Neural basis of impaired safety signaling in Obsessive Compulsive Disorder

Annemieke M. Apergis-Schoute[a,b,c,1], Claire M. Gillan[c,d], Naomi A. Fineberg[c,e,f], Emilio Fernandez-Egea[b,c], Barbara J. Sahakian[b,c], and Trevor W. Robbins[a,c]

[a]Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom; [b]Department of Psychiatry, University of Cambridge, Cambridge CB2 0SZ, United Kingdom; [c]Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge CB2 3EB, United Kingdom; [d]Department of Psychology, Trinity College Dublin, Dublin 2, Ireland; [e]Hertfordshire Partnership University NHS Foundation Trust, University of Hertfordshire, Welwyn Garden City AL8 6HG, United Kingdom; and [f]Postgraduate Medical School, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom

Edited by Ahmad R. Hariri, Duke University, Durham, NC and accepted by Editorial Board Member Marlene Behrmann December 19, 2016 (received for review July 27, 2016)

The ability to assign safety to stimuli in the environment is integral to everyday functioning. A key brain region for this evaluation is the ventromedial prefrontal cortex (vmPFC). To investigate the importance of vmPFC safety signaling, we used neuroimaging of Pavlovian fear reversal, a paradigm that involves flexible updating when the contingencies for a threatening (CS+) and safe (CS–) stimulus reverse, in a prototypical disorder of inflexible behavior influenced by anxiety, Obsessive Compulsive Disorder (OCD). Skin conductance responses in OCD patients (n = 43) failed to differentiate during reversal compared with healthy controls (n = 35), although significant differentiation did occur during early conditioning and amygdala BOLD signaling was unaffected in these patients. Increased vmPFC activation (for CS+ > CS–) during early conditioning predicted the degree of generalization in OCD patients during reversal, whereas vmPFC safety signals were absent throughout learning in these patients. Regions of the salience network (dorsal anterior cingulate, insula, and thalamus) showed early learning task-related hyperconnectivity with the vmPFC in OCD, consistent with biased processing of the CS+. Our findings reveal an absence of vmPFC safety signaling in OCD, undermining flexible threat updating and explicit contingency knowledge. Although differential threat learning can occur to some extent in the absence of vmPFC safety signals, effective CS– signaling becomes crucial during conflicting threat and safety cues. These results promote further investigation of vmPFC safety signaling in other anxiety disorders, with potential implications for the development of exposure-based therapies, in which safety signaling is likely to play a key role.

Obsessive Compulsive Disorder | vmPFC | Pavlovian | fMRI | safety signals

Current behavioral therapies in anxiety-related disorders are based on Pavlovian fear extinction models. As fear extinction relies on revaluation of threatening stimuli as safe, it is critical to address how the brain processes the safety of stimuli in the environment. The ventromedial prefrontal cortex (vmPFC) is known to play a multifaceted role in integrating affective evaluative processes while mediating flexible behavior and is implicated in fear learning and anxiety-related disorders (1–7). Prefrontal inflexibility in Obsessive Compulsive Disorder (OCD) suggests rigidity in threat estimation alongside a persistent urge to perform compulsive behaviors, yet only one study has examined the neural correlates of fear learning and extinction in this disorder, implicating a maladaptive vmPFC (8).

Human fear learning studies usually involve contrasting a threatening (CS+) stimulus that is occasionally paired with a shock with a stimulus that is never paired with a shock and thus safe (CS–). When using the CS+ > CS– contrast, the vmPFC consistently exhibits negative activation values in healthy controls, indicating stronger activation to the CS– than to the CS+ in this region (1–3, 7, 9). Fear reversal (Fig. 1*A*), in which a once-threatening stimulus turns safe and a once-safe stimulus becomes threatening, provides the ideal model for determining the interaction between threat versus safety valuation and prefrontal flexibility, as it indicates an especially important role for the vmPFC in the updating of the safe stimulus to provide "relief relabeling," providing a cognitive categorization of safety to a previously threatening cue (7).

The standard behavioral therapy for OCD is exposure response prevention (ERP), which involves repeatedly exposing the patient to fear-evoking stimuli while the patient is prevented from performing any compulsions that usually provide the patient with a temporary feeling of relief (10). This type of therapy can be very challenging for patients, with many of them unable to complete or engage in ERP, which has led to the idea of modifying ERP to allow for certain safety behaviors (11). Therefore, it remains a critical question how the processing of safety could be altered in OCD.

Based on the significance of the role of the vmPFC to estimate values to guide goal-directed behavior (12) and the interference of maladaptive vmPFC functioning in a variety of tasks in OCD (5, 8, 13–17), we postulated that impaired vmPFC valuation would be central to the disorder. We sought to determine how patients with OCD perform on a task that is dependent on accurate and flexible value signaling by this prefrontal region.

We hypothesized that OCD patients would show inflexibility in updating threat and safety expectancies associated with maladaptive vmPFC safety signaling in a previously validated fear reversal paradigm (7). We compared fear reversal learning in 43 OCD patients and 35 matched healthy controls by assessing threat expectancy with skin conductance responses (SCRs) and its neural correlates with functional magnetic resonance imaging (7). Our goals were to (*i*) relate any differences in threat learning and reversal to group differences on the whole brain level corrected for multiple comparisons,

Significance

Assigning safety to stimuli and situations forms an important part of everyday functioning. Obsessive Compulsive Disorder (OCD) is a prototypical disorder of inflexible behavior influenced by anxiety. Here we show, using neuroimaging of fear reversal learning, which involves a previously threatening stimulus becoming safe while a previously safe stimulus becomes threatening, that OCD patients fail to differentiate a safe from a threatening stimulus only after such a reversal, as measured by skin conductance responses. This OCD impairment in safety signaling, a likely prominent component of successful exposure therapy, is predicted and mediated by altered activity in a neural salience network including the ventromedial prefrontal cortex.

Author contributions: A.M.A.-S. and T.W.R. designed research; A.M.A.-S. and C.M.G. performed research; N.A.F. and E.F.-E. handled patient recruitment; A.M.A.-S. analyzed data; and A.M.A.-S., B.J.S., and T.W.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.R.H. is a Guest Editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. Email: aa545@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1609194114/-/DCSupplemental.

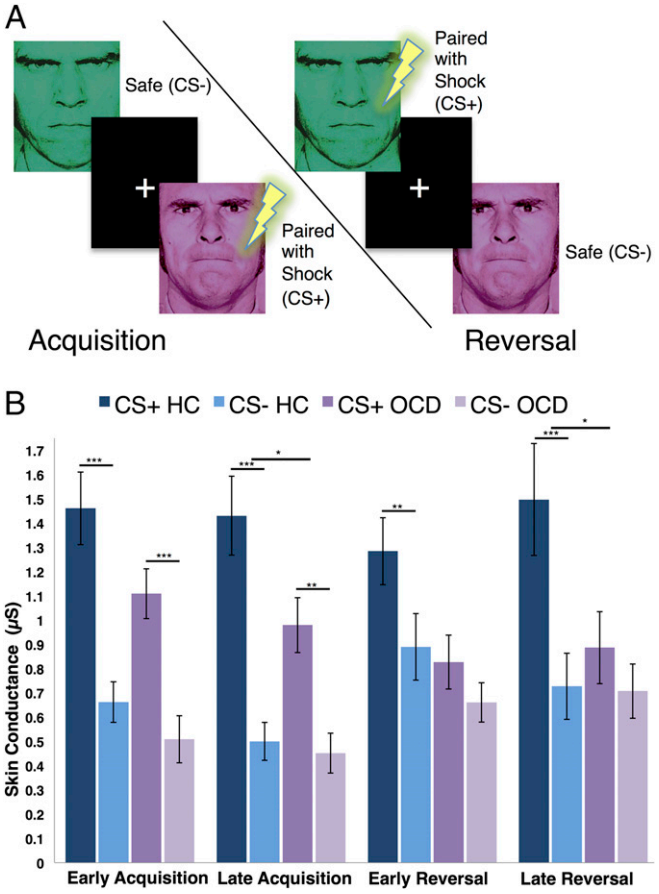

Fig. 1. Threat reversal paradigm and learning as reflected by skin conductance measurements. (*A*) One of the faces coterminated with a shock to the right wrist on one-third of the trials (eight CS+ US trials, which were excluded from analyses) and was therefore associated (16 CS+ trials) whereas the other face was safe (16 CS– trials), and during reversal, these contingencies were reversed so that the old CS+ became the new CS– and vice versa with the same number of trials. (*B*) Measured SCRs demonstrated that both OCD patients and controls acquired early ($t_{42} = 5.808$, $P = 0.0000007$; $t_{34} = 5.704$, $P = 0.000002$) and late ($t_{42} = 5.196$, $P = 0.000006$; $t_{34} = 6.630$, $P = 0.0000001$) threat learning. A between-group repeated-measures analysis with two factors (acquisition and reversal) and two stages each (early and late) revealed an significant main effect of group overall reflecting stronger differential learning in controls ($F_{1,\ 76} = 5.666$, $P = 0.02$) but also a significant Group × Stage interaction ($F_{1,\ 76} = 5.87$, $P = 0.018$), driven by a stronger differentiation deficit in OCD patients during late acquisition ($F_{1,\ 76} = 5.635$, $P = 0.02$) and late reversal ($F_{1,\ 76} = 5.129$, $P = 0.026$) when the presence of safety signaling is required. Asterisks in figure denote level of significance (*$P < 0.05$; ***$P < 0.0001$). Images of faces used with permission from Paul Ekman, PhD/Paul Ekman, LLC.

(*ii*) assess the importance of the vmPFC in accurate threat reversal in OCD, and (*iii*) perform region of interest analyses (ROIs) for key areas in fear learning (amygdala and striatum) to investigate possible group differences. Learning was quantified as the difference between SCRs to the CS+ and CS– during acquisition and reversal, which were divided into early (first half) and late (second half) phases (Fig. 1*B*). The data analyses mainly used the contrast between the CS+ and CS–, which refers to the original CS+ and CS– during acquisition and the reversed CS+ and CS– during reversal.

Results

Skin Conductance Analyses Revealed Weaker Differential Learning in OCD, Driven by Significant Reductions in Differentiation During Late Acquisition and Late Reversal. To directly compare threat learning in the OCD group with controls, a between-group repeated-measures analysis with two factors (acquisition and reversal) and two stages each (early and late) revealed a significant main effect of group overall reflecting stronger differential learning in controls ($F_{1,\ 76} = 5.666$, $P = 0.02$) but also a significant Group × Stage interaction ($F_{1,\ 76} = 5.87$, $P = 0.018$), driven by a stronger differentiation deficit in OCD patients during late acquisition ($F_{1,\ 76} = 5.635$, $P = 0.02$) and late reversal ($F_{1,\ 76} = 5.129$, $P = 0.026$), when the presence of safety signaling is required.

Direct Comparisons Between CS+ and CS– SCRs in Each Group Show That OCD Patients Can Differentiate the CS+ Versus the CS– During Acquisition but Fail to Differentiate During Reversal. To measure contingency knowledge, we measured differential SCRs, where significantly greater responding to the CS+ compared with the CS– reflects intact contingency learning. OCD patients and healthy controls both showed highly significant early ($t_{42} = 5.808$, $P = 0.0000007$; $t_{34} = 5.812$, $P = 0.000002$) and late ($t_{42} = 5.196$, $P = 0.000006$; $t_{34} = 6.630$, $P = 0.0000001$) fear learning. However, during early and late reversal, only healthy controls differentiated the CS+ and CS– ($t_{35} = 3.274$, $P = 0.002$; $t_{35} = 3.836$, $P = 0.001$), whereas OCD patients failed to differentiate between stimuli during both early and late stages of reversal ($t_{42} = 1.562$, $P = 0.126$; $t_{42} = 1.056$, $P = 0.297$).

Our study only included OCD patients without comorbidities and was balanced in terms of males versus females compared with controls. Our study was not designed with the purpose of looking at gender differences in conditioning, but it is interesting to note that females had stronger differential conditioning than males (see *SI Materials and Methods* for statistics), irrespective of the group difference.

Whole-Brain Family Wise Error-Corrected CS+ > CS– fMRI Results. For these same learning stages, we tested for whole-brain group differences at $P < 0.05$ family wise error (FWE) corrected using a CS+ > CS– contrast. This contrast therefore always depicts the CS+ > CS– activation, which is the original CS+ > original CS– during acquisition and the updated CS+ (previous CS–) > updated CS– (original CS+) during reversal. Therefore, positive values depict stronger signaling to the CS+ and negative values depict stronger signaling to the CS–. As seen in Fig. 2*A*, our whole-brain results showed that the vmPFC (–3, 26, –8) contrast was significantly more positive in OCD patients compared with controls during early acquisition at the whole-brain level ($t_{76} = 7.35$, $P < 0.0001$ FWE, 112 voxels). Differential CS+ > CS– activity plotted for these same voxels for the following stages revealed that the vmPFC flexibly tracked the CS– in controls with below baseline activity for this contrast, whereas sustained positive contrast activity in OCD indicated persistent increased CS+ signaling for both the initial CS+ and the reversed CS+ (Fig. 2*B*).

These differential CS+ > CS– negative vmPFC values were significantly different from zero in controls for early acquisition ($t_{34} = –3.674$, $P = 0.001$), late acquisition ($t_{34} = –2.901$, $P = 0.006$), and late reversal ($t_{34} = –4.222$, $P = 0.00017$), indicating stronger vmPFC CS– signaling in controls during all stages apart from early reversal ($t_{34} = –0.126$, $P = 0.9$). The vmPFC in OCD patients exhibited positive vmPFC values that were significantly different from zero during all stages: early acquisition ($t_{42} = 2.381$, $P = 0.022$), late acquisition ($t_{42} = 2.599$, $P = 0.013$), early reversal ($t_{42} = 2.075$, $P = 0.044$), and late reversal ($t_{42} = 3.365$, $P = 0.002$), indicating biased responding to the CS+ in this disorder.

Additionally, our whole-brain group analyses revealed significant differences during late reversal in both the left insula (–30, 23, 1; $t_{76} = 6.55$, $P < 0.0001$ FWE, 260 voxels) and the left globus pallidus (–12, 2, 1; $t_{76} = 6.04$, $P < 0.0001$ FWE, 34 voxels), which both exhibited increased CS+ > CS– activation in controls compared with OCD patients (Fig. S1). These increased whole-brain signals in both the globus pallidus and insula for CS+ > CS– during late reversal in healthy controls further showed that

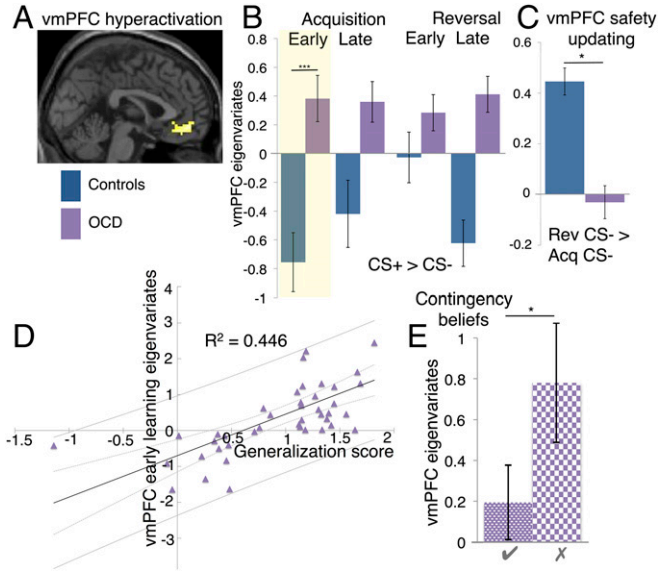

Fig. 2. vmPFC hyperactivation in OCD patients for CS+ > CS– reveals the absence of a safety signal exists from early learning and its level predicts the amount of generalization during reversal. (*A*) Sagittal view of the early vmPFC (–3, 26, –8) hyperactivation in OCD. (*B*) First eigenvariates were extracted for this vmPFC cluster ($t_{76} = 7.35$, $P < 0.0001$, FWE, 112 voxels) for all stages of learning, which showed persistence of vmPFC hyperactivation in OCD for the CS+ > CS– contrast. (*C*) Reversal CS– > acquisition CS– revealed a significant signal for CS– updating in controls' vmPFC (-2, 26, -2) only ($t_{76} = 5.01$, $P = 0.023$, FWE, 16 voxels). (*D*) vmPFC hyperactivation predicted the level of generalization in OCD patients, a score based on CS+ versus CS– differentiation during reversal at a highly significant level ($r = 0.668$, $n = 43$, $P = 0.000001$). (*E*) A subgroup of OCD patients (14/43) with incorrect contingency beliefs showed higher early vmPFC hyperactivation ($F_{1, 41} = 5.395$, $P = 0.025$). Asterisks in figure denote level of significance (*$P < 0.05$; ***$P < 0.0001$).

effective safety updating by the vmPFC results in a whole-brain detectable signature of accurate differential shock expectancy in brain regions preparing responses to possible threat in healthy controls but not in OCD patients (18).

Additional CS+ and CS– > Baseline fMRI Results. In addition to our main analysis investigating predictive vmPFC signaling comparing the CS+ versus the CS– to indicate relative bias of expectancies about shock expectation, we also used a mask of the voxels of the initial CS+ > CS– contrast to investigate vmPFC CS+ and CS– responding separately versus baseline [average of intertrial intervals (ITIs) when the cross was presented], as depicted in Fig. S2. These results confirmed an absence of vmPFC CS– processing in OCD (Early Acq, $t_{42} = 0.710$, $P = 0.481$; Late Acq, $t_{42} = 0.174$, $P = 0.862$; Early Rev, $t_{42} = 1.261$, $P = 0.214$; Late Rev, $t_{42} = 0.588$, $P = 0.56$) compared with mostly significant vmPFC CS+ processing (Early Acq, $t_{42} = 2.579$, $P = 0.013$; Late Acq, $t_{42} = 1.731$, $P = 0.091$; Early Rev, $t_{42} = 2.366$, $P = 0.023$; Late Rev, $t_{42} = 2.673$, $P = 0.011$). In contrast, controls sustained highly significant vmPFC CS– signaling (Early Acq, $t_{34} = 3.545$, $P = 0.001$; Late Acq, $t_{34} = 4.105$, $P = 0.000239$; Early Rev, $t_{34} = 3.696$, $P = 0.001$; Late Rev, $t_{34} = 4.034$, $P = 0.000294$) versus weak vmPFC CS+ signaling that was only marginally significant during early acquisition (Early Acq, $t_{34} = 2.055$, $P = 0.048$; Late Acq, $t_{34} = 1.73$, $P = 0.092$; Early Rev, $t_{34} = 1.144$, $P = 0.261$; Late Rev, $t_{34} = 1.251$, $P = 0.22$).

Whole-Brain FWE New CS– > Original CS– Contrast. To additionally examine updating of the safety response, we used the new CS– (both reversal stages) > original CS– (both acquisition stages) whole-brain contrast (Fig. 2*C*), which revealed a significantly stronger vmPFC (–2, 26, –2) signal in healthy controls compared with the absence of such a signal in OCD patients ($t_{76} = 5.01$, $P = 0.023$, FWE, 16 voxels), indicating that updated vmPFC safety signaling is also lacking in OCD patients as expected.

vmPFC CS+ > CS– Contrast Predicts Reversal Deficit in OCD. As the vmPFC was also the only region showing whole-brain differences during early learning and is critical for safety signaling and reversal, we investigated the correlation between this difference in OCD patients and their subsequent inability to update threat estimation as reflected by their SCRs. This highly significant correlation ($r = 0.668$, $n = 43$, $P = 0.000001$) as seen in Fig. 2*D* showed that the OCD positive activation contrast during early fear acquisition, when learning was still intact, predicted the level of differentiation (or, inversely, generalization) between the reversed CS+ and CS–. Moreover, further examination of the correlation of early vmPFC apparent hyperactivation with CS+ and CS– SCRs for each stage revealed a significant correlation specifically with CS– SCRs during early learning ($r = 0.330$, $n = 43$, $P = 0.031$) and early reversal ($r = 0.361$, $n = 43$, $P = 0.018$), indicating a particular role for vmPFC hyperactivation in maladaptive safety responding in OCD. Furthermore, in controls, early vmPFC hypoactivation for CS+ > CS– significantly correlated with the strength of differentiation during early learning ($r = 0.462$, $n = 35$, $P = 0.005$), further supporting the notion of the important role of vmPFC safety signaling in differential threat learning. It is important to note that the common convention of using the terms "hyperactivation" and "hypoactivation" in task-related fMRI is only reflective of activation in regards to the contrast chosen. For the purposes of conventional description, we also refer to this positive CS+ > CS– contrast in OCD patients as hyperactivation.

Explicit Knowledge of the CS+ and CS– Following Reversal and Ratings of Shock Aversion. We also measured subjects' explicit knowledge of the task contingencies postexperiment. Remarkably, 14/43 OCD patients reported getting shocks to both the new CS– (old CS+) and the new CS+ (old CS–) following reversal, compared with only 2/35 of controls, corresponding to a significant difference across groups, $\chi^2(df = 1, n = 78) = 8.527$, $P = 0.0035$. In a post hoc test, we found that early vmPFC activity (Fig. 2*E*) was significantly higher ($F_{1, 41} = 5.395$, $P = 0.025$) in that subgroup of OCD patients with incorrect contingency knowledge ($n = 14$) compared with those OCD patients reporting the correct contingencies ($n = 29$). Participants were also asked to rate how aversive they perceived the shock to be on a scale from 1–4, and OCD patients and controls showed remarkably similar ratings (mildly aversive at averages of 2.4 and 2.5, respectively; NS, $t_{76} = –0.788$, $P = 0.433$).

Task-Related CS+ > CS– Functional Connectivity of the VmPFC. To elucidate the role of early vmPFC hyperactivation for CS+ > CS– signaling within its associated neuroanatomical circuitry, we used a psychophysiological interaction (PPI) to test for whole-brain differences between OCD patients and controls with a vmPFC ROI for the CS+ > CS– contrast for all stages (Fig. 3). OCD patients had significantly stronger task-related connectivity of the vmPFC with areas of the salience network (19) comprising the dorsal anterior cingulate (–6, 17, 31; $t_{76} = 6.81$, $P < 0.0001$, FWE, 393 voxels), left insula (–33, 20, 7; $t_{76} = 7.47$, $P < 0.0001$, FWE, 506 voxels), right insula (42, 23, 4; $t_{76} = 9.04$, $P < 0.0001$, FWE, 549 voxels), and right thalamus (12, 2, 4; $t_{76} = 6.29$, $P < 0.0001$, FWE, 56 voxels) during early acquisition only, indicating that OCD patients solely assigned salience to the CS+, whereas the CS+ and CS– were likely of similar saliency in controls.

ROIs of the Amygdala and Caudate. Previous work has established important roles for the amygdala-striatal network in aversive learning in humans (2, 20, 21), including for this paradigm (7). Hence we used ROIs (according to the original threat reversal

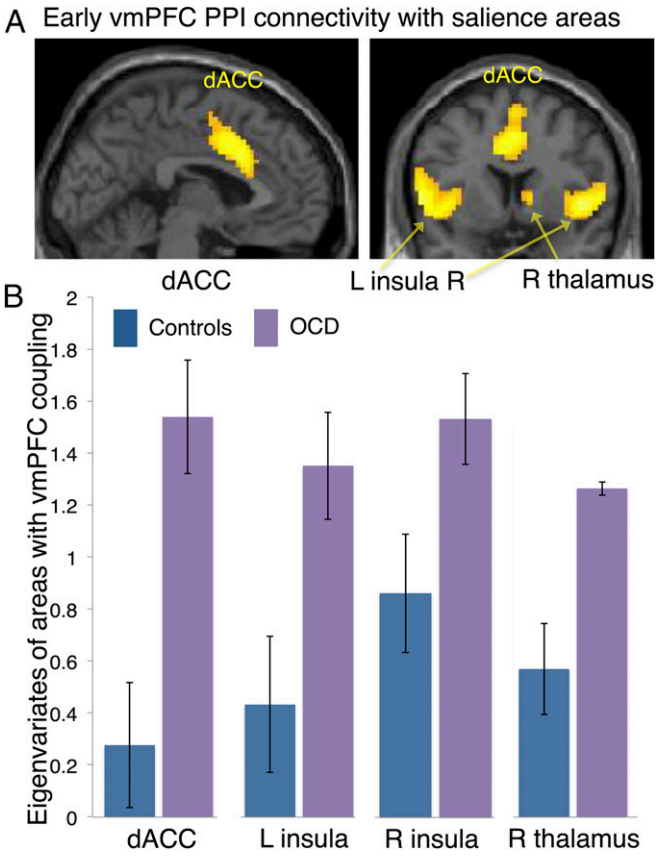

Fig. 3. OCD patients show increased early learning-related coupling of the vmPFC with salience areas. (*A*) Areas involved in salience processing showed increased coupling with the vmPFC during early conditioning in OCD patients. (*B*) First eigenvariates of these salience network regions depict significant increased coupling with the vmPFC in OCD patients during early acquisition dACC (–6, 17, 31; $t_{76} = 6.81$, $P < 0.0001$, FWE, 393 voxels), left insula (–33, 20, 7; $t_{76} = 7.47$, $P < 0.0001$, FWE, 506 voxels), right insula (42, 23, 4; $t_{76} = 9.04$, $P < 0.0001$, FWE, 549 voxels), and right thalamus (12, 2, 4; $t_{76} = 6.29$, $P < 0.0001$, FWE, 56 voxels).

findings in healthy controls) (7) for the left amygdala and left caudate (striatum), two areas known to play important roles in fear learning and reversal. We compared CS+ > CS– activation between OCD patients and controls for each stage (Fig. S3). Left amygdala (–24, –1, –17) differentiation (Fig. S3*A*) was comparable in OCD patients and controls for all stages ($F_{1, 76} < 0.4$, $P > 0.5$), consistent with intact learning of the CS+ to unconditioned stimulus (US) association. The left caudate (–8, 2, 10) tracked differentiation of the CS+ and CS– during acquisition similarly (Fig. S3*B*) for both controls and OCD patients ($F_{1, 76} < 2.6$, $P > 0.1$), but this difference was absent in OCD patients compared with controls during early reversal ($F_{1, 76} = 17.782$, $P < 0.0001$) and late reversal ($F_{1, 76} = 26.372$, $P < 0.0001$), reflecting flexible tracking of updated CS+ versus CS– contingencies by the controls' caudate compared with an absence in updating in OCD patients (7, 22).

Discussion

These findings show a failure to learn about threat versus safety values during reversal in OCD due to a lack of vmPFC safety signaling, demonstrated by their nondifferential SCRs during this stage. A recent meta-analysis highlighted the importance of vmPFC safety signals, expressed as consistent deactivations for the CS+ > CS– contrast in this region, resulting from stronger signaling to the CS– (23), and emphasized how effective CS– processing is an active process. The absence of such a safety signal in the vmPFC during both reversal and acquisition in OCD combined with the ability to differentiate the CS+ versus CS– during basic conditioning suggests that OCD patients acquire threat conditioning mainly by labeling the valence of the CS+, rather than additionally assigning a "safe" label to the CS–. This account is further supported by five findings: (*i*) During early learning, vmPFC connectivity was enhanced within the salience network, indicating that OCD patients relied on the salience of the CS+ alone to acquire threat estimation during learning; (*ii*) positive values for the vmPFC activation contrast for CS+ > CS– during early fear learning was predictive of the generalization deficit found during reversal; (*iii*) this positive vmPFC contrast for CS+ > CS– in OCD during early learning was correlated with CS– responding during early learning and early reversal; (*iv*) the amount of vmPFC hypoactivation for CS+ > CS– in healthy controls was positively correlated with the strength of CS+ versus CS– differentiation during early learning; and (*v*) group comparisons of skin conductance revealed significantly stronger differentiation in controls during late acquisition and late reversal, which indicates that contrast learning about both the threat of the CS+ and safety of the CS– supports stronger differentiation between the two. Previously, it has been shown that the vmPFC has roles in extinction learning (2) and retrieval (1, 3), as well as in emotional regulation (24) and (re)valuation (6, 25), and is implicated in impaired extinction learning in adults with OCD (8). Results found in youth with OCD highlighted impaired inhibitory learning and contingency awareness during extinction (26), whereas findings from a sample of lifelong OCD patients showed impaired extinction recall (27), which are in line with impaired learning about newly acquired safety. However, a critical role in safety signaling (7, 28) has hitherto received insufficient attention in anxiety-related disorders.

Hare et al. have postulated that the vmPFC originally evolved to compute the immediate value of stimuli (12), which renders this brain region central to many psychiatric disorders in which valuation has gone awry. In the case of OCD, it is easy to conceive the vmPFC to be excessively involved in internal valuation of idiosyncratic goals that constitute the compulsions seen in these patients. Even though OCD patients are generally not considered delusional, the value of their compulsions and the urge to keep performing them can severely disrupt their everyday functioning and impedes achievements of long-term goals.

Although we cannot conclude from task-related fMRI that the underlying neural deficit for these findings in OCD is vmPFC hyperactivation, findings from other task-independent studies using resting-state fMRI and positron emission tomography (PET) measures indicate this to be likely (29, 30), emphasizing that the vmPFC is a key node in an overactive thalamo-cortico-striatal pathway. We hypothesize that this sustained vmPFC hypermetabolism impedes safety learning in OCD.

Considering possible explanations for the apparently sustained vmPFC hyperactivation in OCD for the CS+ > CS– contrast, this structure is a central component of the default mode network and is thus implicated in self-referential thinking, normally becoming deactivated during externally directed attention (31). Such deactivation is well known to occur during early differential threat conditioning (1, 2, 7, 8) as a consequence of stronger signaling to the CS–. Furthermore, several studies have shown a failure of OCD patients to deactivate the vmPFC during task-related externally directed attention (13, 32), supporting the hypothesis that OCD patients' vmPFC is overrecruited by self-referential thoughts and thus unavailable for effective valuation (13). Similarly, the vmPFC also exhibited significantly enhanced activation in OCD patients during initial learning of responding to a CS+ versus a CS– in a shock avoidance paradigm compared with controls (33). Safety signaling is one of the aspects of the vmPFC valuation system compromised in OCD but might be of specific relevance, as ERP therapy, the first line of behavioral treatment in OCD, depends on

effective revaluation of stimuli and situations to be considered safe. During ERP, OCD patients are repeatedly confronted with tailored stimuli that normally trigger the urge to perform a compulsion and are instructed to refrain from performing their safety behavior (34). To make this therapy successful, a safety memory has to become established; otherwise, a stressful situation might trigger the return of the urge to perform such irrational yet compulsive safety behavior. The urge to perform an experimentally trained habit has been shown to be associated with hyperactivity in the caudate in OCD patients (33), whereas the vmPFC is thought to influence the caudate for goal selection (35, 36). Therefore, an aberrant vmPFC valuation system as demonstrated across several tasks in OCD (5, 17, 37) is of critical relevance to the maintenance of the disorder.

The differentiation failure during threat reversal in OCD patients was also reflected in their nondiscriminative striatal responses during this stage, likely due to the absence of the vmPFC safety signal (7). Although anxiety is often considered central to OCD, suggesting a putative role for the amygdala (38), OCD patients and controls exhibited strikingly similar amygdala processing during acquisition and reversal learning, indicating that generalization was not due to amygdala impairment and that learning of the CS+ to US association was intact in OCD. Moreover, SCRs were overall lower in OCD, likely due to the absence of the sharpening of the CS+ versus CS– contrast through safety learning, ruling out generalized anxiety related to receiving shocks, additionally confirmed by very similar shock aversion ratings as controls. These results further query the precise relationship of anxiety symptoms to obsessions and compulsions as acknowledged by the new designation of OCD without the Anxiety Disorders categorization in DSM5 and within its own category of OCDs (39). Although fear generalization may play an important role in generalized anxiety disorder (40), fear generalization in OCD only manifested itself during reversal when successful updating became contingent on an instructive vmPFC safety signal. Not only did OCD patients fail to differentiate between the threatening and updated safe stimulus during reversal, a third of them believed that they were receiving shocks to both faces, further underscoring the importance of safety valuation. Only one study to date has investigated neural correlates of fear learning and extinction in OCD (8), finding enhanced retention of fear after extinction. Our threat reversal paradigm highlights instead detrimental effects of the absence of safety signaling by the vmPFC in OCD that normally enables flexible responding to changing threats. In light of a recent study revealing that many fMRI results could reflect false positives (41), it is important to emphasize that our main findings were based on whole-brain group comparisons FEW-corrected at a significance level of $P < 0.0001$, indicative of convincing differences in vmPFC activation between OCD patients and controls. Moreover, a PET study had already implied a prominent role for the vmPFC in OCD, showing that deep brain stimulation of the subthalamic nucleus resulted in a strong decrease in vmPFC metabolism that correlated with an improvement in Yale Brown Obsessive Compulsive (Y-BOCS) scores (30).

To summarize, our findings indicate that a maladaptive vmPFC combined with increased connectivity with areas involved in salience processing undermines accurate safety learning in OCD patients, resulting in inflexible threat beliefs. Further research into safety learning could help in the development of novel exposure-based therapies.

Materials and Methods

Participants. Forty-three OCD patients and 35 matched controls were included in our analyses. Data were collected for 46 OCD patients and 40 controls, of which we had to exclude 4 due to excessive motion artifacts (2 controls and 2 OCD patients) and 4 due to absence of skin conductance measurements (3 controls and 1 OCD patient). The majority of participants were right-handed, but we also included five left-handed participants for each group. Eligible participants reported no history of head trauma, neurological disease, substance dependence, or contraindications for MRI. Participants provided informed written consent before participation, and the Cambridge Central Research Ethics Committee approved the study. Participants' demographics are presented in Table S1 and Fig. S4. Further details about the participants are provided in *SI Materials and Methods*.

Threat Reversal Paradigm and Skin Conductance Measurements. In the acquisition phase, face (face A) occurred 16 times without the shock (CS+) and 8 times with the shock (CS+ US), and the other face (face B) was never paired with the shock (16 trials). In the reversal phase, these contingencies reversed, so that now face B was paired with the US on 8 trials (CS+ US), 16 times without the shock (CS+), and face A was now never paired with the US (new CS–, 16 trials). Reversal immediately and continuously followed acquisition as part of the same scanning run and was unsignaled. The order of the different trial types was pseudorandomized (no consecutive shocks and no more than two consecutive trials of any kind), and the designation of face A and face B was counterbalanced. The analyses only included the trials without a shock, meaning trials that were either a CS+, when a shock might be expected, and CS–, when a shock would never occur. SCRs were defined as the baseline to peak difference within a 7-s interval following the presentation of a CS. SCR data were normalized per participant by dividing all responses to their peak amplitude. Specifics about the paradigm, task instructions, and SCR analyses can be found in *SI Materials and Methods*.

Neuroimaging. All fMRI data were acquired in a single session at the Wolfson Brain Imaging Institute at Addenbrooke's Hospital using a 3 Tesla Siemens Magnetom Trio scanner. All neuroimaging acquisition, preprocessing, and analyses details (data processed in SPM8) are presented in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank A.C. Roberts, A.B. Brühl, and S. Morein-Zamir for their comments. All fMRI data were collected at the Wolfson Brain Imaging Institute. This work was funded by Wellcome Trust Senior Investigator Award 104631/z/14/z (to T.W.R.) and a joint award from the Medical Research Council and the Wellcome Trust supporting the Behavioural and Clinical Neuroscience Institute (G0001354).

1. Kalisch R, et al. (2006) Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. *J Neurosci* 26(37):9503–9511.
2. Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron* 43(6):897–905.
3. Milad MR, et al. (2007) Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol Psychiatry* 62(5):446–454.
4. Cha J, et al. (2014) Circuit-wide structural and functional measures predict ventromedial prefrontal cortex fear generalization: Implications for generalized anxiety disorder. *J Neurosci* 34(11):4043–4053.
5. Stern ER, et al. (2011) Hyperactive error responses and altered connectivity in ventromedial and frontoinsular cortices in obsessive-compulsive disorder. *Biol Psychiatry* 69(6):583–591.
6. Roy M, Shohamy D, Wager TD (2012) Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn Sci* 16(3):147–156.
7. Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA (2008) From fear to safety and back: Reversal of fear in the human brain. *J Neurosci* 28(45):11517–11525.
8. Milad MR, et al. (2013) Deficits in conditioned fear extinction in obsessive-compulsive disorder and neurobiological changes in the fear circuit. *JAMA Psychiatry* 70(6):608–618.
9. Apergis-Schoute AM, Schiller D, LeDoux JE, Phelps EA (2014) Extinction resistant changes in the human auditory association cortex following threat learning. *Neurobiol Learn Mem* 113:109–114.
10. Foa EB, et al. (2005) Randomized, placebo-controlled trial of exposure and ritual prevention, clomipramine, and their combination in the treatment of obsessive-compulsive disorder. *Am J Psychiatry* 162(1):151–161.
11. Rachman S, Shafran R, Rasdomsky AS, Zysk E (2011) Reducing contamination by exposure plus safety behaviour. *J Behav Ther Exp Psychiatry* 42(3):397–404.
12. Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324(5927):646–648.
13. Stern ER, et al. (2013) Subjective uncertainty and limbic hyperactivation in obsessive-compulsive disorder. *Hum Brain Mapp* 34(8):1956–1970.
14. Haber SN, Heilbronner SR (2013) Translational research in OCD: Circuitry and mechanisms. *Neuropsychopharmacology* 38(1):252–253.
15. Gillan CM, Robbins TW (2014) Goal-directed learning and obsessive-compulsive disorder. *Philos Trans R Soc Lond B Biol Sci* 369(1655):20130475.
16. Cavanagh JF, Gründler TOJ, Frank MJ, Allen JJB (2010) Altered cingulate sub-region activation accounts for task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia* 48(7):2098–2109.

17. Fitzgerald KD, et al. (2010) Altered function and connectivity of the medial frontal cortex in pediatric obsessive-compulsive disorder. *Biol Psychiatry* 68(11):1039–1047.
18. Paulus MP, Stein MB (2006) An insular view of anxiety. *Biol Psychiatry* 60(4):383–387.
19. Hermans EJ, et al. (2011) Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science* 334(6059):1151–1153.
20. Jensen J, et al. (2003) Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron* 40(6):1251–1257.
21. Seymour B, et al. (2004) Temporal difference models describe higher-order learning in humans. *Nature* 429(6992):664–667.
22. Goldstein BL, et al. (2012) Ventral striatum encodes past and predicted value independent of motor contingencies. *J Neurosci* 32(6):2027–2036.
23. Fullana MA, et al. (2016) Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Mol Psychiatry* 21(4):500–508.
24. Etkin A, Egner T, Kalisch R (2011) Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci* 15(2):85–93.
25. Rudebeck PH, Saunders RC, Prescott AT, Chau LS, Murray EA (2013) Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nat Neurosci* 16(8):1140–1145.
26. McGuire JF, et al. (2016) Fear conditioning and extinction in youth with obsessive-compulsive disorder. *Depress Anxiety* 33(3):229–237.
27. McLaughlin NCR, et al. (2015) Extinction retention and fear renewal in a lifetime obsessive-compulsive disorder sample. *Behav Brain Res* 280:72–77.
28. Kong E, Monje FJ, Hirsch J, Pollak DD (2014) Learning not to fear: Neural correlates of learned safety. *Neuropsychopharmacology* 39(3):515–527.
29. Whiteside SP, Port JD, Abramowitz JS (2004) A meta-analysis of functional neuroimaging in obsessive-compulsive disorder. *Psychiatry Res* 132(1):69–79.
30. Le Jeune F, et al.; French Stimulation dans le trouble obsessionnel compulsif (STOC) study group (2010) Decrease of prefrontal metabolism after subthalamic stimulation in obsessive-compulsive disorder: A positron emission tomography study. *Biol Psychiatry* 68(11): 1016–1022.
31. Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network: Anatomy, function, and relevance to disease. *Ann N Y Acad Sci* 1124(1):1–38.
32. Maltby N, Tolin DF, Worhunsky P, O'Keefe TM, Kiehl KA (2005) Dysfunctional action monitoring hyperactivates frontal-striatal circuits in obsessive-compulsive disorder: An event-related fMRI study. *Neuroimage* 24(2):495–503.
33. Gillan CM, et al. (2015) Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am J Psychiatry* 172(3):284–293.
34. Rachman S, Hodgson R, Marks IM (1971) The treatment of chronic obsessive-compulsive neurosis. *Behav Res Ther* 9(3):237–247.
35. Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci* 27(15):4019–4026.
36. de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC (2009) Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J Neurosci* 29(36):11330–11338.
37. Fitzgerald KD, et al. (2005) Error-related hyperactivity of the anterior cingulate cortex in obsessive-compulsive disorder. *Biol Psychiatry* 57(3):287–294.
38. van den Heuvel OA, et al. (2004) Amygdala activity in obsessive-compulsive disorder with contamination fear: A study with oxygen-15 water positron emission tomography. *Psychiatry Res* 132(3):225–237.
39. American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Arlington, VA), 5th Ed.
40. Lissek S, et al. (2014) Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biol Psychiatry* 75(11):909–915.
41. Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci USA* 113(28): 7900–7905.
42. Goodman WK, et al. (1989) The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry* 46(11):1006–1011.
43. Montgomery SA, Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* 134(4):382–389.
44. Spielberger CD (1989) State-Trait Anxiety Inventory: Bibliography (Consulting Psychologists Press, Palo Alto, CA), 2nd Ed.
45. Nelson HE, O'Connell A (1978) Dementia: The estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex* 14(2):234–244.
46. Dunsmoor JE, Bandettini PA, Knight DC (2007) Impact of continuous versus intermittent CS-UCS pairing on human brain activation during Pavlovian fear conditioning. *Behav Neurosci* 121(4):635–642.
47. Ekman P, Friesen WV (1976) Pictures of Facial Affect (Consulting Psychologists Press, Palo Alto, CA).
48. Critchley HD, Mathias CJ, Dolan RJ (2002) Fear conditioning in humans: The influence of awareness and autonomic arousal on functional neuroanatomy. *Neuron* 33(4): 653–663.
49. Tzourio-Mazoyer N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1):273–289.
50. Friston KJ, et al. (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6(3):218–229.