

## Test-retest reliability of amygdala response to emotional faces

COLIN L. SAUDER,<sup>a</sup> GREG HAJCAK,<sup>a</sup> MIKE ANGSTADT,<sup>b</sup> AND K. LUAN PHAN<sup>b,c,d</sup>

<sup>a</sup>Department of Psychology, Stony Brook University, Stony Brook, New York, USA

<sup>b</sup>Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA

<sup>c</sup>Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA

<sup>d</sup>Mental Health Service Line, Jesse Brown VA Medical Center, Chicago, Illinois, USA

### Abstract

In the current study, we evaluated the test-retest reliability of amygdala response using an emotional face-matching task that has been widely used to examine pathophysiology and treatment mechanisms in psychiatric populations. Activation within the fusiform face area (FFA) was also examined. Twenty-seven healthy volunteers completed a variation of the face-matching paradigm developed by Hariri et al. (2000) at two time points approximately 90 days apart. Estimates of test-retest reliability of amygdala response to fearful faces were moderate, whereas angry and happy faces showed poor reliability. Test-retest reliability of the FFA was moderate to strong, regardless of facial affect. Collectively, these findings indicate that the reliability of the BOLD MR signal in the amygdala varies substantially by facial affect. Efforts to improve measurement precision, enlarge sample sizes, or increase the number of assessment occasions seem warranted.

**Descriptors:** fMRI, Psychometrics, Emotion, Amygdala, Fusiform face area

Over the past decade, there has been increasing emphasis on research explicating the neural basis of psychopathology. This trend is consistent with major funding initiatives such as the Research Domain Criteria project, part of the National Institute of Mental Health's strategic plan to develop a neuroscience-based system of classification for psychiatric disorders (Insel et al., 2010). Likewise, several researchers have suggested that the evaluation of neural markers is a necessary step toward accurate diagnosis and subsequent treatment (Beauchaine, Neuhaus, Brenner, & Gatzke-Kopp, 2008; Cicchetti & Toth, 2009; Sroufe & Rutter, 1984). This focus on translational neuroscience is evident in the growing literature that uses functional magnetic resonance imaging (fMRI) to better understand psychopathology, as researchers increasingly use this neuroimaging method to both identify neural markers of disease-related traits and characterize treatment response (Matthews, Honey, & Bullmore, 2006).

Although the field has progressed in identifying some of the neural mechanisms of psychopathology and related processes, research on the psychometric properties (i.e., reliability and validity) of fMRI measures is sparse. This is in stark contrast to other domains of research, in which validation of a measure is a prerequisite for its clinical utility. For example, the internal consistency and test-retest reliability of the Beck Depression Inventory-II (Beck, Steer, & Brown, 1996) have been repeatedly demonstrated prior to the release of the measure (Beck, Steer, Ball, & Ranieri,

1996) and over the course of its use (Dozois, Dobson, & Ahnberg, 1998; Sprinkle et al., 2002; Storch, Roberti, & Roth, 2004). Despite nearly two decades of fMRI research examining the neural response to emotional stimuli, evidence supporting the reliability of this measurement is sparse, and the data is mixed.

Classic statistical theory suggests that reliability places an upper limit on the predictive validity of any measure, and that the product of the reliability coefficients of any two measures limits the correlation between them (Shrout, 1998). Low measure reliability also reduces power, particularly to detect relationships between measures (e.g., the ability to correlate brain response and self-report measures). Thus, sufficient data on measure reliability is a prerequisite for both study design and interpretation, as reliability estimates directly impact the number of participants or scans needed (as a function of power) while limiting potential outcomes (i.e., a measure cannot correlate better with another measure than it correlates with itself).

Adequate measure reliability is especially important in intervention studies, which typically assess dependent measures multiple times over the course of several weeks or months. Clinical trials for pharmacotherapy and psychotherapy, as well as psychopharmacological challenge studies, increasingly include pre- and posttreatment fMRI assessments. However, the concurrent and predictive validity of fMRI-based measures of neural activity are necessarily limited by reliability. In the current study, we focus specifically on the test-retest reliability of the amygdala response to emotional faces. As part of the limbic system, the amygdala has been implicated in vigilance, fear processing, pain processing, and the identification and regulation of emotions (Davis, 1992; Davis & Whalen, 2001; Neugebauer, Li, Bird, & Han, 2004). A large corpus of studies have consistently reported augmented amygdala

This study was funded in part by National Institutes of Health—National Institute of Mental Health Grant MH076198 (PI: K. L. Phan).

Address correspondence to: Colin Sauder, Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500. E-mail: colin.sauder@stonybrook.edu

response to emotional faces, regardless of valence (Costafreda, Brammer, David, & Fu, 2008; Phan, Wager, Taylor, & Liberzon, 2002; Sergerie, Chochol, & Armony, 2008).

Moreover, substantial translational research has focused on amygdala activation in response to emotional faces in relation to psychopathology. Altered amygdala response to emotional faces has been implicated in several psychiatric disorders including depression, anxiety disorders, borderline personality disorder, antisocial personality disorder, and bipolar disorder (Etkin & Wager, 2007; Leppänen, 2006; Lis, Greenfield, Henry, Guilé, & Dougherty, 2007; McCloskey, Phan, & Coccaro, 2005; Rauch, Shin, & Wright, 2003; Strakowski, DelBello, & Adler, 2004). Across these studies, abnormal amygdala response to salient emotional stimuli is taken to reflect trait-like abnormalities in emotion processing or its regulation. As a result, several researchers have proposed that amygdala response to faces, in particular exaggerated response to emotional faces, may be a predictive neural marker of disorder pathogenesis (Benes, 2007; Goodman, New, Triebwasser, Collins, & Siever, 2010; Hasler & Northoff, 2011; Rauch et al., 2003).

The utility of this putative biomarker of psychiatric diseases is dependent upon the assumption that amygdala response to emotional faces is trait-like; that is, that this measure is reliable over time. To date, four studies have examined the reliability of amygdala response to emotional faces (Johnstone et al., 2005; Manuck, Brown, Forbes, & Hariri, 2007; Plichta et al., 2012; van den Bulk et al., 2013), with mixed results. The first study by Johnstone et al. (2005) suggested moderate to strong reliability for fearful faces in comparison to fixation across three sessions that were 0, 2, and 8 weeks apart. Reliability estimates varied significantly according to the contrast examined (fear vs. fixation or fear vs. neutral), area of interest (functional vs. anatomically defined amygdala), test-retest period (2 weeks vs. 8 weeks), and hemisphere. Reliability was highest for the functionally defined left amygdala when calculated across all three sessions, intraclass correlation ( $ICC_{(*,1)}$ ) = 0.66, but much lower for the anatomical amygdala,  $ICC_{(*,1)}$  = 0.28. Across longer test-retest periods, the left amygdala response was reliable,  $ICC_{(*,1)}$  = 0.63, whereas the right amygdala showed poor reliability,  $ICC_{(*,1)}$  = 0.27.<sup>1</sup> In comparison, Manuck and colleagues (2007) found moderate test-retest reliability over a 1-year period for threatening faces (i.e., fearful and angry faces combined) in the right amygdala only,  $ICC_{(*,*)}$  = 0.59, despite significant bilateral activation at both time points. However, a recent study by Plichta et al. (2012) suggested strong reliability,  $ICC_{(3,1)} > 0.60$ , bilaterally to fearful and angry faces across a 2-week retest period. Finally, van den Bulk and colleagues (2013) examined amygdala reliability in an adolescent sample, but found poor amygdala reliability in response to fearful, happy, and neutral faces (combined) that was inconsistent across three time points approximately 3 months apart ( $ICC_{(3,1)}$  ranging from  $-0.02$  to  $0.34$ , depending on time period assessed, ROI definition, and hemisphere).

Both the Manuck and Plichta studies used variants of a popular task in which participants match faces based on expression (Hariri, Bookheimer, & Mazziotta, 2000)—and these studies demonstrate higher reliability estimates than the other studies. Moreover, higher reliability using this task was apparent at shorter (e.g., Plichta et al.) as compared to longer (e.g., Manuck et al.) test-retest intervals, although the latter study included only 13 participants. It is difficult to directly compare these studies, however, as the Manuck

et al. study failed to find significant reliability for the functionally defined left amygdala, and did not report reliability associated with anatomical regions of interest. Thus, there is a need to further assess the stability of the amygdala using larger samples and over longer test-retest periods.

Additionally, extant studies have either combined reliability results across multiple emotional expressions or examined fearful faces in isolation. Although fearful faces are likely the most widely utilized facial expression, the use of angry, disgusted, and happy faces is not uncommon in clinical neuroscience research (Leppänen, 2006; Monk, Klein et al., 2008; Monk, Telzer et al., 2008; Somerville, Kim, Johnstone, Alexander, & Whalen, 2004; Weng et al., 2011). Thus, it is difficult to ascertain whether findings from previous studies might generalize to alternative facial expressions.

To address these gaps in the literature, the current study evaluated the test-retest reliability of the amygdala response to fearful, happy, and angry faces using a variant of the emotional face-matching paradigm over a period of approximately 90 days (Hariri et al., 2000; Phan et al., 2008). This face-matching task has been utilized across a variety of disorders including autism spectrum disorders, schizophrenia, depression, bipolar disorder, and Williams-Beuren syndrome (Fakra, Salgado-Pineda, Delaveau, Hariri, & Blin, 2008; Matthews, Strigo, Simmons, Yang, & Paulus, 2008; Meyer-Lindenberg et al., 2005; Robinson et al., 2008; Wang, Dapretto, Hariri, Sigman, & Bookheimer, 2004); it has been linked to serotonin transporter genetic variation (Hariri, Mattay et al., 2002), and has been used in both treatment (Phan et al., 2012) and psychopharmacological challenge studies (Arce, Simmons, Lovero, Stein, & Paulus, 2008; Paulus, Feinstein, Castillo, Simmons, & Stein, 2005; Phan et al., 2008). This task has also been associated with the best reliability estimates in previous studies (Manuck et al., 2007; Plichta et al., 2012). Using a larger sample and longer test-retest period than previous studies, we sought to evaluate the stability of amygdala response over a 90-day period and determine whether it differed by emotion expression. We predicted superior reliability for fearful and angry faces relative to happy faces, which is consistent with previous research indicating greater specificity of amygdala response to threatening stimuli (Costafreda et al., 2008; Phan et al., 2002).

As a point of comparison, we concurrently evaluated the reliability of the fusiform face area (FFA), an area of the brain critical to face processing that has been shown to respond differentially to affective faces in comparison to neutral faces (Kanwisher, McDermott, & Chun, 1997; Vuilleumier & Pourtois, 2007). The FFA is both anatomically and functionally linked to the amygdala (Amaral & Price, 1984; George, Driver, & Dolan, 2001; Sabatinelli, Bradley, Fitzsimmons, & Lang, 2005), and lesion studies have demonstrated that amygdala damage impairs the processing of affective faces within the FFA (Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004). In concordance with previous research demonstrating that areas of the visual cortex respond consistently to affective faces regardless of valence (Phan et al., 2002), we predicted that the reliability of MR response in the FFA would not vary across facial expression.

## Materials and Method

### Participants

The current study consisted of 27 right-handed subjects (14 males, 13 females; age range 18–46,  $M = 27.5$ ,  $SD = 8.3$ ) without a history

1. \* indicates insufficient data to determine ICC type.

of head injury who were recruited as part of a healthy control group in two larger treatment outcome studies. All participants were given the research version of the Structured Clinical Interview for DSM Disorders, nonpatient edition (First et al., 2002) to rule out individuals with previous or current psychiatric illness, substance abuse, or learning disability. Participants also underwent a health screening by a physician to rule out neurologic or major medical problems and completed a urine drug test immediately prior to scanning, and all were free of psychoactive/psychotropic medications (lifetime) or drugs at time of scanning. Consistent with local Institutional Review Board procedures, written consent was acquired from all individuals after explanation of the experimental protocol.

### Design and Procedure

Participants underwent fMRI scanning at two separate time points approximately 90 days apart ( $M = 88.9$ ,  $SD = 6.4$ ). During both visits, participants completed a face-matching task (Hariri et al., 2000) that has been shown to robustly activate the amygdala (Klumpp, Angstadt, & Phan, 2012; Labuschagne et al., 2010; Sripada, Angstadt, McNamara, King, & Phan, 2011). In this task, participants match one of two cues (presented at the bottom of the screen) to a corresponding target (centered at the top of the screen) via a left or right button press using their dominant hand. In face-matching blocks, participants indicated which of the two facial cues (one of which is always neutral) matched the emotion of the target face (always happy, angry, or fearful). In the control condition, participants matched simple geometric shapes. The current study paradigm consisted of three 20-s blocks for each angry, happy, and fearful face-matching condition interleaved with shape-matching control blocks. The entire task lasted approximately 6 min.

### fMRI Acquisition

Imaging data were acquired using a 3 Tesla GE Signa System (General Electric, Milwaukee, WI). A T2\*-weighted reverse spiral sequence was used for all participants with temporal resolution of 2,000 ms, although there was some variation in acquisition across different study control group participants that were combined in the current sample. For most participants, temporal echo (TE) = 25 ms, flip angle = 77°, field of view (FOV) = 240 × 240 mm, slice thickness = 5 mm, 64 × 64 matrix with an in-plane resolution of 3.75 × 3.75 mm, and a total of 30 axially acquired slices. Nine of the participants had slightly different parameters, TE = 30 ms, flip angle = 90°, FOV = 220 × 220 mm, slice thickness = 3 mm, 64 × 64 matrix with a 3.44 × 3.44 mm in-plane resolution and a total of 32 slices. A reverse-spiral acquisition sequence specifically designed to reduce susceptibility artifacts (Stenger, Boada, & Noll, 2000, 2002, 2003) was used for all participants. Remaining scanning parameters were selected to be consistent with the work of Noll and colleagues. To ensure that the small differences in scan procedures within a small subgroup of our sample did not unfairly bias the results, scan subgroup was initially included as a covariate in second level analyses; however, the covariate had no significant effect on BOLD signal, and thus was dropped from the final model.

### fMRI Data Analysis

Data from the face-matching task were analyzed using Statistical Parametric Mapping software, version 8 (SPM 8; [\[www.fil.ion.ucl.ac.uk/spm/software/spm8\]\(http://www.fil.ion.ucl.ac.uk/spm/software/spm8\)\). Data preprocessing for both scanning sessions followed SPM standard procedures, including motion and field warping correction, spatial normalization to the echo-planer imaging template, and 8 mm Gaussian smoothing. A fixed-effects general linear model \(GLM\) was then created for the first and second scanning session of each individual participant separately to compare face-matching blocks to baseline condition \(shape-matching blocks\). Contrast images for the first scanning session only were then entered into a mixed-effects second-level GLM with faces minus shapes as a fixed effect and subject as a random effect, and a one-sample  \$t\$  test of the face – shape contrast was created.](http://</a></p>
</div>
<div data-bbox=)

The resulting parametric  $t$  map was analyzed using a region-of-interest (ROI) approach. Briefly, areas of functional activation for faces in contrast to shapes were extracted from the amygdala and fusiform gyrus, bilaterally, as defined by the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) in conjunction with the Wake Forest University PickAtlas (Maldjian, Laurienti, Kraft, & Burdette, 2003). Parametric maps were corrected for multiple comparisons using a false-discovery rate correction, thresholded at  $p < .05$ , and areas of significant activation within the anatomical labels for the amygdala and fusiform were used to create the functional ROIs. Consistent with previous studies, functional ROIs were created using data from the first time point only, collapsed across emotional expression. These ROIs did not differ substantially from regions of activation for individual emotional expressions, at both time points (i.e., the area activated within the amygdala by angry faces at T1 & T2 showed > 90% overlap with seed ROI). Finally, anatomical ROIs were created using anatomical labels from the AAL atlas.

Functional and anatomical ROIs were then analyzed using MARSBAR toolbox (Brett, Anton, Valabregue, & Poline, 2002), which estimates neural response for each participant individually, averaging across all voxels within the functional/anatomical ROI. Contrast estimates were extracted separately for fearful, angry, and happy face-matching blocks in contrast to shape-matching blocks for both the first and second sessions. To assess differences in signal quality, signal-to-noise ratio (SNR), which is defined as the mean activation across time divided by the standard deviation, was calculated for each functional ROI at both time points.

To evaluate test-retest reliability, ICCs were calculated using extracted contrast estimates from anatomical and functionally defined ROIs for all participants across both scan sessions. Consistent with the most recent study of reliability (Plichta et al., 2012), we calculated measures of absolute agreement,  $ICC_{(2,1)}$  (which reflects both measurement consistency and mean level differences), as well as consistency,  $ICC_{(3,1)}$  (which does not take mean level differences into account). Qualitative descriptions of ICC values are based on standards in psychometric research (Cicchetti & Sparrow, 1981; McGraw & Wong, 1996), and are consistent with recent studies of amygdala reliability (Plichta et al., 2012; van den Bulk et al., 2013):  $< 0.4 =$  poor,  $0.41–0.59 =$  fair,  $0.60–0.74 =$  good, and  $> 0.75 =$  excellent. ICCs were calculated in SPSS using a random effects two-way model, consistent with the work of McGraw and Wong (1996).

### Results

Fearful, happy, and angry faces all produced significantly greater activation in contrast to shapes within both the amygdala and FFA, bilaterally, at both time points (see Table 1). Moreover, patterns of activation were consistent spatially across both time points, with

**Table 1.** Areas of Significant Activation Within the Anatomical Boundaries of the Fusiform Gyrus and Amygdala, Bilaterally

Region	Coordinates	Z score	Cluster volume mm <sup>3</sup>	Overlap %
Left FFA				
T1	-33 -60 -15	7.04	8,640	78
T2	-34 -62 -16	6.00	7,236	
Right FFA				
T1	33 -59 -15	7.02	10,449	82
T2	34 -60 -15	7.20	8,991	
Left amygdala				
T1	-23 -2 -16	5.75	1,350	87
T2	-24 -2 -18	5.58	1,782	
Right amygdala				
T1	26 -1 -17	6.93	1,566	94
T2	27 0 -19	3.52	1,971	

Notes. Coordinates are in Montreal Neurologic Institute (MNI) coordinate space and reflect the center mass of each cluster of activation. Overlap reflects the percentage of active voxels at T1 that are also active at T2. T1 = first assessment; T2 = second assessment.

cluster overlap ranging from 78–94% (see Figure 1). To assess potential differences in activation across hemisphere, time, and facial expression, a repeated measures analysis of variance (ANOVA) was performed for the amygdala and FFA, separately (see Figure 2). For the amygdala, there was a significant main effect of expression,  $F(2,52) = 3.55$ ,  $p < .05$ , with angry faces producing more amygdala activation than both fearful faces,  $t(26) = 2.27$ ,  $p < .05$ , and happy faces,  $t(26) = 2.13$ ,  $p < .05$ ; and no difference in activation between fearful faces and happy faces,  $t(26) < 1$ . Amygdala activation was comparable across hemispheres,  $F(1,26) < 1$ ; though amygdala activation was greater at the first compared to second testing session overall,  $F(1,26) = 7.62$ ,  $p < .05$ . There were no significant interaction effects (all  $p$  values  $> .14$ ).

Overall, the FFA response was equivalent across facial expressions,  $F(2,52) < 1$ ; however, this was qualified by a significant Hemisphere  $\times$  Expression interaction,  $F(2,52) = 4.52$ ,  $p < .05$ . Within the right FFA, angry faces produced greater activation than both fearful,  $t(26) = 2.30$ ,  $p < .05$ , and happy faces,  $t(26) = 2.09$ ,  $p < .05$ , with no difference in activation between fearful and happy faces,  $t(26) < 1$ ; in the left FFA, angry faces produced significantly

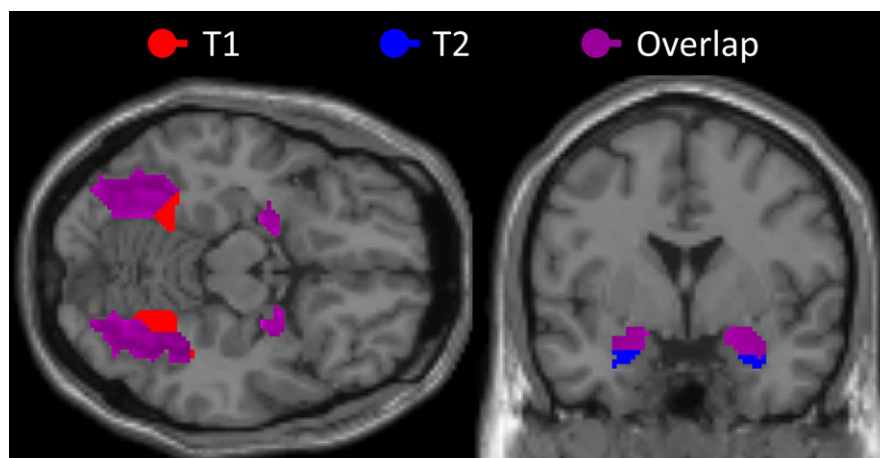
greater activation than happy faces,  $t(26) = 2.13$ ,  $p < .05$ , but not fearful faces,  $t(26) = 1.79$ ,  $p > .05$ , while activation for fearful and happy faces was comparable,  $t(26) = 1.37$ ,  $p > .15$ . Like the amygdala, FFA activation was greater at the first time point,  $F(1,26) = 13.00$ ,  $p < .01$ . There was no overall difference in activation across hemispheres,  $F(1,26) < 1$ . All other two-way and three-way interaction effects were nonsignificant ( $p > 0.12$ ).

A repeated measures ANOVA was also calculated to assess differences in SNR across region, time, and hemisphere. In addition, to account for the potential impact of differences in scanning procedure among a subset of the sample, scan type was included as a covariate. There was a significant main effect of region, such that the amygdala had significantly lower SNR ( $M = 160.31$ ,  $SD = 10.54$ ) than the FFA ( $M = 186.36$ ,  $SD = 8.34$ ),  $F(1,25) = 4.35$ ,  $p < .05$ . There was no main effect of hemisphere or time, and interaction effects on SNR did not reach significance (all  $p$  values  $> .30$ ). Likewise, differences in scanning procedures had no significant impact on SNR ( $p > .15$ ).

Estimates of the test-retest reliability of both the FFA and amygdala are presented in Table 2. For the functionally defined ROI, amygdala response to fearful faces was more reliable than response to happy or angry faces. Only fearful faces produced reliability estimates that were statistically different from zero for both the right and left hemisphere. Happy faces were less reliable, with significant ICC values for the left amygdala only. Reliability estimates for angry faces did not differ from zero in either hemisphere (see Table 2). In contrast, activation within the FFA was much more consistent (see Figure 3). FFA response to faces was reliable over time regardless of facial affect or hemisphere, and reliability estimates in the FFA were generally higher as compared to the amygdala.<sup>2</sup>

In line with the work of Johnstone and colleagues (2005), we also evaluated the reliability of response within the anatomically

2. To be comprehensive, we conducted ICC reliability estimates of amygdala and FFA within each session. As seen in online supporting information Table S1, the FFA, but not amygdala, shows a reliable response to emotional faces within session. However, it should be noted that our study design was not optimized to test reliability within session, which would have necessitated having multiple blocks of each emotion type across multiple functional runs. In other words, an ideal test of within-session reliability involves measuring ICCs across functional runs and not across blocks.



**Figure 1.** Sagittal and axial views of areas of significant activation within the fusiform gyrus and amygdala anatomical boundaries.

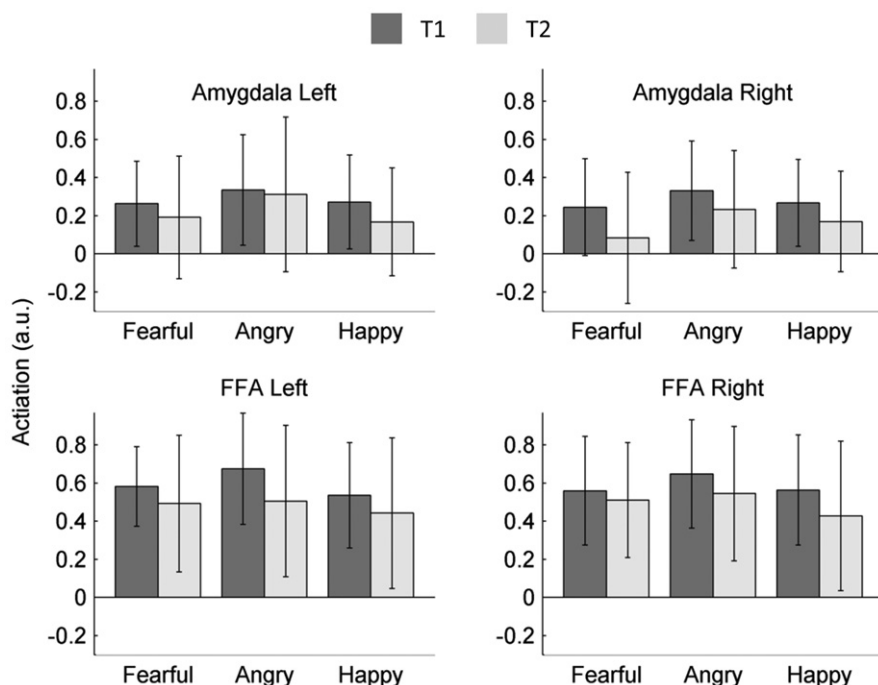


Figure 2. Estimates of amygdala and FFA response to fearful, angry, and happy faces at both time points. Error bars reflect standard deviation.

defined amygdala. Results were largely consistent, suggesting that response within the amygdala is reliable for fearful faces, while neither happy nor angry expressions produced significant ICC estimates.<sup>3</sup>

Discussion

In the current study, we assessed longer term test-retest reliability of MR response to emotional faces within the amygdala. Results

3. As a point of comparison, within-session reliability was also assessed. Reliability estimates were generally lower, with neither fearful, angry, nor happy faces producing consistent response across three 20-s testing blocks (see on-line supporting information Table 1). However, it should be noted that the task was not optimized for the examination of within-session reliability, and thus results should be interpreted with caution.

Table 2. Intraclass Correlation Coefficients for the Functionally Defined Left and Right FFA and Amygdala and the Anatomically Defined Amygdala

		Left			Right		
		Fearful	Angry	Happy	Fearful	Angry	Happy
Amyg-FX	ICC <sub>(2,1)</sub>	0.32*	-0.12	0.35*	0.40*	0.11	0.23
	ICC <sub>(3,1)</sub>	0.34*	-0.12	0.37*	0.43*	0.11	0.24
Amyg-Anat	ICC <sub>(2,1)</sub>	0.42*	-0.11	0.19	0.36*	-0.24	0.21
	ICC <sub>(3,1)</sub>	0.41*	-0.10	0.20	0.38*	-0.24	0.21
FFA-FX	ICC <sub>(2,1)</sub>	0.50*	0.46*	0.62*	0.50*	0.37*	0.49*
	ICC <sub>(3,1)</sub>	0.63*	0.58*	0.68*	0.50*	0.42*	0.53*

Note. ICC<sub>(2,1)</sub> refers to absolute agreement measures of reliability, while ICC<sub>(3,1)</sub> coefficients reflect consistency measures. FX refers to the functionally defined ROI, while Anat refers to anatomical definition. \*ICC values that are significantly different from zero.

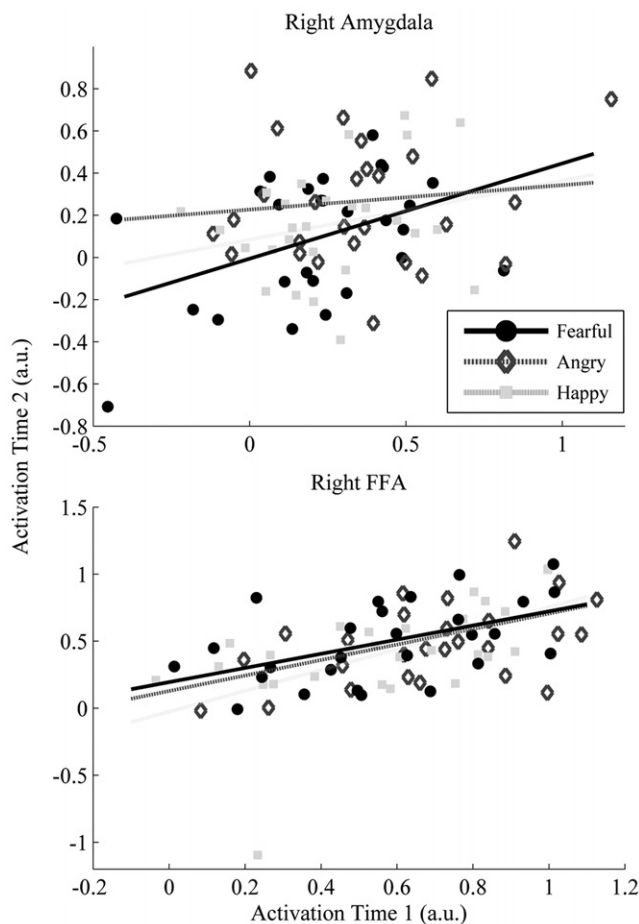


Figure 3. Scatter plots and linear fits of participant responses to happy, fearful, and angry faces within the right amygdala and FFA at initial (x-axis) and final (y-axis) scan acquisition.

indicate moderate reliability over a 90-day interval, but only in response to fearful faces. Although angry faces produced stronger activation than both fearful and happy faces in the amygdala, this response was not reliable over time. These results were consistent for the left and right amygdala, and regardless of whether the amygdala was defined functionally or anatomically. Neural response to happy faces was reliable within the left (but not right) amygdala, and only when functionally defined.

As a point of contrast, the current study also evaluated reliability within the fusiform gyrus, an area of the brain with strong functional and anatomical connectivity with the amygdala that has been shown to respond consistently to facial stimuli (George et al., 2001; Kanwisher et al., 1997). Results suggest that MR signal from this region shows good reliability irrespective of facial affect. Specifically, reliability estimates for the FFA ranged from fair to good for all facial expressions, and this was true across both left and right hemispheres.

Studies of amygdala reliability typically evaluate reliability using functional ROIs, a subset of voxels within the amygdala that exceeds a certain statistical threshold (i.e., a subset of the anatomical region). While this approach may lead to greater reliability estimates, it may result in less generalization of study findings to other emotional paradigms, which may not activate the same subset of voxels within the amygdala. To date, only one other study has reported reliability for the whole, atlas-defined amygdala (Johnstone et al., 2005). In that study, reliability was much lower for the anatomically defined amygdala than the functionally defined amygdala. However, reliability estimates in the current study were largely the same, whether evaluated for the whole anatomical amygdala or smaller regions based on functional activation.

Interestingly, estimates of BOLD signal in the amygdala were greater during the initial assessment as compared to follow-up, despite the fact that clusters of activation were larger at the second time point. While these results may simply reflect attenuation of amygdala signal as a result of habituation (Breiter et al., 1996; Fischer et al., 2003; Wright et al., 2001), it is also important to note that, consistent with previous studies of amygdala reliability, functional ROI definition was dependent on the first testing session only. Thus, despite substantial overlap in active voxels across testing sessions (see Figure 1, Table 1), attenuated signal seen at the second time point may reflect the inclusion of nonsignificant voxels. A similar effect was also seen within the FFA, with greater BOLD signal seen at the first assessment. While the same habituation effects are not typically seen within the FFA, numerous studies indicate a functional link between the FFA and amygdala, with some studies suggesting that amygdala response directly modulates FFA response, and vice versa (George et al., 2001; Pujol et al., 2009; Robinson, Laird, Glahn, Lovallo, & Fox, 2010; Sabatinelli et al., 2005). Finally, the similarity of reliability estimates taken from the functional ROI and anatomical amygdala suggest that these potential confounds have little to no effect on reported reliability.

Results of the current study suggest that amygdala response to emotional faces may be less reliable than response in other regions such as the FFA. This finding is consistent with previous studies that have found greater consistency in whole brain and regional activation as compared to amygdala response (Plichta et al., 2012; van den Bulk et al., 2013). In addition, we found significantly lower SNR within the amygdala compared to the FFA. Low SNR limits measurement precision, can lead to an underestimation of BOLD response (LaBar, Gitelman, Mesulam, & Parrish, 2001), and negatively impacts estimates of fMRI reliability (Bennett & Miller, 2010). Thus, lower reliability estimates for amygdala

response to emotional faces could reflect lower SNR within this region.

Several studies have evaluated reliability of fearful faces, either alone or in conjunction with angry or both angry and happy faces. The first of these studies demonstrated good reliability of amygdala response to passively viewed fearful faces across three scanning sessions over an 8-week period (Johnstone et al., 2005). When reliability was assessed across sessions at 0, 2, and 8 weeks, estimates were higher than in the current study; however, estimates taken across the first and third scanning session were much more similar to the current study. Thus, the number of assessment points may be less important than test-retest interval. For instance, the recent study by Plichta and colleagues (2012) reported good to excellent reliability over a 2-week period. Likewise, estimates of reliability over 2 weeks in the Johnstone et al. (2005) paper were also strong. In both studies, reliability over a 2-week test-retest period was higher than over the longer test-retest period used in the current study. Attenuation of reliability estimates over long intervals may in part reflect the context-dependent nature of amygdala activation. For example, previous research suggests that amygdala response is vulnerable to the effects of state anxiety (Somerville et al., 2004), life stress (Lemogne et al., 2011), sleep and perceived stress (Prather, Bogdan, & Hariri, 2013), and reappraisal capacity (Drabant, McRae, Manuck, Hariri, & Gross, 2009). Thus, increased temporal delay between testing sessions may result in greater variation in environmental context—an effect that may differentially impact amygdala reactivity and reliability.

The current study used very similar task procedures as in the recent paper by Plichta and colleagues (2012). While the Plichta study suggests that reliability of amygdala response to threatening faces (i.e., angry and fearful) over a short test-retest period is strong, our results suggest more modest estimates over a much longer period of time. Reliability estimates in the current study may be particularly relevant to intervention researchers who seek to include fMRI measures of amygdala response as an index of treatment outcome, as the typical duration to test therapeutic effects in pharmacotherapy and psychotherapy clinical trials is approximately 90 days.

In the current study, fearful, happy, and angry faces all produced large areas of significant activation within the FFA and amygdala, at both time points. However, estimates of amygdala reliability for happy and angry faces were generally poor. One exception was that the left functionally defined amygdala response to happy faces was reliable, and on par with that observed for fearful faces. In contrast, FFA reliability estimates were largely equivalent across facial affect and stronger, ranging from fair to good. Thus, facial expression appears to have a substantial impact on amygdala, but not FFA reliability.

These findings have important implications in applied neuroimaging research, as many studies have linked psychopathology to abnormalities of amygdala response. Psychopathology is trait-like and generally stable over time (Krueger, Caspi, Moffitt, & Silva, 1998)—and neural correlates and biomarkers of psychopathology should themselves have high stability and test-retest reliability (i.e., if a disease is trait-like, then a biomarker of that disease should be trait-like as well). However, results from the current study suggest that neither happy nor angry faces produce consistent activation of the amygdala over time, and that reliability in response to fearful faces is at best fair. If amygdala response demonstrates poor reliability over time, it is not clear how this neural measure could reflect trait-like differences that underlie psychopathology.

In the current study, measures of amygdala consistency,  $ICC_{(3,1)}$ , did not exceed 0.41 for the amygdala, and ranged from 0.42–0.68 for the FFA. While this may be considered insufficient reliability in psychometric research (Nunnally, 1978), especially in the case of the amygdala, studies of fMRI reliability typically find ICCs in the 0.33–0.66 range (Bennett & Miller, 2010). However, as pointed out by Bennett and Miller, there is no consensus value of what constitutes an acceptable level of reliability in fMRI research. In other words, it is difficult to determine what constitutes a “good” fMRI measure. In the current study, we attempted to address this question by concurrently evaluating the test-retest reliability of more basic visual processing of facial stimuli in the FFA. Reliability estimates were much higher within this region, suggesting that the maximum long-term reliability of fMRI measures of facial processing in the widely used face-matching task (Hariri et al., 2000) may exceed 0.60, which in behavioral research is considered good (McGraw and Wong, 1996). In comparison, amygdala activation is less reliable, indicating that amygdala response to emotional faces is more varied and perhaps subject to greater sources of measurement error.

Although the current study suggests that amygdala response may not be as stable and trait-like as previously believed, there remains significant evidence that the amygdala response to threatening stimuli is an index of psychopathology and intervention. The results of this study do, however, necessitate caution in the interpretation of studies that examine amygdala response over time. For example, previous studies clearly indicate that amygdala response is sensitive to psychopharmacological agents (Arce et al., 2008; Paulus et al., 2005; Phan et al., 2012). However, treatment studies have not consistently demonstrated intervention-related changes in amygdala activation (Fu, Steiner, & Costafreda, 2012), which may be partially due to low reliability of amygdala response over time—amygdala response cannot correlate with another measure (e.g., symptoms) better than it correlates with itself.

Just as reliability of measurement affects the necessary sample size to detect differences, the accurate assessment of the true reliability of a measure is dependent on the number of participants included in the study (Walter, Eliasziw, & Donner, 1998). The current study is the largest to date ( $N = 27$ ), allowing us to more accurately assess the reliability of amygdala response to emotional faces. The inclusion of both happy and angry faces in addition to fearful faces also allowed us to contrast the stability of amygdala response across expressions, and further confirmed that fearful faces most reliably and accurately recruit the amygdala (Costafreda et al., 2008; Phan et al., 2002). Further, the task used in the current study has shown consistent findings in many populations and has been used in other recent studies of amygdala reliability, allowing for a more direct comparison of study results.

However, there are several limitations in the current study. The inclusion of three task conditions necessarily reduced the number of assessments of neural response to these stimuli, and there is evidence to suggest that reliability of fMRI data may be reduced as a result (Bennett & Miller, 2010). Likewise, several factors including magnet strength, length of MR acquisition, type of stimuli, MR parameters, and preprocessing methodology all influence the stability and strength of amygdala response (Bennett & Miller, 2010; Hariri, Tessitore et al., 2002; LaBar et al., 2001; Robinson, Windischberger, Rauscher, & Moser, 2004), and will differ substantially across studies. Although scan acquisition parameters were specifically designed to reduce signal artifact within the

amygdala (Stenger et al., 2000, 2002, 2003), previous research has demonstrated that small variations in parameters such as TE, voxel size, slice orientation, z-shim, and spiral sequence can all have significant effects on BOLD signal contrast, susceptibility artifacts, and SNR (Chen, Dickey, Yoo, Guttman, & Panych, 2003; Du, Dalwani, Wylie, Claus, & Tregellas, 2007; Glover & Law, 2001; Gorno-Tempini et al., 2002; Merboldt, Fransson, Bruhn, & Frahm, 2001; Morawetz et al., 2008). In the current study, we did not assess how variations in these parameters may impact reliability estimates of BOLD response in the amygdala and FFA. As such, the results of the current study may not apply equally to other fMRI studies of amygdala response.

The current study includes only healthy individuals and does not necessarily reflect expected reliability estimates in clinical groups, which may be more variable than healthy individuals (Krueger et al., 1998). Furthermore, although the current study is the largest such study to date, our sample still included fewer than 30 individuals. Reliability studies typically include hundreds of participants, thus the current study may not accurately assess the limits of amygdala reliability. Future studies of amygdala reliability that include more individuals are needed, allowing researchers to determine more specifically the effects of task conditions, scanning methodology, and sample characteristics.

There is an increasing emphasis on research that seeks to identify biomarkers of psychiatric illness, and neural markers evaluated using fMRI are of particular interest (Insel et al., 2010; Singh & Rose, 2009). However, the utility of amygdala response as a potential neural marker of psychiatric illness is by necessity linked to its reliability. Low reliability reduces power, and in turn requires increases in scanning acquisition time and number of participants, thus increasing utilization cost (Perkins, Wyatt, & Bartko, 2000). More importantly, low reliability limits the validity of a measure. A measure cannot correlate better with another measure than it correlates with itself, and measurement reliability is a necessary precondition for measurement validity (Shrout, 1998). The current study suggests poor to fair reliability of amygdala response to fearful faces over a test-retest period of several months. Although within the range expected based on a recent review of studies of fMRI reliability (Bennett & Miller, 2010), clinical utility of this measure may require multiple assessments, shorter test-retest periods, or longer scanning sessions. The degree to which an individual activates their amygdala in response to fearful faces is somewhat stable over time—but not nearly as stable as other individual difference measures (e.g., self-report measures of anxiety or depression). The current results suggest caution in the trait-like interpretation of amygdala response to fearful faces, and even more so with regard to happy and angry faces.

## Conclusions

Findings from the current study suggest that the test-retest reliability of amygdala response to fearful faces is at best fair over a 12-week period. In contrast, FFA reliability was fair to good, irrespective of facial affect. Future studies should evaluate potential differences in amygdala consistency within clinical populations and assess reliability of amygdala response to other types of stimuli frequently utilized in affective and clinical neuroscience. Meanwhile, further assessment of the impact of scan acquisition parameters on reliability within the amygdala and other regions of the brain is necessary.

## References

- Amaral, D. G., & Price, J. L. (1984). Amygdalo-cortical projections in the monkey (*Macaca fascicularis*). *The Journal of Comparative Neurology*, *230*, 465–496. doi: 10.1002/cne.902300402
- Arce, E., Simmons, A. N., Lovero, K. L., Stein, M. B., & Paulus, M. P. (2008). Escitalopram effects on insula and amygdala BOLD activation during emotional processing. *Psychopharmacology*, *196*, 661–672. doi: 10.1016/j.psychres.2010.02.003
- Beauchaine, T. P., Neuhaus, E., Brenner, S. L., & Gatzke-Kopp, L. (2008). Ten good reasons to consider biological processes in prevention and intervention research. *Development and Psychopathology*, *20*, 745–774. doi: 10.1017/S0954579408000369
- Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *Journal of Personality Assessment*, *67*, 588–597. doi: 10.1207/s15327752jpa6703\_13
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory Manual* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Benes, F. M. (2007). Searching for unique endophenotypes for schizophrenia and bipolar disorder within neural circuits and their molecular regulatory mechanisms. *Schizophrenia Bulletin*, *33*, 932–936. doi: 10.1093/schbul/sbm064
- Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*, *1191*, 133–155. doi: 10.1111/j.1749-6632.2010.05446.x
- Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., . . . Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, *17*, 875–887. doi: 10.1016/S0896-6273(00)80219-6
- Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002). *Region of interest analysis using an SPM toolbox* (Abstract). Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan. Retrieved from <http://marsbar.sourceforge.net/about.html# citing-marsbar>
- Chen, N.-K., Dickey, C. C., Yoo, S.-S., Guttman, C. R. G., & Panych, L. P. (2003). Selection of voxel size and slice orientation for fMRI in the presence of susceptibility field gradients: Application to imaging of the amygdala. *NeuroImage*, *19*, 817–825. doi: 10.1016/S1053-8119(03)00091-0
- Cicchetti, D., & Toth, S. L. (2009). The past achievements and future promises of developmental psychopathology: The coming of age of a discipline. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *50*, 16–25.
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*, 127–137.
- Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. Y. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews*, *58*, 57–70. doi: 10.1016/j.brainresrev.2007.10.012
- Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience*, *15*, 353–375. doi: 10.1146/annurev.neuro.15.1.353
- Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*, 13–34.
- Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment*, *10*, 83–89. doi: 10.1037/1040-3590.10.2.83
- Drabant, E. M., McRae, K., Manuck, S. B., Hariri, A. R., & Gross, J. J. (2009). Individual differences in typical reappraisal use predict amygdala and prefrontal responses. *Biological Psychiatry*, *65*, 367–373. doi: 10.1016/j.biopsych.2008.09.007
- Du, Y. P., Dalwani, M., Wylie, K., Claus, E., & Tregellas, J. R. (2007). Reducing susceptibility artifacts in fMRI using volume-selective z-shim compensation. *Magnetic Resonance in Medicine*, *57*, 396–404. doi: 10.1002/mrm.21150
- Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American Journal of Psychiatry*, *164*, 1476–1488. doi: 10.1176/appi.ajp.2007.07030504
- Fakra, E., Salgado-Pineda, P., Delaveau, P., Hariri, A. R., & Blin, O. (2008). Neural bases of different cognitive strategies for facial affect processing in schizophrenia. *Schizophrenia Research*, *100*, 191–205. doi: 10.1016/j.schres.2007.11.040
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured clinical interview for DSM-IV-TR Axis I disorders, research version, non-patient edition (SCID-I/NP)*. New York, NY: Biometrics Research, New York State Psychiatric Institute.
- Fischer, H., Wright, C. I., Whalen, P. J., McInerney, S. C., Shin, L. M., & Rauch, S. L. (2003). Brain habituation during repeated exposure to fearful and neutral faces: A functional MRI study. *Brain Research Bulletin*, *59*, 387–392. doi: 10.1016/S0361-9230(02)00940-1
- Fu, C. H., Steiner, H., & Costafreda, S. G. (2012). Predictive neural biomarkers of clinical response in depression: A meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiology of Disease*, *52*, 75–83. doi: 10.1016/j.nbd.2012.05.008
- George, N., Driver, J., & Dolan, R. J. (2001). Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing. *NeuroImage*, *13*, 1102–1112. doi: 10.1006/nimg.2001.0769
- Glover, G. H., & Law, C. S. (2001). Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magnetic Resonance in Medicine*, *46*, 515–522. doi: 10.1002/mrm.1222
- Goodman, M., New, A. S., Triebwasser, J., Collins, K. A., & Siever, L. (2010). Phenotype, endophenotype, and genotype comparisons between borderline personality disorder and major depressive disorder. *Journal of Personality Disorders*, *24*, 38–59.
- Gorno-Tempini, M. L., Hutton, C., Josephs, O., Deichmann, R., Price, C., & Turner, R. (2002). Echo time dependence of BOLD contrast and susceptibility artifacts. *NeuroImage*, *15*, 136–142. doi: 10.1006/nimg.2001.0967
- Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. *NeuroReport*, *11*, 43–48.
- Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., . . . & Weinberger, D. R. (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science Signaling*, *297*, 400–403.
- Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The amygdala response to emotional stimuli: A comparison of faces and scenes. *NeuroImage*, *17*, 317–323. doi: 10.1006/nimg.2002.1179
- Hasler, G., & Northoff, G. (2011). Discovering imaging endophenotypes for major depression. *Molecular Psychiatry*, *16*, 604–619. doi: 10.1038/mp.2011.23
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*, 748–751. doi: 10.1176/appi.ajp.2010.09091379
- Johnstone, T., Somerville, L. H., Alexander, A. L., Oakes, T. R., Davidson, R. J., Kalin, N. H., & Whalen, P. J. (2005). Stability of amygdala BOLD response to fearful faces over multiple scan sessions. *NeuroImage*, *25*, 1112–1123. doi: 10.1016/j.neuroimage.2004.12.016
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*, 4302–4311.
- Klumpp, H., Angstadt, M., & Phan, K. L. (2012). Insula reactivity and connectivity to anterior cingulate cortex when processing threat in generalized social anxiety disorder. *Biological Psychology*, *89*, 273–276. doi: 10.1016/j.biopsych.2011.10.010
- Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM-III-R): A longitudinal-epidemiological study. *Journal of Abnormal Psychology*, *107*, 216–227. doi: 10.1037/0021-843x.107.2.216
- LaBar, K. S., Gitelman, D. R., Mesulam, M.-M., & Parrish, T. B. (2001). Impact of signal-to-noise on functional MRI of the human amygdala. *NeuroReport*, *12*, 3461–3464.
- Labuschagne, I., Phan, K. L., Wood, A., Angstadt, M., Chua, P., Heinrichs, M., . . . Nathan, P. J. (2010). Oxytocin attenuates amygdala reactivity to fear in generalized social anxiety disorder. *Neuropsychopharmacology*, *35*, 2403–2413. doi: 10.1038/npp.2010.123
- Lemogne, C., Gorwood, P., Boni, C., Pessiglione, M., Lehericy, S., & Fossati, P. (2011). Cognitive appraisal and life stress moderate the



- effects of the 5-HTTLPR polymorphism on amygdala reactivity. *Human Brain Mapping*, 32, 1856–1867. doi: 10.1002/hbm.21150
- Leppänen, J. M. (2006). Emotional information processing in mood disorders: A review of behavioral and neuroimaging findings. *Current Opinion in Psychiatry*, 19, 34–39.
- Lis, E., Greenfield, B., Henry, M., Guilé, J. M., & Dougherty, G. (2007). Neuroimaging and genetics of borderline personality disorder: A review. *Journal of Psychiatry & Neuroscience*, 32, 162–173.
- Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, 19, 1233–1239. doi: 10.1016/s1053-8119(03)00169-1
- Manuck, S. B., Brown, S. M., Forbes, E. E., & Hariri, A. R. (2007). Temporal stability of individual differences in amygdala reactivity. *American Journal of Psychiatry*, 164, 1613–1614. doi: 10.1176/appi.ajp.2007.07040609
- Matthews, P. M., Honey, G. D., & Bullmore, E. T. (2006). Applications of fMRI in translational medicine and clinical practice. *Nature Review Neuroscience*, 7, 732–744. doi: 10.1038/nrn1929
- Matthews, S. C., Strigo, I. A., Simmons, A. N., Yang, T. T., & Paulus, M. P. (2008). Decreased functional coupling of the amygdala and supragenual cingulate is related to increased depression in unmedicated individuals with current major depressive disorder. *Journal of Affective Disorders*, 111, 13–20. doi: 10.1016/j.jad.2008.05.022
- McCloskey, M., Phan, K., & Coccaro, E. (2005). Neuroimaging and personality disorders. *Current Psychiatry Reports*, 7, 65–72. doi: 10.1007/s11920-005-0027-2
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46.
- Merboldt, K.-D., Fransson, P., Bruhn, H., & Frahm, J. (2001). Functional MRI of the human amygdala? *NeuroImage*, 14, 253–257. doi: 10.1006/nimg.2001.0802
- Meyer-Lindenberg, A., Hariri, A. R., Munoz, K. E., Mervis, C. B., Mattay, V. S., Morris, C. A., & Berman, K. F. (2005). Neural correlates of genetically abnormal social cognition in Williams syndrome. *Nature Neuroscience*, 8, 991–993. doi: 10.1038/nn1494
- Monk, C. S., Klein, R. G., Telzer, E. H., Schroth, E. A., Mannuzza, S., Moulton, J. L., III, . . . Ernst, M. (2008). Amygdala and nucleus accumbens activation to emotional facial expressions in children and adolescents at risk for major depression. *American Journal of Psychiatry*, 165, 90–98. doi: 10.1176/appi.ajp.2007.06111917
- Monk, C. S., Telzer, E. H., Mogg, K., Bradley, B. P., Mai, X., Louro, H., . . . Pine, D. S. (2008). Amygdala and ventrolateral prefrontal cortex activation to masked angry faces in children and adolescents with generalized anxiety disorder. *Archives of General Psychiatry*, 65, 568–576. doi: 10.1001/archpsyc.65.5.568
- Morawetz, C., Holz, P., Lange, C., Baudewig, J., Weniger, G., Irle, E., & Dechent, P. (2008). Improved functional mapping of the human amygdala using a standard functional magnetic resonance imaging sequence with simple modifications. *Magnetic Resonance Imaging*, 26, 45–53. doi: 10.1016/j.mri.2007.04.014
- Neugebauer, V., Li, W., Bird, G. C., & Han, J. S. (2004). The amygdala and persistent pain. *The Neuroscientist*, 10, 221–234. doi: 10.1177/1073858403261077
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Paulus, M. P., Feinstein, J. S., Castillo, G., Simmons, A. N., & Stein, M. B. (2005). Dose-dependent decrease of activation in bilateral amygdala and insula by lorazepam during emotion processing. *Archives of General Psychiatry*, 62, 282–288.
- Perkins, D. O., Wyatt, R. J., & Bartko, J. J. (2000). Penny-wise and pound-foolish: The impact of measurement error on sample size requirements in clinical trials. *Biological Psychiatry*, 47, 762–766. doi: 10.1016/s0006-3223(00)00837-4
- Phan, K. L., Angstadt, M., Golden, J., Onyewuanyi, I., Popovska, A., & de Wit, H. (2008). Cannabinoid modulation of amygdala reactivity to social signals of threat in humans. *The Journal of Neuroscience*, 28, 2313–2319.
- Phan, K. L., Coccaro, E. F., Angstadt, M., Kreger, K. J., Mayberg, H. S., Liberzon, I., & Stein, M. B. (2012). Corticolimbic brain reactivity to social signals of threat before and after sertraline treatment in generalized social phobia. *Biological Psychiatry*, 73, 329–336. doi: 10.1016/j.biopsych.2012.10.003
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, 16, 331–348. doi: 10.1006/nimg.2002.1087
- Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., . . . Meyer-Lindenberg, A. (2012). Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *NeuroImage*, 60, 1746–1758. doi: 10.1016/j.neuroimage.2012.01.129
- Prather, A. A., Bogdan, R., & Hariri, A. R. (2013). Impact of sleep quality on amygdala reactivity, negative affect, and perceived stress. *Psychosomatic Medicine*, 75, 350–358. doi: 10.1097/PSY.0b013e31828ef15b
- Pujol, J., Harrison, B. J., Ortiz, H., Deus, J., Soriano-Mas, C., López-Solà, M., . . . Cardoner, N. (2009). Influence of the fusiform gyrus on amygdala response to emotional faces in the non-clinical range of social anxiety. *Psychological Medicine*, 39, 1177–1187. doi: 10.1017/S003329170800500X
- Rauch, S. L., Shin, L. M., & Wright, C. I. (2003). Neuroimaging studies of amygdala function in anxiety disorders. *Annals of the New York Academy of Sciences*, 985, 389–410. doi: 10.1111/j.1749-6632.2003.tb07096.x
- Robinson, J. L., Laird, A. R., Glahn, D. C., Lohvallo, W. R., & Fox, P. T. (2010). Metaanalytic connectivity modeling: Delineating the functional connectivity of the human amygdala. *Human Brain Mapping*, 31, 173–184. doi: 10.1002/hbm.20854
- Robinson, J. L., Monkul, E. S., Tordesillas-Gutiérrez, D., Franklin, C., Bearden, C. E., Fox, P. T., & Glahn, D. C. (2008). Fronto-limbic circuitry in euthymic bipolar disorder: Evidence for prefrontal hyperactivation. *Psychiatry Research: Neuroimaging*, 164, 106–113. doi: 10.1016/j.psychresns.2007.12.004
- Robinson, S., Windischberger, C., Rauscher, A., & Moser, E. (2004). Optimized 3T EPI of the amygdalae. *NeuroImage*, 22, 203–210. doi: 10.1016/j.neuroimage.2003.12.048
- Sabatinelli, D., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2005). Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. *NeuroImage*, 24, 1265–1270. doi: 10.1016/j.neuroimage.2004.12.015
- Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 32, 811–830. doi: 10.1016/j.neubiorev.2007.12.002
- Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, 7, 301–317.
- Singh, I., & Rose, N. (2009). Biomarkers in psychiatry. *Nature*, 460, 202–207. doi: 10.1038/460202a
- Somerville, L. H., Kim, H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biological Psychiatry*, 55, 897–903. doi: 10.1016/j.biopsych.2004.01.007
- Sprinkle, S. D., Lurie, D., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., & Bissada, N. N. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *Journal of Counseling Psychology*, 49, 381–385. doi: 10.1037/0022-0167.49.3.381
- Sripada, C. S., Angstadt, M., McNamara, P., King, A. C., & Phan, K. L. (2011). Effects of alcohol on brain responses to social signals of threat in humans. *NeuroImage*, 55, 371–380. doi: 10.1016/j.neuroimage.2010.11.062
- Sroufe, L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development*, 55, 17–29.
- Stenger, V. A., Boada, F. E., & Noll, D. C. (2000). Three-dimensional tailored RF pulses for the reduction of susceptibility artifacts in T<sub>2</sub>-weighted functional MRI. *Magnetic Resonance in Medicine*, 44, 525–531. doi: 10.1002/1522-2594(200010)44:4<525::aid-mrm5>3.0.co;2-1
- Stenger, V. A., Boada, F. E., & Noll, D. C. (2002). Multishot 3D slice-select tailored RF pulses for MRI. *Magnetic Resonance in Medicine*, 48, 157–165. doi: 10.1002/mrm.10194
- Stenger, V. A., Boada, F. E., & Noll, D. C. (2003). Variable-density spiral 3D tailored RF pulses. *Magnetic Resonance in Medicine*, 50, 1100–1106. doi: 10.1002/mrm.10623
- Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-II in a sample of college students. *Depression and Anxiety*, 19, 187–189. doi: 10.1002/da.20002
- Strakowski, S. M., DelBello, M. P., & Adler, C. M. (2004). The functional neuroanatomy of bipolar disorder: A review of neuroimaging findings. *Molecular Psychiatry*, 10, 105–116.

- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289. doi: 10.1006/nimg.2001.0978
- van den Bulk, B. G., Koolschijn, P. C., Meens, P. H., van Lang, N. D., van der Wee, N. J., Rombouts, S. A., . . . Crone, E. A. (2013). How stable is activation in the amygdala and prefrontal cortex in adolescence? A study of emotional face processing across three measurements. *Developmental Cognitive Neuroscience*, *4*, 65–76. doi: 10.1016/j.dcn.2012.09.005
- Vuilleumier, P., & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia*, *45*, 174–194. doi: 10.1016/j.neuropsychologia.2006.06.003
- Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*, 1271–1278. doi: 10.1038/nn1341
- Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine*, *17*, 101–110. doi: 10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e
- Wang, A. T., Dapretto, M., Hariri, A. R., Sigman, M., & Bookheimer, S. Y. (2004). Neural correlates of facial affect processing in children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*, 481–490. doi: 10.1097/00004583-200404000-00015
- Weng, S.-J., Carrasco, M., Swartz, J. R., Wiggins, J. L., Kurapati, N., Liberzon, I., . . . Monk, C. S. (2011). Neural activation to emotional faces in adolescents with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *52*, 296–305. doi: 10.1111/j.1469-7610.2010.02317.x
- Wright, C. I., Fischer, H., Whalen, P. J., McInerney, S. C., Shin, L. M., & Rauch, S. L. (2001). Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. *NeuroReport*, *12*, 379–383.

(RECEIVED February 19, 2013; ACCEPTED May 29, 2013)

### Supporting Information

Additional supporting information may be found in the online version of this article:

**Table S1:** Within-session reliability for the functionally defined regions of interest at both assessment points for fearful, angry, and happy faces separately.