

## Neural mechanisms of persistent avoidance in OCD: A novel avoidance devaluation study



Henry W. Chase<sup>a,\*</sup>, Simona Graur<sup>a</sup>, Amelia Versace<sup>a</sup>, Tsafrir Greenberg<sup>a</sup>, Lisa Bonar<sup>a</sup>, Robert Hudak<sup>a</sup>, Gregory J. Quirk<sup>b</sup>, Ben D. Greenberg<sup>c</sup>, Steven A. Rasmussen<sup>c</sup>, Suzanne N. Haber<sup>d</sup>, Mary L. Phillips<sup>a</sup>

<sup>a</sup> Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

<sup>b</sup> Departments of Psychiatry and Anatomy & Neurobiology, University of Puerto Rico School of Medicine, San Juan, PR, USA

<sup>c</sup> Department of Psychiatry and Human Behavior, Brown Medical School, Butler Hospital and Providence, VA Medical Center, Providence, RI, USA

<sup>d</sup> Department of Pharmacology and Physiology, University of Rochester Medical Center, Rochester, NY, USA

### ARTICLE INFO

#### Keywords:

Obsessive-Compulsive Disorder  
Goal-directed behavior  
Compulsivity  
Inhibition  
fMRI

### ABSTRACT

Obsessive-Compulsive Disorder (OCD) is characterized by repetitive avoidance behavior which is distressing and associated with marked impairment of everyday life. Recently, paradigms have been designed to explore the hypothesis that avoidance behavior in OCD is consistent with a formal conception of habit. Such studies have involved a devaluation paradigm, in which the value of a previously rewarded cue is altered so that avoidance is no longer necessary. We employed a rule-based avoidance task which included a devaluation, examining behavioral performance on the task and their neural correlates using functional MRI in groups of participants with OCD ( $n = 44$ ) and healthy control participants ( $n = 46$ ). Neuroimaging data were analyzed using a general linear model (GLM), modelling valued, devalued and control cues, as well as feedback events. First, while no overall effect of OCD was seen on devaluation performance, patients with longer illness duration showed poorer devaluation performance ( $\chi^2 = 13.84$ ,  $p < 0.001$ ). Reduced devaluation was related to impaired learning on the overtraining phase of the task, and to enhanced feedback activation in the caudate and parietal lobe during within-scanner retraining ( $T = 5.52$ ,  $p_{FWE} = 0.003$ ), across all participants. Second, a significant interaction effect was observed in the premotor cortex ( $F = 29.03$ ,  $p_{FWE} = 0.007$ ) coupled to the devalued cue. Activations were divergent in participants with OCD (lower activation) and healthy controls (higher activation) who did not change responding to the devalued cue following devaluation, and intermediate in participants who did change responding ( $T = 5.39$ ,  $p_{FWE} = 0.003$ ). Finally, consistent with previous work, medial orbitofrontal cortex activation coupled to valued cues was reduced in OCD compared to controls ( $T = 3.49$ ,  $p_{FWE} = 0.009$ ). The findings are discussed in terms of a prediction error-based model of goal-directed and habitual control: specifically, how goal-directed control might be diminished in OCD in favor of habits. They suggest that illness duration might be significant determinant of variation in impaired goal-directed learning in OCD, and be a factor relevant for understanding discrepancies across studies. Overall, the study shows the potential of conceptual replication attempts to provide complementary insights into compulsive behavior and its associated neural circuitry in OCD.

### 1. Introduction

Obsessive Compulsive Disorder (OCD) is a chronic and debilitating psychiatric disorder affecting 2–3% of the world's population (Rasmussen and Eisen, 1992), characterized by repetitive, compulsive behaviors including cleaning and checking, as well as obsessions. The formation of habits – actions that are persistently performed, independently of their consequences (Balleine and Dezfouli, 2019;

Balleine and Dickinson, 1998) – has been proposed to underlie increased compulsive behavior in OCD (Gillan et al., 2016). Broadly, there is considerable experimental support for this idea, with numerous demonstrations of abnormally inflexible behavior and altered correlates of flexible learning in OCD (Apergis-Schoute et al., 2017; Gillan et al., 2011; Remijnse et al., 2006, 2009). In particular, two recent studies demonstrated greater levels of persistent avoidance behavior following overtraining in OCD relative to healthy participants. These studies

\* Corresponding author at: Western Psychiatric Institute and Clinic, Loeffler Building, 121 Meyran Avenue, Pittsburgh, PA 15213, USA.

E-mail address: [chaseh@upmc.edu](mailto:chaseh@upmc.edu) (H.W. Chase).

<https://doi.org/10.1016/j.nicl.2020.102404>

Received 14 June 2020; Received in revised form 24 August 2020; Accepted 25 August 2020

Available online 28 August 2020

2213-1582/ © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

employed a devaluation paradigm: in an avoidance scenario, this involves training the participant to make an avoidance response when a cue which predicts an aversive outcome appears. After this cue-response behavior is established and the cue becomes 'valued', the aversive outcome is then 'devalued': the aversive properties of the expected outcome are diminished or removed. Participants should stop responding on a cue associated with a devalued outcome. A proportion of participants with OCD continued to respond to avoid an aversive outcome (shock) despite the removal of the apparatus required to elicit the shock (Gillan et al., 2015, 2014) i.e., these participants did not show a complete devaluation response in these contexts by continuing to exhibit habitual-like responding to the otherwise devalued cue. Reduction in devaluation is thus a potential component process relevant to understanding the development of habitual behavior in OCD.

To date, few studies examined the neural basis of a bias toward habitual behavior in OCD (Banca et al., 2015). One study (Gillan et al., 2015) examined neural activity during an avoidance devaluation paradigm. Individuals with OCD showed altered activity in the medial orbitofrontal cortex (OFC), as well as other default mode network (DMN) regions, during responding to valued versus control cues, which switched from greater (OCD > healthy) to lower (healthy > OCD) activity throughout the course of cue-outcome contingency training. This finding is consistent with the role of the OFC in value and cue-outcome encoding (Bartra et al., 2013; Sadacca et al., 2018).

In their study, Gillan and colleagues evaluated neural activity coupled to valued, but not devalued, cues during the devaluation phase. They reasoned that such activity might reflect habitual behavior, as responding to both devalued and valued cues should become habitual in non-devaluers. However, given that extended training with the cues may render cue-elicited responding automatic, successful non-responding to devalued cues may also require inhibitory processes to be engaged (Wessel et al., 2014). The premotor cortex (Aron et al., 2007; Picton et al., 2007), inferior frontal gyrus (IFG: Aron et al., 2014) and subthalamic nucleus (Aron and Poldrack, 2006; Eagle et al., 2008) support the capacity to inhibit prepotent behaviors. In particular, previous studies reported abnormal activity in the IFG during behavioral inhibition, set shifting and task switching (Britton et al., 2010; Gu et al., 2008; Roth et al., 2007) in individuals with OCD, while alterations of the structure and function of the premotor cortex and IFG have also been consistently identified within the OCD literature via meta-analysis (Norman et al., 2016). Finally, functional connectivity of the subthalamic nucleus is related to individual differences in compulsivity (Morris et al., 2017). Thus, altered goal-directed control in OCD might be associated with alterations in neural activation in regions involved in inhibitory control, but to our knowledge little evidence testing this proposal using fMRI exists within the context of stimulus-response learning.

In the present study, we considered two potential neural mechanisms underlying persistent responding to all cues in OCD: 1. impaired inhibition of responding to devalued cues, involving lower premotor cortical and inferior frontal gyrus activity to these cues; 2. altered representation of value to valued vs. control cues during the retraining and devaluation phases of the task, as indexed by aberrant patterns of orbitofrontal cortex activity. We thus examined neural activity to devalued and valued cues during the devaluation phase, to valued cues throughout the retraining and devaluation phases during an avoidance devaluation task in OCD and healthy participants.

In summary, we tested the following hypotheses:

**Hypothesis 1:** A greater proportion of OCD than healthy individuals would continue to respond to devalued cues, i.e., would not show a robust devaluation effect, consistent with habit formation, and in support of Gillan et al. (2015, 2011). Devaluing participants would be expected to show a strong preference not to respond for the devalued stimulus, whereas non-devaluing participants would be expected to continue responding for this stimulus or respond randomly.

We then examined the three potential neural mechanisms

underlying reduced devaluation, which we hypothesized to be evident more in individuals with OCD, given Hypothesis 1.

**Hypothesis 2:** In support of impaired inhibition of responding to devalued cues: non-devaluers relative to devaluers would show lower activity in frontal lobe and subthalamic regions important for inhibition. Given this hypothesis is highly inferentially dependent on the location of the activations (e.g. inferior frontal gyrus), and more importantly, the capacity of these activations to support the inference about psychological function, we performed a *post hoc* behavioral decoding analysis to determine the specificity of the observed activations for inhibition (see [supplemental information](#)).

**Hypothesis 3:** In support of impaired encoding of valued versus control cues, and the previous study (Gillan et al., 2015): non-devaluers relative to devaluers would show decreased value encoding-related activity in the orbitofrontal cortex to valued cues across the retraining and devaluation phases of the task.

These three hypotheses were derived from Gillan and colleagues' previous work (Gillan et al., 2015), although no evidence was obtained in that study supporting the equivalent of Hypothesis 2. A further hypothesis, Hypothesis 4, relating to the neural correlates of feedback learning, is included in the [supplemental information](#).

## 2. Methods

### 2.1. Participants

48 participants with obsessive compulsive disorder (OCD) were recruited for the study, as were 50 healthy control participants. All participants were 18–35 years of age and right-handed. Five participants were excluded due to motion (> 5 mm: 1 healthy, 4 OCD), while three further participants were excluded due to behavioral data which did not reflect task rules (e.g., not responding correctly to cues during the retraining phase). Thus, the analyzed data included 46 healthy and 44 participants with OCD (see [Table 1](#)).

During a screening session prior to the fMRI data acquisition session, all participants were interviewed using the Structured Clinical Interview for DSM-5. The following clinical variables were also obtained during this session: the Yale-Brown Obsessive Compulsive Scale (YBOCS), Hamilton Depression Scale (HAMD: Hamilton, 1960), Hamilton Anxiety Scale (HAMA: Hamilton, 1959), the Obsessive-Compulsive Inventory-Revised (OCI-R: Foa et al., 2002) and Pathological Obsessive Compulsive Personality Scale (POPS: Pinto, 2011) scales. Premorbid IQ was measured using the NART scale (Blair and Spreen, 1989). Clinical and demographic information regarding participants is displayed in [Table 1 and 2](#). All participants signed informed consent, with a protocol approved by the University of Pittsburgh Institutional Review Board.

Exclusion criteria for all participants were personal/family history (1st/2nd degree relatives) of schizophrenia/schizoaffective disorder, other primary psychotic disorders, bipolar disorder; present posttraumatic stress disorder; present psychotic symptoms; personal history of head injury, neurological, neurodevelopmental (e.g., autism), tic disorders, systemic medical (metabolic, endocrine, chronic inflammatory, vascular, autoimmune) disease from medical records and self-report, all of which may confound interpretation of neuroimaging measures; MMSE score < 24; premorbid IQ estimate < 85; visual disturbance (< 20/40 corrected Snellen visual acuity); left/mixed handedness; current, and history in the last 3 months of, alcohol and illicit substance abuse/dependence, determined by urine screen and clinical assessment of alcohol and substance use; current suicidal ideation, as assessed using the Hamilton Rating Scale for Depression, contraindications to MRI: metallic foreign objects, e.g., aneurysm clips/pacemakers, or questionable history of metallic fragments, prone to panicking in enclosed spaces; positive pregnancy test for females/self-reporting of pregnancy.

Additional exclusion criteria for participants with OCD were Yale-

**Table 1**

Demographic and clinical information for all included participants. Aside from count data (Gender/Task version/Successful Devaluation), means are presented with standard deviations in parentheses. Further notes: S/W = summer or winter: the two versions of the paradigm. POPS: data missing for one control participant. Illness duration: ‘As long as I can remember’ response categorized as 3 years of age for illness onset, given 4 was the lowest stated. Antidepressant usage as follows: Citalopram (n = 1), Fluoxetine (n = 4), Escitalopram (n = 4), Sertraline (n = 7). Diagnosis data includes any lifetime diagnosis. Eating disorders include anorexia, bulimia and binge eating. Co-morbid anxiety disorders include general anxiety disorder, social anxiety disorder, panic disorder, phobias and other anxiety disorders. Other diagnoses within the OCD group included ADHD (n = 3), psychosis (n = 3), and trauma (n = 3).

	Healthy Controls	OCD	Statistic
Gender (M/F)	16/30	16/28	$\chi^2 < 1$
Age	23.63 (4.11)	23.60 (4.39)	T < 1
Educational Level	5.96 (1.38)	5.84 (1.40)	T < 1
NART	112.32 (6.36)	110.79 (7.076)	T(88) = 1.078, p = 0.28
Task Version (S/W)	23/23	22/22	$\chi^2 = 0$
YBOCS	0.11 (0.74)	20.64 (3.29)	T(47.14) = 40.80, p < 0.001
HAMD	1.26 (1.12)	9.98 (5.29)	T(46.71) = 10.70, p < 0.001
HAMA	0.96 (1.17)	10.89 (6.45)	T(45.72) = 10.055, p < 0.001
OCI	3.20 (4.51)	28.75 (12.90)	T(52.93) = 12.44, p < 0.001
POPS	102.80 (32.061)	175.16 (40.058)	T(87) = 9.42, p < 0.001
Illness Duration (years)		13.80 (7.37)	-
Devaluation (Y/N)	34/12	31/13	$\chi^2 < 1$
Within scanner motion (framework displacement)	0.25 (0.089)	0.28 (0.11)	T(88) = 1.23, p = 0.22
Antidepressant medication (Y/N)	0/46	16/28	-
Major Depressive Disorder (Y/N)	0/46	24/20	-
Anxiety disorders (Y/N)	0/46	30/14	-
Eating Disorder (Y/N)	0/46	8/36	-

Brown Obsessive Compulsive Scale (Y-BOCS) score < 16 and predominant hoarding symptoms (Van Ameringen et al., 2014). Additional exclusion criteria for healthy control participants were personal history of Axis I disorder, or substance abuse/dependence, family history (1st/2nd degree relatives) of OCD, neurodevelopmental disorder, schizophrenia and/or schizoaffective disorder, primary psychotic disorder, bipolar disorder, and present PTSD.

2.2. Paradigm

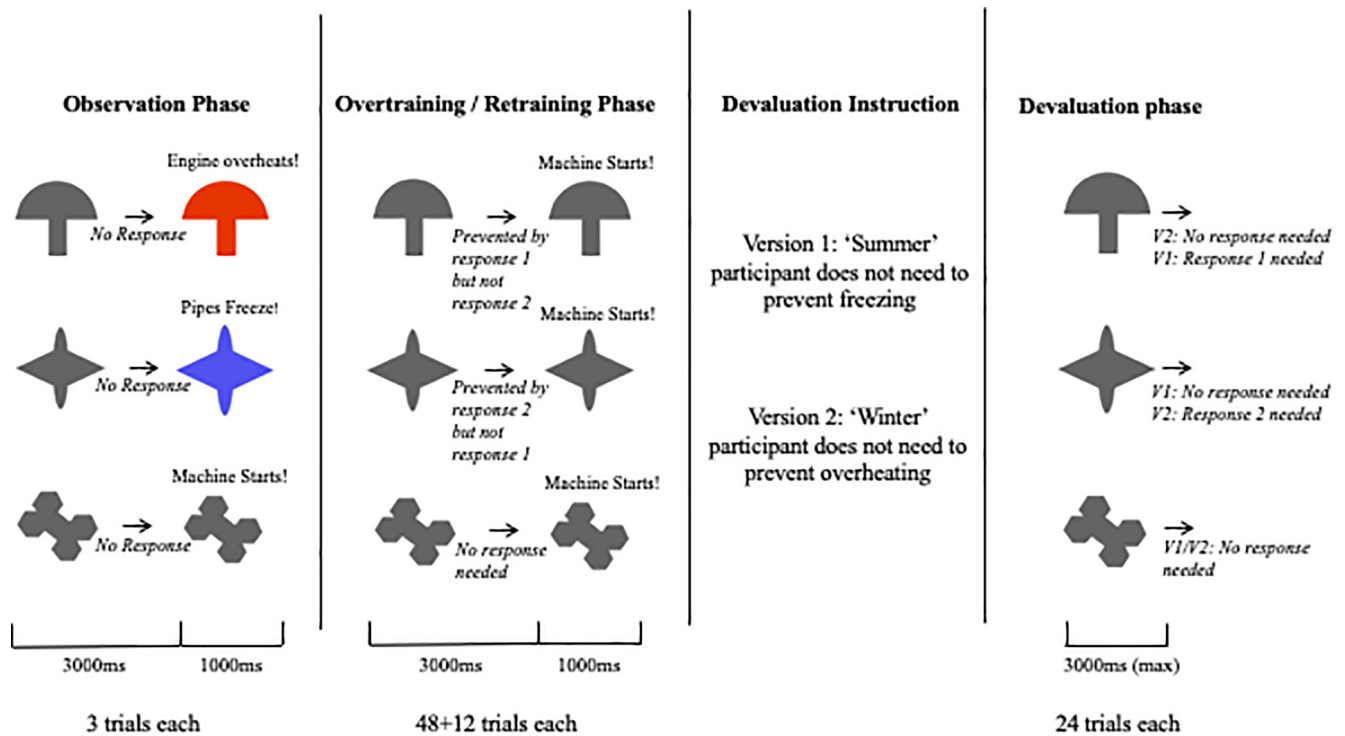
The avoidance paradigm was designed to be analogous to that of

Gillan and colleagues (Gillan et al., 2015), but without the shock. A cover story was provided to frame the requirement for avoidance. Briefly, there were three cues (A/B/C), which represented machines whose function might need to be maintained by responding on one of two buttons (Fig. 1). The buttons could prevent the machines from overheating or freezing, and allow them to run normally. Two versions of the task were generated, employing different cues to represent the machines adapted from those used in Chase et al. (2008) and Shohamy et al. (2006), different response/outcome mappings, and counterbalancing the relationship between the to-be-devalued cue (‘B’) and the overheating/freezing outcome (see below).

**Table 2**

Table describing clinical, demographic and behavioral data for groups, dichotomized by devaluation performance (DV/nDV = devaluer/non-devaluer) and group (OCD/HC). Means are reported, with SDs in parentheses, for all participants. Statistics for the behavioral data (accuracy) are presented in the main text. For the F statistic’s degrees of freedom, all F’s F(1,86), except for POPS due to one missing data point. D = main effect of devaluation; C = main effect of cohort (OCD/HC); D\*C = interaction effect. % cor = percent correct.

	HC DV	HC nDV	OCD DV	OCD nDV	Statistics
Sex (M/F)	11/23	5/7	12/19	4/9	$\chi^2 < 1$
Age at Scan	23.92 (4.018)	22.78 (4.44)	21.91 (3.00)	27.61 (4.64)	D*C F = 14.14, p < 0.001
Level of Education	6.21 (1.30)	5.25 (1.42)	5.71 (1.35)	6.15 (1.52)	D*C F = 4.75, p = 0.032
Predicted Full IQ	112.13 (6.086)	107.00 (8.51)	113.56 (6.12)	109.38 (6.16)	D F = 9.34, p = 0.003
HRSD17 Total Score	1.12 (1.038)	1.67 (1.30)	10.39 (5.59)	9.00 (4.55)	C F = 86.18, p < 0.001
HAMA Total	0.82 (1.086)	1.33 (1.37)	11.10 (6.82)	10.38 (5.68)	C F = 78.50, p < 0.001
YBOCS Total	0.15 (0.86)	0.00 (0.00)	21.03 (3.46)	19.69 (2.72)	C F = 1351.81, p < 0.001
OCTCDQ Harm Avoidance Total Score	1.79 (3.13)	6.67 (7.062)	21.32 (8.91)	22.92 (6.69)	C F = 130.94, p < 0.001 D F = 4.28, p = 0.041
OCTCDQ Incompleteness Total Score	3.47 (3.89)	6.58 (6.61)	21.74 (9.67)	18.85 (10.015)	C F = 72.39, p < 0.001
OCIR Total	2.59 (3.82)	4.92 (5.90)	27.74 (12.29)	31.15 (12.075)	C F = 129.34, p < 0.001
POPS Total	104.27 (30.87)	98.75 (36.25)	178.10 (39.61)	168.15 (41.87)	C F = 69.22, p < 0.001
Difficulty with Change	18.30 (6.61)	17.67 (7.88)	34.77 (8.35)	30.23 (8.40)	C F = 64.073, p < 0.001
Emotional Overcontrol	16.67 (6.30)	14.33 (5.18)	24.68 (8.62)	22.08 (10.029)	C F = 18.93, p < 0.001
Rigidity	28.79 (10.21)	27.67 (13.72)	45.23 (14.41)	44.69 (13.034)	C F = 31.24, p < 0.001
Maladaptive Perfectionism	24.39 (8.78)	23.17 (7.31)	46.68 (12.15)	45.00 (11.94)	C F = 80.74, p < 0.001
Reluctance to Delegate	17.79 (6.70)	17.42 (7.20)	30.58 (9.60)	29.54 (9.43)	C F = 40.54, p < 0.001
Framework displacement	0.25 (0.098)	0.25 (0.062)	0.27 (0.12)	0.30 (0.075)	F’s < 1
OCD Illness Duration (years)	-	-	11.11 (5.91)	20.22 (6.61)	T(42) = 4.51, p < 0.001
Overtraining AB (% cor)	0.92 (0.096)	0.86 (0.13)	0.93 (0.077)	0.77 (0.20)	-
Overtraining C (% cor)	0.98 (0.032)	0.92 (0.15)	0.97 (0.057)	0.94 (0.079)	-
Retraining AB (% cor)	0.99 (0.017)	0.96 (0.072)	0.99 (0.033)	0.97 (0.054)	-
Retraining C (% cor)	1.00 (0)	0.99 (0.024)	1.00 (0)	0.99 (0.023)	-
Devaluation A (% cor)	0.99 (0.018)	0.95 (0.094)	1.00 (0.017)	0.99 (0.018)	-
Devaluation errors B (% cor)	0.98 (0.060)	0.063 (0.13)	0.98 (0.048)	0.048 (0.13)	-
Devaluation errors C (% cor)	0.98 (0.10)	0.98 (0.041)	0.94 (0.23)	0.99 (0.034)	-
Retraining phase RT (AB)	659.88 (144.35)	704.98 (180.55)	697.98 (299.26)	840.53 (526.091)	n.s.
Devaluation phase RT (A)	519.59 (104.28)	630.40 (187.54)	590.74 (261.28)	751.10 (440.86)	D F = 5.56, p = 0.021



**Fig. 1.** Figure describing task structure, with the four phases of the task. Within the observational and training phases, participants learn to avoid two potentially negative outcomes (machines overheating or freezing) associated with two cues (A/B). A third cue (C) does not signal a negative outcome and no response is needed. Durations of the cue and outcome phases are presented, as are the number of trials for each session per cue. Following successful acquisition of the paradigm, a devaluation instruction is then presented: the instruction describes a season change, which means that one machine does not overheat (winter: V2) or one machine does not freeze (summer: V1), depending on the version of the task the participant is given. During the devaluation phase itself, participants are then required to respond or withhold responses to each of the three cues (A/B/C) in light of these new task requirements, during a 3000 ms window. No feedback is provided during this stage.

The paradigm had four phases (see Fig. 1). First, there was an observational phase, in which participants passively observed outcome feedback for each cue: one machine cue 'overheating' (cue A or B, depending on the version), one machine cue 'freezing' (cue B or A, respectively, depending on the version), and the final machine cue functioning normally (cue C). Overheating was reflected by the cue turning red, a bubbling sound occurring and text above reading 'engine overheats!'; freezing was reflected by the machine turning blue, a crackling sound, and text stating 'pipes freeze!'; normal function was reflected by the cue staying the same color (grey) and the engine purring into life, with text above reading 'machine starts!'. No additional reward was provided for the machine working correctly - the malfunctioning of the machines was designed to be more salient than their normal working function. The observation phase included 3 trials of each cue. Cues were displayed for 3000 ms each, before the outcome (i.e. change of color, sound and text) was presented. Outcome feedback lasted 1000 ms; a fixation cross before the machine cue lasted 500 ms and a jittered inter-trial interval between 500 and 1500 ms.

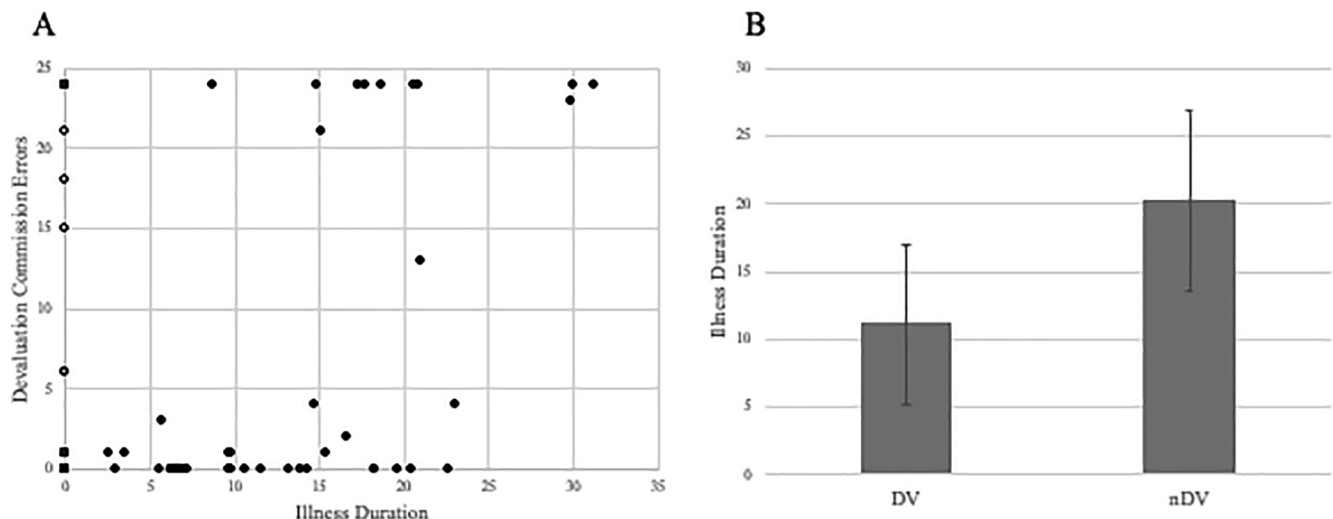
After the observational phase, participants were instructed to try responding on one of two buttons per trial to prevent the malfunction outcome from occurring on each trial, and to keep the machines running. They were not instructed which of the buttons would prevent the malfunction for each stimulus, nor which stimulus did not need a response. Thus, they learned the correct response via trial and error. They were also informed that there was a nominal cost to responding, which was reflected by a '\$\$\$' sign straight after a button was pressed. The cost clarified that the overall task objective was to prevent the machines from malfunctioning, but that responding should not be performed unless necessary to keep the machine running. If participants selected the correct button for the overheating/freezing cues (A/B), the machine would function normally; if they picked the wrong button or did not

respond, it would malfunction as before. This relationship was deterministic: responding correctly always prevented the malfunction from occurring. Participants performed 8 blocks of trials (overtraining), with each block including 6 trials of each cue (A/B/C). The trial structure was the same as the observational phase, with the response needing to be made within the 3000 ms cue presentation duration.

Participants were then transferred to the scanner. After structural and resting state scans, participants performed the task while neuroimaging data were collected. This included a block of reacquisition of the avoidance task ('retraining phase': 12 trials per cue). Next, participants received instructions stating that there was a season change (summer or winter, depending on which version of the task participants performed). In the winter version, participants did not need to respond to stop the relevant machine overheating (cue B), but would need to stop the freezing machine from doing so (cue A); and vice versa for the summer version (with A being the cue for which it was still necessary to respond, and B for which it was not still necessary). After the instruction presentation, participants performed 24 trials on each cue. The trial structure was the same as previous phases, although feedback was omitted at this stage, and the cue disappeared if a response was made.

### 2.3. Neuroimaging data acquisition parameters

Functional neuroimaging data were collected using a multiband sequence, of 400 images per participant (TR = 1500 ms; TE = 31 ms; flip angle = 55°; multiband factor 4; Bandwidth = 1594 Hz/Px; FoV = 220 mm; 60 slices; slice thickness 2 mm/voxel size 2 mm<sup>3</sup>). Distortions were corrected using a fieldmap (TR = 554; TEs = 4.38/6.84 ms; flip angle = 60°; FoV = 220 mm; 60 slices; slice thickness 2 mm/voxel size 2 mm<sup>3</sup>). Structural MRI data were collected using an MP-RAGE sequence (TR = 1520 ms; TE = 3.17 ms; TI = 800;



**Fig. 2.** Figure describing relationships between illness duration and devaluation performance. A) Increasing commission errors on the devalued cue related to OCD illness duration. The distribution of HC group data is displayed via open circles at an illness duration of 0. B) Difference in the mean illness duration between devaluers and non-devaluers, within the OCD group. Error bars reflect standard deviation.

FoV = 258 mm; 176 slices; slice thickness 1 mm/voxel size 1 mm<sup>3</sup>).

#### 2.4. fMRI preprocessing

Neuroimaging data preprocessing was performed with Nipype (Gorgolewski et al., 2011). Preprocessing included realignment, coregistration of structural and functional data, segmentation of the anatomical scan, distortion correction using fieldmaps (missing for 1 participant), normalization of the functional imaging data to MNI space using the DARTEL method, despiking and data smoothing using SUSAN (6 mm kernel).

#### 2.5. fMRI subject-level model

A first level model was constructed using SPM8 including the following components. ‘Valued cues’ included the A/B cues during the retraining phase, and A after devaluation (i.e. cues for which it was necessary to respond). These were modeled to be as long as the associated reaction time for each event. ‘Feedback’ events were the positive feedback following all cues during the retraining phase (unless an error had been made). These were modeled to be 1000 ms long. ‘Devalued cues’ were the B cues, following devaluation. These were modeled to be 3000 ms, unless a response was made – in which case they were modeled to be as long as the reaction time. ‘Control cues’ were the C cues, both during retraining after devaluation. These were modeled to be 3000 ms, unless a response was made. ‘Errors’ were incorrect responses to the valued cues during the retraining phase, and also to the valued cue on the devaluation phase. These were modeled to be as long as the entire trial. A second first level model was constructed to test hypotheses H3 (see below), in which neural responses to valued cues were modelled separately for the retraining phase and devaluation phase.

Analysis focused on model components which had been convolved with the canonical hemodynamic response function (HRF). Temporal and dispersion derivatives were included for each of these factors, as were 6 motion parameters and three noise components obtained using the CompCor method (Behzadi et al., 2007). A 60-second high pass filter was also included.

#### 2.6. fMRI group level model and hypothesis testing

To test our main hypotheses, we focused our analysis on three main

contrasts: devalued versus control cues during the devaluation phase (H2); valued versus control cues during the devaluation phase alone, valued versus control cues during the retraining phase alone, and across both phases (H3); feedback events alone (i.e. versus implicit baseline).

All analyses were performed with an ANOVA model, including group (healthy/OCD) and devaluation performance (devaluer/non-devaluer) as between-subjects factors. All analyses employed a family wise error rate (FWE), peak-level correction of  $p < 0.05$ , but across different search spaces depending on the hypothesis. For H2, we used a combined frontal lobe and subthalamic nucleus region of interest (ROI) defined by the WFU Pick atlas (Maldjian et al., 2003). For H3, we focused on the medial OFC region identified by Gillan and colleagues (Gillan et al., 2015), using an 8 mm sphere around the coordinates (6, 23, -11). Post-hoc t-tests were performed to evaluate significant F tests. Post-hoc tests of the interaction term employed the [1, -1, -1, 1] structure for each of the four subgroups (healthy/non-devaluer; healthy/devaluer; OCD/non-devaluer; OCD/devaluer), and its inverse, to probe a significant interaction. All between-subjects neuroimaging analyses included age, gender, years of education, motion (framewise displacement) and task version as covariates of no interest.

#### 2.7. Behavioral analysis

The key behavioral measure was the number of (erroneous) responses made during devalued cue presentation. In Table 2, this is represented as a ‘percent correct’ measure: the number of correct non-responses out of 24 trials. As this measure was strongly bimodal (see Fig. 2), a dichotomized measure of ‘devaluer’ and ‘non-devaluer’ was generated for analysis. Other retraining phase measures, including number of correct responses on the valued cues, and non-responses on the control cues, were also examined to ensure adequate acquisition of the task rules and contingencies. Both of these analyses were conducted using Chi-squared tests. In addition, we examined performance during the overtraining phase, by computing mean accuracy across all of the 3 cues (A/B/C). These data were arcsin transformed before analysis. An analysis of variance (ANOVA) was performed to examine the effect of group (OCD/HC), while relationships with devaluation performance were conducted using t-tests.

Exploratory analyses were conducted on any significant neuroimaging findings relating to our hypotheses to examine the potential impact on findings of: demographic variables; primary OCD symptom severity (YBOCS/OCL/POPS/age of onset/illness duration); medication, co-

morbid depression and anxiety symptoms (HAMD/HAMA). All correlational analysis employed Spearman's rho ( $\rho$ ), or logistic regression for multivariate prediction of devaluation status.

### 3. Results

#### 3.1. Behavioral data

##### 3.1.1. Overtraining performance

Overtraining phase performance was evaluated by finding the mean of the proportion of correct responses on the cues associated in an avoidance response (i.e. A and B) and the proportion of correct non-responses for C. Three participants did not complete overtraining due to computer errors, and three further participants achieved < 60% performance accuracy on one or more of the cues (3 healthy/3 OCD). These six individuals were not included in the analysis of overtraining data: in the remaining participants ( $n = 84$ ), no main effect of group (HC/OCD) was seen ( $F(1,78) = 1.14, p = 0.29$ ), nor group by cue interaction ( $F < 1$ ). Overtraining performance was related to IQ ( $\rho = 0.37, p = 0.001, n = 84$ ). Descriptive statistics for all 90 participants are reported in Table 2.

##### 3.1.2. Retraining performance

All included participants showed good performance on the within-scanner retraining phase, making few errors on retraining performance (A/B commission errors: 1 participant made 6, all remaining made < 5; C commission errors all < 2: see Table 2). Thus, all included participants aligned their responses with the basic requirements of the paradigm at the retraining phase.

##### 3.1.3. Devaluation phase performance: Hypothesis 1

All included participants continued to respond correctly on cue A during the devaluation phase (1 participant made 8 errors, all others made < 3). Three participants started responding on C after devaluation (> 13 out of 24 responses), despite showing low rates of responding on C during the retraining phase. The remaining participants continued to withhold responding on C (all remaining participants < 4 commission errors: see Table 2).

In terms of the key devaluation phase, responding on cue B was strongly bimodal (see Fig. 2). Individuals categorized as 'devaluers' responded < 7 out of 24 times to the devalued cue, whereas 'non-devaluers' responded > 12 out of 24 times (no participants responded between 7 and 12 times). It should be noted that binomial theorem indicates that 18 or more correct responses out of 24 is associated with a low probability (1.1%) if responding is random (i.e. probability of response and non-response are both 50%), so all individuals within the devaluer group reveal evidence of a specific preference against responding for the devalued cue. Thus, non-devaluers either show a preference for the devalued cue, or are responding around chance. Using this categorization, the ratio of those who did and did not show devaluation within each group was very similar ( $\chi^2 < 1, p = 0.71$ : see Table 2).

Devaluers with OCD were younger than non-devaluers with OCD ( $t(42) = 4.85, p < 0.001$ ), but no such difference was seen in healthy participants ( $t < 1$ : see also Table 2/Fig. 2). We hypothesized that this might be related to duration of OCD illness, and therefore performed a *post hoc* analysis by splitting the groups into healthy participants, participants with OCD with a relatively short (< 14 years) or long illness duration (> 13 years). There was a significant illness duration group by devaluation interaction ( $\chi^2 = 13.84, p < 0.001$ ). Only one of the 22 short illness duration participants did not devalue, whereas 12 of the 22 long illness duration participants did not devalue. The long and short illness duration groups did not differ on clinical metrics (all  $T$ 's < 1) or antidepressant medication use ( $\chi^2 < 1$ ).

Illness duration was strongly correlated with age in the OCD group ( $\rho = 0.62, p < 0.001$ ), making it difficult to rule out age *per se* as a

determinant of devaluation performance. The lack of an age effect in the HC group implies however that OCD illness duration is a critical factor determining devaluation performance. Devaluers of both groups showed a higher IQ than non-devaluers (see below/ Table 2). In a logistic regression analysis including age of onset, illness duration and IQ, duration (Wald = 8.83,  $p = 0.003$ ) was a better predictor of devaluation performance than age of onset (Wald = 3.28,  $p = 0.070$ ).

##### 3.1.4. Relationships among task performance metrics, and with other individual differences measures

Across all participants, overtraining phase performance ( $T(82) = 3.54, p = 0.001$ : not including the 6 low overtraining performers) and IQ ( $T(88) = 3.012, p = 0.003$ ) were lower in individuals who did not devalue compared to those who did, with generally similar patterns of findings across both groups ( $T$ 's = 1.61–2.55). However, the severity of OCD symptoms or secondary comorbidities such as anxiety or depression did not differ between participants with OCD who did not devalue compared to those who did (all  $T$ 's < 1.1). In OCD participants, age of illness onset ( $\rho = 0.33, p = 0.035$ ) and illness duration ( $\rho = -0.39, p = 0.013$ ) were related to overtraining phase performance.

Remarkably, given the very low error rates of retraining phase performance (see Table 2), individuals who made one or more errors on the retraining phase (AB trials) were significantly more likely to be non-devaluers than individuals who made no errors ( $\chi^2 = 7.42, p = 0.006$ ).

#### 3.2. Neuroimaging data

##### 3.2.1. Devalued vs control during devaluation: Hypothesis 2

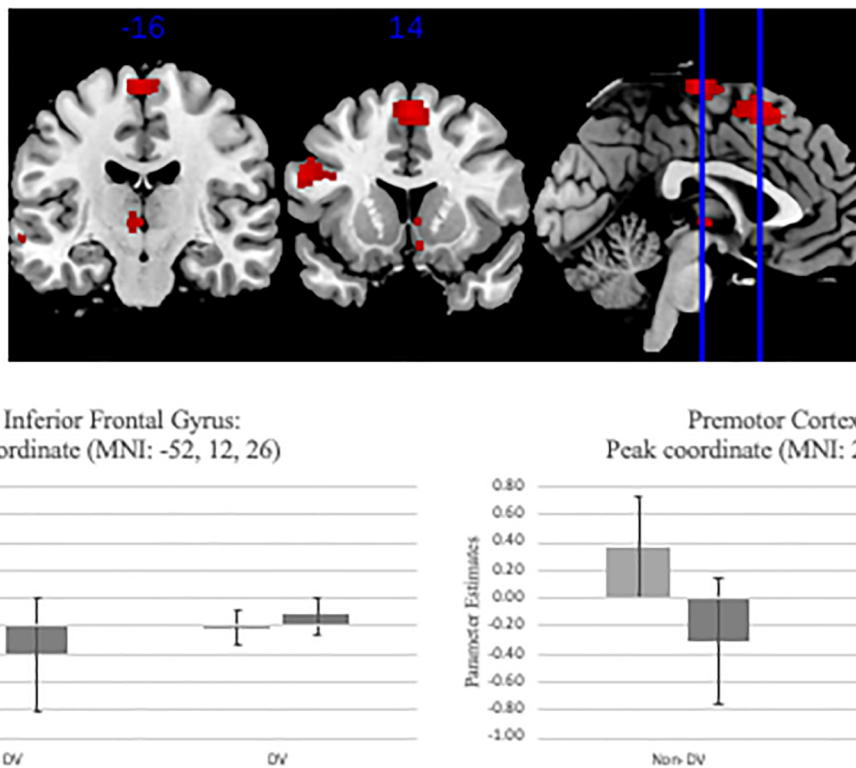
A  $2 \times 2$  ANOVA model revealed a main effect of group (OCD/healthy:  $xyz = 2, 20, 56; F = 28.53, p_{FWE} = 0.008$ ) as well as a group\*devaluer interaction ( $xyz = 2, 20, 56, F = 29.03, p_{FWE} = 0.007$ ) in the premotor cortex. An additional peak associated with the interaction was seen in the motor cortex ( $xyz = -2, -18, 70, F = 23.28, p_{FWE} = 0.046$ ). *Post hoc* interaction tests (see supplement) were implemented reflecting crossover interactions: these confirmed the presence of a significant interaction in premotor cortex ( $xyz = 2, 20, 56; T = 5.39, p_{FWE} = 0.003$ ) and motor cortex ( $xyz = -2, -18, 70, T = 4.82, p_{FWE} = 0.024$ ), as well as the left inferior frontal gyrus ( $xyz = -52, 12, 26; T = 4.69, p_{FWE} = 0.037$ ). The pattern of this interaction (Fig. 3) was as follows: healthy participants and participants with OCD who devalued showed similar activity, but participants with OCD who did not devalue showed lower activation relative to devaluers, while healthy participants who did not devalue showed higher activation relative to devaluers. Significant interaction terms with an opposite pattern were not observed. *Post hoc* comparison of all healthy participants and OCD yielded significant differences in the premotor cortex (healthy > OCD:  $xyz = 2, 20, 56, T = 5.34, p_{FWE} = 0.004$ ) and motor cortex (healthy > OCD:  $xyz = -2, -18, 72, T = 4.69, p_{FWE} = 0.036$ ).

A main effect of devaluer group (non-devaluers > devaluers) was observed in the cingulate cortex, left motor cortex and frontal operculum (Table 3). These activations reflected greater activity in the non-devaluer group compared to the devaluer group, and were consistent with expected neural correlates of higher motor output for devalued cues in non-devaluers.

A *post-hoc* analysis of the premotor and left IFG activations revealed no significant relationships with antidepressant medications. In addition, no significant relationships were observed with symptoms of depression or anxiety in the whole sample, or HC/OCD separately (all  $p$ 's > 0.05).

##### 3.2.2. Further *post-hoc* tests for Hypothesis 2

In contrast to the tests reported in the main part of the manuscript, none of the pairwise tests between OCD and healthy within the devaluer and non-devaluer groups were significant within the frontal lobe



**Fig. 3.** Figure showing regions showing significant group by devaluation interaction effects on the devalued cue versus control cue contrast (thresholded at  $p < 0.0005$  uncorrected,  $k = 10$ , no masking used for display purposes). The inserts display the pattern of findings in regions of interest: premotor cortex and left inferior frontal gyrus for non-devaluers (non-DV) and devaluers (DV), healthy (light grey) and OCD (dark grey). Error bars reflect standard deviation.

**Table 3**  
Coordinates of the main effect of Devaluer/Non-Devaluer on devalued versus control cues.

Region	MNI coordinates (x y z)	F	p_FWE
Left Postcentral Gyrus	-48-24 50	54.15	$P < 0.001$
Left Precentral Gyrus	-36-14 54	49.92	
Left Postcentral Gyrus	-42-22 44	48.37	
Anterior/Mid Cingulate Cortex	-8 0 52	24.05	$P = 0.036$
Left Insula (Frontal Operculum)	-40 20 6	23.20	$P = 0.048$

for the devalued versus control cue contrast. However, within the devaluer group alone, OCD showed greater activity within the left inferior frontal gyrus than healthy controls ( $xyz = -48, 6, 26$ ;  $T = 3.38$ ,  $p = 0.001$  uncorrected). In the non-devaluer group alone, healthy controls showed greater activity in the left inferior frontal gyrus compared to OCD ( $xyz = -50, 12, 26$ ;  $T = 4.30$ ,  $p < 0.001$  uncorrected), and also the premotor cortex ( $xyz = 2, 20, 56$ ;  $T = 5.47$ ,  $p_{FWE} = 0.002$ ) and motor cortex ( $xyz = -2, -18, 70$ ;  $T = 4.85$ ,  $p_{FWE} = 0.021$ ).

**3.2.3. Valued vs control during devaluation, retraining, and across both phases: Hypothesis 3**

Using ANOVA, no significant main effects of group (healthy/OCD), devaluer status, nor interactions were observed using any contrasts. However, using planned T-tests, the medial orbitofrontal cortex ( $xyz = 8, 26, -6$ ,  $T = 3.49$ ,  $p_{FWE} = 0.009$ ) showed a group difference (healthy > OCD) in the retraining phase consistent with previous work (Gillan et al., 2015). Across all participants, no significant relationships were found between these regions' activity and clinical variables. In addition, non-devaluers showed higher activity than devaluers in medial OFC to valued cues during the devaluation phase ( $xyz = 10, 26, -16$ ;  $T = 3.13$ ,  $p_{FWE} = 0.022$ ) and across both phases ( $xyz = 10, 22, -16$ ;  $T = 2.92$ ,  $p_{FWE} = 0.036$ ).

**4. Discussion**

In the present study, we examined the neural correlates of habitual avoidance behavior in participants with OCD and healthy participants. We used a novel avoidance devaluation task involving rule-based rather than shock avoidance, which was otherwise similar to a previous design (Gillan et al., 2015). We tested four hypotheses regarding the behavioral and neural basis of persistent avoidance in OCD. Our primary hypothesis was not directly confirmed: similar performance was observed across both groups during the devaluation phase of the task. However, within participants with OCD, a *post hoc* analysis revealed that participants with a longer illness duration/earlier illness onset showed relatively poorer performance during the devaluation phase. There were no significant relationships with OCD symptom severity or medication, further suggesting that illness duration rather than illness severity is related to responding to devalued cues. In the OCD group, age and illness duration were strongly correlated, and thus both measures predicted devaluation performance in the OCD group. The lack of an age effect in the healthy group implies that OCD illness duration is probably the relevant factor influencing devaluation performance. This is largely consistent with prior work: previous studies of Gillan and colleagues (2015, 2014, 2011) recruited older participants (mean age: ~37–43) than were tested in the present study. A study of adolescents with OCD obtained evidence of impairment of goal-directed behavior (Gottwald et al., 2018), although this was primarily through a slips of action test and differential outcome learning. Thus, across the present study and prior work, the age of the participants, and the duration of OCD illness, may be an important determinant of devaluation performance, and might explain discrepancies across studies. Specifically, the devaluation test may be relatively insensitive to goal-directed deficits seen in earlier stages of the illness, compared to other tests of cue-outcome learning, but these deficits may become more severe and apparent across different metrics later in the illness. It should be noted that the association of devaluation performance with illness duration

was not predicted, and emerged via *post hoc* follow-up analysis of a demographic confound (age). Thus, further studies are necessary to confirm its importance, but, regardless of this observation, illness duration remains a potential account of discrepancies between studies given the substantial age differences between the present study and prior studies of Gillan and colleagues.

Recent evidence has suggested that predicted learning impairments in compulsivity are not seen across all patients with OCD, but are specifically related to a dimension of compulsivity which is enriched in OCD (Gillan et al., 2019). Our findings might be consistent with this proposal, if symptoms of compulsivity are also greater in high illness duration individuals and/or individual who show goal-directed deficits. Indeed, the OCTCDQ harm avoidance and OCI-R scales showed a small numerical tendency to be higher in individuals (both HC and OCD) who did not devalue, which may represent preliminary support for such a proposal. Overall, if OCD symptoms do change with illness duration, this may have treatment implications, including necessitating regular monitoring, and tailored interventions depending on the time since symptom onset.

Participants with OCD who did not devalue successfully showed significantly reduced activity compared to other groups in several regions, including premotor cortex and left inferior frontal gyrus. Although previous work has suggested that these regions are important for cognitive and response inhibition (Aron et al., 2007; Hung et al., 2018; Picton et al., 2007), a decoding analysis of the activated clusters regions did not provide strong evidence supporting the inhibition hypothesis, perhaps due to the left hemispheric focus of the IFG activation. Instead, stronger relationships were observed with language and general cognitive processes including reasoning, attention and working memory. Thus, our findings do not strongly support the inhibition hypothesis, and could reflect an altered representation and implementation of explicit rule-based decision making (Ashby and Maddox, 2005) across OCD and HC.

Neural activity to devalued cues in these regions showed a pattern that might correspond to the classical inverted-U function (Cools and D'Esposito, 2011), insofar as optimal performance (i.e. successful devaluation) was generally associated with intermediate levels of activity in OCD and healthy participants. A meta-analysis of Norman and colleagues (2016) of volumetric and inhibition-related alterations in OCD identified differences in both of these regions (left IFG, premotor cortex). These regions were shown to have reduced grey matter in OCD compared to controls, and a mixed pattern of decreased and increased inhibition-related activation. A further mega-analysis by Norman and colleagues (2019) identified *heightened* activation in the premotor cortex during error processing and inhibitory control, which was accompanied by performance deficits (RT/error rate). Together, these findings suggest that the premotor cortex may be relatively inefficient in OCD and susceptible to both over- and under-activation, perhaps due to reduced grey matter integrity. Efforts to intervene in this region's function in OCD have demonstrated some therapeutic benefit (Berlim et al., 2013), and our findings are broadly consistent with this strategy.

Another hypothesis was tested in which impaired encoding of valued cues would be related to a reduced devaluation in participants with OCD. Gillan and colleagues (2015) demonstrated a cross-over interaction in regions of medial orbitofrontal cortex relating to overtraining. Our findings were consistent with Gillan et al.'s findings, insofar as activity to valued cues during retraining in medial orbitofrontal cortex was lower in participants with OCD relative to healthy participants. Gillan et al. observed a similar difference towards the end of their overtraining procedure. In our data, however, and contrary to Hypothesis 3, non-devaluers showed greater activity in this region coupled to valued cues during the devaluation phase compared with devaluers. Together, our findings, and those of Gillan et al. (2015), do not support a simple relationship between orbitofrontal activity and goal directed behavior, but are consistent with the proposal that alterations in medial orbitofrontal cortex value encoding characterize OCD.

A limitation of our study was that some participants were medicated, although none of our main behavioral or neural findings were associated with psychotropic medication in participants with OCD. Illness duration was the only measure of continuous clinical symptoms that was clearly associated with neural or behavioral measures, independent of group (healthy/OCD)-related confounding. Future studies might aim to replicate our findings in independent samples to establish the extent to which goal-directed avoidance learning might reflect clinical differences between early and late onset OCD (Taylor, 2011). Our data suggests that the key factor is the duration of experiencing OCD symptoms, rather than the time of their onset, so a design would be needed which could orthogonalize illness duration from age of onset. In addition, our study lacked a specific self-report measure of compulsivity symptoms (Chamberlain and Grant, 2018), which might have helped interpret individual differences more specifically.

In terms of limitations of the paradigm, previous work suggests the importance of reinforcer magnitude for habit formation (Nelson and Killcross, 2006; Nordquist et al., 2007), and this was an obvious difference between our paradigm and that employed by Gillan et al. (2015) which might benefit from further investigation. Another limitation is that the devaluer and non-devaluer groups were confounded with respect to motor responses on the devalued versus control contrast. This was reflected in a differential pattern of motor cortex activations, which was consistent with what would be expected from such a confound (i.e. heightened left motor cortex activation in the non-devaluer group). We do not believe that this confound affected group comparisons or group-related interactions effects, however, due to the balancing of non-devaluers across both groups, but further work can examine this point in more detail. Further limitations of the paradigm are discussed in the [supplement](#).

In summary, we show that reduced devaluation in the context of an avoidance paradigm is associated with longer illness duration in OCD, and alterations in activation within premotor and inferior frontal regions coupled to devalued cues. We also observed reduced medial orbitofrontal cortex activation in OCD compared to HC coupled to valued cues, consistent with previous work. Our findings also highlight increased feedback-related responses in the caudate associated with reduced devaluation, and reveal a potential pathway by which an impairment of goal-directed learning might be manifest in learning performance and feedback-related neural activity. Overall, our findings demonstrate how conceptual replications can provide complementary information which may yield further insights into the nature of the relationship between habitual behavior and compulsivity in OCD.

### Conflicts of interest

None of the authors declare any financial or other conflicts of interest that might have biased the work.

### Funding

This study was supported by National Institute of Mental Health grant P50 MH106435 (PI: Dr. Haber).

### CRediT authorship contribution statement

**Henry W. Chase:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Supervision, Writing - original draft, Writing - review & editing. **Simona Graur:** Investigation, Methodology. **Amelia Versace:** Investigation, Methodology, Writing - review & editing. **Tsafrir Greenberg:** Investigation, Methodology. **Lisa Bonar:** Data curation. **Robert Hudak:** Investigation, Methodology. **Gregory J. Quirk:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing - review & editing. **Ben D. Greenberg:** Conceptualization, Funding acquisition, Project administration, Supervision. **Steven A. Rasmussen:** Conceptualization, Funding



acquisition, Project administration, Supervision. **Suzanne N. Haber:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision. **Mary L. Phillips:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing - original draft, Writing - review & editing.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2020.102404>.

## References

- Apergis-Schoute, A.M., Gillan, C.M., Fineberg, N.A., Fernandez-Egea, E., Sahakian, B.J., Robbins, T.W., 2017. Neural basis of impaired safety signaling in Obsessive Compulsive Disorder. *Proc. Natl. Acad. Sci. U S A* 114, 3216–3221.
- Aron, A.R., Behrens, T.E., Smith, S., Frank, M.J., Poldrack, R.A., 2007. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* 27, 3743–3752.
- Aron, A.R., Poldrack, R.A., 2006. Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. *J. Neurosci.* 26, 2424–2433.
- Aron, A.R., Robbins, T.W., Poldrack, R.A., 2014. Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn Sci* 18, 177–185.
- Ashby, F.G., Maddox, W.T., 2005. Human category learning. *Annu. Rev. Psychol.* 56, 149–178.
- Balleine, B.W., Dezfouli, A., 2019. Hierarchical action control: adaptive collaboration between actions and habits. *Front. Psychol.*
- Balleine, B.W., Dickinson, A., 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419.
- Banca, P., Voon, V., Vestergaard, M.D., Filipiak, G., Almeida, I., Pochinho, F., Relvas, J., Castelo-Branco, M., 2015. Imbalance in habitual versus goal directed neural systems during symptom provocation in obsessive-compulsive disorder. *Brain* 138, 798–811.
- Bartra, O., McGuire, J.T., Kable, J.W., 2013. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76, 412–427.
- Behzadi, Y., Restom, K., Liu, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37, 90–101.
- Berlim, M.T., Neufeld, N.H., Van den Eynde, F., 2013. Repetitive transcranial magnetic stimulation (rTMS) for obsessive-compulsive disorder (OCD): an exploratory meta-analysis of randomized and sham-controlled trials. *J. Psychiatr. Res.* 47, 999–1006.
- Blair, J.R., Spreen, O., 1989. Predicting premorbid IQ: a revision of the national adult reading test. *Clin. Neuropsychol.* 3, 129–136.
- Britton, J.C., Rauch, S.L., Rosso, I.M., Killgore, W.D., Price, L.M., Ragan, J., Chosak, A., Hezel, D.M., Pine, D.S., Leibenluft, E., Pauls, D.L., Jenike, M.A., Stewart, S.E., 2010. Cognitive inflexibility and frontal-cortical activation in pediatric obsessive-compulsive disorder. *J. Am. Acad. Child Adolesc. Psychiatry* 49, 944–953.
- Chamberlain, S.R., Grant, J.E., 2018. Initial validation of a transdiagnostic compulsivity questionnaire: the Cambridge-Chicago Compulsivity Trait Scale. *CNS Spectr.* 23, 340–346.
- Chase, H.W., Clark, L., Myers, C.E., Gluck, M.A., Sahakian, B.J., Bullmore, E.T., Robbins, T.W., 2008. The role of the orbitofrontal cortex in human discrimination learning. *Neuropsychologia* 46, 1326–1337.
- Cools, R., D'Esposito, M., 2011. Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol. Psychiatry* 69, e113–125.
- Eagle, D.M., Baunez, C., Hutcherson, D.M., Lehmann, O., Shah, A.P., Robbins, T.W., 2008. Stop-signal reaction-time task performance: role of prefrontal cortex and subthalamic nucleus. *Cereb. Cortex* 18, 178–188.
- Foa, E.B., Huppert, J.D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., Salkovskis, P.M., 2002. The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychol. Assess.* 14, 485–496.
- Gillan, C.M., Apergis-Schoute, A.M., Morein-Zamir, S., Urcelay, G.P., Sule, A., Fineberg, N.A., Sahakian, B.J., Robbins, T.W., 2015. Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am. J. Psychiatry* 172, 284–293.
- Gillan, C.M., Kalanithroff, E., Evans, M., Weingarden, H.M., Jacoby, R.J., Gershkovich, M., Snorrason, I., Campeas, R., Cervoni, C., Crimarco, N.C., Sokol, Y., Garnaat, S.L., McLaughlin, N.C.R., Phelps, E.A., Pinto, A., Boisseau, C.L., Wilhelm, S., Daw, N.D., Simpson, H.B., 2019. Comparison of the association between goal-directed planning and self-reported compulsivity vs obsessive-compulsive disorder diagnosis. *JAMA Psychiatry* 1–10.
- Gillan, C.M., Morein-Zamir, S., Urcelay, G.P., Sule, A., Voon, V., Apergis-Schoute, A.M., Fineberg, N.A., Sahakian, B.J., Robbins, T.W., 2014. Enhanced avoidance habits in obsessive-compulsive disorder. *Biol. Psychiatry* 75, 631–638.
- Gillan, C.M., Pappmeyer, M., Morein-Zamir, S., Sahakian, B.J., Fineberg, N.A., Robbins, T.W., de Wit, S., 2011. Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am. J. Psychiatry* 168, 718–726.
- Gillan, C.M., Robbins, T.W., Sahakian, B.J., van den Heuvel, O.A., van Wingen, G., 2016. The role of habit in compulsivity. *Eur. Neuropsychopharmacol.* 26, 828–840.
- Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* 5, 13.
- Gottwald, J., de Wit, S., Apergis-Schoute, A.M., Morein-Zamir, S., Kaser, M., Cormack, F., Sule, A., Limmer, W., Morris, A.C., Robbins, T.W., Sahakian, B.J., 2018. Impaired cognitive plasticity and goal-directed control in adolescent obsessive-compulsive disorder. *Psychol. Med.* 48, 1900–1908.
- Gu, B.M., Park, J.Y., Kang, D.H., Lee, S.J., Yoo, S.Y., Jo, H.J., Choi, C.H., Lee, J.M., Kwon, J.S., 2008. Neural correlates of cognitive inflexibility during task-switching in obsessive-compulsive disorder. *Brain* 131, 155–164.
- Hamilton, M., 1959. The assessment of anxiety states by rating. *Br. J. Med. Psychol.* 32, 50–55.
- Hamilton, M., 1960. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 23, 56–62.
- Hung, Y., Gaillard, S.L., Yarmak, P., Arsalidou, M., 2018. Dissociations of cognitive inhibition, response inhibition, and emotional interference: voxelwise ALE meta-analyses of fMRI studies. *Hum. Brain Mapp.* 39, 4065–4082.
- Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239.
- Morris, L.S., Baek, K., Voon, V., 2017. Distinct cortico-striatal connections with subthalamic nucleus underlie facets of compulsivity. *Cortex* 88, 143–150.
- Nelson, A., Killcross, S., 2006. Amphetamine exposure enhances habit formation. *J. Neurosci.* 26, 3805–3812.
- Nordquist, R.E., Voorn, P., de Mooij-van Malsen, J.G., Joosten, R.N., Pennartz, C.M., Vanderschuren, L.J., 2007. Augmented reinforcer value and accelerated habit formation after repeated amphetamine treatment. *Eur. Neuropsychopharmacol.* 17, 532–540.
- Norman, L.J., Carlisi, C., Lukito, S., Hart, H., Mataix-Cols, D., Radua, J., Rubia, K., 2016. Structural and functional brain abnormalities in attention-deficit/hyperactivity disorder and obsessive-compulsive disorder: a comparative meta-analysis. *JAMA Psychiatry* 73, 815–825.
- Norman, L.J., Taylor, S.F., Liu, Y., Radua, J., Chye, Y., De Wit, S.J., Huyser, C., Karahanoglu, F.I., Luks, T., Manoach, D., Mathews, C., Rubia, K., Suo, C., van den Heuvel, O.A., Yucel, M., Fitzgerald, K., 2019. Error processing and inhibitory control in obsessive-compulsive disorder: a meta-analysis using statistical parametric maps. *Biol. Psychiatry* 85, 713–725.
- Picton, T.W., Stuss, D.T., Alexander, M.P., Shallice, T., Binns, M.A., Gillingham, S., 2007. Effects of focal frontal lesions on response inhibition. *Cereb. Cortex* 17, 826–838.
- Pinto, A., 2011. Introducing the POPS (pathological obsessive-compulsive personality scale): Derivation and exploratory factor analysis. Paper presented at the annual meeting of the Society for Personality Assessment, Cambridge, MA.
- Rasmussen, S.A., Eisen, J.L., 1992. The epidemiology and differential diagnosis of obsessive compulsive disorder. *J. Clin. Psychiatry* 53 (Suppl.), 4–10.
- Remijne, P.L., Nielen, M.M., van Balkom, A.J., Cath, D.C., van Oppen, P., Uylings, H.B., Veltman, D.J., 2006. Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. *Arch. Gen. Psychiatry* 63, 1225–1236.
- Remijne, P.L., Nielen, M.M., van Balkom, A.J., Hendriks, G.J., Hoogendijk, W.J., Uylings, H.B., Veltman, D.J., 2009. Differential frontal-striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive-compulsive disorder. *Psychol. Med.* 39, 1503–1518.
- Roth, R.M., Saykin, A.J., Flashman, L.A., Pixley, H.S., West, J.D., Mamourian, A.C., 2007. Event-related functional magnetic resonance imaging of response inhibition in obsessive-compulsive disorder. *Biol. Psychiatry* 62, 901–909.
- Sadacca, B.F., Wied, H.M., Lopatina, N., Saini, G.K., Nemirovsky, D., Schoenbaum, G., 2018. Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task. *Elife* 7.
- Shohamy, D., Myers, C.E., Gekhman, K.D., Sage, J., Gluck, M.A., 2006. L-dopa impairs learning, but spares generalization, in Parkinson's disease. *Neuropsychologia* 44, 774–784.
- Taylor, S., 2011. Early versus late onset obsessive-compulsive disorder: evidence for distinct subtypes. *Clin. Psychol. Rev.* 31, 1083–1100.
- Van Ameringen, M., Patterson, B., Simpson, W., 2014. DSM-5 obsessive-compulsive and related disorders: clinical implications of new criteria. *Depress Anxiety* 31, 487–493.
- Wessel, J.R., O'Doherty, J.P., Berkebile, M.M., Linderman, D., Aron, A.R., 2014. Stimulus devaluation induced by stopping action. *J. Exp. Psychol. Gen.* 143, 2316–2329.