiness level was found in the right amygdala. The generated statistical parametric map was superimposed on anatomical sections of the standardized MNI T1-weighted brain template. Images are in neurological format with participant left on image left. R = right. (B) The linear response was not significantly different across stimulus ranges. (C) The linear response was not significantly different between social contexts. (D) Parameter estimate for each stimulus range and social context. Error bars denote 1 *SEM* across participants. (E) Increasing quadratic response to trustworthiness level was primarily found in the right amygdala. (F) The quadratic response was not significantly different across stimulus ranges. (G) The quadratic response was not significantly different between social contexts. (H) Parameter estimate for each stimulus range and social context.

response for both most trustworthy and least trustworthy faces but lowest response for the stimuli in the middle.

Note that when adding RT as another orthogonalized parametric modulator, we derived qualitatively the same results. Thus, our observed quadratic response could not be attributed simply to different RTs. Furthermore, the above results remain qualitatively the same when using an anatomical ROI of the entire amygdala.

In conclusion, we found both linear and quadratic responses to trustworthiness levels in the amygdala, especially in the right amygdala.

The Amygdala's Response to Trustworthiness Adapted to Stimulus Ranges

We have shown above that (a) participants behaviorally adapt to trustworthiness ranges and (b) the amygdala shows both linear and quadratic responses to trustworthiness level. We next examined whether the amygdala's response to trustworthiness also adapted to stimulus ranges. First, we found that the amygdala encoded linear response similarly for each range (Figure 3B; average of all voxels of the functional ROI; two-way ANOVA of Range × Context; main effect of Range: $F(2, 47) = 0.33$, $p = .72$, $\eta^2 = .014$). Furthermore, we found that the amygdala encoded quadratic response similarly for each range as well (Figure 3F; main effect of Range: $F(2, 47) = 1.86$, $p = .17$, $\eta^2 = .072$). Together, these results suggested that the amygdala's response to trustworthiness (both linear and quadratic) adapted to stimulus ranges.

The Amygdala's Response to Trustworthiness Adapted to Social Contexts

We next examined whether the amygdala's response to trustworthiness was modulated by social context. We found that the amygdala encoded linear response similarly for each social context (Figure 3C; two-way ANOVA of Range × Context; main effect of Context: $F(1, 47) = 0.41$, $p = .52$, $\eta^2 = .0083$). Furthermore, we found that the amygdala encoded quadratic response similarly for each social context as well (Figure 3G; main effect of Context: $F(1, 47) = 0.64$, $p = .43$, $\eta^2 = .012$). Therefore, even though social context modulated the approach behavior, it did not modulate the amygdala's response to trustworthiness (neither linear nor quadratic).

No Interaction Was Found between Stimulus Range and Social Context

In addition, we found that there was no significant interaction between stimulus range and social context for both linear response (Figure 3D; two-way ANOVA of Range × Context; interaction: $F(2, 47) = 0.74$, $p = .48$, $\eta^2 = .030$) and quadratic response (Figure 3H; interaction: $F(2, 47) = 0.13$, $p = .88$, $\eta^2 = .0050$). We further confirmed our results and found that, for both linear and quadratic responses, there was no significant difference between context of approach and context of avoid for each stimulus range (two-tailed t test: all ps $> .05$).

Lastly, a whole-brain analysis of neural adaptation is shown in Supplementary Table S1.

Together, our data indicate that the amygdala adapts dynamically to the range of faces as well as social context.

Quadratic Response to Trustworthiness Levels in the Amygdala

Like this study, many previous studies have observed a nonlinear (quadratic) response in the amygdala to trustworthiness levels (Mende-Siedlecki, Said, et al., 2013; Said et al., 2010). We here further conducted an additional analysis to confirm this quadratic response. Specifically, we first fitted a model for each face (each individual trustworthiness level), extracting the mean regression response to each face (see Said et al., 2010) in an anatomically defined ROI of the amygdala. Second, for each participant, we fitted a second-order polynomial regression to these responses. Finally, we performed group analyses on the quadratic coefficients extracted from the polynomial regression and compared the coefficients across stimulus ranges and social contexts.

As shown in Figure 4, we found that the overall response in the amygdala demonstrated a quadratic relationship to trustworthiness levels (Figure 4A; quadratic coefficients: 0.0013 ± 0.0034 (mean ± *SD*); one-sample two-tailed t test against 0: $t(51) = 2.72$, $p = .0090$). This was the case in both the left amygdala (Figure 4B; 0.0013 ± 0.0032, $t(51) = 2.94$, $p = .0049$) and right amygdala (Figure 4C; 0.0013 ± 0.0041, $t(51) = 2.23$, $p = .030$), and there was no significant difference between the left versus right amygdala (two-tailed paired t test: $t(51) = 0.074$, $p = .94$).

We also separately analyzed the early half (Runs 1–4; Figure 4D; 0.0012 ± 0.0050, $t(51) = 1.72$, $p = .091$) versus late half (Runs 5–8; Figure 4E; 0.0016 ± 0.0041, $t(51) = 2.73$, $p = .0088$) of the experiment and found similar results ($t(51) = 0.40$, $p = .69$), suggesting that the amygdala adapted to the trustworthiness levels rapidly and that such adaptation was sustained.

Finally, confirming the above results, the amygdala had a similar quadratic response to trustworthiness levels for different ranges (Figure 4F; −3 to 0 *SD*: 0.0022 ± 0.0029, −3 to +3 *SD*: 0.0011 ± 0.0029, 0 to +3 *SD*: 0.00062 ± 0.0042; one-way ANOVA: $F(2, 49) = 0.95$, $p = .40$). Therefore, our data are inconsistent with the hypothesis that the commonly observed nonlinear response in the amygdala is a combination of two linear responses from two smaller composing ranges of the stimuli. Furthermore, the amygdala had a similar quadratic response for different social contexts (Figure 4G; approach: 0.00090 ± 0.0039, avoid: 0.0017 ± 0.0028; two-tailed unpaired t test: $t(50) = 0.84$, $p = .40$).

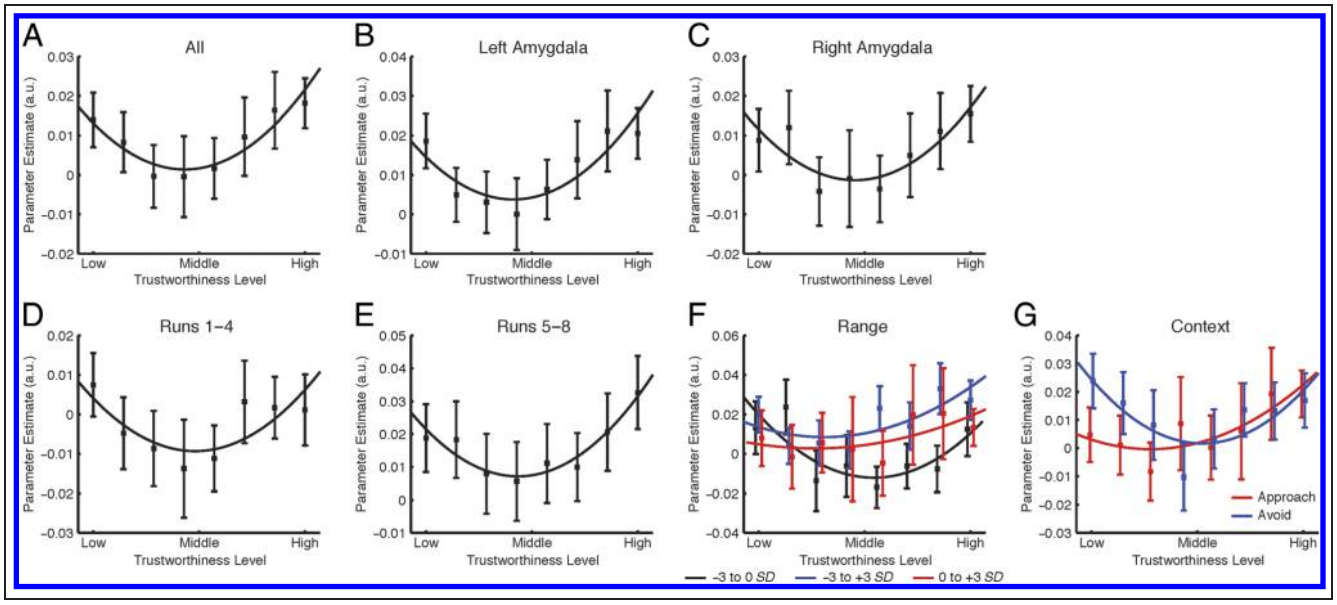

Figure 4. Nonlinear (quadratic) response in the anatomical amygdala. Each plot shows the parameter estimate (beta values) for each trustworthiness level. Error bars denote 1 *SEM* across participants. The line denotes the best quadratic fit. (A) The entire amygdala for all runs. (B) The left amygdala for all runs. (C) The right amygdala for all runs. (D) The entire amygdala for Runs 1–4. (E) The entire amygdala for Runs 5–8. (F) The entire amygdala for different stimulus ranges. (G) The entire amygdala for different social contexts.

The results were qualitatively the same when repeating the analysis using the functional ROI of the amygdala (Figure S3; see figure legend for statistics).

In conclusion, our data not only confirmed the nonlinear (quadratic) response as observed in previous studies but importantly showed that such response was similar for different stimulus ranges and contexts, which in turn suggested a dynamic neural adaptation in the amygdala.

DISCUSSION

In this study, we examined the behavioral and neural adaptation to different ranges of trustworthiness, as well as the role of social context. We specifically tested two hypotheses: whether the amygdala adapts to different trustworthiness ranges and whether the amygdala is modulated by task instructions inducing different evaluative goals. Depending on the overall trustworthiness range, a particular face could be regarded as trustworthy in one setting but untrustworthy in another. However, participants rapidly adapted to the trustworthiness ranges and dynamically adjusted approach behavior. Within each range of trustworthiness, half of the participants had a social context to look for faces to approach, and half had a social context to look for faces to avoid. As expected, participants approached faces more in the social context of approach. The fMRI findings showed both linear and quadratic responses to trustworthiness levels in the amygdala, and such responses also adapted to stimulus range and social context. Together, our data reveal both behavioral and neural adaptations to facial trustworthiness.

Behavioral Adaptation to Faces

Perception of faces is shaped by both the long-term average of experienced faces and immediate learning of faces. That is, this perception is influenced by the set of faces that observers are exposed to over a life time and in the immediate context (see Webster & MacLeod, 2011, for a review). In this study, participants adapted to the ranges of trustworthiness rapidly and adjusted approach behavior according to the statistically sampled face range. In line with this finding, adaptation to a consistent distortion in faces shifts what looks most normal and what looks most attractive toward that distortion, suggesting that perceptual adaptation can rapidly recalibrate people's preferences to fit the faces they see (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003). Furthermore, these adaptation effects are pronounced for both natural variations in faces and for natural categorical judgments about faces, indicating that adaptation may routinely influence face perception (Webster, Kaping, Mizokami, & Duhamel, 2004). In the context of a computationally derived "face space," adaptation specifically shifts perception along a trajectory passing through the adapting and average faces, and identification of a face is facilitated by adapting to its computationally opposite identity (Leopold, O'Toole, Vetter, & Blanz, 2001). Notably, a recent study has shown that statistical learning of the distribution of faces can bias face evaluation—faces are judged more negatively when they deviate further from a learned central tendency (Dotsch, Hassin, & Todorov, 2016). Lastly, in addition to the effect of explicit social context, the extraction of face range and computation of face norm might also indicate that the implicit

context of faces (i.e., the presence of other faces) can influence trustworthiness judgment, consistent with prior research that faces with ambiguous valence (i.e., surprised faces) are interpreted as more positive when they are presented within the context of positive faces (i.e., happy faces), whereas they are interpreted as more negative when they are presented within the context of negative faces (i.e., angry faces; Neta, Davis, & Whalen, 2011). In conclusion, our present finding of adaptation to facial trustworthiness fits a broader framework of face adaptation, and these findings together suggest that people not only adapt to various aspects of faces but also adjust their behavior accordingly.

Neural Adaptation to Faces

In this study, we found that the amygdala not only adapted to bottom–up stimulus ranges but also had similar responses for different top–down social contexts. Consistent with this study, it has been shown that the amygdala encodes the social value of faces independent of the task (i.e., approach-avoidance vs. one-back recognition decisions), suggesting that the amygdala encodes bottom–up stimulus independent of top–down instructions (Todorov et al., 2011). In addition, the amygdala has been shown to dynamically adapt to different top–down evaluative goals using the same bottom–up stimulus. For example, when evaluating positive aspects of famous people, the amygdala activation is associated with positive names, whereas when evaluating negative aspects of famous people, the amygdala activation is then associated with negative names (Cunningham et al., 2008).

The properties of face adaptation suggest that this adaptation in part reflects response changes at high and possibly face-specific levels of visual processing, although the form of the adaptation and the norm-based codes also show many parallels with the adaptations and functional organization underlying the encoding of perceptual attributes like color (Webster & MacLeod, 2011). Prior studies of neural adaptation reveal a response proportional to the distance from a stored prior: The response to a stimulus is proportional to its dissimilarity from the immediately preceding stimulus (Jiang et al., 2006), and studies of norm-based coding have found that, when facial geometry (head shape, hair line, internal feature size and placement) is varied, the fMRI signal increases with increasing distance from the mean face (Loffler, Yourganov, Wilkinson, & Wilson, 2005). Such norm-based face encoding has also been shown at the level of single neurons in the monkey inferotemporal cortex: These neurons are tuned around the average, identity-ambiguous face and reflect structural differences between an incoming face and an internal reference or norm (Leopold, Bondar, & Giese, 2006). Together, these prior findings are consistent with our present result: Given a different range of stimuli, participants rapidly extracted the average trustworthy face (i.e., the norm), and regardless of the ranges, the amygdala was tuned around the average face and showed both linear and nonlinear responses.

A recent study shows that variation in timescale can account for both neural adaptation and norm-based effects and that there is a smooth gradient of temporal integration across the ventral pathway of the visual cortex (Mattar, Kahn, Thompson-Schill, & Aguirre, 2016). Interestingly, face adaptation is modulated by emotion, as neural adaptation decreases linearly with negative valence, with the least adaptation to fearful expressions (Gerlicher, van Loon, Scholte, Lamme, & van der Leij, 2014), but the adaptation to face identities in the fusiform face-responsive regions is invariant to spatial scales (low vs. high spatial frequency; Eger, Schyns, & Kleinschmidt, 2004). A clear future direction is to test beyond facial trustworthiness and investigate whether similar neural adaptation across ranges also exists for other facial attributes such as emotion and identity.

Amygdala's Response to Trustworthiness

We found that the amygdala increased activity to more trustworthy faces. However, some studies have observed increased responses in the amygdala for untrustworthy faces (Winston et al., 2002), and during implicit evaluations of trustworthiness of faces, the amygdala has been shown to increase response to decreasing facial trustworthiness (Engell, Haxby, & Todorov, 2007). This is likely due to different tasks involved and different face stimuli (synthetic vs. natural faces). A similar dissociation has been observed in the amygdala for emotional faces: The response increases for fearful faces in a task using a face mask whereas it decreases for the same fearful faces in a task using a pattern mask (Kim et al., 2010). Therefore, the sign of the amygdala's BOLD response to faces may largely depend on the task and stimuli. It is worth noting that both directions of response have been extensively observed in previous studies (see Mende-Siedlecki, Said, et al., 2013, for a comprehensive review).

On the other hand, we found that the amygdala increased activity to faces at the extremes of the face continuum, consistent with other studies (Freeman et al., 2014; Todorov et al., 2008). Similarly, extensive nonlinear responses have been observed in the fusiform gyri and dorsal amygdala, with greater responses to faces at the extremes of the face valence continuum than faces in the middle, and these responses correlated with participants' avoidance decisions, with faces more likely to be avoided evoking stronger responses (Todorov et al., 2011). However, another possible explanation of the amygdala's quadratic response to trustworthiness levels could be that the amygdala encodes the distance from the average face, consistent with a previous finding that the amygdala has a stronger response to the extremes of

the dimensions than to faces near the average face (Said et al., 2010; also see above for norm-based face coding).

A third explanation, not mutually exclusive with the explanations above, is that the quadratic response may track the intensity/magnitude of the stimulus. Whereas the valence of the stimulus increases monotonically from the most untrustworthy face to the most trustworthy face, both the most trustworthy and the most untrustworthy faces have a high intensity/magnitude to express trustworthiness and vice versa for faces in the middle of the continuum. Trustworthiness intensity/magnitude was in turn associated with how easily (RT; Figure 1F) and consistently (choice; Figure 1C) a participant could judge the trustworthiness of the face—for each range, faces at the extremes of the range (the most and least trustworthy faces) were judged faster and had more consistent judgments, whereas faces in the middle of the range were judged more slowly and had more variable judgments.

We found that both linear and quadratic responses in the amygdala adapted to stimulus range and social context. Adaptation of the linear response indicates that the amygdala always differentiates the valence of the face from a given population. Adaptation of the quadratic response, on the other hand, indicates that the amygdala keeps track of the stimulus intensity/magnitude and encodes a relative distance to the average face, which together helps to define the boundary and scope of the stimulus space. It is worth noting that we instructed participants to make approach/avoid decisions rather than explicitly rate the perceived trustworthiness of a face (i.e., trustworthiness level or trustworthiness intensity/magnitude). However, as an implicit measure of stimulus intensity/magnitude, we found that RT had a similar pattern across stimulus ranges (Figure 2D) and did not increase for smaller ranges, where the physical difference in stimuli between adjacent trustworthiness levels was smaller and thus faces were less discriminable and more difficult to compare. Therefore, these findings suggest that trustworthiness intensity/magnitude could also be readjusted even within a smaller stimulus range, consistent with the amygdala's neural adaptation. An important future study is needed to directly link amygdala's adaptation to behavioral adaptation and investigate to what extent the amygdala's response contributes to behavioral adaptation.

Possible Caveats and Future Directions

Trustworthiness judgments are highly correlated with valence and attractiveness judgments (Mende-Siedlecki, Said, et al., 2013; Oosterhof & Todorov, 2008). Therefore, the present results might also be explained by general responses to stimulus valence. This is consistent with the amygdala's response profile to emotional faces—the amygdala shows both linear and quadratic responses to emotion levels (Wang et al., 2017). Future studies will be needed to study the specificity of the amygdala's response.

It is worth noting that we employed a between-subject design to avoid any order or carryover effects of different conditions, resulting in a relatively large number of participants (53 in total). To further increase statistical power, we pooled participants of the same social context to study the effect of stimulus range (Figure 3B, F; each range had around 18 participants), and we pooled participants of the same stimulus range to study the effect of social context (Figure 3C, G; each context had around 27 participants). However, it is notable that our conclusion of neural adaptation in the amygdala was based on the similarity between conditions, which was in turn indicated by the lack of a statistical difference. Before we discuss the issue of statistical power, we should point out that we observed both reliable behavioral and neural responses to the perceived trustworthiness of faces. It is indeed possible that these responses may vary as a function of the range of faces and the social context, but we lacked sufficient statistical power to detect such effects. A power analysis based on the present data indicated that we would need an extremely large sample size to detect these effects, suggesting that if the effects exist they are rather small.

We found that even when restricting our analysis to trials from the first run only, there was a similar behavioral adaptation (Supplementary Figure S1), suggesting that participants could rapidly adapt to the stimulus ranges. Although participants needed to sample a series of faces to find out the norm and range, an important question will be to reveal the time point at which the adaptation first occurs. Furthermore, it is also important to understand the factors that can influence behavioral and neural adaptation, such as the spacing between stimulus levels and the distance to the face norm. The adaptation effect observed in this study should also be generalized to emotions and other complex facial attributes.

Conclusion

In this study, we employed faces from different trustworthiness ranges to investigate whether people use the statistically sampled face norms and ranges to guide approach behavior and whether the amygdala dynamically computes the face norms and ranges from a population of faces. Comparing a full range of trustworthiness (−3 to +3 *SD* away from the average face) to two smaller composing ranges, one negative (−3 to 0 *SD*) and one positive (0 to +3 *SD*), we also tested whether the commonly observed nonlinear response in the amygdala is composed of two linear responses from the composing ranges. Furthermore, with the identical stimuli, we employed two social contexts, which were framed as "approaching people for help" versus "looking out for people to avoid" in a dangerous environment, to test whether the amygdala would be modulated by top–down

instructions. Behaviorally, we found that participants dynamically adjusted approach behavior and adapted to the trustworthiness ranges. The neuroimaging data confirmed the parametric coding of trustworthiness, but surprisingly, such parametric coding was independent of the trustworthiness range of the faces or the social context, suggesting that the amygdala also adapted dynamically to the range of faces as well as the social context. Together, our data reveal a robust behavioral adaptation to different trustworthiness ranges as well as a neural substrate underlying approach behavior based on perceived facial trustworthiness.

Acknowledgments

We thank Rongjun Yu, Sai Sun, and Joel Martinez for help with data analysis. This research was supported by the PNI Innovation Award (to A. T.) and the Dana Foundation Clinical Neuroscience Research Award (to S. W.). A. T. designed and supervised experiments. V. F., J. P., and C. S. performed research. All authors analyzed data. S. W. and A. T. wrote the paper.

Reprint requests should be sent to Shuo Wang, Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, WV 26506, or via e-mail: wangshuo45@gmail.com or Alexander Todorov, Department of Psychology, Princeton University, Princeton, NJ 08544, or via e-mail: atodorov@princeton.edu.

REFERENCES

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current Opinion in Neurobiology, 18,* 166–172.

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences, 1191,* 42–61.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393,* 470–474.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., et al. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia, 37,* 1111–1117.

Aguirre, G. K. (2007). Continuous carry-over designs for fMRI. *Neuroimage, 35,* 1480–1494.

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.

Broks, P., Young, A. W., Maratos, E. J., Coffey, P. J., Calder, A. J., Isaac, C. L., et al. (1998). Face processing impairments after encephalitis: Amygdala damage and recognition of fear. *Neuropsychologia, 36,* 59–70.

Calder, A. J. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology, 13,* 699–745.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science, 21,* 54–59.

Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activity. *Psychological Science, 19,* 152–160.

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour, 1,* 0001.

Eger, E., Schyns, P. G., & Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. *Neuroimage, 22,* 232–242.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience, 19,* 1508–1519.

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *Neuroimage, 30,* 1441–1448.

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., & Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. *Journal of Neuroscience, 34,* 10573–10581.

Fried, I., MacDonald, K. A., & Wilson, C. L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron, 18,* 753–765.

Gerlicher, A. M. V., van Loon, A. M., Scholte, H. S., Lamme, V. A. F., & van der Leij, A. R. (2014). Emotional facial expressions reduce neural adaptation to face identity. *Social Cognitive and Affective Neuroscience, 9,* 610–614.

Harrison, L. A., Hurlemann, R., & Adolphs, R. (2015). An enhanced default approach bias following amygdala lesions in humans. *Psychological Science, 26,* 1543–1555.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4,* 223–233.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron, 50,* 159–172.

Kim, M. J., Loucks, R. A., Neta, M., Davis, F. C., Oler, J. A., Mazzulla, E. C., et al. (2010). Behind the mask: The influence of mask-type on amygdala response to fearful faces. *Social Cognitive and Affective Neuroscience, 5,* 363–368.

Kling, A. S., & Brothers, L. A. (1992). The amygdala and social behavior. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction* (pp. 353–377). New York: Wiley-Liss.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature, 442,* 572–575.

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.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience, 8,* 1386–1391.

Mattar, M. G., Kahn, D. A., Thompson-Schill, S. L., & Aguirre, G. K. (2016). Varying timescales of stimulus integration unite neural adaptation and prototype formation. *Current Biology, 26,* 1669–1676.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience, 8,* 285–299.

Mende-Siedlecki, P., Verosky, S. C., Turk-Browne, N. B., & Todorov, A. (2013). Robust selectivity for faces in the human amygdala in the absence of expressions. *Journal of Cognitive Neuroscience, 25,* 2086–2106.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383,* 812–815.

Neta, M., Davis, F. C., & Whalen, P. J. (2011). Valence resolution of ambiguous facial expressions using an emotional oddball task. *Emotion, 11,* 1425–1433.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 11087–11092.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London, Series B, Biological Sciences, 265,* 1809–1817.

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

Said, C. P., Baron, S. G., & Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience, 21,* 519–528.

Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimensions. *Neuropsychologia, 48,* 3596–3605.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience, 3,* 119–127.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308,* 1623–1626.

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

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27,* 813–833.

Todorov, A., Said, C. P., Oosterhof, N. N., & Engell, A. D. (2011). Task-invariant brain responses to the social value of faces. *Journal of Cognitive Neuroscience, 23,* 2766–2781.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273–289.

Wang, S., Tudusciuc, O., Mamelak, A. N., Ross, I. B., Adolphs, R., & Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. *Proceedings of the National Academy of Sciences, U.S.A., 111,* E3110–E3119.

Wang, S., Yu, R., Tyszka, J. M., Zhen, S., Kovach, C., Sun, S., et al. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications, 8,* 14821.

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

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences, 366,* 1702–1725.

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science, 306,* 2061.

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

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5,* 277–283.