Guilt-Specific Processing in the Prefrontal Cortex

Ullrich Wagner[1,2,3,4], Karim N'Diaye[1], Thomas Ethofer[1] and Patrik Vuilleumier[1,2]

[1]Department of Neuroscience, University Medical School, University of Geneva, 1211 Geneva, Switzerland, [2]Swiss Center for Affective Sciences, University of Geneva, 1205 Geneva, Switzerland, [3]School of Psychology, University of Bangor, Gwynedd LL57 2AS, UK and [4]Department of Psychiatry and Psychotherapy, Division of Mind and Brain Research, Charité—University Medicine Berlin, 10117 Berlin, Germany

Address correspondence to Dr Ullrich Wagner, Department of Psychiatry and Psychotherapy, Division of Mind and Brain Research, Charité—University Medicine Berlin, Campus Charité Mitte, Charitéplatz 1, D-10117 Berlin, Germany. Email: ullrich.wagner@charite.de.

Guilt is a central moral emotion due to its inherent link to norm violations, thereby affecting both individuals and society. Furthermore, the nature and specificity of guilt is still debated in psychology and philosophy, particularly with regard to the differential involvement of self-referential representations in guilt relative to shame. Here, using functional magnetic resonance imaging (fMRI) in healthy volunteers, we identified specific brain regions associated with guilt by comparison with the 2 most closely related emotions, shame and sadness. To induce high emotional intensity, we used an autobiographical memory paradigm where participants relived during fMRI scanning situations from their own past that were associated with strong feelings of guilt, shame, or sadness. Compared with the control emotions, guilt episodes specifically recruited a region of right orbitofrontal cortex, which was also highly correlated with individual propensity to experience guilt (Trait Guilt). Guilt-specific activity was also observed in the paracingulate dorsomedial prefrontal cortex, a critical "Theory of Mind" region, which overlapped with brain areas of self-referential processing identified in an independent task. These results provide new insights on the unique nature of guilt as a "self-conscious" moral emotion and the neural bases of antisocial disorders characterized by impaired guilt processing.

Keywords: fMRI, guilt, moral, PFC, self-conscious emotions, ToM

Introduction

Human behavior is potently guided by emotional processes. Recent psychological and neuroscientific studies indicate that this may even be true for moral judgment and decision making, traditionally regarded as purely cognitive processes based on rational thinking (Damasio 1994; Greene et al. 2001; Haidt 2001). The same realization has occurred in the field of economics following findings that human economic decisions are not purely rational (as predicted by traditional theories) but also frequently depend on emotional and motivational processing (Sanfey et al. 2003; Camerer and Fehr 2006). However, although recent work in neuroscience has clearly shown that the adherence to moral and social norms is closely linked to emotional processes and despite tremendous advance on the neural bases of basic emotions (such as fear and disgust; e.g., LeDoux 2000; Calder et al. 2001), it remains unknown how the more complex emotions that are crucially implicated in moral and social behavior are represented in the brain.

The most relevant emotion in this context is guilt because it is intimately linked to social and moral norm violations (Kugler and Jones 1992; Wallbott and Scherer 1995; Teroni and Deonna 2008). Elucidating the exact neural circuits implicated in guilt feelings is crucial to better understand the role of guilt-related emotions in moral decisions and moral behavior. However, it is currently unknown which brain regions mediate the self-conscious guilt feelings generated by one's own social norm violations. Clinical studies suggest a specific involvement of the orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (VMPFC) in affective processes guiding social conduct, which might therefore also be more specifically implicated in guilt-related affective processing. Patients with OFC/VMPFC dysfunction (due to developmental brain anomalies or externally caused injuries) are remarkably insensitive to social norms and frequently display patterns of antisocial or psychopathic behavior (Anderson et al. 1999; Blair 2007; Yang and Raine 2009). A recent analysis mathematically modeling performance of these patients during interactive economic games suggests that their reduced sensitivity to social norms and fairness might be best explained by a defective parameter akin to guilt (Krajbich et al. 2009). In fact, the lack of guilt or remorse is one of the most striking characteristics and even a defining feature of psychopathy (Hare 1991; Lykken 1995), possibly representing a causal factor for the disregard of social and moral norms in these individuals. However, brain lesions in such patients always encompass relatively large areas within OFC and VMPFC, such that their deficits generally affect other emotions than guilt as well, depending on the exact extent of damage (Rolls 2004; Zald 2009). Any conclusion about the role of specific prefrontal areas in guilt processing would therefore require demonstrating a selective recruitment when healthy subjects experience guilt feelings but not when they experience other negative emotions, such as shame and sadness, which are less directly connected to decisions of own norm violations (Teroni and Deonna 2008).

In addition, guilt inherently requires the anticipation of thoughts and intentions of other persons (i.e., the victim of one's misconduct; Baumeister et al. 1994), an ability that is the hallmark of "Theory of Mind" (ToM; Vogeley et al. 2001; Gallagher and Frith 2003) and recruits distinct brain areas in dorsomedial prefrontal cortex (DMPFC), together with more posterior regions in superior temporal sulcus (STS) and temporoparietal junction (TPJ; Gallagher and Frith 2003; Saxe et al. 2004; Ciaramidaro et al. 2007). Parts of this network related to ToM might therefore also be implicated in the appropriate processing of guilt feelings.

Here, we use functional magnetic resonance imaging (fMRI) to pinpoint the involvement of specific prefrontal brain areas in guilt- and other-related social emotions. To induce reliable individual guilt feelings, we designed an autobiographical memory paradigm that takes advantage of the fact that intense emotions can efficiently be elicited by reliving strong emotional memories from the individual past (Damasio et al. 2000; Kross

et al. 2009). Prior to fMRI, our participants first specified (in a questionnaire) several events from their past that were accompanied by strong personal guilt feelings and gave some keywords as reminders for this event. The same was done for the control emotions (shame and sadness), as well as a neutral condition. During fMRI, participants were later prompted by their own keywords and asked to relive vividly the emotion experienced during the target event. Unlike other paradigms targeting more evaluative moral processes by asking participants to judge hypothetical scripts of social or moral actions (e.g., Takahashi et al. 2004; Moll et al. 2007; Kedia et al. 2008; Takahashi et al. 2008; Burnett et al. 2009), this procedure allows the induction of a genuine, personally relevant feeling of guilt.

As an additional means to ascertain the specificity of guilt-related activity in the brain, we also determined interindividual differences in Trait Guilt (using the Guilt Inventory; Jones et al. 2000), which measures stable individual propensity to experience guilt in various situations. We predicted that any area in OFC and VMPFC selectively activated by guilt feelings in our group analysis may also parametrically vary in relation to the intensity of Trait Guilt at the individual level. Such finding would support the specific involvement of these areas in guilt, indicating that their role is not only to react generally to emotional events associated with guilt but directly related to the propensity to experience guilt in corresponding situations.

Furthermore, by using shame as 1 of the 2 control emotions, our study also aimed at contributing from a neuroscientific perspective to a fundamental debate in psychology and philosophy concerning the idiosyncratic differences between guilt and shame (Tangney et al. 1996; Teroni and Deonna 2008). Both of these emotions are not only thought to represent prototypes of the "moral" or "self-conscious" emotions (Leary 2007; Tangney et al. 2007) but also appear phenomenologically and functionally very similar. Moreover, guilt and shame typically tend to co-occur in many situations (Eisenberg 2000; Olthof et al. 2000). However, a critical distinction has been proposed between these 2 emotions with respect to the role of self-related representations (Tangney et al. 2007). According to this view, shame is an emotion characterized by a subjective devaluation of the whole self, whereas guilt refers to consequences of one's own behavior that caused damage to another person. That is, although both guilt and shame are regarded as self-conscious in the sense of implying self-awareness of the social and moral impact of own actions, shame has been suspected to entail a stronger self-focus than guilt, whereas the latter would instead rely on a representation of the other (i.e., the victim of own misbehavior) more strongly than shame. Alternatively, because guilt more than shame is related to (morally bad) own decisions for which oneself bears responsibility and is therefore experienced to a greater extent as caused by the self (Wallbott and Scherer 1995; Teroni and Deonna 2008), it may be regarded as the more self-relevant emotion. To address this issue at the neurobiological level, we determined brain regions differentially recruited during self-related processing in each individual participant, by applying a "functional localizer" task (Saxe et al. 2006) of self- versus other-related processing in addition to the emotional induction scanning session. Using these functional networks as inclusive masks for the comparison between emotion conditions, we were able to directly identify any overlap of activation in shame-specific and guilt-specific networks with those brain areas recruited by either self-referential or other-referential processing.

Materials and Methods

Participants
Eighteen healthy female participants (25–30 years) without any history of psychiatric or neurological disorders participated in the study. Three participants were excluded from analysis due to data loss resulting from technical problems during fMRI scanning. The study was approved by the local ethics committee at the University of Geneva, and all participants gave informed written consent prior to participation.

Prescanning Questionnaire
About 2–3 weeks prior to scanning, participants filled in a questionnaire to specify events from their past that were associated with strong feelings of guilt, shame, and sadness (2 events of each type). To avoid remote childhood experiences, instructions stated that all events should have occurred after the age of 16-year-old. Importantly, participants were told that they should remember different "emotional events" from their life, but the labels of "guilt," "shame," or "sadness" were not explicitly mentioned in these instructions. Instead, we gave broad 3-sentence descriptions of situations in which one of the target emotions (guilt, shame, sadness) typically occurs and then required the participants to remember 2 events corresponding to each of these situations that were highly emotional. These descriptions (for details, see Supplementary Material) were chosen on the basis of theoretical considerations (Teroni and Deonna 2008) and behavioral pilot testing that confirmed that each situation description induced the intended target emotion more than other emotions.

After retrieving a specific event corresponding to a situation description, participants rated on a list of emotion words (including anger, disgust, fear, guilt, happiness, pride, relief, sadness, shame, surprise) how strongly they had felt each of these emotions during this event (on a scale from 0 to 10). To guarantee privacy, no information about the content of the specific event had to be given, but participants provided a few keywords to be later used as a reminder for each event in the fMRI session. (In 3 participants, questionnaire data indicated that one event remembered for the sadness situation description led to higher shame rating than one event remembered for the shame situation description, and vice versa. In these cases, the target emotions were exchanged for the respective event pair in the subsequent fMRI session in order to yield the strongest possible induction of each target emotion during scanning in the respective condition.)

For all events (emotional and neutral), participants provided some general context information and additional keywords, which were later used as reminder cues during the fMRI session (for details, see Supplementary Material).

Experimental Procedure during fMRI Scanning
Immediately before scanning, participants were presented again with their own reminder cues from the questionnaire (i.e., context information and keywords for each event) and had to confirm that reading the respective cues would remind them of the corresponding event that had been rated in the questionnaire. This allowed us to ensure that the personal event information provided in the previous interview session could indeed work as an efficient reminder for all target events, even 2–3 weeks after filling in the questionnaire.

Within the scanner, 2 runs were performed, during which participants had to remember and mentally relive the emotions from all the 12 events that they had specified in the questionnaire (2 guilt events, 2 shame events, 2 sadness events, 2 neutral events, and 4 positive filler events). Within a run, the 12 events were relived in a pseudorandom order that was predetermined according to the constraint that there was always a positive filler event or a neutral event between 2 negative events, and that events referring to the same target emotion (guilt, shame, sadness) were separated by at least 4 other intervening events.

The time line for each trial is depicted in Figure 1. Each trial began with a slide showing for 2.5 s the target emotion label, written in upper case letters in the center of the screen. (In contrast to the prescanning questionnaire, it was not necessary to avoid explicit reference to emotion terms in this phase because individual emotion ratings for each event had already been obtained. Participants were told that we
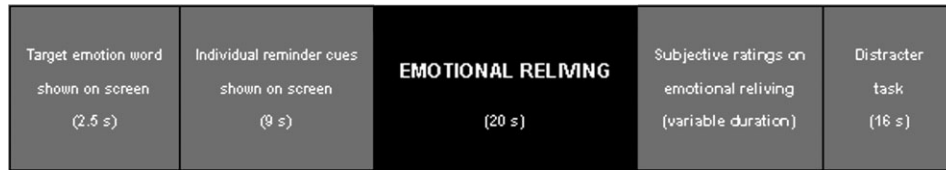
Figure 1. Time line for each trial during fMRI scanning. Our analyses compared emotion conditions (guilt, shame, sadness, neutral) in the 20 s interval of reliving (black background). During these intervals, visual input from the monitor was exactly the same for all conditions, so that any difference in brain activation between conditions was entirely determined by differential processes of memory-induced emotional experience. Note, width of sections is not proportional to duration of respective intervals.

had selected one of their strongest emotions for each event from their questionnaire, and that they should focus on this specific emotion when reliving the event in the scanner. This procedure was chosen to further ensure that the target emotion was relived as strongly as possible during scanning.) Following the target emotion word, all reminder cues of the corresponding event were presented for 9 s on a single slide, followed by a 20-s reliving phase, during which a slide was shown on the screen with the instruction "Relive the memory and in particular the emotion."

After this reliving interval, participants judged the vividness of the memory and intensity of the target emotion during reliving (on a 4-point scale, by pressing 1 of 4 keys on a response box held in their right hand, according to verbal instruction shown on the screen: "very low" = key1, "low" = key2, "high" = key3, "very high" = key4). An additional rating screen asked whether the participant had been able to maintain the target emotion during the whole reliving interval, and if not, for how much time from the beginning of the interval they actually maintained the emotion (according to the following instruction shown on the screen: "0–25%" = key1, "25–50%" = key2, "50–75%" = key3, "75–100%" = key4). This information was subsequently used to model the duration of reliving individually in the fMRI analysis (see below). All these ratings after the reliving phase were given in a self-paced manner but with a maximum of 9 s for intensity and vividness and 15 s for the questions on reliving duration estimates, which was sufficient time to exclude occurrence of missing answers in all participants.

After the ratings, a fixation cross was shown for 3 s, followed by a simple number detection task used as a cognitively and emotionally undemanding baseline task, which also served as a distracter task to clear the participant's mind before the next trial began. In this task, 5 single digits randomly chosen from 1 to 9 were presented successively at a 2-s pace in the middle of the screen, and participants had to press a key whenever the digit "3" appeared. At the end of the task, a fixation cross was shown again for 3 s at the screen center to announce the beginning of the next trial.

Participants received detailed instructions about the procedure and the successive intervals of each trial before scanning started. Instructions emphasized that the 20-s reliving phase was the most critical one, and that they should focus specifically on the target emotion experienced during this event. To additionally familiarize them with the exact timing of the procedure, an initial practice trial was performed that did not refer to their personal events from the questionnaire (using the terror attacks of 9/11 as an event, with the target emotion "Fear," and as reminder cues: "World Trade Center, New York"; "11 September 2001"; "terrorists, victims"; "airplane," "skyscraper," "impact," "fire").

Self-Referential Versus Other-Referential Task

In a separate fMRI run, participants performed an additional "localizer" task (Saxe et al. 2006) that served to determine individual brain regions devoted to self- versus other-referential processing. This task was derived from a well-established experimental paradigm that allows a comparison of patterns of brain activity associated with self-related versus other-related representations, previously used in several neuroimaging studies investigating the neurobiological underpinnings of self-referential processing (Craik et al. 1999; Kelley et al. 2002; Macrae et al. 2004). Participants read a variety of trait adjectives (taken from a standard adjective list; Anderson 1968) and had to indicate for each of them, in 3 separate conditions, either how well it described themselves (="self" condition); how well it described their best friend (="other"

condition); or how many syllables the adjective contained (=non-personal control condition). Each task condition was given in blocks of 20-s duration, with 5 adjectives presented successively in each block. There were 30 blocks altogether (10 blocks per condition), with random order of conditions. (Before the start of each block, the cue word "ME," "FRIEND," or "SYLLABLES" was shown for 3 s in the middle of the screen to announce the condition for the block to the participant.) For the purpose of the present study, the 2 critical contrasts of self > other and other > self were used to create, respectively, a "self-related mask" and an "other-related mask," which were then used to determine brain regions within emotion-specific contrasts that overlapped with self- versus other-referential processing (see below: "MRI acquisition and analysis"). One of the 15 participants did not perform this task and was therefore not included in the mask contrasts.

Trait Guilt Questionnaire

After fMRI scanning, participants filled in the "Trait Guilt" scale of the "Guilt Inventory" (Jones et al. 2000), which assess individual propensity to experience guilt in various situations. This personality characteristic was used to additionally specify guilt-specific activation on the basis of stable interindividual personality differences.

MRI Acquisition and Analysis

MRI data were acquired on a 3 T whole-body scanner (Siemens TRIO), using standard head–coil configuration. For each participant, a structural image was obtained with a T_1-weighted sequence (3D-GR/IR, repetition time [TR] = 2300 ms, echo time [TE] = 2.89 ms, flip angle = 9°). Functional images, covering the whole brain, were obtained with a T_2-weighted echo-planar imaging sequence (2D-EP, TR = 2200 ms, TE = 30 ms, flip angle = 85°, voxel size = 2 × 2 × 2 mm³). For correction of image distortions, a fieldmap (36 slices, slice thickness 3 mm + 1 mm gap, TR = 400 ms, TE [1] = 5.19 ms, TE [2] = 7.65 ms, flip angle = 60°, voxel size = 3 × 3 × 4 mm³) was acquired prior to the experimental runs.

Images were analyzed with statistical parametric mapping software SPM5 (Wellcome Department of Imaging Neuroscience; www.fil.ion.ucl.ac.uk/spm). Image preprocessing comprised realignment, unwarping (Andersson et al. 2001), coregistration and normalization into Montreal Neurological Institute (MNI) stereotaxic space (Collins et al. 1994), and smoothing with an 8 mm full-width at half-maximum Gaussian kernel. A high-pass frequency filter (cutoff 128 s) and correction for autocorrelation between scans were applied to the time series.

Statistical analysis was performed using the general linear model implemented in SPM5, with a canonical hemodynamic response function convolved with each modeled event. For the main experiment, separate regressors for each emotion category (including the neutral condition as a category) during the 20-s reliving periods and an additional regressor for the 10-s distracter task (number detection) between the reliving phases were defined. The duration of each reliving intervals was individually modeled according to the participants' ratings of how long they could maintain the target emotion during this interval (see above). Although these ratings were generally high (average 90.0 ± 2.8% of the interval duration [corresponding to 18.0 ± 0.6 s duration out of 20 s], without differences between emotion conditions, $P > 0.69$), this individualized duration modeling was employed to increase test power because any fading or disruption of the target emotion in this phase

would blur the main process of interest (i.e., the specific emotion feeling). For the self- versus other-processing task, a standard block design was applied, with 3 separate regressors modeling each condition (self, other, and syllable processing), with a fixed duration of 20 s for each block.

Statistical parametrical maps of blood oxygen level–dependent signal changes were generated from linear contrasts between the different conditions in each participant. Each emotion type (guilt, shame, sadness) was contrasted against the neutral condition. Furthermore, to determine brain regions of guilt-specific processing, the guilt condition was also directly contrasted with the control emotions shame and sadness. A second-level random effect analysis was then performed for the entire group, using one-sample t-tests for each comparison of interest across the whole brain. Unless stated otherwise, the standard threshold criterion of significant activation at a voxel level of $P = 0.001$ or smaller (uncorrected) was applied, with a cluster size of at least 10 voxels (Worsley et al. 1996). In addition, the critical regions in the OFC and DMPFC identified as guilt-specific in the contrast of guilt against the 2 control emotions, survived correction for multiple comparisons when small volume correction (SVC) was applied with regard to activation peaks reported in previous social-emotional and tactical processing in relation to actual own interpersonal behavior (Eisenberger et al. 2003; Fukui et al. 2006; 10 mm sphere, $P < 0.05$, family-wise error correction). To specify commonalities between emotion-specific processing and self-referential and other-referential processing, the emotion contrasts were additionally overlaid with a self-related mask or an other-related mask obtained from the self > other and the other > self contrast, respectively (thresholded at $P < 0.05$ uncorrected, inclusive masking; Ritchey et al. 2008; Pourtois et al. 2009). Furthermore, brain activation related to interindividual differences in Trait Guilt was modeled by including individual scores on this scale (Jones et al. 2000) as a parametric regressor into the second-level analysis for the relevant contrasts.

Results

Behavioral Results

Mean emotion ratings for the different events reported in the prescanning questionnaire (targeting guilt, shame, and sadness, respectively) are shown in Table 1. These data confirm that each of the 3 situation types strongly elicited the corresponding target emotion (means ± standard error of the mean [SEM] on a scale from 0 to 10: for "guilt in 'guilt' situations" 8.1 ± 0.3; for "shame in 'shame' situations" 7.8 ± 0.3; and for "sadness in 'sadness' situations" 7.7 ± 0.6). In addition, for all situations types, the subjective strength of the target emotion was significantly higher in comparison to any other emotion felt in the same events and in comparison to the strength of the same emotion in the other situation types (all $P < 0.05$; see Table 1).

Table 1

Mean ratings (±SEM) of emotions for 3 different emotion conditions (situations relived from personal autobiographical memories)

Emotion	Guilt situation	Shame situation	Sadness situation
Anger	4.3 ± 0.7	4.6 ± 0.6	4.8 ± 0.7
Disgust	2.6 ± 0.7	3.0 ± 0.8	1.9 ± 0.5
Fear	4.4 ± 0.6	5.5 ± 0.6	3.6 ± 0.8
Guilt	**8.1 ± 0.3**[*]	6.0 ± 0.6	3.0 ± 0.7
Happiness	0.2 ± 0.2	0.2 ± 0.2	0.1 ± 0.1
Pride	0.4 ± 0.3	0.5 ± 0.3	0.2 ± 0.1
Relief	0.8 ± 0.4	1.0 ± 0.5	0.6 ± 0.3
Sadness	4.6 ± 0.6	3.5 ± 0.7	**7.7 ± 0.6**[*]
Shame	5.9 ± 0.5	**7.8 ± 0.3**[*]	1.7 ± 0.6
Surprise	1.9 ± 0.5	2.2 ± 0.7	3.2 ± 0.7

Note: Rating scale ranging from 0 to 10. Target emotions are in bold.

[*]$P < 0.05$, for pairwise comparisons with all other values in the same column and in the same row.

These ratings for target emotions in each situation were also confirmed by the subsequent ratings of vividness and intensity of reliving in the scanner (see Materials and Methods), which were generally evaluated as high (means ± SEM on a scale from 1 to 4, for vividness: guilt 3.1 ± 0.1, shame 3.0 ± 0.1, sadness 3.2 ± 0.1 and for intensity: guilt 3.0 ± 0.2, shame 2.8 ± 0.2, sadness 3.2 ± 0.1), without significant differences between the emotions (all $P > 0.10$). Regarding nontarget emotions additionally rated in the prescanning questionnaire, the 3 situation types did not differ, except for fear, which was judged as stronger in shame situations than in guilt or sadness ($P < 0.05$).

We also analyzed the time points when each event had occurred. This analysis showed that all emotional events had occurred ~3 to 4 years before experimental testing, with no significant difference between the 3 critical emotion conditions (means ± SEM: guilt 3.5 ± 0.7 years, shame 3.2 ± 0.7 years, sadness 4.2 ± 0.7 years; $P > 0.29$). However, neutral events were generally more recent than emotional events (often occurring a few days or weeks before the study), consistent with our instructions that the corresponding memories had to refer to specific events and be as vivid as emotional events. Thus, vividness ratings for neutral events in the scanner confirmed a high vividness (3.2 ± 0.1), which did not significantly differ from the vividness of emotional events ($P > 0.21$).

fMRI Results

Separate Contrasts of Guilt and Control Emotions against the Neutral Condition

Although the direct comparison between guilt and control emotions was the primary topic of the present study, we first contrasted for explorative purposes each of the 3 target emotions guilt, shame, and sadness, separately against the neutral condition, allowing us to identify commonalities between the 3 emotion conditions (Supplementary Table S1). Consistent with task requirements, all 3 contrasts revealed a common network (statistically confirmed by conjunction analysis performed on these 3 contrasts) comprising brain regions critically implicated in emotional and social processing, as well as memory retrieval and mental imagery, including in particular the right and left retrosplenial cortex (extending into posterior cingulate cortex [PCC] and precuneus) but also left anterior insula, bilateral temporal poles, and lingual gyri (medial occipital cortex), as well as the cerebellum. Notably, this shared network also included the TPJ and anterior STS, suggesting a common involvement of these components of the ToM system outside the prefrontal cortex (PFC) across all 3 emotions.

Apart from these areas similarly recruited by all 3 emotions, guilt and sadness (but not shame) showed additional activations in several other brain regions in comparison to the neutral condition, mostly in prefrontal and temporal areas (for details, see Supplementary Table S1).

No brain regions were more strongly activated in the neutral than in any emotional condition (contrasts neutral > guilt, neutral > shame, and neutral > sadness).

Direct Contrasts between Guilt and Control Emotions

With regard to our main goal, that is, to identify brain regions specifically recruited when participants feel guilt as compared with other closely related negative emotions, the most critical test was the direct contrast between guilt and the 2 control

emotions (Guilt > Shame + Sadness). This comparison revealed guilt-specific activation in 2 prefrontal areas, namely the right lateral OFC (*xyz* peak 36/32/–4) and the left paracingulate region of the DMPFC (peak –10/42/34; Table 2).

Extraction of parameter estimates of activity (beta values) from the cluster peak in the OFC showed that it was indeed strongly selective in its responsiveness to guilt, with no signal change in the other conditions (Fig. 2*A*; see also Supplementary Table S2, for all separate pairwise contrasts between emotion conditions). Furthermore, guilt specificity of this region was additionally confirmed by parametric analyses of activation patterns between subjects, testing for any proportional increase in this contrast in relation to the individual Trait Guilt scores obtained from each participant (across the whole brain). Again, the same right lateral OFC area was found as the only brain region that was parametrically correlated with the degree of individual Trait Guilt (peak 30/32/–10; Fig. 2*B*).

In contrast to the lateral OFC, the paracingulate DMPFC did not show such modulation by Trait Guilt (*P* = 0.43). Furthermore, beta estimates and direct pairwise contrasts between emotions for the paracingulate DMPFC cluster showed that activation here was primarily driven by stronger guilt-related recruitment of this region in comparison to shame rather than in comparison to sadness (Fig. 3; Supplementary Table S2).

Emotion Contrasts Overlaid with Self- and Other-Related Masks

Consistent with previous work using the same or similar paradigms (Craik et al. 1999; Kelley et al. 2002; Macrae et al. 2004), the self-related mask obtained from the separate self-referential versus other-referential task (see Materials and Methods) basically covered the DMPFC, rostral and dorsal anterior cingulate cortex (ACC) anterior medial frontal gyrus, anterior insula, precentral gyrus, postcentral gyrus (somatosensory cortex), PCC, precuneus, medial occipital gyrus, lateral occipital gyrus, anterior mesencephalon (all bilateral), and left thalamus, whereas the other-related mask basically included the bilateral retrosplenial cortex, medial OFC, right inferior frontal gyrus, bilateral superior frontal sulcus, subgenual ACC, parts of dorsal ACC, left TPJ/angular gyrus, bilateral anterior STS, bilateral perirhinal cortex, and right dorsal amygdala. These masks were overlaid on the above-mentioned networks activated by the different conditions in the emotion reliving task, allowing us to determine whether the regions recruited by each of these emotions (i.e., guilt, shame, sadness) were also related to self-referential processing or to other-referential processing (indicated by an "S" label or "O" label, respectively, in Table 2 and Supplementary Tables S1 and S2), or unrelated to self versus other processing. Correlation analyses of beta values

extracted in each subject for the self-condition and for the guilt condition (with reference to the baseline) confirmed a positive association between guilt and self-referential processing in this area (*r* = 0.54, *P* < 0.05).

Only 2 regions involved in self-related processing were shared by all 3 emotions (compared with the neutral condition), namely the left anterior insula and the medial frontal pole (Supplementary Table S1), consistent with a more general role of these regions in self-awareness of emotional and pain-related processing (Price 2000; Craig 2003; Gilbert et al. 2006). Most interestingly, an extended DMPFC area recruited by self-referential processing was found to overlap with the paracingulate DMPFC region that showed a selective activation to guilt in the direct contrast between guilt and the 2 control emotions, shame and sadness (Table 2).

In contrast to areas within the self-related network, there was no region associated with other-referential processing that was not shared by all 3 emotions (Supplementary Table S1). These shared regions included the left TPJ/angular gyrus and anterior STS, that is, 2 critical components of the ToM network (Gallagher and Frith 2003; Saxe 2006; Bedny et al. 2009), as well as the retrosplenial cortex/PCC and lingual gyrus, generally involved in emotional memory retrieval and mental imagery (Maddock 1999; Maratos et al. 2001; Vann et al. 2009; Burianova et al. 2010). In addition, when specifically compared with shame, activation by guilt was found to overlap with 2 areas that were selective for other-related processing, namely, the right dorsolateral prefrontal cortex/anterior superior frontal gyrus and the right amygdala, a brain structure implicated in a variety of processes of emotional and social evaluation (Adolphs et al. 1998; LeDoux 2000; Zald 2003). Conversely, shame evoked no specific increases compared with guilt, neither in self- nor in other-related networks (Supplementary Table S2).

Discussion

The present study was designed to identify the neural substrates of guilt feelings in healthy individuals and to determine the specificity of guilt relative to other negative or self-conscious emotions. By using an autobiographical memory paradigm, in which participants were prompted by private keywords to relive highly emotional experiences from their past (Damasio et al. 2000; Cabeza and St Jacques 2007; Kross et al. 2009), we were able to induce strong personal guilt feelings during fMRI. The closely related emotions shame and sadness, likewise successfully elicited by the autobiographical memory procedure, served as critical control conditions to identify guilt-specific activity in the brain.

Consistent with our prediction, the results demonstrate a crucial involvement of the OFC in guilt-related emotional processing. Specifically, a right lateral area in OFC was activated by guilt in comparison to both sadness and shame and therefore appears to selectively mediate those processes that are inherent to guilt but not the other closely related emotions shame and sadness. Furthermore, the same region was also parametrically activated in relation to individual scores of Trait Guilt, as measured in a separate personality questionnaire. That is, the more participants reported being prone to experience guilt in everyday life, the more they recruited the right OFC in situations conceived to elicit guilt compared with other emotional situations. Thus, a specific role of the right lateral

Table 2
Guilt-specific regions identified by direct contrast with control emotions

Anatomical definition	BA	Hem.	S/O	MNI coordinates	*t* value	Cluster size
Guilt > Shame + Sadness						
Lateral OFC	47	R		36 32 −4	4.38	12
DMPFC/paracingulate cortex	9/32	L	S	−10 42 34	4.63	10

Note: S = included in self-related mask, O = included in other-related mask. No brain regions were activated in the opposite contrast (Shame + Sadness > Guilt).
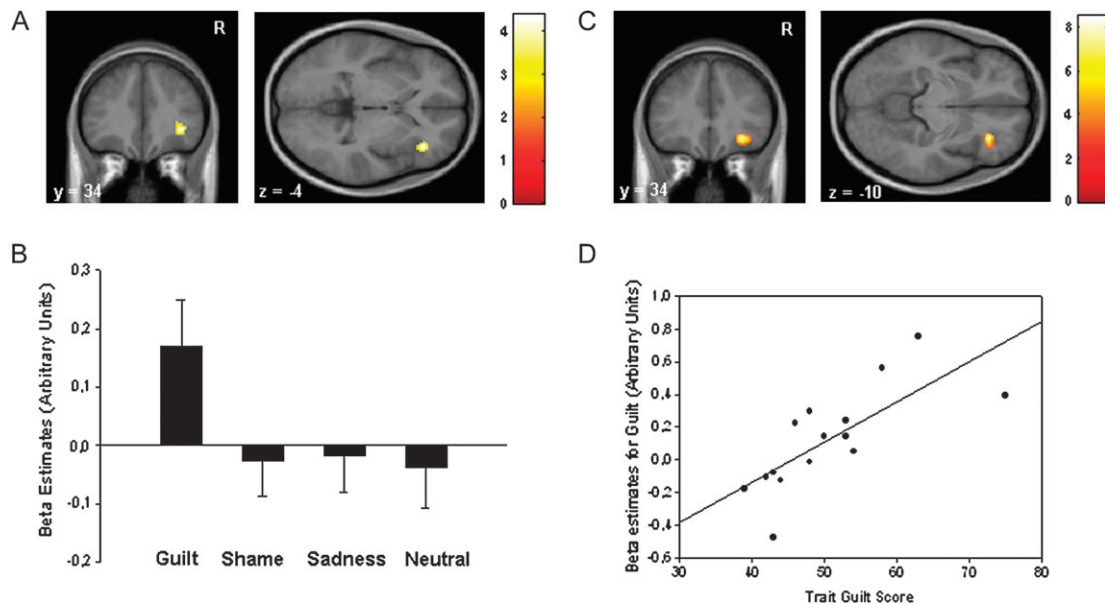*P* < 0.001, uncorrected; cluster size *k* ≥ 10. BA, Brodmann area; Hem., Hemisphere.

Figure 2. Guilt-specific processing in right OFC. (*A*) Brain activation for guilt compared with the 2 control emotions, shame and sadness (Guilt > Shame + Sadness). (*B*) Parameter estimates of activation (betas) extracted from the OFC cluster peak (36/32/−4) for all experimental conditions. Analysis of variance indicates significantly higher activation for guilt in comparison to all other conditions ($P < 0.05$). Shame, sadness, and neutral conditions did not differ from each other ($P > 0.90$). (*C*) Additional parametric whole-brain analysis correlating the individual propensity to experience guilt (Trait Guilt) in the contrast Guilt > Shame + Sadness across subjects. This analysis independently reveals guilt-specific processing in the same right lateral OFC region, now as a function of interindividual differences ($t = 8.56$, $P < 0.0001$). No additional region was parametrically modulated by Trait Guilt in this whole-brain analysis. (*D*) Graphical depiction of the relationship between individual Trait Guilt scores obtained from each subject and the corresponding extent of OFC activation in the guilt condition (beta estimates at peak voxel).

OFC in guilt experience was confirmed by both the overall group analysis and the correlation analysis of dispositional differences between individuals. A second region in the DMPFC/paracingulate cortex was also activated by guilt compared with the control emotions but was not correlated with the dispositional measure of Trait Guilt. In fact, apart from the right lateral OFC, no other region in the whole-brain analysis showed any correlation with Trait Guilt. Importantly, these results cannot be explained by general differences in emotional strength because vividness and intensity ratings did not differ between the 3 emotions.

Our finding of a crucial function of the lateral OFC in experiencing guilt converges with other observations suggesting that this region is particularly involved in negative emotional processing, unlike the more medial parts of OFC that preferentially relate to positive, reward-related affect (O'Doherty et al. 2001). However, the lateral OFC appears to encode not simply negative valence in general but more specifically the negative affect associated with particular social contexts or expected outcomes. For example, this region was found to be activated when participants experience social rejection (Eisenberger et al. 2003), and results from game-theoretical paradigms suggest that it may be responsible for negative feeling states that determine social decision making (Rilling et al. 2008). In the context of emotional processing, lateral frontal areas including lateral OFC have also been described as regions involved in inhibitory control or suppression of emotions (Beauregard et al. 2001; Ochsner et al. 2004). However, it is very unlikely that emotion suppression could have played any substantial role here, since emotional ratings and direct contrasts between emotions showed no evidence for reduced emotional responses in guilt as compared with other conditions.

Thus, the guilt-specific activation of lateral OFC suggests a regulatory process that is inherent to this emotion, presumably related to the control of behavior, which is necessary to anticipate and compensate for the harm inflicted to another person due to wrongdoing. Consistent with this interpretation, a neuroimaging study by Windmann et al. (2006) found that lateral OFC is also specifically activated when behavioral changes are required to maximize long-term benefits. Thus, in line with the theoretical claim that guilt primarily serves to maintain interpersonal relationships (Baumeister et al. 1994), our results suggest that such control processes may be an integral part of guilt feelings. Although inhibition is usually regarded as a deliberate and effortful activity, behavioral control by the lateral OFC could be more automatically activated as a central component of the emotional experience of guilt, serving to inhibit transgressions of social norms and/or anticipate their negative outcome (for a similar account of guilt based on developmental data, see Kochanska et al. 2009).

Our results provide an important missing link in the clinical findings in patients with OFC lesion or dysfunction. These patients exhibit striking abnormalities in social judgment and behavior (Damasio 1994; Stone et al. 2002; Beer et al. 2006), and formal mathematical models of their performance during economic games point to a deficit in guilt-related signals (Krajbich et al. 2009). However, the latter data alone cannot establish a selective role for the OFC in guilt feelings because brain lesions are seldom restricted to the OFC, and a variety of other behavioral abnormalities unrelated to guilt processing are typically present in these patients (Bechara et al. 2000; Rolls 2004). Against this background, our data suggest that certain antisocial features observed after OFC lesion or dysfunction may specifically arise from an impairment of normal guilt experience. A disturbed sense of guilt could be causally linked to an
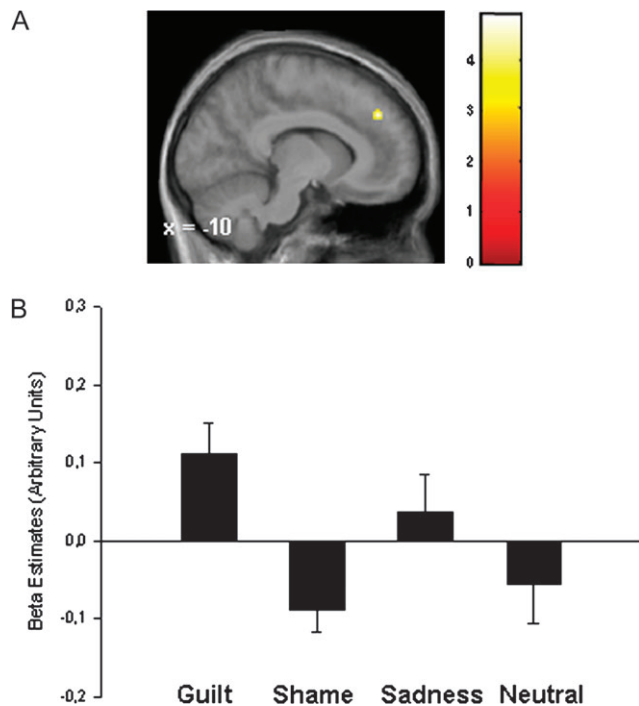
A

B

Figure 3. Guilt-specific processing in the paracingulate region of the DMPFC. (*A*) Activation for guilt compared with the 2 control emotions, shame and sadness (Guilt > Shame + Sadness). (*B*) Parameter estimates of activation (betas) extracted from the paracingulate DMPFC cluster peak (−10/42/34) for all experimental conditions. Analysis of variance indicates significantly higher activation for guilt in comparison to shame and the neutral condition ($P < 0.01$). Sadness produced intermediate effects between neutral and guilt conditions, differing from both by trend only ($P = 0.10$ and $P = 0.12$, respectively). This region also overlapped with areas recruited during self-referential processing, as identified by a separate functional localizer task (see text).

insensitivity to own norm violations, thus leading to social conflict and transgressions. Our data also support the notion that it is indeed the OFC proper rather than the VMPFC that is specifically involved in the affective experience of guilt, a conclusion that would be difficult to draw from clinical studies alone because prefrontal lesions or anomalies in many cases extend beyond OFC into more ventrally and medially neighboring areas. Moreover, although we also found evidence for guilt-associated activation in the VMPFC/rostral ACC, this was similarly observed for sadness, suggesting a more general role for the VMPFC in social-affective processing than for the OFC proper.

Apart from the OFC, the only other brain area showing guilt-specific activity was the paracingulate region of the DMPFC, although in this case independent of the magnitude of individual propensity to this emotion. As this region is known to represent the primary prefrontal component of the ToM network (Walter et al. 2004; Saxe 2006), this finding is consistent with our hypothesis that guilt should recruit regions within the ToM network more strongly than the control emotions. Due to its direct link to social transgression causing harm to the other(s), guilt may be inherently associated with reflecting and understanding other people's thoughts. Interestingly, however, other areas associated with ToM (TPJ, STS) were similarly activated by guilt, shame, and sadness in comparison with the neutral condition, probably reflecting the common occurrence of these emotions in interpersonal contexts that require monitoring others' thoughts and beliefs. In addition, only the paracingulate cortex overlapped with self-related processing,

whereas the other ToM regions commonly activated by all emotions (TPJ, STS) overlapped with other-related processing. Other authors have similarly described the DMPFC as a key region where self-referential processing and perspective-taking interact (D'Argembeau et al. 2007) or where self-relevance in interpersonal contexts is represented (Schilbach et al. 2006). These findings therefore add to previous attempts to disentangle the differential contributions of subregions within the ToM network to different psychological processes involved in ToM capabilities (Saxe 2006; Ciaramidaro et al. 2007; Hampton et al. 2008; Jenkins and Mitchell 2010), suggesting that the temporal and parietal parts of this network may represent the mental states, attributes, and/or intentions of others, while the frontal part in the paracingulate cortex may connect these representations with those related to the self, consistent with evidence for a particular involvement of this area in social tactics (Fukui et al. 2006). This interpretation is in line with the recent proposal by Saxe (2006) that the TPJ supports the human ability to reason about the content of mental states, while the DMPFC is involved when the self must coordinate his/her own current goal or focus of attention with another person. Such integration between one's assumption about others' thoughts and one's own goals is of particular relevance in guilt feelings, where wrong actions of the self are harmful or damaging to another person. If the DMPFC mediates the representation of wrongdoings in relation to the relevance of inflicted harm for the self or the other, an important question to address in future studies would be whether its activation depends on personal values or ideals endorsed by the self (more than on conventional societal norms). Consistent with this idea, we found that for the paracingulate DMPFC area, unlike lateral OFC, guilt specificity was mainly driven by higher activation to guilt situations in comparison to shame but with an intermediate activation to sadness. As sadness is typically most strongly felt after the loss of a personally valued person, while shame typically occurs as a consequence of a conflict with societal conventions, these data indeed suggest that the DMPFC is involved in social–emotional processing to the extent that self-relevant ideals are affected.

One major advantage of the autobiographical memory paradigm used here is that it allowed us to induce strong individual guilt feelings genuinely linked to own norm transgressions that had actually occurred. Although emotions were induced indirectly by the reliving of respective affective experiences from the past, without referring to the specific emotion's name, the subjective ratings clearly indicate that this method of inducing the target emotions during scanning was highly efficient. In this way, our study goes beyond previous neuroimaging experiments that focused primarily on judgmental rather than affective aspects of moral processing, including guilt-related processing, by presenting subjects during scanning with scripts of hypothetical scenarios of prototypical social or moral transgressions (Takahashi et al. 2004; Moll et al. 2007; Kedia et al. 2008; Burnett et al. 2009). These studies found activations in several brain regions associated with social cognition, ToM, and emotional processing, generally shared with other moral conditions, but critically, they did not report an involvement of the 2 specific prefrontal regions related to guilt here, confirming our interpretation that these regions are linked to an individual affective experience of guilt that probably would not be induced in sufficient intensity by reading hypothetical scenarios.

An additional novel aspect of our study is that we aimed to determine the specificity of guilt-related affective processing by comparison with control emotions that are closely related to guilt but less strongly linked to own norm violations. In this way, our data critically extend the results from the only previous neuroimaging study that used an autobiographical memory paradigm to elicit guilt feelings during positron emission tomography scanning but without any comparison with other control emotions (Shin et al. 2000). These authors showed a predominant activation in anterior insula and temporal poles for guilt in comparison to a neutral condition but due to the close relatedness of guilt with shame and sadness, reliving guilt memories in their study also induced strong feelings of shame and sadness (as confirmed by emotional ratings in their own study as well as in our study), so that this pattern of brain activation for guilt relative to the neutral condition could also reflect activation elicited by these other emotions. Our imaging data clearly support this conclusion. When we compared guilt with the neutral condition, we also found, among others, activation in anterior insula and the temporal poles, as reported by Shin et al. (2000). However, these activations were also found when we compared shame or sadness with the neutral condition but not in the direct contrast of guilt versus shame and sadness. Thus, consistent with other work (Singer et al. 2004; Olson et al. 2007; Zahn et al. 2007), our results confirm that the anterior insula and temporal pole are critically involved in social-affective processing but do not indicate any specificity of guilt-related processing for these regions.

Another theoretical issue concerning guilt specificity, to which the present study contributes for the first time from a neuroscientific perspective, pertains to a current debate within psychology and philosophy as to what uniquely distinguishes the 2 emotions guilt and shame from each other (Tangney et al. 1996; Olthof et al. 2000; Teroni and Deonna 2008). Because a key aspect of this debate focuses on the notion of a differential involvement of self- versus other-referential processing in guilt and shame, we tested whether emotional circuits recruited by guilt versus shame showed distinct patterns of overlap with self- or other-related representations in the brain. Although neuroscientific findings cannot directly answer such theoretical questions concerning the nature of guilt and shame, they can contribute to the debate by providing an additional set of information pertinent to the issue. On the basis of the predominant theoretical assumption that self-related representations may be more strongly engaged in shame than in guilt (Tangney et al. 2007), less overlap with brain networks activated by self-related processing for feelings of guilt than for feelings of shame might have been expected, but we found no support for this expectation. In fact, while both guilt and shame, when compared with the neutral condition, shared activations in a number of other-related areas (including TPJ and STS within the ToM network), several of the brain regions that were activated in the direct comparison between guilt versus shame overlapped with self-related processing (e.g., rostral ACC and anterior insula) or were unrelated to the self/other distinction.

There was likewise no evidence in the opposite contrast that shame relies more than guilt on circuits of self-related processing. In fact, no single brain region was more strongly activated in shame than guilt. Even in comparison to the neutral condition, all brain areas activated by shame were also activated by guilt. We cannot entirely exclude that the latter findings reflected limited power or that self-related representations

engaged by shame involve different brain areas than those activated during our "self-localizer" task. Nevertheless, our results strongly suggest that, at the brain level, shame does not depend on a functionally distinguishable process but rather relies on subcircuits of the same network that is involved in the affective processing of guilt, while guilt additionally involves distinctive self-related representations that are not implicated in shame. This interpretation fits with a recent theoretical analysis describing the critical difference between guilt versus shame as emotions evoked by norm- versus value-oriented violations, respectively (Teroni and Deonna 2008). Because social norms are generally formalized rules on social values and behaviors, emotional processing of social norms (as in guilt) are likely to include emotional processing of values (as in shame) but not necessarily vice versa.

There is to our knowledge, no similar discussion concerning the differences between guilt and sadness, or their relationship with self- and other-referential processing, although—as demonstrated here and previously (Shin et al. 2000)—guilt-eliciting situations tend to simultaneously trigger not only shame but also sadness to a certain degree. The phenomenological difference between guilt and sadness is not debated, and unlike guilt and shame, sadness is usually not regarded as a self-conscious emotion. However, we found more neurobiological similarities between guilt and sadness than between guilt and shame. In fact, the right OFC was the only brain region activated by guilt in the direct contrast with sadness, while many other brain regions were activated when comparing guilt with shame. Moreover, many of these brain activations in the guilt versus shame contrast (such as the self-related areas in rostral ACC and insula) were similarly found in the contrast of sadness versus shame. Thus, sadness rather than shame turned out to be a tighter control emotion for guilt, allowing us to draw even stronger conclusions with regard to guilt-specific brain activations than with shame alone as comparison condition. Thus, we suggest that future studies investigating affective guilt processing should include sadness as a relevant comparison condition as well, rather than shame only.

In sum, we identified 2 regions in the PFC, the OFC and the paracingulate DMPFC, as most specifically involved in experiencing guilt as an emotion critically connected to own actual norm violations causing damage to other persons. Apart from their theoretical importance within social neuroscience, our results may ultimately also contribute in a clinical and forensic context to a better understanding of antisocial and psychopathic disorders, where the affective processing of norm violations is impaired, frequently with legal consequences. On a broader perspective, this research therefore also converges with recent efforts to strengthen the links between neuroscience, forensic psychology, and the law (Mobbs et al. 2007; Gazzaniga 2008; Schleim et al. 2011).

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

References

Adolphs R, Tranel D, Damasio AR. 1998. The human amygdala in social judgment. Nature. 393:470–474.

Anderson N. 1968. Likableness ratings of 555 personality-trait words. J Pers Soc Psychol. 9:272–279.

Anderson SW, Bechara A, Damasio H, Tranel D, Damasio AR. 1999. Impairment of social and moral behavior related to early damage in human prefrontal cortex. Nat Neurosci. 2:1032–1037.

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K. 2001. Modeling geometric deformations in EPI time series. Neuroimage. 13:903–919.

Baumeister RF, Stillwell AM, Heatherton TF. 1994. Guilt: an interpersonal approach. Psychol Bull. 115:243–267.

Beauregard M, Levesque J, Bourgouin P. 2001. Neural correlates of conscious self-regulation of emotion. J Neurosci. 21:RC165.

Bechara A, Damasio H, Damasio AR. 2000. Emotion, decision making and the orbitofrontal cortex. Cereb Cortex. 10:295–307.

Bedny M, Pascual-Leone A, Saxe RR. 2009. Growing up blind does not change the neural bases of Theory of Mind. Proc Natl Acad Sci U S A. 106:11312–11317.

Beer JS, John OP, Scabini D, Knight RT. 2006. Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. J Cogn Neurosci. 18:871–879.

Blair RJ. 2007. Dysfunctions of medial and lateral orbitofrontal cortex in psychopathy. Ann N Y Acad Sci. 1121:461–479.

Burianova H, McIntosh AR, Grady CL. 2010. A common functional brain network for autobiographical, episodic, and semantic memory retrieval. Neuroimage. 49:865–874.

Burnett S, Bird G, Moll J, Frith C, Blakemore SJ. 2009. Development during adolescence of the neural processing of social emotion. J Cogn Neurosci. 21:1736–1750.

Cabeza R, St Jacques P. 2007. Functional neuroimaging of autobiographical memory. Trends Cogn Sci. 11:219–227.

Calder AJ, Lawrence AD, Young AW. 2001. Neuropsychology of fear and loathing. Nat Rev Neurosci. 2:352–363.

Camerer CF, Fehr E. 2006. When does "economic man" dominate social behavior? Science. 311:47–52.

Ciaramidaro A, Adenzato M, Enrici I, Erk S, Pia L, Bara BG, Walter H. 2007. The intentional network: how the brain reads varieties of intentions. Neuropsychologia. 45:3105–3113.

Collins DL, Neelin P, Peters TM, Evans AC. 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr. 18:192–205.

Craig AD. 2003. Interoception: the sense of the physiological condition of the body. Curr Opin Neurobiol. 13:500–505.

Craik FIM, Moroz TM, Moscovitch M, Stuss DT, Winocur G, Tulving E, Kapur S. 1999. In search of the self: a positron emission tomography study. Psychol Sci. 10:26–34.

Damasio AR. 1994. Descartes' error: emotion, reason, and the human brain. New York: Grosset/Putnam.

Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LL, Parvizi J, Hichwa RD. 2000. Subcortical and cortical brain activity during the feeling of self-generated emotions. Nat Neurosci. 3:1049–1056.

D'Argembeau A, Ruby P, Collette F, Degueldre C, Balteau E, Luxen A, Maquet P, Salmon E. 2007. Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. J Cogn Neurosci. 19:935–944.

Eisenberg N. 2000. Emotion, regulation, and moral development. Annu Rev Psychol. 51:665–697.

Eisenberger NI, Lieberman MD, Williams KD. 2003. Does rejection hurt? An FMRI study of social exclusion. Science. 302:290–292.

Fukui H, Murai T, Shinozaki J, Aso T, Fukuyama H, Hayashi T, Hanakawa T. 2006. The neural basis of social tactics: an fMRI study. Neuroimage. 32:913–920.

Gallagher HL, Frith CD. 2003. Functional imaging of 'theory of mind'. Trends Cogn Sci. 7:77–83.

Gazzaniga MS. 2008. The law and neuroscience. Neuron. 60:412–415.

Gilbert SJ, Spengler S, Simons JS, Frith CD, Burgess PW. 2006. Differential functions of lateral and medial rostral prefrontal cortex (area 10) revealed by brain-behavior associations. Cereb Cortex. 16:1783–1789.

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. 2001. An fMRI investigation of emotional engagement in moral judgment. Science. 293:2105–2108.

Haidt J. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol Rev. 108:814–834.

Hampton AN, Bossaerts P, O'Doherty JP. 2008. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U S A. 105:6741–6746.

Hare RD. 1991. The hare psychopathy checklist—revised. Toronto (ON): Multi-Health Systems.

Jenkins AC, Mitchell JP. 2010. Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. Cereb Cortex. 20:404–410.

Jones WH, Schratter AK, Kugler K. 2000. The guilt inventory. Psychol Rep. 87:1039–1042.

Kedia G, Berthoz S, Wessa M, Hilton D, Martinot JL. 2008. An agent harms a victim: a functional magnetic resonance imaging study on specific moral emotions. J Cogn Neurosci. 20:1788–1798.

Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF. 2002. Finding the self? An event-related fMRI study. J Cogn Neurosci. 14:785–794.

Kochanska G, Barry RA, Jimenez NB, Hollatz AL, Woodard J. 2009. Guilt and effortful control: two mechanisms that prevent disruptive developmental trajectories. J Pers Soc Psychol. 97:322–333.

Krajbich I, Adolphs R, Tranel D, Denburg NL, Camerer CF. 2009. Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. J Neurosci. 29:2188–2192.

Kross E, Davidson M, Weber J, Ochsner K. 2009. Coping with emotions past: the neural bases of regulating affect associated with negative autobiographical memories. Biol Psychiatry. 65:361–366.

Kugler K, Jones WH. 1992. On conceptualizing and assessing guilt. J Pers Soc Psychol. 62:318–327.

Leary MR. 2007. Motivational and emotional aspects of the self. Annu Rev Psychol. 58:317–344.

LeDoux JE. 2000. Emotion circuits in the brain. Annu Rev Neurosci. 23:155–184.

Lykken DT. 1995. The antisocial personalities. Hillsdale (NJ): Erlbaum.

Macrae CN, Moran JM, Heatherton TF, Banfield JF, Kelley WM. 2004. Medial prefrontal activity predicts memory for self. Cereb Cortex. 14:647–654.

Maddock RJ. 1999. The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. Trends Neurosci. 22:310–316.

Maratos EJ, Dolan RJ, Morris JS, Henson RN, Rugg MD. 2001. Neural activity associated with episodic memory for emotional context. Neuropsychologia. 39:910–920.

Mobbs D, Lau HC, Jones OD, Frith CD. 2007. Law, responsibility, and the brain. PLoS Biol. 5:e103.

Moll J, de Oliveira-Souza R, Garrido GJ, Bramati IE, Caparelli-Daquer EM, Paiva ML, Zahn R, Grafman J. 2007. The self as a moral agent: linking the neural bases of social agency and moral sensitivity. Soc Neurosci. 2:336–352.

O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci. 4:95–102.

Ochsner KN, Ray RD, Cooper JC, Robertson ER, Chopra S, Gabrieli JD, Gross JJ. 2004. For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. Neuroimage. 23:483–499.

Olson IR, Plotzker A, Ezzyat Y. 2007. The enigmatic temporal pole: a review of findings on social and emotional processing. Brain. 130:1718–1731.

Olthof T, Schouten A, Kuiper H, Stegge H, Jennekens-Schinkel A. 2000. Shame and guilt in children: differential situational antecedents and experiential correlates. Br J Develop Psychol. 18:51–60.

Pourtois G, Schwartz S, Spiridon M, Martuzzi R, Vuilleumier P. 2009. Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. Cereb Cortex. 19:1806-1819.

Price DD. 2000. Psychological and neural mechanisms of the affective dimension of pain. Science. 288:1769-1772.

Rilling JK, Goldsmith DR, Glenn AL, Jairam MR, Elfenbein HA, Dagenais JE, Murdock CD, Pagnoni G. 2008. The neural correlates of the affective response to unreciprocated cooperation. Neuropsychologia. 46:1256-1266.

Ritchey M, Dolcos F, Cabeza R. 2008. Role of amygdala connectivity in the persistence of emotional memories over time: an event-related FMRI investigation. Cereb Cortex. 18:2494-2504.

Rolls ET. 2004. The functions of the orbitofrontal cortex. Brain Cogn. 55:11-29.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. 2003. The neural basis of economic decision-making in the ultimatum game. Science. 300:1755-1758.

Saxe R. 2006. Uniquely human social cognition. Curr Opin Neurobiol. 16:235-239.

Saxe R, Brett M, Kanwisher N. 2006. Divide and conquer: a defense of functional localizers. Neuroimage. 30:1088-1096.

Saxe R, Carey S, Kanwisher N. 2004. Understanding other minds: linking developmental psychology and functional neuroimaging. Annu Rev Psychol. 55:87-124.

Schilbach L, Wohlschlaeger AM, Kraemer NC, Newen A, Shah NJ, Fink GR, Vogeley K. 2006. Being with virtual others: neural correlates of social interaction. Neuropsychologia. 44:718-730.

Schleim S, Spranger TM, Erk S, Walter H. 2011. From moral to legal judgment: the influence of normative context in lawyers and other academics. Soc Cogn Affect Neurosci. 6:48-57.

Shin LM, Dougherty DD, Orr SP, Pitman RK, Lasko M, Macklin ML, Alpert NM, Fischman AJ, Rauch SL. 2000. Activation of anterior paralimbic structures during guilt-related script-driven imagery. Biol Psychiatry. 48:43-50.

Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD. 2004. Empathy for pain involves the affective but not sensory components of pain. Science. 303:1157-1162.

Stone VE, Cosmides L, Tooby J, Kroll N, Knight RT. 2002. Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. Proc Natl Acad Sci U S A. 99:11531-11536.

Takahashi H, Matsuura M, Koeda M, Yahata N, Suhara T, Kato M, Okubo Y. 2008. Brain activations during judgments of positive self-conscious emotion and positive basic emotion: pride and joy. Cereb Cortex. 18:898-903.

Takahashi H, Yahata N, Koeda M, Matsuda T, Asai K, Okubo Y. 2004. Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. Neuroimage. 23:967-974.

Tangney JP, Miller RS, Flicker L, Barlow DH. 1996. Are shame, guilt, and embarrassment distinct emotions? J Pers Soc Psychol. 70:1256-1269.

Tangney JP, Stuewig J, Mashek DJ. 2007. Moral emotions and moral behavior. Annu Rev Psychol. 58:345-372.

Teroni F, Deonna JA. 2008. Differentiating shame from guilt. Conscious Cogn. 17:725-740.

Vann SD, Aggleton JP, Maguire EA. 2009. What does the retrosplenial cortex do? Nat Rev Neurosci. 10:792-802.

Vogeley K, Bussfeld P, Newen A, Herrmann S, Happe F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K. 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. Neuroimage. 14:170-181.

Wallbott HG, Scherer K. 1995. Cultural determinants in experiencing shame and guilt. In: Tangney JP, Fischer KW, editors. Self-conscious emotions. New York: Guilford Press. p. 465-487.

Walter H, Adenzato M, Ciaramidaro A, Enrici I, Pia L, Bara BG. 2004. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. J Cogn Neurosci. 16:1854-1863.

Windmann S, Kirsch P, Mier D, Stark R, Walter B, Gunturkun O, Vaitl D. 2006. On framing effects in decision making: linking lateral versus medial orbitofrontal cortex activation to choice outcome processing. J Cogn Neurosci. 18:1198-1211.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp. 4:58-73.

Yang Y, Raine A. 2009. Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis. Psychiatry Res. 174:81-88.

Zahn R, Moll J, Krueger F, Huey ED, Garrido G, Grafman J. 2007. Social concepts are represented in the superior anterior temporal cortex. Proc Natl Acad Sci U S A. 104:6430-6435.

Zald DH. 2003. The human amygdala and the emotional evaluation of sensory stimuli. Brain Res Brain Res Rev. 41:88-123.

Zald DH. 2009. Orbitofrontal cortex contributions to food selection and decision making. Ann Behav Med. 1(38 Suppl):S18-S24.