Intersubject Variability in Fearful Face Processing: The Link Between Behavior and Neural Activation

Tracy J. Doty[1,2], **Shruti Japee**[1], **Martin Ingvar**[2], and **Leslie G. Ungerleider**[1]

[1]Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, USA

[2]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Abstract

Stimuli that signal threat show considerable variability in the extent to which they enhance behavior, even among healthy individuals. However, the neural underpinning of this behavioral variability is not well understood. By manipulating expectation of threat in an fMRI study of fearful vs. neutral face categorization, we uncovered a network of areas underlying variability in threat processing in healthy adults. We explicitly altered expectation by presenting face images at three different expectation levels: 80%, 50%, and 20%. Subjects were instructed to report as fast and as accurately as possible whether the face was fearful (signaled threat) or not. An uninformative cue preceded each face by 4 seconds (s). By taking the difference between response times (RT) to fearful compared to neutral faces, we quantified an overall fear RT bias (i.e. faster to fearful than neutral faces) for each subject. This bias correlated positively with late trial fMRI activation (8 s after the face) during unexpected fearful face trials in bilateral ventromedial prefrontal cortex, the left subgenual cingulate cortex, and the right caudate nucleus and correlated negatively with early trial fMRI activation (4 s after the cue) during expected neutral face trials in bilateral dorsal striatum and the right ventral striatum. These results demonstrate that the variability in threat processing among healthy adults is reflected not only in behavior but also in the magnitude of activation in medial prefrontal and striatal regions that appear to encode affective value.

Keywords

emotion; fearful faces; ventromedial prefrontal cortex; subgenual cingulate cortex; striatum; affective value

INTRODUCTION

Threat has a privileged status in the brain, which is reflected in enhanced neural processing of threatening stimuli relative to neutral stimuli (Davidson, Maxwell, & Shackman, 2004; Ohman & Mineka, 2001). From an evolutionary perspective, identifying and reacting both quickly and efficiently to stimuli that signal threat is an important skill for survival. This

Corresponding Author: Tracy Jill Doty, Translational Neuroscience Branch, Human Research and Engineering Directorate, United States Army Research Laboratory, Aberdeen Proving Ground, MD 21005, Telephone: (410) 278-0722, tracy.doty@alumni.duke.edu, Fax: (410) 278-8828.

privileged status has been shown in the laboratory setting using fearful and angry faces, both of which signal threat. Fearful faces represent an indirect threat to the viewer, because the person displaying the fearful expression is afraid of something in the surrounding environment. Angry faces indicate a direct threat to the viewer, because the person displaying the angry expression is upset with the viewer. Angry faces are found more quickly during visual search than faces displaying positive emotions (Fox et al., 2000; Hansen & Hansen, 1988). Also, the detection of a target is enhanced when presented immediately after an angry or fearful face (Fox, 2002; Wilson & MacLeod, 2003). Additionally, there is enhanced evoked brain activation to angry and fearful faces relative to neutral faces in both the amygdala and the network of areas that selectively process faces (Morris et al., 1998; Pessoa, McKenna, Gutierrez, & Ungerleider, 2002).

More recently, the behavioral enhancement for threatening faces has been shown to vary with anxiety levels. Large speed advantages for detecting targets in the location preceded by threatening faces have been associated with anxiety in patients [(Bradley, Mogg, White, Groom, & de Bono, 1999), also see review (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007)]. Additionally, the effects of anxiety are evident even among the normal population. Healthy individuals with higher levels of state and trait anxiety show an attentional bias towards targets presented following an angry face (Mogg & Bradley, 1999). Healthy subjects with greater anxiety are also better able to correctly identify a fearful face when less emotional intensity is shown in that face (Richards et al., 2002), and they are more accurate at categorizing fearful faces (Surcinelli, Codispoti, Montebarocci, Rossi, & Baldaro, 2006; Winton, Clark, & Edelmann, 1995). Thus, while threatening faces enjoy enhanced behavioral processing, this enhancement is not uniform across subjects and instead appears to be related to subclinical levels of anxiety.

One can also see differences in the processing of threatening faces without using a personality measure, such as anxiety, to pre-classify healthy subjects. Pessoa and colleagues used signal detection theory to classify their subjects into one of two groups: subjects who were better than chance at detecting briefly presented fearful faces and those who were no better than chance (Japee, Crocker, Carver, Pessoa, & Ungerleider, 2009; Pessoa, Japee, Sturman, & Ungerleider, 2006; Pessoa, Japee, & Ungerleider, 2005). They found that subjects who were better than chance, compared to those who were not, showed increased activity in two preselected areas: the amygdala and ventral visual cortex. These studies demonstrate that there is a difference in neural activation in face and emotion processing areas between healthy adults who can efficiently detect the presence of fearful faces and healthy adults who cannot, but it is still unclear what neural mechanisms underlie behavioral variability to process fearful faces in healthy adults.

To better understand the neural correlates of variability in threat processing in healthy adults, we combined fMRI with a simple decision-making task in which subjects categorized faces as either fearful or neutral under varying expectation conditions. We built upon previous research in four ways. First, we used behavioral data to guide our fMRI analysis, by correlating the RT bias in responding to threatening faces with magnitude of brain activation. This enabled us to isolate areas involved in hypervigilance towards threatening faces. Second, we employed an analysis method that allowed us to estimate the fMRI

response over time. This approach allowed us to probe for variability in fMRI responses across an entire trial by investigating each timepoint in the hemodynamic response. Third, we manipulated the predictability of fearful faces in our decision-making task. As expectation of threat defines anxiety (Barlow, 2004; Epstein, 1972), we hoped that this manipulation would elicit a spectrum of fearful face processing among healthy adults and build upon previous work investigating intersubject variability in fearful face processing without changes in expectation. Fourth, we collected personality measures related to anxiety to determine if there would be a relationship between enhanced fearful face detection and anxiety. By employing these four advancements, we hoped to identify the sources of threat processing variability in the healthy human brain.

EXPERIMENTAL PROCEDURES

Subjects

Twenty-four right-handed healthy volunteers (15 females; mean age 28 years ±6.6 SD; see Table S1 for more demographic information) with no neurological or psychiatric history participated in this study. All subjects gave written informed consent in accordance with protocols approved by the National Institute of Mental Health Institutional Review Board.

Personality Assessments

All subjects completed the Spielberger State Trait Anxiety Inventory (STAI) at the beginning of the experiment (Spielberger, 1983). A subset of these subjects completed two additional questionnaires measuring personality types related to anxiety: 16/24 completed the NEO Five Factor Inventory (NEO-FFI), and 13/24 completed the Harm Avoidance Subscale of the Tridimensional Personality Questionnaire (TPQ-HA) (Cloninger, Przybeck, & Svrakic, 1991; Costa & McCrae, 1992).

Study Design

All stimuli were presented with the Presentation software package (Neurobehavioral Systems Inc., Albany, CA). The stimuli consisted of 40 fearful and 40 neutral faces chosen from the Karolinska Directed Emotional Faces (KDEF) set (Lundqvist, Flykt, & Öhman, 1998), as well as from a set of faces developed and validated at NIMH (Ishai, Pessoa, Bikle, & Ungerleider, 2004). A white fixation cross was displayed centrally during the experiment. Faces were presented on top of this fixation cross and subtended 4° of visual angle.

Subjects were instructed to categorize each face as fearful or neutral as accurately and quickly as possible by pressing one of two buttons. To manipulate expectation of fear, we presented fearful and neutral faces in runs containing different proportions of the two face types. There were three different run types: 1) runs containing 80% fearful faces and 20% neutral faces (80F:20N); 2) runs containing 20% fearful faces and 80% neutral faces (20F: 80N); and 3) runs containing 50% fearful faces and 50% neutral faces (50F:50N). This created six different stimulus types: fearful and neutral faces presented in 80% (expected), 20% (unexpected), and 50% expectation conditions. Subjects were verbally and visually notified of the run type (i.e. 80F:20N, 50F:50N, or 20F:80N) at the beginning of each run. Each run contained 50 trials (for schematic of a single trial see Figure 1A). Within each run

of trials, faces were pseudo-randomized and a given face was not repeated. At the beginning of each trial, the white fixation cross turned red for 250 ms (cue). Four seconds (s) later a face was presented centrally for 250 ms. The cue itself indicated only the upcoming presentation of a face, and not its emotional content.

Each run type was repeated three times and was presented in a pseudo-randomized fashion such that subjects always completed each of the three different expectation run types before receiving a repeated one. The experiment consisted of nine 6-minute runs. An event-related design was used and the time between the end of the response period (2 s after the appearance of the face) and the beginning of the next trial was jittered 0–8 s (averaging 3 s) using optimal stimulus sequences generated by the program *optseq* (Dale, 1999). Subjects responded to each face with their right index or middle finger (counterbalanced across subjects). No feedback was given to the subjects after their response. Before entering the scanner, subjects completed a short practice session that included a 50F:50N run of ten trials.

Behavioral Analysis

Trials in which subjects chose the wrong face type or did not respond were labeled incorrect and removed from the analysis. Accuracy ratings were high (>85% correct for each stimulus type for each subject), so the focus of our behavioral analysis was reaction time (RT). Statistical results were computed using a repeated measures ANOVA (Expectation Condition x Valence) and post-hoc analyses were conducted using a Fisher Least Significant Differences (LSD) Test. Additionally, in order to directly compare effect sizes between experiments, we included the partial eta squared ($p\eta^2$) measure of magnitude for results from all behavioral ANOVAs.

During runs in which a face type was presented for 80% or 50% of the trials, there were multiple repetitions in a row of that face type. Because repetition effects have been shown to differ between fearful and neutral faces (Ishai et al., 2004), we confined our analyses to trials in which faces were presented immediately after the alternate face type. In addition, only correct trials were included in the final analyses (please see Supplementary Methods for details on the number of trials for each stimulus type).

To quantify RT differences across our subjects, we created a Valence Bias measure for each subject which was calculated from the RTs to fearful faces in all three expectation conditions subtracted from the RTs to neutral faces in all three expectation conditions. The Valence Bias measure used was calculated as:

Valence Bias = RT [Neutral (80% + 50% + 20%)] − RT [Fearful (80% + 50% + 20%)]

This measure quantified RT differences as a single number to reflect each subject's overall bias to respond more quickly or more slowly to fearful relative to neutral faces. In our group analyses, we created two Valence Bias groups: 1) Fear Fast Responders: Those subjects who were faster to respond to fearful faces compared to neutral faces, 2) Fear Slow Responders: Those subjects who were faster to respond to neutral faces compared to fearful faces.

fMRI Data Acquisition

Whole brain MR images were collected on a 3T GE Signa scanner (GE Medical Systems) using an 8-channel GE head coil. Standard parameters for echoplanar imaging data were used, including: FOV 200 mm, 64×64 matrix, 25 axial slices of 5-mm thickness, 3.125 mm in-plane resolution, 2.0 s TR, 30 ms TE, 90° flip angle. MP-RAGE scans, collected in the same session, were acquired for anatomical comparison using the following parameters: FOV 22.0 mm, 256×256 matrix, minimum full TE, 1.2-mm slice thickness.

fMRI Data Analysis

Imaging data were preprocessed, analyzed, and displayed with the AFNI software package (Cox, 1996). Individual subject data were preprocessed as follows: slice timing correction, volume registration, smoothing via a 6-mm full width half max filter, normalization, and applying a six-parameter rigid motion correction.

Next, a variable shape deconvolution model was computed for each subject individually. In order to model the fixed length expectation (cue) period that preceded every face stimulus as well as the response period that followed every face, individual subject imaging data were deconvolved using a 16-s tent function. This tent function was time-locked to the TR (i.e. a stick function with 8 sticks), started 4 s before the face in order to model the onset of the cue, and concluded 12 s after the presentation of the face. This analysis method allowed us to estimate the BOLD signal at each timepoint in the trial individually. The six different stimulus types (i.e. 3 different expectation conditions $\times$ 2 face types) were modeled separately. Additionally, we modeled the 2nd, 3rd, and 4th occurrences of each face type (i.e. repetitions in a train of stimuli), but these data were not included in the current statistical analyses (See Supplementary Methods for more information). In preparation for the group analysis, each subject's individual beta weights were resampled to a $3 \times 3 \times 3$ mm voxel size and transformed to Talairach space (Talairach & Tournoux, 1988) using AFNI.

For the group analysis, the beta weights for each subject at each timepoint were entered into a correlation analysis in order to determine which brain areas (across subjects) had activations that correlated with Valence Bias (i.e. the measure of RT bias towards either face type, fearful or neutral). Correlations were calculated for each of the six stimulus types (3 expectation conditions $\times$ 2 face types) at each of the 8 timepoints in the trial separately. Data were cluster-corrected for multiple comparisons across all voxels and for the multiple tests performed (8 timepoints $\times$ 3 expectation conditions $\times$ 2 face types = 48 tests). This was achieved using a Monte Carlo simulation (via AFNI's AlphaSim program) with an individual voxel threshold p-value of 0.001 and corrected to a p-value of 0.01, resulting in a cluster threshold of 33 voxels (cluster volume of 891 mm^3).

In order to better understand the interaction between Valence Bias and time course of fMRI signal, regions-of-interest (ROIs) were created by using 5-mm radius spheres centered around the peaks from clusters in the correlation analysis. Timecourses for the two Valence Bias groups (i.e. Fear Fast and Fear Slow Responders) were extracted from the 20% fearful and 80% neutral trials. We performed post-hoc pair-wise t-tests for each timepoint in the

waveforms extracted from the ROIs. These t-tests were Bonferroni corrected for the multiple tests performed (i.e. 8 timepoints $\times$ 2 face types = 16 tests).

RESULTS

Overall Results

A 3×2 repeated measures ANOVA (3 expectation conditions × 2 face types) of RT data showed a main effect of expectation [$F(2,46) = 34.14$, $p = 8.12 \times 10^{-10}$; $p\eta^2 = 0.60$; Figure 1B]. A post-hoc Fisher LSD Test revealed a significant difference between expected faces (i.e. presented 80% of the time in a run) and unexpected faces (i.e. presented only 20% of the time in a run), where subjects were faster to categorize expected faces ($p = 1.38 \times 10^{-10}$). Additionally, the expectation effect was found to be parametric, as subjects were also significantly faster to categorize faces expected 80% of the time compared to 50% ($p = 1.75 \times 10^{-5}$) and 50% of the time compared to 20% ($p = 0.001$). However, there was no main effect of valence [$F(1,23) = 0.28$, $p = 0.60$; $p\eta^2 = 0.01$]. Consistent with behavioral results, a similar ANOVA of fMRI data showed the main effect of expectation was manifested in the fronto-parietal network later in the trial (2–10 s post-face), including the medial frontal gyrus and inferior parietal lobule, implicated in visual oddball and novelty detection (Table S2). No main effect of valence or the expectation by valence interaction was found.

An analysis of accuracy data showed a similar main effect of expectation, with subjects being more accurate to categorize the 80% expected face type than the 20% expected face type.

Since our behavioral results did not show an overall RT advantage for fearful faces as compared to neutral faces, we next evaluated whether the results were due solely to an expectation effect (independent of emotional valence). Accordingly, 13 of our original 24 subjects also completed a gender categorization task. For the gender task, we replaced all the neutral faces from the original task with new neutral male faces and all the fearful faces from the original task with new neutral female faces. Although subjects completed the gender task outside the scanner and the valence task inside the scanner, all other aspects were conserved between the two tasks. RT data from this task did not show a significant effect of expectation [$F(2,24) = 0.50$, $p = 0.06$; $p\eta^2 = 0.20$; Figure 1C]. Although a slight trend toward significance was seen, for this same subset of 13 subjects we saw a very large main effect of expectation during the valence categorization task [$F(2,24) = 28.44$, $p = 2.85 \times 10^{-7}$; $p\eta^2 = 0.73$], similar to what was seen in the whole group analysis (compare Figure 1D to Figure 1B). Thus, for this subset of 13 subjects, the parametric effect of expectation nearly disappeared when all faces were neutral and the task focused on a non-emotional aspect of the faces. Indeed, when data from both tasks were combined into one analysis, a main effect of task type (i.e. valence versus gender categorization) [$F(1,12) = 11.11$, $p = 0.006$] as well as an interaction between expectation and task were significant [$F(2,24) = 13.18$, $p=1.37 \times 10^{-4}$], such that the effect of expectation in all three conditions was significantly different between the two tasks (80% condition, $p = 1.10 \times 10^{-4}$; 50% condition, $p = 7.59 \times 10^{-9}$; 20% condition, $p = 1.59 \times 10^{-11}$). This result suggests that the presence of fearful faces strongly contributed to the highly significant main effect of expectation, and that this effect was not solely due to expectation of a face-type category

independent of valence *per se*. However, despite the strength of this effect, because of the small number of subjects, it is possible that there may be factors other than valence that contributed to the expectation effect.

Intersubject Variability: Behavioral Results

While at the group level fearful faces did not have an overall RT advantage over neutral faces among our subjects, upon closer inspection of the data, we discovered that subjects showed significant variability in their RTs, in particular to fearful faces. To quantify RT differences across the group, we created a Valence Bias measure for each subject which was calculated from the RTs to fearful faces in all three expectation conditions subtracted from the RTs to neutral faces in all three expectation conditions. This measure quantified RT differences for each subject as a single number to reflect their overall bias toward responding quickly or slowly to fearful relative to neutral faces. The average Valence Bias for our 24 subjects was 22.22 ± 204.11 (SD) and ranged from -369.34 to $+366.04$; this range indicated that overall some subjects were faster to respond to fearful faces but others were slower (Figure 2A).

Intersubject Variability: Behavior and Personality

As we expected correlation analyses between Valence Bias scores and personality measures showed that Valence Bias correlated positively with measures of Harm Avoidance for the 13 subjects who completed this questionnaire ($r = 0.79$, $p < 0.0014$; see Figure 2B and Table S3). Those subjects who responded faster to fearful faces were more harm avoidant. Valence Bias scores did not correlate with State Anxiety, Trait Anxiety, or with the five personality traits measured by the NEO-FFI. Although Valence Bias did not correlate with State Anxiety, Trait Anxiety, or Neuroticism, these three personality measures correlated positively with Harm Avoidance [State Anxiety ($r = 0.84$, $p < 0.0003$), Trait Anxiety ($r = 0.85$, $p < 0.0002$), and Neuroticism ($r = 0.85$, $p < 0.002$)]. Therefore, subjects scoring higher on the overall Harm Avoidance scale also scored higher on other anxiety-related scales.

Intersubject Variability: Behavior and fMRI

In order to explore the large amount of variability in Valence Bias in our group of healthy subjects, we correlated whole brain fMRI activity at each timepoint with Valence Bias scores. For this, we correlated the whole brain fMRI responses for each of the six stimulus types (i.e. fearful and neutral faces presented in the 80%, 50%, and 20% expectation conditions) for each subject for each timepoint with the Valence Bias scores. This analysis allowed us to see which brain areas, across subjects, showed activations associated with RT differences.

The results showed that Valence Bias scores correlated positively with fMRI signal intensity 8 s after the presentation of unexpected fearful faces (i.e. 12 s post-cue in 20% fear trials) in two clusters containing three regions of the brain. The first cluster (118 voxels) was located in the left medial prefrontal cortex. The strongest correlation in this cluster was located in a peak voxel in the first brain region: left ventromedial prefrontal cortex (VMPFC, Figure 3A–D, $r = 0.78$; see Table 1 for coordinates and statistics for each cluster). Additionally, within this cluster, a second peak was found in a second brain region: the left subgenual

cingulate cortex (SGC, Figure 2A, B and F, r = 0.75). The second cluster (44 voxels) that correlated positively with Valence Bias scores at this timepoint contained two peaks: one located in the right VMPFC (Figure 3C and E, r = 0.72) and the other located in a third brain region: the right head of the caudate nucleus (Cd, Figure 3B and G, r = 0.72). No other timepoints displayed a significant correlation between Valence Bias scores and percent signal change for unexpected fearful faces.

In contrast to the late positive correlation with unexpected fearful faces, Valence Bias scores correlated negatively with fMRI signal intensity evoked by expected neutral faces (i.e. 80% neutral faces) 4 s after the cue was presented (i.e. coincident with the presentation of the neutral face). This negative correlation was found in a large cluster (82 voxels) located medially, which included peak voxels in the dorsal and ventral striatum of the basal ganglia: left head of the Cd (Figure 4A,B, r = −0.70), right nucleus accumbens (NAcc, Figure 4A,C, r = −0.71), and right putamen (Put, Figure 4A, D, r = −0.72). No other timepoints displayed a significant correlation between Valence Bias scores and percent signal change for expected neutral faces. Additionally there were no significant correlations between Valence Bias scores and percent signal change for timepoints in expected fear, unexpected neutral, and 50% fear and neutral trials.

Comparing Valence Bias Groups: Behavioral Results

To further explore the relationship between Valence Bias and fMRI signal changes found for unexpected fearful faces and expected neutral faces, we created two separate groups: Subjects with a Valence Bias score greater than zero were categorized as "Fear Fast Responders", being faster overall to categorize fearful compared to neutral faces (N = 15), and subjects with a Valence Bias score less than zero were categorized as "Fear Slow Responders", being slower overall to categorize fearful compared to neutral faces (N = 9) (see Figure 2A for illustration of the Valence Bias variability).

Figure 5 shows the RT data for the three different expectation conditions for the two groups. While the Expectation × Valence × Group interaction was not significant [F(2,44) = 1.21, p = 0.31; $p\eta^2$ = 0.05], we did find a significant interaction of Valence × Group [F(1,22) = 51.10, p = 3.62×10^{-7}; $p\eta^2$ = 0.70]. Post-hoc analyses revealed that the RTs for the Fear Fast and Fear Slow Responders were significantly different for fearful faces (p = 0.04); Fear Fast Responders were faster to categorize fearful faces and Fear Slow Responders were slower to categorize fearful faces. However, RTs for neutral faces were not significantly different between the two groups (p = 0.72). Therefore, while the two groups were established based on the overall difference in RT between fearful and neutral faces [Fear Slow Responders were faster to respond to neutral than fearful faces overall (p = 4.5×10^{-5}) and Fear Fast Responders were faster to respond to fearful than neutral faces overall (p = 3.9×10^{-5}) (See Table S4)], only the response to fearful faces was significantly different between the two groups.

Comparing Valence Bias Groups: fMRI Results

We used peak voxels from significant clusters found in the correlation analysis (Table 1) to create ROIs, and then extracted expected fear and unexpected neutral waveforms for Fear

Fast and Fear Slow Responders within these ROIs. This allowed for better visualization of the complex relationship between the BOLD signal and the various experimental conditions for each group. The four ROIs showing a late trial positive correlation between Valence Bias scores and activation to unexpected fearful faces displayed a pattern of activation consistent with the correlation FMRI results (Figure 6). Timecourses within all four ROIs (right and left VMPFC, left SGC, and right Cd) showed a significant difference between the two groups for unexpected fearful faces late in the trial (i.e. fear face related) where Fear Fast Responders had greater activation compared to Fear Slow Responders. Indeed, whereas Fear Fast Responders showed a strong evoked response to the presentation of the unexpected fearful faces, for the Fear Slow Responders the evoked fMRI response fell below baseline. While these ROIs were chosen because their late trial activation to unexpected fearful faces significantly correlated with Valence Bias scores, all areas also displayed an early trial (i.e. cue related) difference for expected neutral faces, where Fear Slow Responders had greater activation than Fear Fast Responders. That is, the Fear Slow Responders showed a strong evoked fMRI response following the cue in expected neutral face trials, while the Fear Fast Responders exhibited an fMRI response below baseline. Thus, although the correlation results only showed effects in these brain regions late in the trial for unexpected fearful faces, by breaking the subjects into two Valence Bias groups we also found an effect early in the trial for expected neutral faces.

The three ROIs that showed an early trial negative correlation between Valence Bias scores and activation to expected neutral faces also displayed a pattern of activation consistent with the fMRI correlation results (Figure 7). Timecourses within all three ROIs (left Cd, right NAcc, and right Put) showed a significant difference between the two groups for expected neutral faces early in the trial (i.e. cue related) where Fear Slow Responders had greater activation compared to Fear Fast Responders. Whereas Fear Slow Responders showed a strong evoked response to the cue in the expected neutral face trials, for the Fear Fast Responders the evoked fMRI response fell below baseline.

In summary, most of the ROIs identified by the correlation analysis demonstrated both early and late trial effects when the data were analyzed by breaking the subjects into two Valence Bias groups; importantly, early trial effects were seen only for expected neutral faces and late trial effects were seen only for unexpected fearful faces.

DISCUSSION

The aim of the current study was to investigate the neural correlates of variability in processing threatening faces among healthy individuals. To probe intersubject variability, we manipulated the core component of anxiety, i.e. the expectation of a threatening stimulus. More specifically, we created a behavioral measure that reflected how rapidly subjects categorized faces as fearful or not. We then used this behavioral index as a measure of fear sensitivity and identified brain regions whose fMRI activity correlated with this index. Additionally, we correlated this behavioral index with personality measures related to anxiety.

Intersubject Variability in Fearful Face Categorization

In line with recent studies (Doty, Japee, Ingvar, & Ungerleider, 2013; Japee et al., 2009), we found that healthy subjects displayed considerable variability in detecting threatening faces. However, unlike studies that have reported behavioral enhancement for threatening faces in the healthy population (e.g. (Fox, 2002; Fox et al., 2000; Hansen & Hansen, 1988; Ishai et al., 2004; Wilson & MacLeod, 2003), over one-third (9/24) of our subjects were actually faster to categorize neutral faces than fearful faces. When we compared the RTs between subjects who were faster to categorize neutral faces with those who were faster to categorize fearful faces, we found that the RT differences were driven exclusively by responses to fearful faces. Therefore, over one-third of our subjects were slower to categorize fearful faces. When RT data for Fear Fast and Fear Slow Responders were combined, there was no overall RT advantage for fearful faces. While we did not find an overall RT enhancement for fearful faces, we did find a main effect of expectation. Subjects were fastest to categorize expected (80%) faces, slower to categorize 50% faces, and slowest to categorize unexpected (20%) faces.

Personality Measures and RT Bias to Fearful Faces

High anxiety levels have been associated with an attentional bias towards threat in clinical and subclinical populations (see review (Bar-Haim et al., 2007)). We found that Valence Bias scores did not significantly correlate with state or trait anxiety, as measured by the Spielberger State Trait Anxiety Inventory. However, Valence Bias scores did significantly correlate with Tridimensional Personality Questionnaire Harm Avoidance (TPQ-HA) scores, which were collected from a subset of our cohort (13/24 subjects), such that a more harm avoidant, cautious person was more likely to be hypersensitive to unexpected threatening stimuli. TPQ Harm Avoidance scores also correlated positively with state and trait anxiety, which is consistent with findings from other groups (Caseras, Àvila, & Torrubia, 2003; Stewart, Ebmeier, & Deary, 2005).

While we did not find a significant correlation directly between anxiety and threatening face detection, other studies have reported correlations between anxiety measures and enhanced behavioral responses to threatening faces. However, those studies only used subjects with the highest and lowest scores on anxiety questionnaires drawn from a larger population (Arrais et al., 2010; Richards et al., 2002; Surcinelli et al., 2006). Our subjects were, by contrast, not preselected based on anxiety scores and represented a range of healthy anxiety scores not typically included in previous research. Even without pre-selection of subjects, we found a positive correlation between Valence Bias scores and an anxiety-related trait, namely, Harm Avoidance.

Medial Prefrontal and Striatal Activity Modulated by RT Bias to Fearful Faces

We found a large amount of variability in how quickly subjects responded to fearful relative to neutral faces. To understand the neural origins of this variability, we correlated Valence Bias scores with fMRI activations across the brain. We found that, in trials where fearful faces were unexpected, late trial activity in the ventromedial prefrontal cortex, subgenual cingulate cortex and caudate correlated positively with Valence Bias scores. Subjects who responded faster overall to fearful than to neutral faces showed higher activity in these

regions, while subjects who responded slower overall to fearful than to neutral faces showed less activity in these regions. Activity in these regions was, thus, predictive of a subject's general bias towards fearful faces.

In contrast to the late, fear face related positive correlation between activation to unexpected fearful faces and Valence Bias scores, we found an early, cue-related negative correlation in expected neutral face trials. The higher the RT bias towards fearful faces in general, the smaller was the activation in the dorsal and ventral striatum of the basal ganglia, including the caudate and putamen, and the nucleus accumbens.

We placed the subjects into two groups (Fear Fast and Fear Slow Responders) to further probe for differences between the groups to unexpected fearful and expected neutral faces over time. The timecourses within the areas where activation significantly correlated with Valence Bias scores reflected the results from the correlation data. However, the group analysis also revealed that both early and late effects were common to most regions. That is, subjects who were faster to respond to fearful faces had an enhanced response late in the trial in these regions following an unexpected fearful face but also had an attenuated preparatory response early in the trial in the same regions preceding an expected neutral face; the opposite was true for the subjects who were slower to respond to fearful faces. These results point to a network of regions in medial prefrontal cortex and dorsal and ventral striatum (including ventromedial prefrontal cortex, subgenual cingulate cortex, caudate, putamen, and nucleus accumbens) that play a role in how individuals process threat.

The co-activation of these regions is not surprising since neuroanatomical studies in nonhuman primates have shown that the ventromedial prefrontal cortex projects to both the dorsal and ventral striatum (Haber, Kunishio, Mizobuchi, & Lynd-Balta, 1995) as well as the subgenual cingulate cortex (Carmichael & Price, 1996). The subgenual cingulate cortex also projects to the ventral striatum (Kunishio & Haber, 1994). The dense connections between these regions have been confirmed in the human brain using diffusion tensor imaging and functional connectivity mapping (Di Martino et al., 2008; Johansen-Berg et al., 2008; Leh, Ptito, Chakravarty, & Strafella, 2007; Lehéricy et al., 2004). This evidence thus provides the anatomical underpinnings for an interconnected network.

What drives the correlation of activity in this network with behavior? We propose that, in our task, the medial prefrontal – striatal network plays a role in encoding the affective value of unexpected fearful faces. Affective value is defined here as the biological relevance of an emotionally charged stimulus to guide behavior. Thus, subjects for whom fearful faces were more valued in general (i.e. Fear Fast Responders) engaged these regions to a greater extent when fearful faces were unexpected, which in turn resulted in faster motor responses to those unexpected valued stimuli. On the other hand, subjects for whom fearful faces were less valued (i.e. Fear Slow Responders) deactivated these regions, which resulted in slower motor responses to the unexpected non-valued stimuli. Thus, activity in this network appears to be directly proportional to the affective value fearful faces hold for a given subject when they are presented in an unexpected condition.

We know from previous work that these regions, particularly the ventromedial prefrontal cortex, are crucial in evaluating the value of stimuli based on context (Blair et al., 2006; Schoenbaum & Roesch, 2005; Schultz & Dickinson, 2000; Tobler, O'Doherty, Dolan, & Schultz, 2006). Human imaging studies have shown that the ventromedial prefrontal cortex is involved in the extinction of conditioned fear (Phelps, Delgado, Nearing, & LeDoux, 2004) and skin conductance changes during a risky decision-making paradigm (Critchley, Elliott, Mathias, & Dolan, 2000)– where the ventromedial prefrontal cortex tracks how the affective representation of a stimulus changes based on context. Still others have shown that patients with lesions of the ventromedial prefrontal cortex have deficiencies in using new information to make advantageous decisions (Bechara, Damasio, Damasio, & Anderson, 1994; Bechara, Tranel, & Damasio, 2000), have difficulty integrating emotion into their decision-making process during moral decisions (Koenigs et al., 2007), and have trouble judging harmful intentions of others (Young et al., 2010). Importantly, a recent neuroimaging study found that the VMPFC encoded changes in subjective emotional ratings when subjects were told to just experience emotional stimuli but not during times when the subjects were told to regulate their emotions (Winecoff et al., 2013). Neuroimaging and postmortem evidence supports the role of the subgenual cingulate cortex in the pathology of social phobia and depression and also its role in negative mood (Furmark et al., 2002; Mayberg et al., 1999). Human imaging and monkey physiological studies have both demonstrated that the dorsal and ventral striatum respond to cues that signal an upcoming salient stimulus (Hikosaka, Sakamoto, & Usui, 1989; Knutson, Adams, Fong, & Hommer, 2001; Samejima, Ueda, Doya, & Kimura, 2005). These wide-ranging studies of the medial prefrontal - striatal network illustrate its involvement in tracking the affective value of stimuli as context changes.

In spite of the many studies pointing to the VMPFC's role in tracking emotional value an alternative interpretation of our results could be that the variability seen in our study is due to individual differences in emotion regulation. Support for this interpretation comes from two studies that have linked differences in connectivity between the VMPFC and the amygdala (a circuit heavily implicated in emotion regulation) to trait anxiety in healthy adults (Kim, Gee, Loucks, Davis, & Whalen, 2011; Kim & Whalen, 2009). Although viable, we do not favor this explanation for our study because better emotional regulation, as indicated by increased VMPFC activation, would not predict faster reaction times only for unexpected fearful faces.

The early trial negative correlation between Valence Bias scores and activation evoked by expected neutral faces trials was not predictive of faster RTs to neutral faces, because the behavioral differences among our subjects were driven by fearful faces exclusively. As this negative correlation occurred early in the trial, we believe that this effect is related to the anticipation of the upcoming stimulus, i.e., the expected neutral face. It is important to note that in order to limit repetition effects we only included in our analysis trials in which a neutral or fearful face was immediately followed by the alternate face type. Thus, on expected neutral trials, subjects were highly certain that the upcoming stimulus would be a neutral face (since they had just received a fearful face on the previous trial). We know from other fMRI studies that expected stimuli can elicit an anticipatory response (Chawla, Rees, & Friston, 1999; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999). In our study,

subjects who valued fearful faces less (i.e. Fear Slow Responders), showed a normal anticipatory effect, while those subjects who valued fearful faces more (i.e. Fear Fast Responders) showed greater suppression of this anticipatory response.

It is important to note that activity in the medial prefrontal cortex and striatum did not encode the affective value of fearful faces in general in our task, but rather represented the combination of the affective value of fearful faces and context, i.e., the likelihood of appearing (expected or unexpected). This is confirmed by the lack of correlation of activity in these regions with Valence Bias scores on trials in which fearful faces were expected and neutral faces were unexpected. The interaction between affect and expectation is further supported by research demonstrating that low probability threat produces a stronger autonomic nervous system response (as measured via heart rate and skin conductance responses) than high probability threat, not only to the anticipation of the upcoming stimulus but also to the stimulus itself (Deane, 1969; Epstein & Roupenian, 1970).

As research has shown that threatening stimuli generally have a behavioral and neural advantage, it is curious that 38% (9/24) of our healthy adults (i.e. the Fear Slow Responders) showed a behavioral disadvantage for fearful relative to neutral faces. These same subjects also showed deactivation for unexpected fearful faces and enhanced activation in anticipation of being presented with an expected neutral face. Additionally these are the subjects who scored lowest on the harm avoidance questionnaire. In a follow up analysis, we found that Fear Slow Responders had significantly lower harm avoidance scores than the population norm, while Fear Fast Responders had harm avoidance score that did not differ from the population norm. This evidence suggests that Fear Slow Responders have a blunted threat response and may not readily register threat in the environment; as a consequence, these individuals may be less likely to avoid harm in their daily lives.

One brain area that was notably absent in our correlation results was the amygdala, which is heavily interconnected with the regions showing significant correlations with Valence Bias scores (Carmichael & Price, 1995; Porrino, Crane, & Goldman-Rakic, 1981; Russchen, Bakst, Amaral, & Price, 1985; Vogt & Pandya, 1987). Given that the amygdala has been shown to activate in response to fearful faces over other face types (Breiter et al., 1996; Morris et al., 1996), this was a surprising finding. There are at minimum three possibilities underlying this finding: 1) it could suggest that the amygdala may respond relatively automatically to threat rather than tracking expectation (Morris et al., 1998; Whalen et al., 1998), 2) the signal quality of the axially acquired scan could have been too poor for meaningful data analyses near the sinus cavity, or 3) the parameters of the task (i.e., an active rather than passive task) were not conducive to modulating signals in the amygdala (Blasi et al., 2009; Costafreda, Brammer, David, & Fu, 2008).

Conclusion

In summary, we have demonstrated that healthy adults vary in their sensitivity to fearful faces, with some adults actually showing a behavioral disadvantage for fearful faces. This variability correlated positively with activation in the ventromedial prefrontal and subgenual cingulate cortices and the dorsal striatum to unexpected fearful faces (late in the trial) and correlated negatively with activation in the dorsal and ventral striatum to the standard start-

of-trial, uninformative cue early in expected neutral face trials. Additionally, subjects who had an RT bias towards fearful faces were also the subjects who were most harm avoidant. These results underscore the importance of viewing healthy adults along a spectrum in emotion-related tasks. They also imply that a network of brain regions, involving the medial prefrontal cortex along with the dorsal and ventral striatum, plays a significant role in the variability of threat processing, potentially through encoding affective value as it changes according to expectation. The variability among our healthy subjects in behavior, fMRI activation patterns, and personality measures suggests that future studies examining threat processing should consider intersubject variability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Citations

Arrais KC, Machado-de-Sousa JP, Trzesniak C, Santos Filho A, Ferrari MCF, Osório FL, Crippa JAS. Social anxiety disorder women easily recognize fearfull, sad and happy faces: the influence of gender. Journal of Psychiatric Research. 2010; 44(8):535–540. [PubMed: 19962717]

Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, van IJzendoorn MH. Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. Psychological Bulletin. 2007; 133(1):1–24. [PubMed: 17201568]

Barlow, DH. Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic. Guilford Press; 2004.

Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition. 1994; 50(1–3):7–15. [PubMed: 8039375]

Bechara, Antoine; Tranel, D.; Damasio, H. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain. 2000; 123(11):2189–2202. [PubMed: 11050020]

Blair K, Marsh AA, Morton J, Vythilingam M, Jones M, Mondillo K, Blair JR. Choosing the Lesser of Two Evils, the Better of Two Goods: Specifying the Roles of Ventromedial Prefrontal Cortex and Dorsal Anterior Cingulate in Object Choice. The Journal of Neuroscience. 2006; 26(44):11379–11386. [PubMed: 17079666]

Blasi G, Hariri AR, Alce G, Taurisano P, Sambataro F, Das S, Mattay VS. Preferential Amygdala Reactivity to the Negative Assessment of Neutral Faces. Biological Psychiatry. 2009; 66(9):847–853. [PubMed: 19709644]

Bradley BP, Mogg K, White J, Groom C, de Bono J. Attentional bias for emotional faces in generalized anxiety disorder. The British Journal of Clinical Psychology / the British Psychological Society. 1999; 38(Pt 3):267–278. [PubMed: 10532148]

Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, Rosen BR. Response and habituation of the human amygdala during visual processing of facial expression. Neuron. 1996; 17(5):875–887. [PubMed: 8938120]

Carmichael ST, Price JL. Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. The Journal of Comparative Neurology. 1995; 363(4):615–641. [PubMed: 8847421]

Carmichael ST, Price JL. Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. The Journal of Comparative Neurology. 1996; 371(2):179–207. [PubMed: 8835726]

Caseras X, Àvila C, Torrubia R. The measurement of individual differences in Behavioural Inhibition and Behavioural Activation Systems: a comparison of personality scales. Personality and Individual Differences. 2003; 34(6):999–1013.

Cloninger CR, Przybeck TR, Svrakic DM. The Tridimensional Personality Questionnaire: U.S. normative data. Psychological Reports. 1991; 69(3 Pt 1):1047–1057. [PubMed: 1784653]

Costa, PT.; McCrae, RR. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual. Odessa, Florida: Psychological Assessment Resources; 1992.

Costafreda SG, Brammer MJ, David AS, Fu CHY. Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. Brain Research Reviews. 2008; 58(1):57–70. [PubMed: 18076995]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, an International Journal. 1996; 29(3):162–173.

Critchley HD, Elliott R, Mathias CJ, Dolan RJ. Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2000; 20(8):3033–3040. [PubMed: 10751455]

Dale AM. Optimal experimental design for event-related fMRI. Human Brain Mapping. 1999; 8(2–3): 109–114. [PubMed: 10524601]

Davidson RJ, Maxwell JS, Shackman AJ. The privileged status of emotion in the brain. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(33):11915–11916. [PubMed: 15304648]

Deane GE. Cardiac activity during experimentally induced anxiety. Psychophysiology. 1969; 6(1):17–30. [PubMed: 5820902]

Di Martino A, Scheres A, Margulies DS, Kelly AMC, Uddin LQ, Shehzad Z, Milham MP. Functional connectivity of human striatum: a resting state FMRI study. Cerebral Cortex (New York, N.Y.: 1991). 2008; 18(12):2735–2747.

Doty TJ, Japee S, Ingvar M, Ungerleider LG. Fearful face detection sensitivity in healthy adults correlates with anxiety-related traits. Emotion (Washington, D.C.). 2013; 13(2):183–188.

Epstein, S. Anxiety: Current Trends in Theory and Research. Vol. Vol. 2. Academic Press; 1972. The nature of anxiety with emphasis upon its relationship to expectancy.

Epstein S, Roupenian A. Heart rate and skin conductance during experimentally induced anxiety: the effect of uncertainty about receiving a noxious stimulus. Journal of Personality and Social Psychology. 1970; 16(1):20–28. [PubMed: 5485942]

Fox E. Processing emotional facial expressions: the role of anxiety and awareness. Cognitive, Affective & Behavioral Neuroscience. 2002; 2(1):52–63.

Fox E, Lester V, Russo R, Bowles RJ, Pichler A, Dutton K. Facial Expressions of Emotion: Are Angry Faces Detected More Efficiently? Cognition & Emotion. 2000; 14(1):61–92. [PubMed: 17401453]

Furmark T, Tillfors M, Marteinsdottir I, Fischer H, Pissiota A, Långström B, Fredrikson M. Common changes in cerebral blood flow in patients with social phobia treated with citalopram or cognitive-behavioral therapy. Archives of General Psychiatry. 2002; 59(5):425–433. [PubMed: 11982446]

Haber SN, Kunishio K, Mizobuchi M, Lynd-Balta E. The orbital and medial prefrontal circuit through the primate basal ganglia. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 1995; 15(7 Pt 1):4851–4867. [PubMed: 7623116]

Hansen CH, Hansen RD. Finding the face in the crowd: an anger superiority effect. Journal of Personality and Social Psychology. 1988; 54(6):917–924. [PubMed: 3397866]

Hikosaka O, Sakamoto M, Usui S. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. Journal of Neurophysiology. 1989; 61(4):814–832. [PubMed: 2723722]

Ishai A, Pessoa L, Bikle PC, Ungerleider LG. Repetition suppression of faces is modulated by emotion. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(26):9827–9832. [PubMed: 15210952]

Japee S, Crocker L, Carver F, Pessoa L, Ungerleider LG. Individual differences in valence modulation of face-selective M170 response. Emotion (Washington, D.C.). 2009; 9(1):59–69.

Johansen-Berg H, Gutman DA, Behrens TEJ, Matthews PM, Rushworth MFS, Katz E, Mayberg HS. Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for treatment-resistant depression. Cerebral Cortex (New York, N.Y.: 1991). 2008; 18(6):1374–1383.

Kim MJ, Gee DG, Loucks RA, Davis FC, Whalen PJ. Anxiety Dissociates Dorsal and Ventral Medial Prefrontal Cortex Functional Connectivity with the Amygdala at Rest. Cerebral Cortex. 2011; 21(7):1667–1673. [PubMed: 21127016]

Kim MJ, Whalen PJ. The Structural Integrity of an Amygdala-Prefrontal Pathway Predicts Trait Anxiety. The Journal of Neuroscience. 2009; 29(37):11614–11618. [PubMed: 19759308]

Knutson B, Adams CM, Fong GW, Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2001; 21(16):RC159. [PubMed: 11459880]

Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature. 2007; 446(7138):908–911. [PubMed: 17377536]

Kunishio K, Haber SN. Primate cingulostriatal projection: limbic striatal versus sensorimotor striatal input. The Journal of Comparative Neurology. 1994; 350(3):337–356. [PubMed: 7533796]

Leh SE, Ptito A, Chakravarty MM, Strafella AP. Fronto-striatal connections in the human brain: a probabilistic diffusion tractography study. Neuroscience Letters. 2007; 419(2):113–118. [PubMed: 17485168]

Lehéricy S, Ducros M, Van de Moortele P-F, Francois C, Thivard L, Poupon C, Kim D-S. Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. Annals of Neurology. 2004; 55(4):522–529. [PubMed: 15048891]

Lundqvist, D.; Flykt, A.; Öhman, A. The Karolinska Directed Emotional Faces. Psychology section, Department of Clinical Neuroscience, Karolinska Institutet; 1998.

Mayberg HS, Liotti M, Brannan SK, McGinnis S, Mahurin RK, Jerabek PA, Fox PT. Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. The American Journal of Psychiatry. 1999; 156(5):675–682. [PubMed: 10327898]

Mogg K, Bradley BP. Some methodological issues in assessing attentional biases for threatening faces in anxiety: a replication study using a modified version of the probe detection task. Behaviour Research and Therapy. 1999; 37(6):595–604. [PubMed: 10372472]

Morris JS, Friston KJ, Büchel C, Frith CD, Young AW, Calder AJ, Dolan RJ. A neuromodulatory role for the human amygdala in processing emotional facial expressions. Brain: a Journal of Neurology. 1998; 121(Pt 1):47–57. [PubMed: 9549487]

Morris JS, Frith CD, Perrett DI, Rowland D, Young AW, Calder AJ, Dolan RJ. A differential neural response in the human amygdala to fearful and happy facial expressions. Nature. 1996; 383(6603): 812–815. [PubMed: 8893004]

Ohman A, Mineka S. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. Psychological Review. 2001; 108(3):483–522. [PubMed: 11488376]

Pessoa L, Japee S, Sturman D, Ungerleider LG. Target visibility and visual awareness modulate amygdala responses to fearful faces. Cerebral Cortex (New York, N.Y.: 1991). 2006; 16(3):366–375.

Pessoa L, Japee S, Ungerleider LG. Visual awareness and the detection of fearful faces. Emotion (Washington, D.C.). 2005; 5(2):243–247.

Pessoa L, McKenna M, Gutierrez E, Ungerleider LG. Neural processing of emotional faces requires attention. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99(17):11458–11463. [PubMed: 12177449]

Phelps EA, Delgado MR, Nearing KI, LeDoux JE. Extinction learning in humans: role of the amygdala and vmPFC. Neuron. 2004; 43(6):897–905. [PubMed: 15363399]

Porrino LJ, Crane AM, Goldman-Rakic PS. Direct and indirect pathways from the amygdala to the frontal lobe in rhesus monkeys. The Journal of Comparative Neurology. 1981; 198(1):121–136. [PubMed: 6164704]

Richards A, French CC, Calder AJ, Webb B, Fox R, Young AW. Anxiety-related bias in the classification of emotionally ambiguous facial expressions. Emotion (Washington, D.C.). 2002; 2(3):273–287.

Russchen FT, Bakst I, Amaral DG, Price JL. The amygdalostriatal projections in the monkey. An anterograde tracing study. Brain Research. 1985; 329(1–2):241–257. [PubMed: 3978445]

Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. Science (New York, N.Y.). 2005; 310(5752):1337–1340.

Schoenbaum G, Roesch M. Orbitofrontal cortex, associative learning, and expectancies. Neuron. 2005; 47(5):633–636. [PubMed: 16129393]

Schultz W, Dickinson A. Neuronal coding of prediction errors. Annual Review of Neuroscience. 2000; 23:473–500.

Spielberger, CD. Manual for the State-trait anxiety inventory (form Y). Palo Alto, CA: Consulting Psychologists Press; 1983.

Stewart ME, Ebmeier KP, Deary IJ. Personality correlates of happiness and sadness: EPQ-R and TPQ compared. Personality and Individual Differences. 2005; 38(5):1085–1096.

Surcinelli P, Codispoti M, Montebarocci O, Rossi N, Baldaro B. Facial emotion recognition in trait anxiety. Journal of Anxiety Disorders. 2006; 20(1):110–117. [PubMed: 16325118]

Talairach, J.; Tournoux, P. Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging. New York, NY, US: Thieme Medical Publishers; 1988.

Tobler PN, O'Doherty JP, Dolan RJ, Schultz W. Human Neural Learning Depends on Reward Prediction Errors in the Blocking Paradigm. Journal of Neurophysiology. 2006; 95(1):301–310. [PubMed: 16192329]

Vogt BA, Pandya DN. Cingulate cortex of the rhesus monkey: II. Cortical afferents. The Journal of Comparative Neurology. 1987; 262(2):271–289. [PubMed: 3624555]

Whalen PJ, Rauch SL, Etcoff NL, McInerney SC, Lee MB, Jenike MA. Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 1998; 18(1):411–418. [PubMed: 9412517]

Wilson E, MacLeod C. Contrasting two accounts of anxiety-linked attentional bias: selective attention to varying levels of stimulus threat intensity. Journal of Abnormal Psychology. 2003; 112(2):212–218. [PubMed: 12784830]

Winecoff A, Clithero JA, Carter RM, Bergman SR, Wang L, Huettel SA. Ventromedial Prefrontal Cortex Encodes Emotional Value. The Journal of Neuroscience. 2013; 33(27):11032–11039. [PubMed: 23825408]

Winton EC, Clark DM, Edelmann RJ. Social anxiety, fear of negative evaluation and the detection of negative emotion in others. Behaviour Research and Therapy. 1995; 33(2):193–196. [PubMed: 7887878]

Young L, Bechara A, Tranel D, Damasio H, Hauser M, Damasio A. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. Neuron. 2010; 65(6):845–851. [PubMed: 20346759]
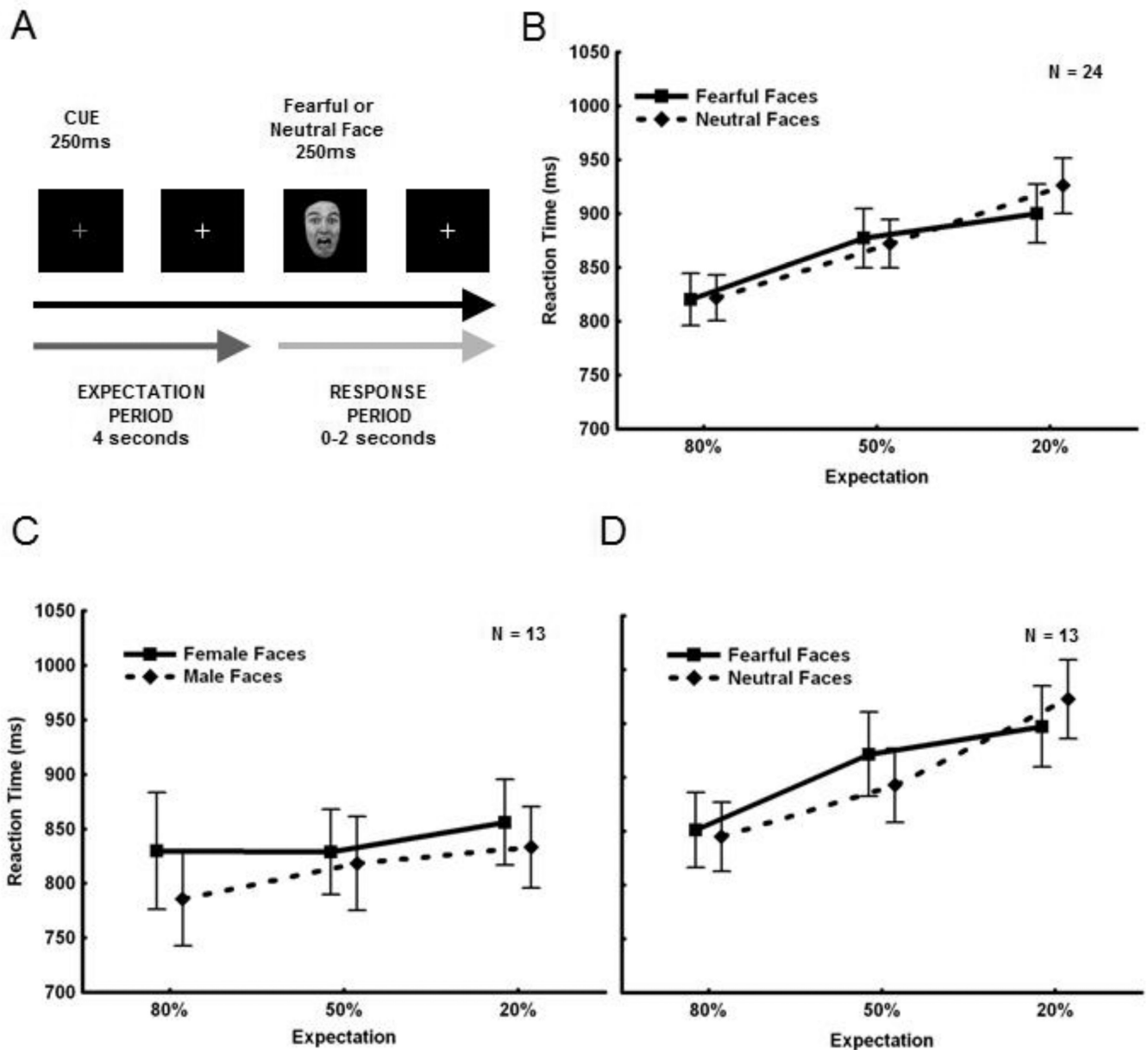
Figure 1. Experimental Design and Behavioral Data

A. Illustration of a single trial. Subjects completed 50 trials in each run of the experiment and received three different run types : 80F:20N, 50F:50N, 20F:80N. The task was a mixed design where run types were blocked by expectation and the face type was event-related. On each trial, the subjects received a cue (i.e. red fixation cross) signaling that a face would appear 4 s later. When the fearful or neutral face appeared, subjects were instructed to respond as fast and as accurately as possible, indicating which face type they had seen. B. RT as a function of expectation. The solid line depicts RT to fearful faces, and the dotted line to neutral faces. Overall, subjects were faster to categorize the face type that was expected (main effect of expectation). Subjects were faster to categorize expected faces (presented 80% of the time) (far left on axis), slower to categorize unexpected faces

(presented 20% of the time) (far right on axis), and RTs to faces presented 50% of the time fell inbetween (middle of axis). Overall, subjects had no RT advantage for fearful faces compared to neutral faces. C. Data from subjects on a gender categorization task. A subset of 13/24 subjects also performed a gender categorization task, in which all the neutral faces were replaced with neutral male faces and all the fearful faces were replaced with neutral female faces. Subjects were asked to categorize each face as either male or female. The results showed only a trend towards a main effect of expectation. These findings indicate that the expectation effect seen in the main experiment (B) was due to the presence of fearful faces and could not be attributed solely to expectation. D. The data from the original valence categorization experiment for the same 13 subjects. Again there was a significant main effect of expectation.
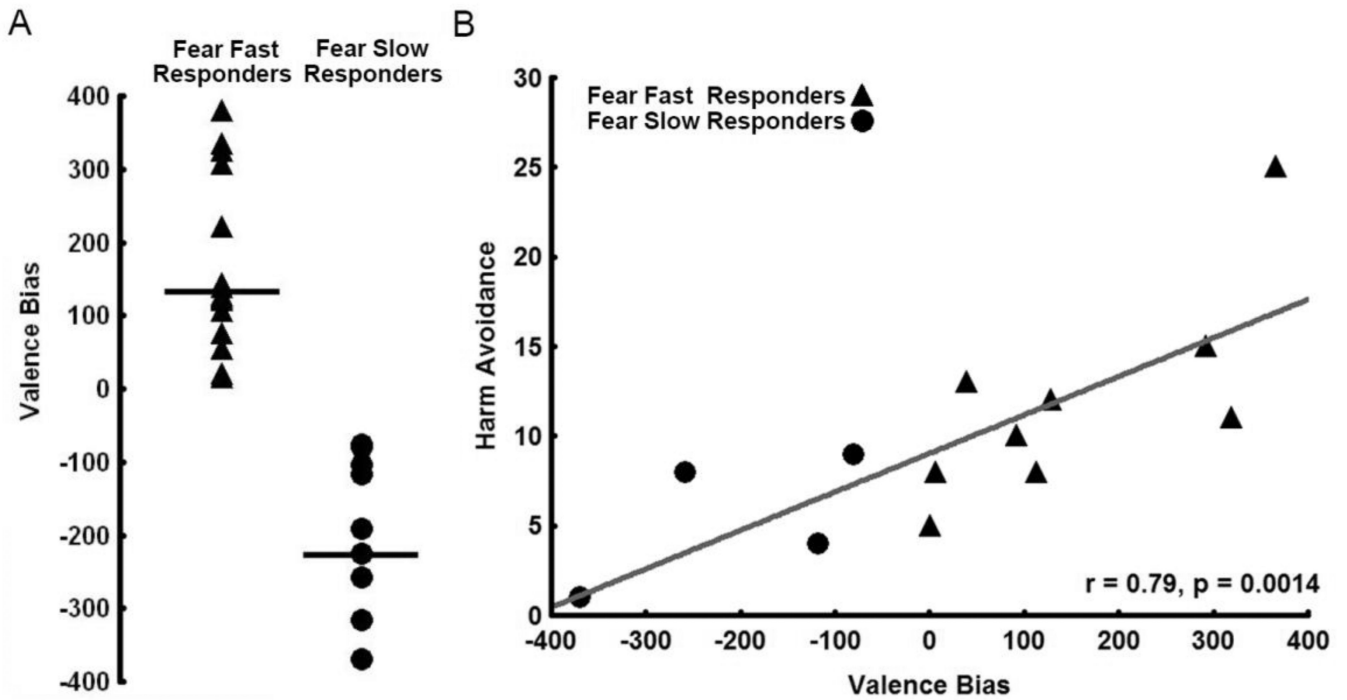
Figure 2. Classification of Valence Bias Groups and Correlation with Harm Avoidance
A. Illustration of the variability in Valence Biases. Subjects were assigned to one of two groups: Those subjects with a Valence Bias score greater than 0 (i.e. those who were faster to categorize fearful faces overall) were placed into the Fear Fast Responders group, while those with a Valence Bias score less than 0 (i.e. those who were slower to categorize fearful faces overall) were placed into the Fear Slow Responders group. Valence Biases for subjects responding faster to fearful faces across all three expectation conditions are represented by triangles, while Valence Biases for subjects responding slower to fearful faces across all three expectation conditions are represented by circles. B. Valence Bias positively correlates with Harm Avoidance in a subset (N=13) of subjects. Subjects who are more harm avoidant have a general bias towards responding faster to fearful faces compared to neutral faces.
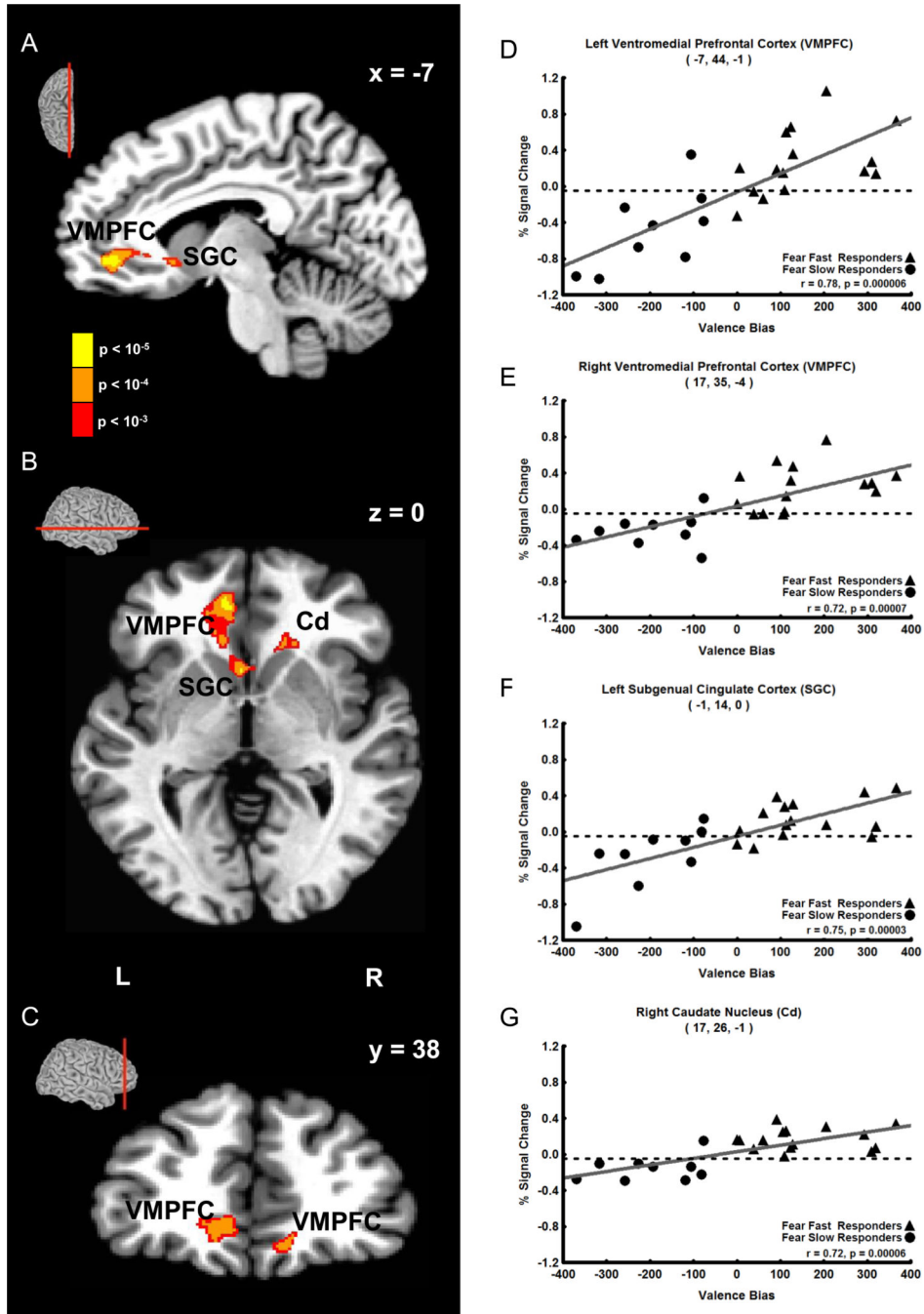
Figure 3. Positive Correlation of Late Trial (8 s after face) Unexpected (20%) Fearful Face Activation with Valence Bias Scores

A–C. Sagittal (A), axial (B), and coronal (C) views showing that late trial activation to unexpected fearful faces (i.e. fearful faces presented only 20% of the time) positively correlated with Valence Bias scores in the following areas (p < 0.01 cluster corrected): left and right ventromedial prefrontal cortex (VMPFC), left subgenual cingulate cortex (SGC), and the right caudate nucleus (Cd). The sagittal, axial, and coronal sections are shown in small brains above the slice views. D–G. Late trial percent signal change evoked by unexpected fearful faces plotted against Valence Bias scores for each subject in the peak

voxel of the left VMPFC (D), right VMPFC (E), left SGC (F), and right Cd (G). Subjects who had a general bias towards fearful faces (i.e. had a higher Valence Bias score) also had increased activation to unexpected fearful faces in these regions.
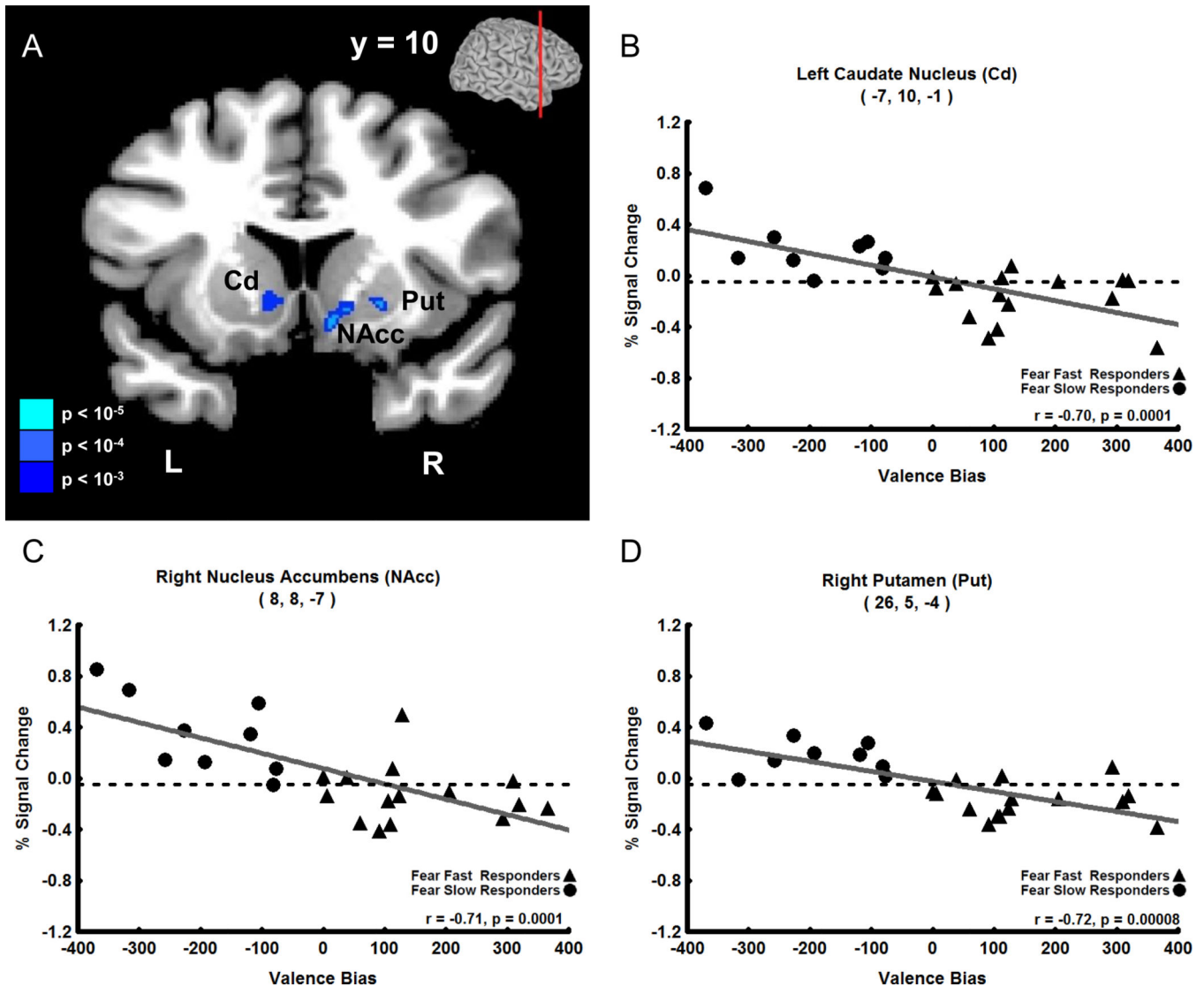
Figure 4. Negative Correlation of Early Trial (4 s after cue) Expected (80%) Neutral Face Activation with Valence Bias Scores

A. Coronal view showing that early activation to expected neutral faces (i.e. neutral faces presented 80% of the time) negatively correlated with the Valence Bias in the following areas (p < 0.01 cluster corrected): left caudate nucleus (Cd), right nucleus accumbens (NAcc), and right putamen (Put); the level of the coronal section is shown in small brain at top. B–D. Early trial percent signal change evoked by expected neutral faces plotted against Valence Bias scores for each subject in the peak voxel of the Cd (B), NAcc (C), and Put (D). Subjects who had a general bias away from fearful faces (i.e. had a lower Valence Bias score) had increased activation to expected neutral faces in these regions.
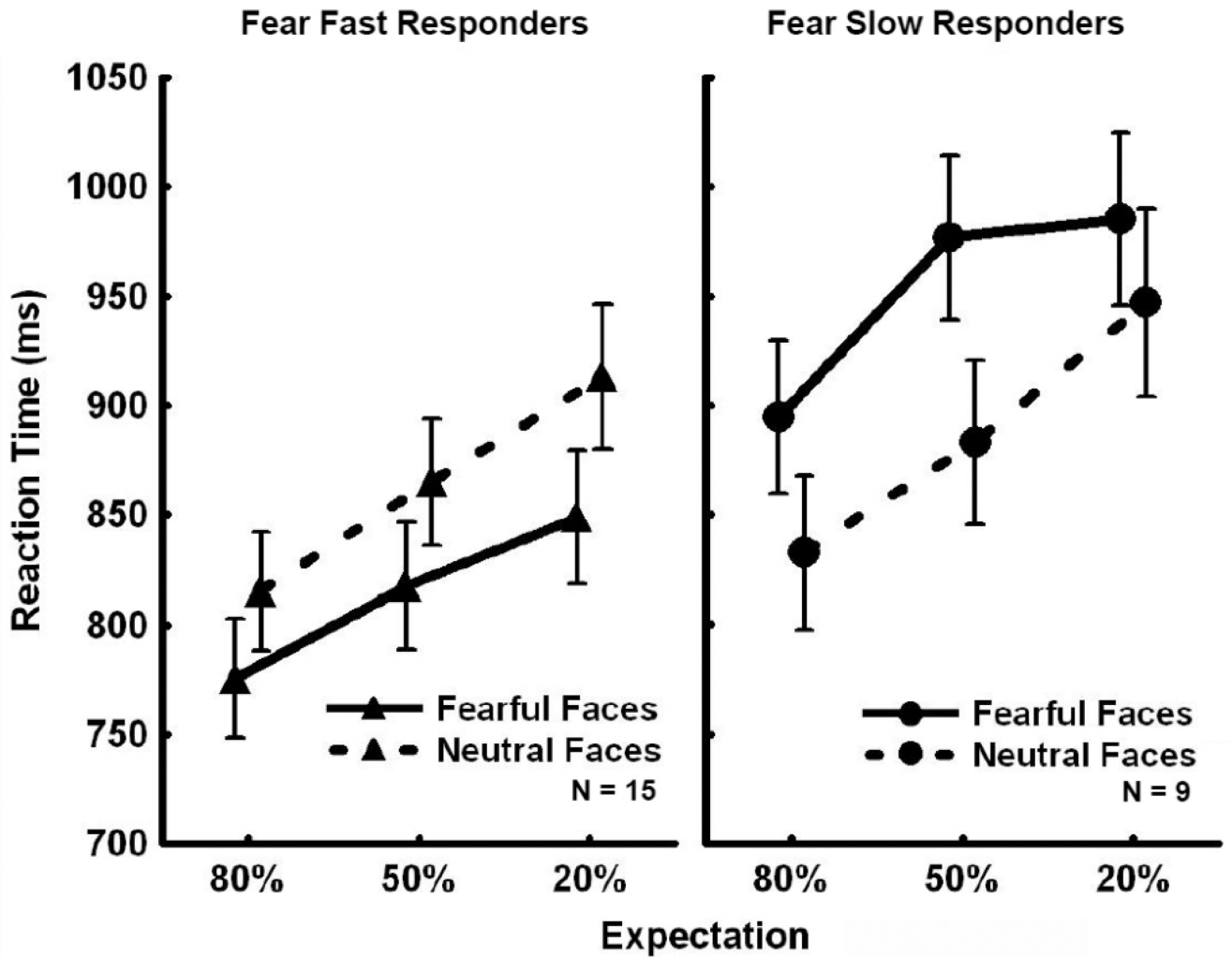
Figure 5. Behavior by Valence Bias Group

Reactions times plotted for Fear Fast and Fear Slow Responders as a function of expectation condition. There was a significant interaction between valence and group: Fear Fast Responders were faster to categorize fearful faces than neutral faces, and Fear Slow Responders were faster to categorize neutral faces than fearful faces across all expectation conditions. Additionally, Fear Fast Responders were faster than Fear Slow Responders to categorize fearful faces, while there was no significant difference in reaction time to neutral faces between the two groups.
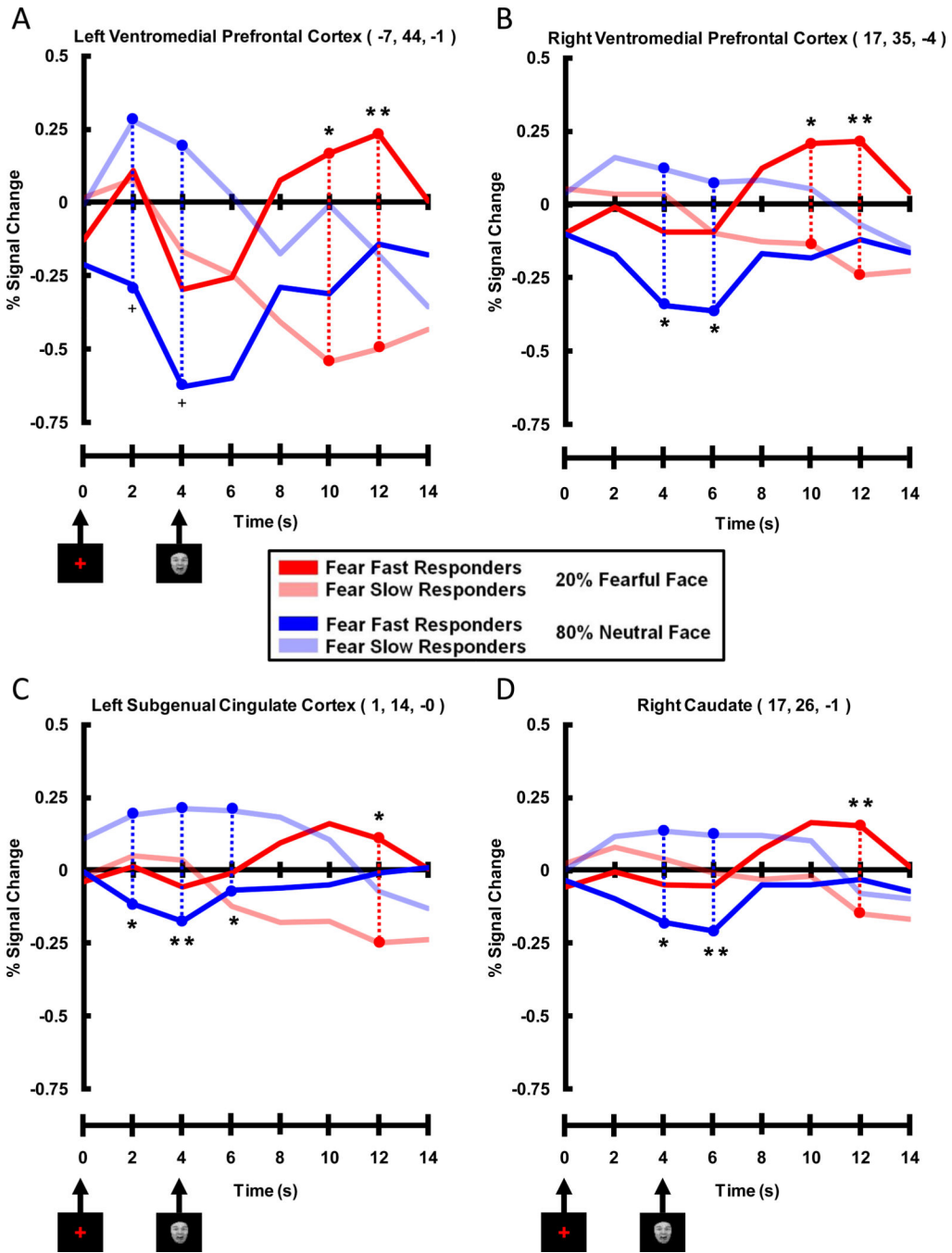
Figure 6. Timecourses Extracted from Peak Voxels Displaying Late Positive Correlation Between Percent Signal Change for Unexpected Fearful Faces and Valence Bias Scores
Prefrontal and subcortical areas included A) left and B) right ventromedial prefrontal cortex, C) left subgenual cingulate cortex, and D) right caudate. These timecourses illustrate that Fear Fast Responders (dark red lines) had more activation than Fear Slow Responders (light red lines) in these ventromedial prefrontal and subcortical areas late (i.e. 10 and 12 s post-cue, 6 and 8 s post-face) in unexpected fearful face trials. It is interesting to note that although these peak voxels were significant for the positive correlation between Valence Bias scores and percent signal change for unexpected fearful faces, these peak voxels also

displayed an early (i.e. 2, 4, and 6 s post-cue) trial enhancement for Fear Slow Responders (light blue lines) compared to Fear Fast Responders (dark blue lines) for expected neutral faces. (** = p < 0.01 corrected, * = p < 0.05 corrected, + = p < 0.10 corrected)
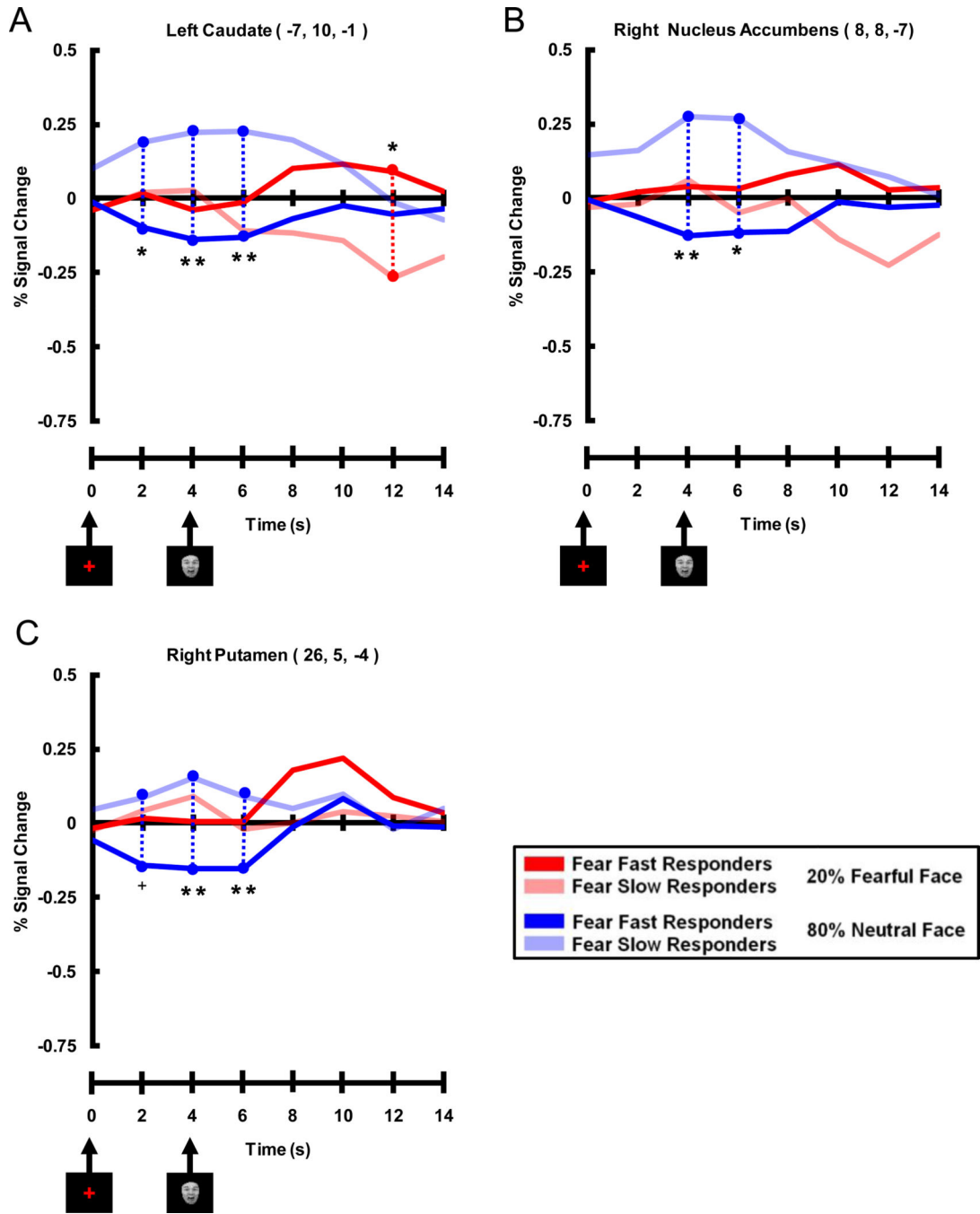
Figure 7. Timecourses Extracted from Peak Voxels Displaying Early Negative Correlation Between Percent Signal Change for Expected Neutral Faces and Valence Bias Scores
Areas of the dorsal and ventral striatum of the basal ganglia included A) left caudate, B) right nucleus accumbens, and C) right putamen. These time courses illustrate that Fear Slow Responders (light blue lines) had more activation than Fear Fast Responders (dark blue lines) early (i.e. 2, 4 and 6 s post-cue, 0 and 2 s post-face) in expected neutral face trials. It is interesting to note that although these peak voxels were significant for the negative correlation between Valence Bias scores and percent signal change for expected neutral faces, the left caudate (A) also displayed a late (i.e. 12 s post-cue) trial enhancement for Fear

Fast Responders (dark red lines) compared to Fear Slow Responders (light red lines) for unexpected fearful faces. (** = p < 0.01 corrected, * = p < 0.05 corrected, + = p < 0.10 corrected)

Table 1

Regions Showing Correlation with Valence Bias

All clusters significant at p < 0.01 corrected 1 voxel = 3 mm × 3 mm × 3 mm = 27 mm^3

Positive Correlation with Valence Bias

Timepoint	Stimulus	Anatomical Location	Peak Voxel Talairach Coordinates			Peak r value	Peak t-statisti	Number of Voxels in Cluster (Volume mm^3)
			x	y	z			
12 s post-cue (8 s post-face)	20% Fearful Face	L Ventromedial Prefrontal Cortex	−7	44	−1	0.78	5.94	118 (3186)
		L Subgenual Cingulate Cortex	−1	14	0	0.75	5.27	
		R Ventromedial Prefrontal Cortex	17	35	−4	0.72	4.89	44 (1188)
		R Head of the Caudate Nucleus	17	26	−1	0.72	4.97	

Negative Correlation with Valence Bias

Timepoint	Stimulus	Anatomical Location	Peak Voxel Talairach Coordinates			Peak r value	Peak t-statisti	Number of Voxels in Cluster (Volume mm^3)
			x	y	z			

Positive Correlation with Valence Bias

Timepoint	Stimulus	Anatomical Location	Peak Voxel Talairach Coordinates			Peak r value	Peak t-statisti	Number of Voxels in Cluster (Volume mm³)
			x	y	z			
4 s post-cue (0 s post-face)	80% Neutral Face	L Head of Caudate Nucleus	−7	10	−1	−0.70	−4.58	
		R Nucleus Accumbens	8	8	−7	−0.71	−4.73	82 (2214)
		R Putamen	26	5	−4	−0.72	−4.85	