Check for updates

Amygdala response predicts clinical symptom reduction in patients with borderline personality disorder: A pilot fMRI study

Dirk E. M. Geurts[1,2], Thom J. Van den Heuvel[2,3],
Quentin J. M. Huys[4], Robbert J. Verkes[2,5] and
Roshan Cools[1,2]

[1]Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, Netherlands, [2]Department of Psychiatry, Radboud University Medical Center, Nijmegen, Netherlands, [3]Department of Scelta, Expert Centre for Personality Disorders, GGNet, Nijmegen, Netherlands, [4]Mental Health Neuroscience Department, Division of Psychiatry and Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Institute of Neurology, University College London, London, United Kingdom, [5]Kairos Center for Forensic Psychiatry, Pro Persona Mental Health, Nijmegen, Netherlands

Borderline personality disorder (BPD) is a prevalent, devastating, and heterogeneous psychiatric disorder. Treatment success is highly variable within this patient group. A cognitive neuroscientific approach to BPD might contribute to precision psychiatry by identifying neurocognitive factors that predict who will benefit from a specific treatment. Here, we build on observations that BPD is accompanied by the enhanced impact of the aversive effect on behavior and abnormal neural signaling in the amygdala. We assessed whether BPD is accompanied by abnormal aversive regulation of instrumental behavior and associated neural signaling, in a manner that is predictive of symptom reduction after therapy. We tested a clinical sample of 15 female patients with BPD, awaiting dialectical behavior therapy (DBT), and 16 matched healthy controls using fMRI and an aversive Pavlovian-to-instrumental transfer (PIT) task that assesses how instrumental behaviors are influenced by aversive Pavlovian stimuli. Patients were assessed 1 year after the start of DBT to quantify changes in BPD symptom severity. At baseline, behavioral aversive PIT and associated neural signaling did not differ between groups. However, the BOLD signal in the amygdala measured during aversive PIT was associated with symptom reduction at 1-year follow-up: higher PIT-related aversive amygdala signaling before treatment was associated with reduced clinical improvement at follow-up. Thus, within the evaluated group of BPD patients, the BOLD signal in the amygdala before treatment was related

to clinical symptom reduction 1 year after the start of treatment. The results suggest that less PIT-related responsiveness of the amygdala increases the chances of treatment success. We note that the relatively small sample size is a limitation of this study and that replication is warranted.

Introduction

Borderline personality disorder (BPD) is a prevalent and devastating psychiatric disorder associated with severe functional impairments and high mortality rates (American Psychiatric Association, 2000; Grant et al., 2008; Bolton and Robinson, 2010). Costs for society are high due to heavy use of expensive health care resources and persistent lack of productivity (Wunsch et al., 2014). Optimizing care for this patient group is of major importance (Gunderson, 2009).

Although several psychotherapeutic treatments exist for BPD, the response is highly variable and treatment effects are modest overall (Stoffers et al., 2012). For example, 27-35% of patients continue to have admissions, harm themselves, and conduct suicidal gestures (Lana and Fernández-San Martín, 2013). Only a few general predictors of outcome have been reported (Barnicot et al., 2012). The discovery of new outcome predictors is essential for the advancement of the field of personalized psychiatry. Neurocognitive mechanistic research might identify key predictors of available treatment outcomes and thus mitigate the large variability in treatment efficacy (Jones et al., 2015; Heinz et al., 2016; Huys et al., 2021). We report a proof-of-principle, pilot study focused on the relation between BPD symptom reduction over 1 year and affect-related neural processing, measured prior to the start of 1 year of dialectical behavior therapy (DBT).

Maladaptive and inflexible behavior in BPD has been argued to reflect derailed interaction between two principle controllers of human behavior, i.e., an instrumental and a Pavlovian controller (Hallquist et al., 2018). Instrumental control allows us to flexibly optimize our chances to achieve specific goals by learning what to do when (based on stimulus–action–outcome learning or operant conditioning). The Pavlovian system regulates inflexible, "automatic," motivational responses in reaction to external and internal emotional stimuli (based on stimulus–outcome learning or classical conditioning). In the context of BPD, the interaction between Pavlovian and instrumental control, the so-called Pavlovian-instrumental transfer (PIT), is particularly worth investigating, as dysregulation of this interaction has been related to heightened impulsivity (e.g., behavioral activation instead

of inhibition by aversive contextual cues) (Breland and Breland, 1961; Guitart-Masip et al., 2014; Hinojosa-Aguayo and González, 2020; Geurts et al., 2022), increased influence of emotional/motivational states (e.g., hampering effective goal pursuit) (Dolan and Dayan, 2013; Watson et al., 2014) and interpersonal hypersensitivity (cf. Hallquist et al., 2018), and a combination of symptomatology lying at the core of BPD (Gunderson, 2009). Here, we will probe whether BPD is indeed characterized by an aberrant influence of the Pavlovian system by assessing PIT in BPD patients and healthy controls. Critically, we will explore within the group of BPD patients whether neurocognitive correlates of PIT are related to symptom reduction over 1 year of DBT.

Relying on the biosocial model of emotion regulation, DBT is one of the leading evidence-based psychotherapies for BPD with the main focus on skillfully regulating impulsive and emotion-driven behavior (Linehan, 1993). DBT teaches how aversive motivational tendencies can be accepted and dealt with skillfully through the training of skills like mindfulness, distress tolerance, emotion regulation, and interpersonal effectiveness. Thus, DBT might help optimize the interaction between aversive motivational (Pavlovian) influences and instrumental behavior. In this manuscript, we assess the 'vulnerability' of instrumental, goal-appropriate behaviors to disruptions by aversive Pavlovian conditioned stimuli (CS). For this purpose, we used a previously validated behavioral PIT task that allows us to quantify the impact of motivational cues on instrumental decision-making.

Specifically, we measure aversive PIT, which refers to the observation that aversive instrumental actions, such as inhibition and withdrawal, are potentiated in the context of aversive Pavlovian CS, i.e., stimuli that predict aversive outcomes. Thus, aversive Pavlovian CS have been shown to inhibit instrumental approach actions (i.e., aversive inhibition) and to enhance instrumental withdrawal actions (Huys et al., 2011; Geurts et al., 2013a). Accumulating evidence from experimental studies with animals and healthy humans (Talmi et al., 2008; Prevost et al., 2012; Geurts et al., 2013a) and patients (Garbusow et al., 2016; van Timmeren et al., 2020) demonstrates the involvement of (prefrontal) limbic circuitry in PIT, including the ventral striatum and amygdala (Cardinal et al., 2002; Talmi et al., 2008; Balleine and Doherty, 2009;

Prevost et al., 2012; Geurts et al., 2013a; Ly et al., 2014). The involvement of the amygdala is particularly relevant in the context of the current study, because the amygdala has also been central to neurocognitive theories and empirical research on BPD (Minzenberg et al., 2007; Hazlett et al., 2012; Soloff et al., 2017; Degasperi et al., 2021). For example, a recent meta-analysis reported functional hyperactivity of the left amygdala during aversive vs. neutral stimuli, as well as smaller gray matter volume of the amygdala in BPD (Schulze et al., 2016, 2019). This amygdala hyperactivation has been proposed to reflect the deviant salience of negative emotional stimuli and to be remediated by psychotropic medication (Schulze et al., 2016) and psychotherapy (Iskric and Barkley-Levenson, 2021) in BPD. We note that it is unclear whether remediation of amygdala hyperactivity is related to specific treatments or whether it is a general prerequisite for recovery from borderline symptomatology. Notwithstanding this ambiguity, evidence shows that effects of DBT are also associated with changes in blood oxygen level-dependent (BOLD) signal in the amygdala (Schnell and Herpertz, 2007; Krause-Utz et al., 2014; Salvador et al., 2016; Iskric and Barkley-Levenson, 2021). Here, we build on these previous findings by assessing the hypothesis that BPD is accompanied by abnormalities in aversive PIT and associated BOLD signal in the amygdala. Moreover, we ask whether aversive PIT and related amygdala signal before the start of therapy is associated with symptom reduction after treatment (Schmitt et al., 2016; cf. Schmitgen et al., 2019).

Thus, we hypothesize that borderline symptomatology might result from an imbalance between two major control systems of behavior: the motivational, reactive Pavlovian system on the one hand and a goal-oriented, instrumental system on the other. We explore this hypothesis by first investigating differences in baseline performance on a behavioral PIT task between healthy controls and BPD patients. Based on the above findings, we hypothesized that, relative to controls, BPD patients exhibit the enhanced impact of aversive Pavlovian CS on instrumental behavior, that is, greater aversive PIT (i.e., increased behavioral inhibition and withdrawal). Furthermore, we expect increased PIT-related BOLD signal in BPD relative to controls in the amygdala. Critically, we expect that the between-subject differences in amygdala response are related to symptom reduction across 1 year of DBT in the BPD group.

Materials and methods

Participants

To maximize external validity, we aimed for a patient sample that would represent patients treated in general mental health practice as closely as possible (Hoertel et al., 2015). Therefore, all patients who were enrolled in the pre-treatment phase of a 1-year DBT program at the Radboud University Medical Centre between March 2012 and March 2013 ($n = 29$) were invited to participate in this study. Twenty-three patients volunteered. Imaging datasets were obtained for 15 patients (all women), and clinical outcome measures after treatment were obtained for 14 of these patients (see **Supplementary materials** for details on inclusion). In addition, 16 healthy (MINI-plus) controls matched for gender and age were recruited per advertisement (for group demographics and questionnaire scores, see **Table 1**, and for comorbidity and medication use of the BPD group, see **Supplementary Table 1**). The local Medical Ethical Committee approved the study (NL36001.091.11), and consent was obtained from all participants.

Procedure

All patients enrolled in the pre-treatment phase of DBT were invited to attend three sessions: the first was a screening session, the second was a pre-treatment scan session just before treatment, and the third was a post-treatment assessment.

Screening session

During the screening session, participants received a full diagnostic structured interview, which included the MINI-plus international neuropsychiatric interview and the Structured Clinical Interview for DSM-IV Axis II disorders (SCID-II), administered by a senior resident in psychiatry (author DG). To familiarize subjects during the first visit with the scanning environment and procedures, we employed a short scan session of about 15 min during which a structural MRI scan was obtained and subjects were familiarized with the instructions and instrumental and Pavlovian training stages in the scanner.

Pre-treatment scan session

During the second visit, before treatment started, subjects completed several questionnaires (**Table 1**), of which the Borderline Personality Disorder Checklist (BPD47) measuring the symptom severity was of primary interest. Before entering the scanner, instructions on the computer task were repeated orally. After receiving the instructions for a third time, now projected on the scanner screen, they started the PIT paradigm (**Figure 1**). After a 15-min break, subjects performed a short neuropsychological test battery (**Table 1**).

Treatment

Participants received a 1-year group version of the standard DBT protocol (Linehan, 1993; Gutteling et al., 2012) divided into

the standard 4 weekly components (DBT group psychotherapy, groups skills training, 24/7 telephone coaching, and a therapist consultation team). The program differed from standard DBT only in that the weekly psychotherapy sessions were offered not individually but in groups. All DBT strategies (dialectics, behavior chain analysis, radical acceptance strategies of validation and mindfulness, contingency management, exposure, cognitive restructuring, and skills training) were used across all components addressing the five functions of DBT (increasing behavioral capabilities, improving motivation for skillful behavior, generalization of skills to the natural environment, reinforcement of functional over dysfunctional behavior, and enhancing therapist effectiveness) and were performed by well-trained DBT therapists and skill trainers. Although more elaborate research is needed to show that scaled versions as described above are as effective as standard DBT, Gutteling et al. (2012) demonstrated evidence that suggests that this scaled version of DBT is as effective as standard DBT for the treatment of borderline patients.

Post-treatment, follow-up session

The third and final follow-up session followed after treatment had ended, approximately 1 year after the pre-training

scan session. Subjects completed the same questionnaires and participated in the same neuropsychological test battery as in the second session (Table 1). In addition, the MINI was administered once again to investigate whether axis I classifications had changed and the BPD47 to measure changes in borderline symptom severity.

Pavlovian-instrumental transfer paradigm

Participants performed a computerized PIT task to assess how instrumental approach and withdrawal actions are influenced by aversive Pavlovian CS, i.e., aversive PIT (Geurts et al., 2013a). The experiment consisted of three stages: (1) instrumental conditioning, (2) Pavlovian conditioning, and (3) PIT (see Figure 1 for a global overview and Table 2 for details on the experimental layout).

Stage 1. Participants performed an instrumental learning task to earn as much money as possible. There were two Action Contexts in this task: (i) One in which the active response led to an approach and (ii) another in which the active response led to a withdrawal. In each context, different instrumental stimuli (mushrooms/shells) were repeatedly presented to the participant (Figure 1A). In

TABLE 1 Demographical and clinical characteristics of the borderline personality disorder and healthy matched control participants.

	Healthy controls		Borderline personality disorder group			
	Baseline				1-year follow-up	
Size of group	N = 16		N = 15		N = 14	
	Mean	SD	Mean	SD	Mean	SD
Age	29.5	8.8	28.5	8.8	–	–
IQ (NLV)	101.8	12.3	100.3	11.5	–	–
Right handedness	16	–	14		13	
BPD47	6.7	6.5	79.7***	33.2	64.8[#]	30.0
OQ – total	42.5	20.6	91.5**	19.2	79.1[##]	22.5
Sympt. distr.	19.1	10.6	56.7**	14.1	50.0[#]	16.5
Inter. pers.	8.8	4.9	20.2**	3.8	17.5	5.4
Social role	8.8	4.2	15.0**	4.5	11.6[#]	3.6
BDI-II	3.6	4.0	33.4***	14.3	28.4	14.0
BIS	18.4	7.1	23.5***	4.1	24.4	3.8
BAS	38.7	14.7	40.1***	5.8	41.4	5.1
Box Completion (s)	85.4	30.7	107.0**	20.9	96.8	27.4
Digit Span	13.2	2.5	16.2	4.0	15.2	4.1
Forward	7.1	1.6	8.3	1.9	7.7	2.2
Backward	6.0	1.3	7.9	2.4	7.5	2.4
Verbal Fluency	44.6	12.1	38.1	11.4	42.3	10.3

SD, standard deviation; NLV, Dutch reading test; BPD47, Borderline personality disorder checklist; OQ, outcome questionnaire; BDI-II, Beck depression index 2nd version; BIS, behavioral inhibition systems; BAS, behavioral activation system.
* Indicate significant differences between the groups (HC vs. BPD: *p < 0.05, **p < 0.01, ***p < 0.001)), # indciate significant differences between baseline and follow-up measurement ([#]p < 0.05, [##]p < 0.01).
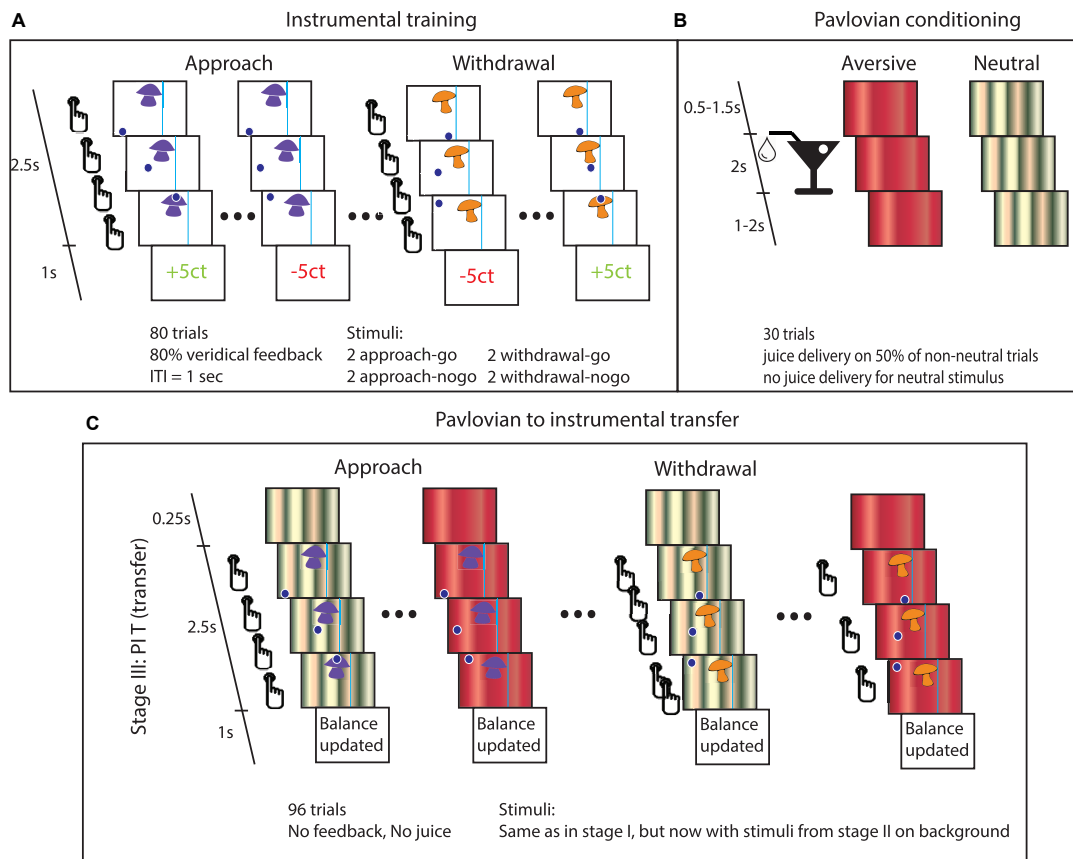
FIGURE 1
Task details. **(A)** Instrumental stage. Trials started with the appearance of the instrumental stimulus at the top center of the screen and a dot at the bottom of the screen. In approach trials, the dot appeared either on the left or on the right bottom of the screen. From left to right: Participants could choose to do nothing (approach-no-go), in which case the dot would move past the instrumental stimulus. Alternatively, they could press the button repeatedly to steer the dot through the instrumental stimulus (approach-go). In withdrawal trials, the dot started centrally beneath the instrumental stimulus. Participants could choose to press the button repeatedly to avoid moving through instrumental stimulus (withdrawal-go) or to do nothing (withdrawal-no-go). If the dot entered the target region, then the instrumental stimulus was 'collected'. The vertical line to one side of the instrumental stimulus could not be crossed by the dot. **(B)** Pavlovian conditioning. Participants were presented with different stimuli that were followed by juice delivery. **(C)** PIT stage. The PIT stage paralleled the instrumental training, except that Pavlovian CS tiled the background. The effect of interest is how the Pavlovian CS changed instrumental behavior (mean proportion of go-actions and the average number of button presses over the go-actions). Note that the trials involving the appetitive CS were omitted from this figure, because this particular paradigm has been shown to be insensitive to detecting appetitive PIT (see **Supplementary material**) and our hypotheses concern aversive PIT.

the approach, Action Context participants learned through monetary feedback (wins and losses) whether to 'collect' the instrumental stimulus (approach-go) or not (approach-no-go). In the withdrawal Action Context, they learned to avoid collecting instrumental stimuli (withdrawal-go) or not (withdrawal-no-go).

In both the approach and withdrawal Action Contexts, there were two go-stimuli, which yielded reward more often (i.e., 80% of the cases) after active responses (and punishment after not responding), and two no-go-stimuli, which yielded reward more often (i.e., also 80% of the cases) after not responding (and punishment after go-responding). Instrumental learning was assessed by calculating the proportion of correct responses (p(correct)) over time.

Stage 2. In this Pavlovian stage, different Pavlovian CS were conditioned (**Figure 1B**). During a classical conditioning procedure, three audiovisual stimuli were presented. The appetitive and aversive conditioned stimuli (CS) were followed, respectively, by appetitive or aversive juice (i.e., the unconditioned stimuli, USs) on 50% of trials. The neutral CS resulted in no outcome. The appetitive juice was based on subjective preference for apple, orange, or strawberry lemonade. The aversive juice was a bitter magnesium sulfate solution (0.3M).

Conditioning was assessed in two ways: (1) participants indicated the degree to which they liked each of the CS (and USs) by using a visual analog scale (VAS), before and after the experiment; and (2) participants chose one of the two presented

Pavlovian stimuli (presented for 2 s; ITI 0.5 s) in extinction on 12 interspersed query trials during the Pavlovian stage.

Stage 3. In the PIT stage, we tested how instrumental approach and withdrawal actions (trained in *stage 1*) are influenced by aversive Pavlovian CS (*conditioned in stage 2.*). Therefore, stimulus presentation was the same as in the instrumental stage, except that Pavlovian CS from the Pavlovian stage 2 tiled the background from 250 ms before and during the trials, and this stage was run in nominal extinction, i.e., no juice or monetary outcomes were presented (**Figure 1C**). Participants were instructed that their choices counted toward the final monetary total and that the juices associated with the Pavlovian outcomes were collected outside the scanner for them to drink afterward. Whether instrumental approach and withdrawal actions were influenced by aversive Pavlovian CS was assessed per Action Context (approach/withdrawal) and CS stimulus (neutral/aversive).

There were two independent runs separated by a 2-min break (each including run-specific stimuli/CS), with each run including all three stages. Each instrumental stimulus was presented 12 times and each Pavlovian CS 32 times. These Pavlovian CS were counterbalanced over the eight instrumental stimuli.

Image acquisition

Whole-brain imaging was performed on a 1.5 Tesla MR scanner (Avanto, Siemens Medical Systems, Erlangen, Germany). Functional data were obtained using a multi-echo gradient T2*-weighted echo-planar (ME-EPI)

scanning sequence (Poser et al., 2006) (see **Supplementary materials** for details).

Analysis

Our primary analysis was restricted to the PIT stage. Analysis and results of the instrumental and Pavlovian training data are presented in the **Supplementary materials**. The analyses presented below consist of two parts: First, we assessed the effects of the group on behavior and fMRI BOLD response during the PIT stage, measured at baseline. Here, we focus on both the behavioral and fMRI analyses in line with our hypothesis on aversive PIT. We discern two aspects of aversive PIT: Action Context-specific aversive PIT and aversive PIT that is independent of Action Context, i.e., aversive PIT across Action Contexts. Action Context-specific aversive PIT quantifies the differential effect of an aversive CS on approach and withdrawal behavior, whereas aversive PIT across Action Contexts quantifies general effects of the Pavlovian CS on instrumental behavior irrespective of whether it is approach or withdrawal behavior. Statistically, aversive PIT across Action Contexts is captured by the main effect of CS Valence (neutral vs. aversive across Action Contexts), while Action Context-specific aversive PIT is captured by the interaction between CS Valence and Action Context. These different aversive PIT effects have been associated in previous studies with different clinical outcomes and neural mechanisms (Geurts et al., 2013a,b; Garbusow et al., 2016; Huys et al., 2016). Specifically, while Action Context might arise from a vmPFC-dependent process (Geurts et al., 2013a) that is predictive of recovery from

TABLE 2 Experimental layout.

	# trials/time per trial(s)	# stimuli	Reinforcement
Instrumental training	80/2.5s		
Blocks of 8 trials per Action Context (approach/withdrawal)	20	2 stimuli requiring approach: $S^I_{1,2}$	80% reward/20% punishment for go
	20	2 stimuli requiring approach-nogo: $S^I_{3,4}$	80% reward/20% punishment for nogo
	20	2 stimuli requiring withdrawal: $S^I_{5,6}$	80% reward/20% punishment for go
	20	2 stimuli requiring withdrawal-nogo: $S^I_{7,8}$	80% reward/20% punishment for nogo
Pavlovian training	60/3s	3	
Each 10th trial a query trial to choose between two CS	20	1 stimulus followed by aversive juice → aversive CS: S^P_1	50% of trials are reinforced
	20	1 stimulus without reinforcement → neutral CS: S^P_2	No reinforcement
	20	1 stimulus followed by appetitive → juice appetitive CS: S^P_3	50% of trials are reinforced
Pavlovian to instrumental transfer	96/2.5s	4/3	
Blocks of 8 trials per Action Context (approach/withdrawal)	32	$S^I_{1,2,3,4,5,6,7,8} \mid S^P_1$	No direct reinforcement
	32	$S^I_{1,2,3,4,5,6,7,8} \mid S^P_2$	No direct reinforcement
	32	$S^I_{1,2,3,4,5,6,7,8} \mid S^P_3$	No direct reinforcement
2 Runs			

depression (Huys et al., 2016), the extent to which Pavlovian CS inhibit ongoing behavior across Action Contexts likely reflects amygdala/striatal activity and changes in serotonergic transmission (Geurts et al., 2013b), and is instead associated with the psychopathic tendency in a sample of violent offenders (unpublished findings, submitted to the current special issue of Frontiers in Behavioral Neuroscience).

Second, within the BPD group, we assessed whether aversive PIT and associated BOLD signals were associated with symptom reduction at the end of the 1-year DBT program.

We note in addition that our previous work in healthy controls, on which the current study builds, revealed that the current paradigm was not sensitive to (and therefore less valid to assess group effects on) appetitive PIT (Geurts et al., 2013a). We therefore only present the data on aversive PIT. In the **Supplementary material**, we confirm that, indeed, the current paradigm is not sensitive to appetitive PIT.

Pavlovian-instrumental transfer

Behavioral analyses

We focused our analyses on aversive PIT, i.e., the effect of aversive Pavlovian CS on instrumental behavior. The effects of Action Context (approach/withdrawal), CS Valence (neutral/aversive), and group (healthy controls/BPD patients) in the critical transfer test were assessed in terms of proportion of go-choices [p(go)] and the average number of button presses (BP, made during these go-choices). Note that our previous work in healthy controls, on which the current study builds, revealed that the current paradigm was not sensitive to (and therefore less valid to assess group effects on) appetitive PIT (Geurts et al., 2013a). We present behavioral data on appetitive PIT in the **Supplementary materials**.

Thus, analyses were targeted at the degree to which aversive CS influenced instrumental behavior. More specifically, we analyzed across Action Context (approach and withdrawal) how much the aversive Pavlovian CS (compared with the neutral CS) inhibited instrumental 'go' responding (i.e., the main effect of CS Valence). In addition, we also assessed the Action Context specificity of aversive PIT, i.e., to what extent the effect of the aversive Pavlovian CS is dependent on Action Context (i.e., interaction CS Valence X Action Context). The dependent variables were first averaged across runs and normality was assessed, before they were submitted to a repeated measures ANOVA (rmANOVA), with Action Context (approach/withdrawal) and CS Valence (neutral/aversive) as within-subject factors and group (healthy controls/BPD patients) as a between-subject factor. Due to non-normal distribution of p(go), we employed non-parametric tests to assess whether there was a significant aversive PIT effect across groups (related-samples Wilcoxon signed-rank test

comparing the difference between p(go) for neutral Valence and p(go) for aversive as a function of Action Context) and whether there was a difference in aversive PIT between groups (independent samples median test comparing the compound measure of Action Context-specific aversive PIT, i.e., [(approach neutral- approach aversive – (withdrawal neutral - withdrawal aversive)] and aversive PIT across Action Contexts [(approach neutral + withdrawal neutral – (approach aversive - withdrawal aversive)] between groups).

fMRI analysis

An fMRI analysis was performed with SPM5 software (Wellcome Trust Centre for Cognitive Neuroimaging, London, United Kingdom). Pre-processing steps and first-level fMRI analysis were identical to those employed by Geurts et al. (2013a): First, realignment parameters were estimated for the images acquired at the first echo time and consequently applied to images resulting from the three other echoes. The echo images were combined by applying a PAID-weight algorithm assessing the signal-to-noise ratio as described by Poser et al. (2006). Thirty volumes, acquired before each instrumental training session, were used as input for this algorithm. Thereafter, the following preprocessing steps were applied: slice-time correction, co-registration, and a segmentation procedure using the tissue probability maps provided by SPM5 for gray matter, white matter, and CSF centered in MNI space to estimate normalization parameters based on the structural image. Structural and functional images were then normalized by applying these estimations. All normalized images were smoothed with an isotropic 8 mm full-width half-maximum Gaussian kernel (Worsley and Friston, 1995). The fMRI analysis was restricted to the PIT stage and was similar to our previous analyses (Geurts et al., 2013a). The general linear model (GLM, **Supplementary Figure 1**) at the participant level consisted of six main regressors representing the onset of the six different PIT trials [Action Context (approach/withdrawal) x CS Valence (appetitive/neutral/aversive)]. For each main regressor, an additional parametric regressor was added (Büchel et al., 1996): The PIT regressor (Talmi et al., 2008; cf. Geurts et al., 2013a) was a parametric modulator of BOLD responses by the number of button presses per trial. Contrasting this regressor between the different CS Valence measures thus reveals "PIT-related regions", i.e., regions where the BOLD signal is associated with valence-dependent coupling between amygdala BOLD signal and instrumental behavior on a trial by trial basis. Note that such a contrast goes beyond simple reactivity of a region to a CS or to instrumental behavior *per se*; it critically captures its interaction, i.e., PIT. A further parametric regressor contained the expectation associated with each instrumental stimulus (the Q-value) per trial as estimated from a model-based analysis of behavior (Huys et al., 2011) applied to the current data. This was done based on prior data showing that the BOLD signal in the prefrontal cortex and striatum, our regions of

interest, covaries with instrumental action value (O'Doherty, 2004; Valentin et al., 2007; Wunderlich et al., 2009; Smith et al., 2010; Jocham et al., 2011; see for meta-analysis, Chase et al., 2015). As such, this approach maximized the degree to which our GLM captured variability in the relevant BOLD signals. Furthermore, realignment parameters were added, high-pass filtering (128s) was applied, and parameter estimates were obtained by maximum-likelihood estimation (AR1).

The parameter estimates for the neutral and aversive parametric PIT regressors were used in a 2 × 2 × 2 rmANOVA at the group level (with random effects) with Action Context (approach/withdrawal) and Valence (neutral/aversive) as within-participant factors and group (healthy controls/BPD) as a between-participants factor. Within this rmANOVA, we assessed Action Context-specific aversive PIT and aversive PIT across Action Context for group differences. Moreover, we also assessed the main effect of Action Context. Based on Geurts et al. (2013a), we expected this analysis to reveal that the BOLD signal in the ventromedial prefrontal cortex would be Action Context-specific (approach > withdrawal). We did not expect a group effect on this contrast.

To capture additional PIT signals related to stable patterns of behavior beyond trial-by-trial variation in instrumental vigor, we contrasted the main regressors (Figure 2) at the participant level to calculate the main effect of Valence [(approach&neutral + withdrawal&neutral) - (approach&aversive + withdrawal&aversive)] (cf. Talmi et al., 2008; cf. Geurts et al., 2013a). The resulting SPM was then used in a two-sample t-test at the group level with aversive PIT in terms of the average number of button presses as a covariate for each group separately enabling comparison between groups. Based on Geurts et al. (2013a), we expected that behavioral aversive PIT across Action Contexts in terms of the average number of button presses [(BP| approach&neutral + BP| withdrawal&neutral) - (BP| approach&aversive + BP| withdrawal&aversive)] would be related to BOLD signal change (neutral-aversive) in the amygdala and nucleus accumbens and that this relationship would differ between the groups (i.e., a stronger correlation within the BPD group).
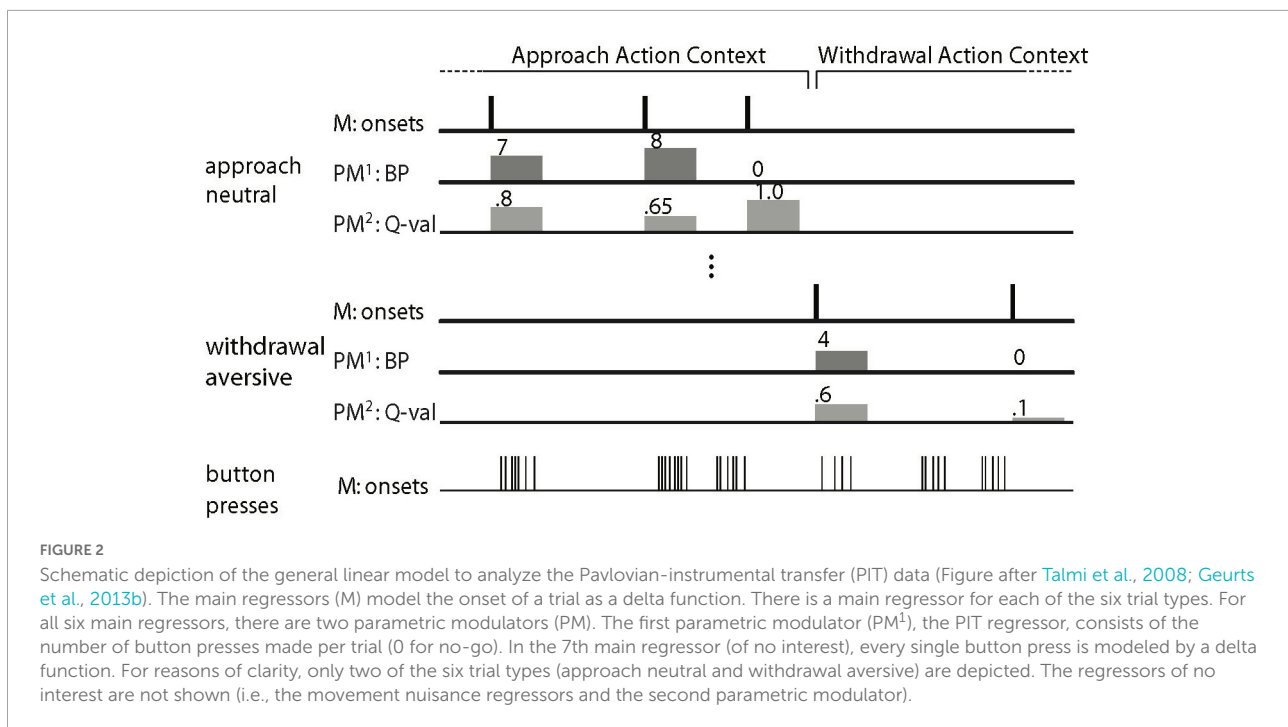
Treatment success and its prediction

Our primary measure of treatment success was the Borderline Personality Disorder Checklist, (BPD47, Bloo et al., 2017) a 47-item self-report questionnaire based on the Borderline Personality Disorder Severity Index (Arntz et al., 2003). Furthermore, as secondary measures, we also assessed the quality of life with the Outcome Questionnaire (OQ, Lambert et al., 1996) and depressive symptoms with the Beck Depression Inventory second edition (BDI-II, Beck et al., 1996). The treatment effect was computed by subtracting the post-treatment scores from those acquired during the first scan session.

Predictive relationship between aversive Pavlovian-to-instrumental transfer and symptom reduction

We assessed the association between aversive PIT and associated BOLD signal [at the whole-brain level and within the predefined amygdala region of interest (ROI)], measured pre-treatment, with clinical symptom reduction 1 year later. A second-level random-effects simple regression analysis was conducted to assess whether PIT-related neural signal was associated with symptom severity at baseline, and/or symptom reduction over 1 year. To this end, we computed Action Context-specific aversive PIT-related BOLD signal [(PITregressor| approach&neutral - PITregressor| approach&aversive) - (PITregressor| withdrawal&neutral - PITregressor| withdrawal&aversive)], aversive PIT-related BOLD signal across Action Contexts [PIT regressor| approach&neutral + PIT regressor| withdrawal&neutral – (PIT regressor| approach&aversive + PIT regressor| withdrawal&aversive)], as well as BPD47 scores at baseline and BPD47 change (before-after). These latter covariates of interest were tested in two simple regression analyses of the aversive PIT statistical parametric maps. Any relationship between the PIT-related BOLD contrasts and the BPD47 change (without a baseline relationship) would indicate that PIT-related signaling is predictive of symptom reduction. In addition, as a sensitivity analysis because of the small sample size, we also performed the non-parametric equivalent of this analysis with SnPM (Winkler et al., 2014) and we employed a leave-one-participant-out procedure (Esterman et al., 2010), in which a single participant is iteratively left out of the second-level correlational analysis. The resulting clusters within the anatomically defined bilateral amygdala (thresholded at $p < 0.001$ uncorrected) were then used to extract the mean beta weights of the left-out participant to calculate the aversive PIT contrast. This procedure was repeated for each participant. The GLM from the remaining participants thus serves as an independent localizer for the participant left out (Esterman et al., 2010).

Statistical thresholding

We report effects that survive family-wise error (FWE) correction for multiple comparisons across the whole brain ($P_{WB} < 0.05$, voxel-level) or in one of the following ROIs: The amygdala (automated anatomical labeling atlas, Tzourio-Mazoyer et al., 2002) was our primary ROI to assess the effect of symptom reduction. Both the amygdala and nucleus accumbens (same as in Geurts et al., 2013a) were chosen as ROIs for the analysis of the main PIT task effects (across and between groups) based on their key role in PIT (Corbit, 2005; Talmi et al., 2008; Corbit and Balleine, 2011; Prevost et al., 2012; Geurts et al., 2013a; Garbusow et al., 2016). Specifically, in our previous study, we found BOLD response in both these regions to be associated with behavioral PIT on a participant-by-participant basis. Following our prior work, we also assessed

FIGURE 2
Schematic depiction of the general linear model to analyze the Pavlovian-instrumental transfer (PIT) data (Figure after Talmi et al., 2008; Geurts et al., 2013b). The main regressors (M) model the onset of a trial as a delta function. There is a main regressor for each of the six trial types. For all six main regressors, there are two parametric modulators (PM). The first parametric modulator (PM[1]), the PIT regressor, consists of the number of button presses made per trial (0 for no-go). In the 7th main regressor (of no interest), every single button press is modeled by a delta function. For reasons of clarity, only two of the six trial types (approach neutral and withdrawal aversive) are depicted. The regressors of no interest are not shown (i.e., the movement nuisance regressors and the second parametric modulator).

action specificity in the ventromedial prefrontal cortex: The region shown to be sensitive to Action Context in our previous PIT study was used as ROI (MNI coordinates of ROI center: xyz = [-8 36-8]) (Geurts et al., 2013a). The left and right elements of each bilateral volume of interest were combined using Marsbar[TM] (Brett et al., 2002).

Results

Baseline behavioral data

Pavlovian-instrumental transfer

Consistent with our previous studies using this paradigm, we observed opposite effects of the aversive Pavlovian CS on approach and withdrawal actions (in terms of choice p(go), **Figure 2**): aversive Pavlovian CS inhibited approach and activated withdrawal actions. Planned contrasts confirmed the statistical significance of this action specificity of the aversive PIT effect (related-samples Wilcoxon signed rank test [p(go| approach&neutral) - p(go| approach&aversive)] > [p(go| withdrawal&neutral)- p(go| withdrawal&aversive)]: p = 0.031, one-tailed). There were no differences between the groups (independent samples median test: p = 0.48), but we note that the action-specific PIT effect was present in healthy controls (p = 0.008), but not in patients (p = 0.860) when examined separately.

There were no main task effects except for the main effect of Action Context in terms of the average number of button presses ($F_{(1,29)}$ = 33.7, p < 0.001, all other F < 1.8 and p > 0.2, **Supplementary Table 2**). There were no group differences.

Performance on the instrumental task and assessments of Pavlovian training also did not differ between the groups (**Supplementary Results**). To be complete, we confirmed the already established insensitivity to detect appetitive PIT with the current paradigm (**Supplementary Results**).
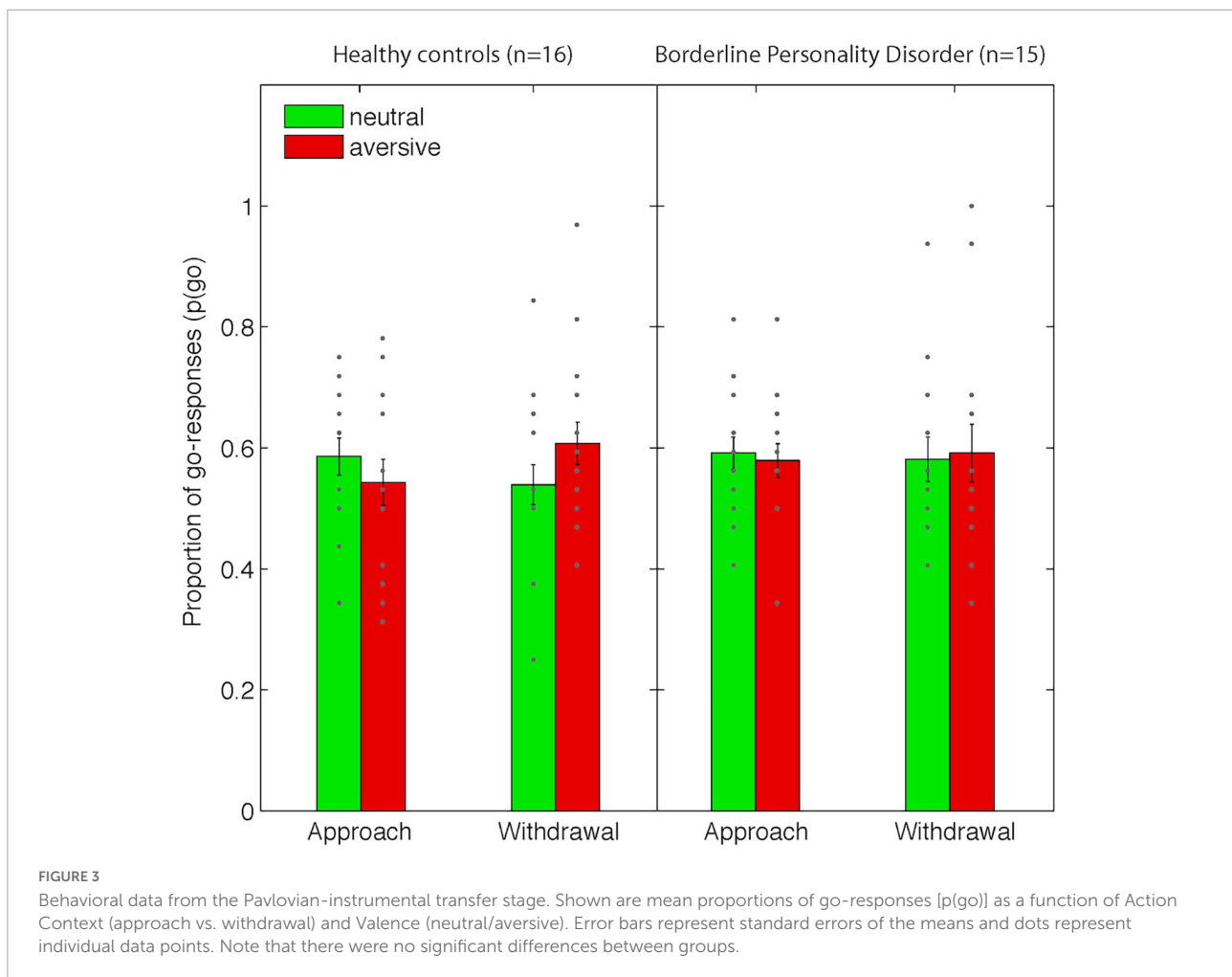
Baseline imaging data

Consistent with our previous fMRI study using this paradigm, trial-by-trial instrumental action-related BOLD signal in the vmPFC varied as a function of Action Context. The BOLD signal was greater during approach than during withdrawal (small volume corrected results for the vmPFC ROI: peak voxel MNI-coordinates [−6 32 −12], k = 45, Z = 3.86, p_{FWE} = 0.021, **Figure 3**).

Conversely, we did not replicate the previously observed correlation between individual differences in behavioral aversive PIT and BOLD signals in the amygdala and nucleus accumbens. Moreover, we did not find significant main effects of or interactions with the factor group.

Aversive Pavlovian-to-instrumental transfer and symptom reduction
Symptom reduction

The 14 patients who were seen at follow-up, 1 year after the start of therapy, showed a significant reduction in symptom severity as measured with the BPD47 (mean difference = −17.3, t_{13} = 2.5, p = 0.027, reliable change index (Jacobson and Truax, 1991): 15.8), OQ (mean difference = −12.4, t_{13} = 3.1, p = 0.009),

FIGURE 3
Behavioral data from the Pavlovian-instrumental transfer stage. Shown are mean proportions of go-responses [p(go)] as a function of Action Context (approach vs. withdrawal) and Valence (neutral/aversive). Error bars represent standard errors of the means and dots represent individual data points. Note that there were no significant differences between groups.

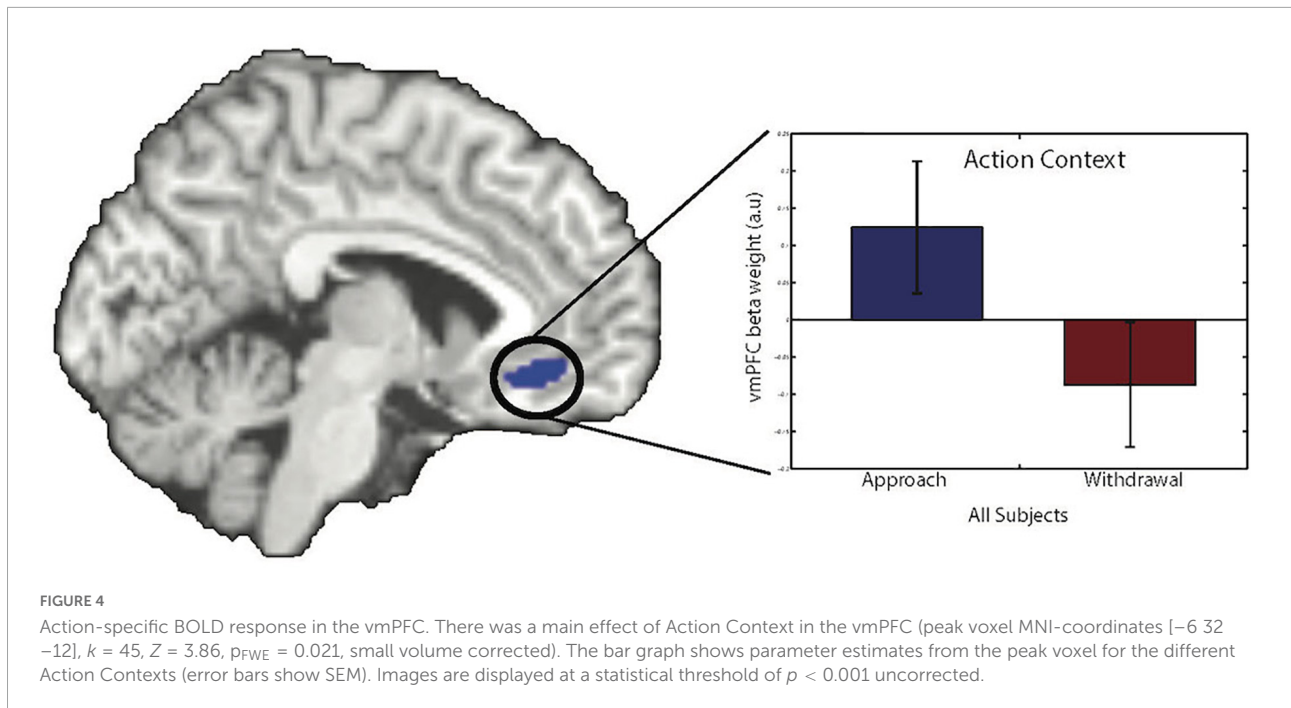and in trend with the BDI-II (mean difference = −4.8, t_{13} = 1.8, p = 0.090).

None of the neuropsychological tests reported in **Table 1** changed significantly from baseline to 1 year after treatment (all $−1.9 > t_{13} < 2.2$, all $p \geq 0.05$).

Pavlovian-to-instrumental transfer-related BOLD signal in the amygdala is related to symptom reduction 1 year later

Pre-treatment PIT-related BOLD signal in the bilateral amygdala was related to BPD symptom reduction after 1 year (**Figure 4**). Higher aversive PIT-related signals across Action Contexts were associated with less symptom reduction 1 year later. This observation was substantiated by using both parametric and non-parametric statistical analyses (small volume corrected effects in the amygdala; parametric tests with SPM: peak voxel MNI-coordinates [−24 0 −16], k = 22, Z = 3.79, p_{FWE} = 0.027; non-parametric test with SnPM: peak voxel MNI-coordinates [−24 0 −18], pseudo-t = 4.22, p_{FWE} = 0.013; and MNI-coordinates [22 4 −18], pseudo t = 3.19, p_{FWE} = 0.06). The robustness of these effects was confirmed by cross-validation

($r_{(14)}$: −0.655, p = 0.011) and by supplementary analyses on mean beta estimates extracted from the anatomically defined bilateral amygdala (Pearson $r_{(14)}$: −0.667, p = 0.009). Note, that no significant relation was observed between baseline BPD47 scores and PIT-related amygdala signal (Pearson $r_{(14)}$:0.33, p = 0.25).

Next, we explored the specificity of this predictive effect with respect to other (more easily acquired) baseline measures, including baseline BPD47, OQ, BDI-II, BIS, BAS, box completion time, verbal fluency, and digit span (**Table 1**). A stepwise linear regression analysis (with criteria probability of F to enter ≤ 0.05 and o F to remove ≥ 0.10) identified two predictors of symptom reduction. Indeed, pre-treatment PIT-related signal in the bilateral amygdala accounted for variance in symptom reduction over and above the other collected baseline measures. Verbal fluency was the only other selected predictor of symptom reduction (Final regression model including PIT-related amygdala signal and verbal fluency: $F_{(2, 13)}$ = 11.6, p = 0.002, standardized coefficients beta for amygdala signal:0.64, p = 0.003; and for the verbal fluency: −0.48, p = 0.016). All other measures did not enter the model

FIGURE 4

Action-specific BOLD response in the vmPFC. There was a main effect of Action Context in the vmPFC (peak voxel MNI-coordinates [−6 32 −12], $k = 45$, $Z = 3.86$, $p_{FWE} = 0.021$, small volume corrected). The bar graph shows parameter estimates from the peak voxel for the different Action Contexts (error bars show SEM). Images are displayed at a statistical threshold of $p < 0.001$ uncorrected.

(all $|$ t $|$ $\leq$ 1.5, all $p > 0.2$). Next, we examined whether the predictive effect of pre-treatment PIT-related amygdala signal was specific to BPD47 change or whether it extended to other changes in clinical or neuropsychological measures. Indeed, stepwise multiple regression analysis with this amygdala signal as a dependent variable revealed that the association of this signal with BPD47 improvement ($F_{(1,12)} = 9.6$, $p = 0.009$) did not extend to any of the other changes in clinical or neuropsychological measures (all $|$ t $|$ < 1.9, all $p > 0.18$). This is relevant because improvement in borderline severity was accompanied by improvement in depressive symptoms as measured with the BDI-II ($r_{14} = -0.67$, $p = 0.008$), as well as improvement in verbal fluency ($r_{14} = 0.91$, $p < 0.001$).

Discussion

Results failed to confirm our prediction that patients with borderline personality disorder exhibit abnormal aversive PIT compared to healthy controls at the group level. However, on an individual level, the results demonstrate that the BOLD signal in the amygdala elicited during the aversive PIT task is related to symptom reduction in these patients across 1 year of follow-up. Greater PIT-related responsiveness of the (bilateral) amygdala was associated with reduced clinical improvement 1 year later. More specifically, this suggests that individual differences in the degree to which amygdala processing relates to trial-by-trial instrumental responding in the context of an aversive Pavlovian CS predict resistance to clinical improvement of (or slower recovery from) BPD. Thus, participants who showed increased

coupling between the amygdala BOLD signal and instrumental behavior during aversive Pavlovian CS presentation showed less clinical improvement. In more general terms, this suggests that individual differences in amygdala response could predict clinical improvement of BPD.

Based on observations that BPD is associated with the abnormal impact of aversive stimuli on behavior (Soloff et al., 2017; Hallquist et al., 2018), we employed an aversive PIT task that measures the degree to which aversive Pavlovian CS alter instrumental behavior. We replicated the previously observed basic behavioral task effects, including the Action Context-specificity of aversive PIT (Huys et al., 2011; Geurts et al., 2013a), with an aversive Pavlovian CS suppressing approach, but potentiating withdrawal actions. These task effects were not modulated by BPD, although when analyzing the groups separately, we only found significant effects in the healthy controls. The absence of a group effect might be due to the relatively stressful scanner environment (Talmi et al., 2008; cf. discussion of Geurts et al., 2013a). Indeed, there are indications that stress reduces behavioral PIT effects (Quail et al., 2016; but see Pool et al., 2015) and patients with BPD might be more sensitive to this stress. It might also be a consequence of the use of psychotropic medication in about two-thirds of our patients, which has been associated with attenuated amygdalar hyperreactivity in BPD (Schulze et al., 2016) and is likely to change PIT through changing monoaminergic signaling (cf. Geurts et al., 2013b; Hebart and Gläscher, 2015; Swart et al., 2017). Moreover, given the small sample sizes, the absence of a group effect on action-specific PIT might also reflect insufficient statistical power to detect such a difference. However, we cannot

exclude that, as a group, BPD patients indeed do not exhibit abnormal aversive PIT.

The key observation of this study is that neural activity of the amygdala in BPD patients is associated with clinical symptom reduction. These results substantiate the promise of neurocognitive strategies for predicting treatment outcomes in various psychiatric disorders (Nitschke et al., 2009; Pizzagalli, 2010; Roiser et al., 2011; Månsson et al., 2015; Perez et al., 2016; Garbusow et al., 2016; Huys et al., 2016; Schmitgen et al., 2019; Westlund Schreiner et al., 2019; Sampedro et al., 2021). The considerable gap between cognitive neuroscience and clinical practice has been the subject of a fruitful ongoing debate (Paulus et al., 2016; Stephan et al., 2016; Huys, 2018). One major problem in the clinical relevance of neurocognitive research is that most studies have compared groups of patients, failing to address individual differences in the treatment efficacy. Future work is required to investigate whether an aversive PIT-related neural signal is associated selectively with DBT efficacy, or rather reflects general treatment efficacy or even BPD symptom change more irrespective of treatment.
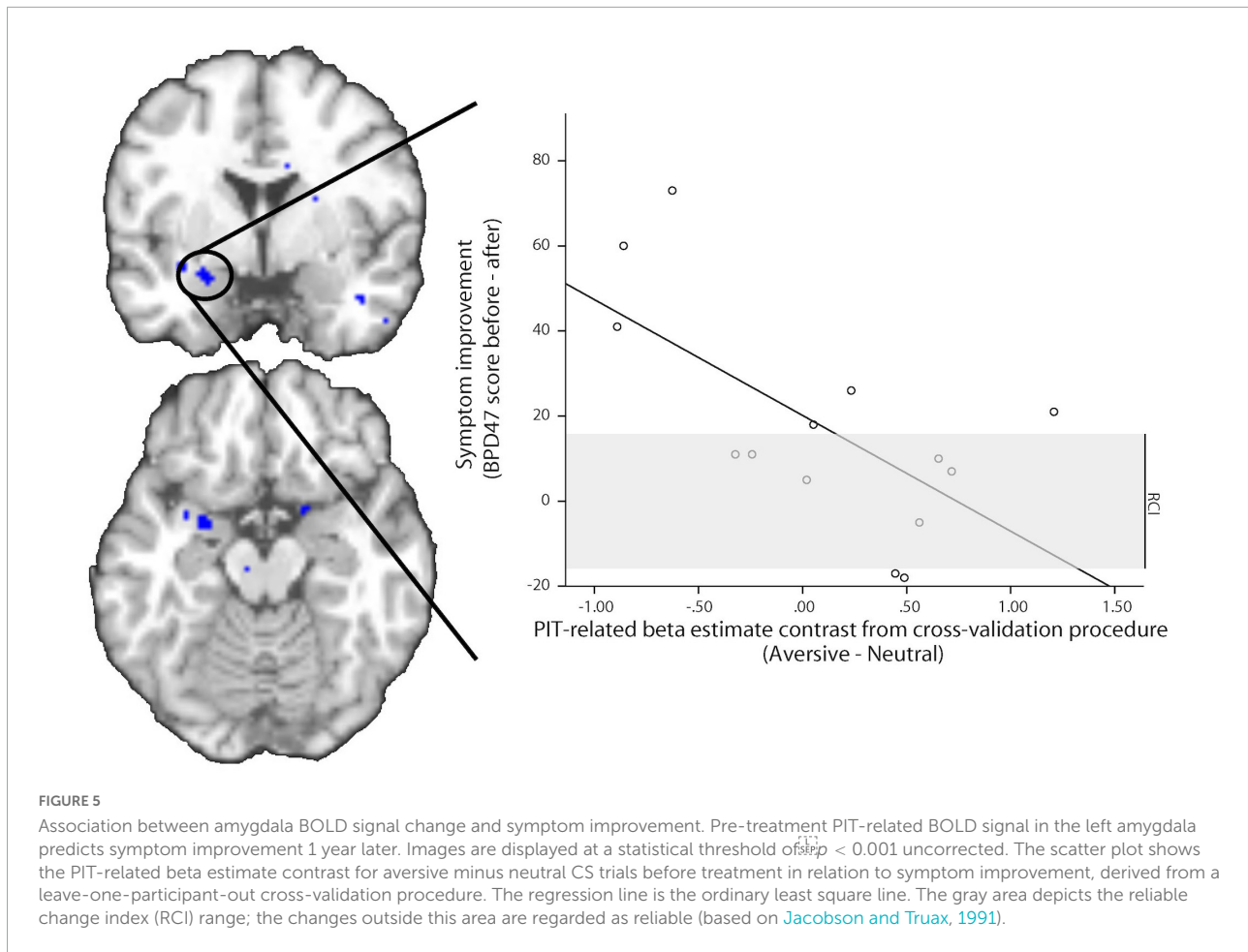
Moreover, our results provide converging evidence for the validity of the PIT paradigm for predicting clinical symptom changes [in depression (Huys et al., 2016) and addiction (Garbusow et al., 2016)]. It should be noted that, here, amygdala signal across Action Contexts was the predictor, whereas in the study of Huys et al. (2016), it was the Action Context specificity of behavior that predicted recovery from depression. We did not find such an association for symptom reduction in patients with borderline personality disorder. Moreover, in the study of Garbusow et al. (2016), it was the PIT effect in the nucleus accumbens that predicted relapse in alcohol use. This suggests that different aspects of the neurocognitive mechanisms underpinning the transfer between Pavlovian CS and instrumental behavior might be disorder and/or treatment specific. We note that these studies, just like the current study, are relatively small in sample size. Nevertheless, these studies make concrete steps in translating hypotheses on mechanistic relevance for clinical treatments and as such are stepping stones for larger future studies making use of their methodology, which are already emerging (e.g., Chen et al., 2021).

The present results suggest that symptom reduction after DBT is greater in BPD patients who show lower amygdala signals during aversive PIT. The finding that the amygdala signal is predictive of symptom reduction in BPD after DBT concurs with empirical findings and neurocognitive theories, implicating a central role for the amygdala in BPD (Schulze et al., 2016, 2019) and DBT (Schnell and Herpertz, 2007; Goodman et al., 2014; Schmitt et al., 2016; Schmitgen et al., 2019). Several recent studies have shown changes in amygdala signaling after DBT (Schnell and Herpertz, 2007; Goodman et al., 2014; Schmitt et al., 2016; Niedtfeld et al., 2017; but see Winter et al., 2017). Schnell et al. employed a pilot study with six BPD patients who received several fMRI scans during 3 months of DBT. The

four patients who responded to DBT all showed decreases in amygdala BOLD responses to emotional pictures. In keeping with this finding, Goodman et al. (2014) reported decreases in amygdala responses to emotional pictures and associated improvement in self-reported emotional regulation in 11 BPD patients after 1 year of DBT treatment. Moreover, Niedtfeld et al. (2017) showed in 28 patients with BPD that a scaled version of 12 weeks of DBT attenuated amygdala deactivation in response to pain. Schmitt et al. (2016) showed that patients who responded well to DBT exhibited reduced activation in, among other regions, the amygdala, during the reappraisal of negative stimuli after DBT.

These studies suggest that the association between amygdala signaling and symptom reduction, observed in the current study, might relate to treatment-induced changes in the amygdala. We stress, however, that we did not collect behavioral or fMRI data after therapy, which precludes us from assessing whether the amygdala signal indeed changed during this treatment, or whether it is a stable trait that indexes the susceptibility to the offered treatment. Moreover, due to the absence of a control condition in our design, we restrict our conclusions to the general case of clinical improvement. Thus, we cannot claim the specificity of our results to DBT. Moreover, PIT-related amygdala signal might also reflect more general, less treatment-specific, process underlying improvement like the ability to (emotionally) engage and/or commit oneself to treatment. We thus restrict our conclusion to the general *predictive* effect of amygdala signal on symptom change.

Although several studies, as mentioned above, assessed pre- to post-therapy changes in neural processing in BPD, so far, only two other studies assessed the value of selectively pre-treatment task-based fMRI signals for predicting treatment success (Perez et al., 2016; Schmitgen et al., 2019). In the study by Perez et al. (2016) including 10 patients with BPD, a greater pre-treatment BOLD signal in the right anterior cingulate cortex during an emotional go/no-go task was associated with reduced improvement after transference-focused psychotherapy (TFP) in terms of the factor 'constraint' of the multidimensional Personality Questionnaire. Moreover, a greater BOLD signal in the left posterior-medial OFC/ventral striatum was associated with reduced improvement in terms of the total score on the Affective Lability Scale. The study of Schmitgen et al. (2019) is of specific interest for the current study, because it explicitly addressed the prediction of clinical DBT effects based on, amongst others, task-based fMRI in a relatively large sample (n = 31) of BPD patients with a sophisticated cross-validation procedure to optimize a random forest prediction algorithm. They employed three emotion regulation tasks, fMRI, and structural MRI before 12 weeks of DBT. They showed that (left) amygdala (and parahippocampus) activation during a cognitive reappraisal task was particularly informative for treatment response prediction. Accuracy of predicting treatment response (base rate 52%) of the model based on solely these fMRI data

FIGURE 5
Association between amygdala BOLD signal change and symptom improvement. Pre-treatment PIT-related BOLD signal in the left amygdala predicts symptom improvement 1 year later. Images are displayed at a statistical threshold of $p < 0.001$ uncorrected. The scatter plot shows the PIT-related beta estimate contrast for aversive minus neutral CS trials before treatment in relation to symptom improvement, derived from a leave-one-participant-out cross-validation procedure. The regression line is the ordinary least square line. The gray area depicts the reliable change index (RCI) range; the changes outside this area are regarded as reliable (based on Jacobson and Truax, 1991).

reached 75%. Of note is that responders, while instructed to look at negative emotional pictures, showed lower left amygdala reactivity before therapy compared to non-responders. Together with these prior data, our findings strengthen the observation that particular limbic circuitry processing during affective action regulation renders BPD patients more resistant to clinical improvement after therapy. Moreover, differences between these studies employing two different treatment regimes (TFP vs. DBT) might speak to the future practical, clinical use of these findings. Future research should investigate how we can make treatment regimes more efficient by allocating specific patients to specific treatment modalities based on their functional neural signature. Thus, combining different neural predictors for treatment success specific to different treatment modalities might help us to reveal which patients should be allocated to which treatment. Before being able to implement this in clinical practice, more large-scale practice-based studies should be carried out to ensure the reliability and clinical usefulness of these predictions. Our data provide proof of principle of such a procedure within a practice-based convenience sample of BPD patients. We note, however, that only about half of the patients that were planned to follow during the DBT

treatment did not volunteer or dropped out of this study. This observation is important for assessing the feasibility of employing these procedures broadly in clinical practice. Future qualitative, implementation research on facilitators and barriers to these procedures is warranted.

Further limitations of our study deserve special attention: First, our main result is based on a small sample size. Although we assessed the robustness of the effect extensively, for example, by cross-validation (leave-one-participant-out procedure) and by permutation-based analyses (SnPM), replication of our data is needed. Second, because we did not include a patient control group, we cannot assess whether the amygdala signal is a general predictor of positive change in symptomatology or whether it specifically moderates treatment outcomes. Third, our paradigm was insensitive to appetitive PIT (**Supplementary Material**) and therefore we cannot make any claims on the valence specificity of the presented results.

Fourth, we set out to include all the patients who were offered DBT during the inclusion period of this study. In this setting, all BPD patients were female, which thus precludes conclusions about male BPD patients. Moreover, the inclusion resulted in a 'real-life' BPD patient group with

the majority of patients being on psychotropic medication and having multiple comorbidities. This choice of patient selection was at the expense of internal validity [e.g., a recent meta-analysis shows that medicated compared with non-medicated patients with BPD show blunted amygdala responses (Schulze et al., 2016)], which we deliberately traded off against enhanced external validity (Hoertel et al., 2015). The majority of patients in normal clinical practice with BPD have multiple comorbidities and, although discouraged in many guidelines, take psychotropic medications, such as selective serotonin inhibitors. We acknowledge that we cannot exclude the possibility that differences in medication use contribute to the observed effect. With the low sample size and the diverse medication regimens of the included patients, we have no means to address this quantitively. To provide as much insight as possible, we report medication use in the supplementary materials for each patient (Supplementary Table 1) and also added a graph, similar to Figure 5, showing bilateral amygdala signals for those with and without medication use (Supplementary Figure 1). Choosing such a sample is in line with our ultimate aim to find useful biobehavioral markers to predict and optimize treatment success in real-life clinical practice.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by METC Oost-Nederland. The patients/participants provided their written informed consent to participate in this study.

Author contributions

DG, QH, TV, and RC contributed to conception and design of the study. DG collected the data and performed the statistical analysis under supervision of RC. DG and RC wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh. 2022.938403/full#supplementary-material

References

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders*, 4th Edn. Washington, DC: American Psychiatric Association.

Arntz, A., van den Hoorn, M., Cornelis, J., Verheul, R., van den Bosch, W. M. C., and de Bie, A. J. H. T. (2003). Reliability and validity of the borderline personality disorder severity index. *J. Pers. Disord.* 17, 45–59.

Balleine, B. W., and Doherty, J. P. O. (2009). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69. doi: 10.1038/npp.2009.131

Barnicot, K., Katsakou, C., Bhatti, N., Savill, M., Fearns, N., and Priebe, S. (2012). Factors predicting the outcome of psychotherapy for borderline personality disorder: A systematic review. *Clin. Psychol. Rev.* 32, 400–412.

Beck, A. T., Steer, R. A., Ball, R., and Ranieri, W. (1996). Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J. Pers. Assess.* 67, 588–597. doi: 10.1207/s15327752jpa6703_13

Bloo, J., Arntz, A., and Schouten, E. (2017). The borderline personality disorder checklist: Psychometric evaluation and factorial structure in clinical and nonclinical samples. *Ann. Psychol.* 20, 311–336. doi: 10.18290/rpsych.2017.20.2-3en

Bolton, J. M., and Robinson, J. (2010). Population-attributable fractions of axis i and axis ii mental disorders for suicide attempts: Findings Froma representative sample of the adult, Noninstitutionalized US Population. *Am. J. Public Health* 100, 2473–2480. doi: 10.2105/AJPH.2010

Breland, K., and Breland, M. (1961). The misbehavior of organisms. *Am. Psychol.* 16, 681–684. doi: 10.1037/h0040090

Brett, M., Anton, J.-L., Valabregue, R., and Poline, J.-B. (2002). "Region of interest analysis using an SPM toolbox," in *Proceedings of the 8th International Conference on Functional Mapping of the Human Brain*, Sendai.

Büchel, C., Wise, R. J., Mummery, C. J., Poline, J. B., and Friston, K. J. (1996). Nonlinear regression in parametric activation studies. *Neuroimage* 4, 60–66. doi: 10.1006/nimg.1996.0029

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehavior. Rev.* 26, 321–352. doi: 10.1016/S0149-7634(02)00007-6

Chase, H. W., Kumar, P., Eickhoff, S. B., and Dombrovski, A. Y. (2015). Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cogn. Affect. Behav. Neurosci.* 15, 435–459. doi: 10.3758/s13415-015-0338-7

Chen, H., Mojtahedzadeh, N., Belanger, M. J., Nebe, S., Kuitunen-Paul, S., Sebold, M., et al. (2021). Model-based and model-free control predicts alcohol consumption developmental trajectory in young adults: A 3-Year Prospective Study. *Biol. Psychiatry* 89, 980–989. doi: 10.1016/j.biopsych.2021.01.009

Corbit, L. H. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J. Neurosci.* 25, 962–970. doi: 10.1523/JNEUROSCI.4507-04.2005

Corbit, L. H., and Balleine, B. W. (2011). The general and outcome-specific forms of pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *J. Neurosci.* 31, 11786–11794. doi: 10.1523/JNEUROSCI.2711-11.2011

Degasperi, G., Cristea, I. A., Di Rosa, E., Costa, C., and Gentili, C. (2021). Parsing variability in borderline personality disorder: A meta-analysis of neuroimaging studies. *Transl. Psychiatry* 11:314. doi: 10.1038/s41398-021-01446-z

Dolan, R. J., and Dayan, P. (2013). Goals and habits in the brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007

Esterman, M., Tamber-Rosenau, B. J., Chiu, Y.-C., and Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: Leave one subject out. *Neuroimage* 50, 572–576. doi: 10.1016/j.neuroimage.2009.10.092

Garbusow, M., Schad, D. J., Sebold, M., Friedel, E., Bernhardt, N., Koch, S. P., et al. (2016). Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addict. Biol.* 21, 719–731. doi: 10.1111/adb.12243

Geurts, D., den Ouden, H., Janssens, L., Swart, J., Fröböse, M. I., Cools, R., et al. (2022). Aversive inhibition in adult ADHD and its restoration by mindfulness-based cognitive therapy: A behavioral pilot study. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/aenr8

Geurts, D. E. M., Huys, Q. J. M., den Ouden, H. E. M., and Cools, R. (2013a). Aversive Pavlovian control of instrumental behavior in humans. *J. Cogn. Neurosci.* 25, 1428–1441. doi: 10.1162/jocn_a_00425

Geurts, D. E. M., Huys, Q. J. M., den Ouden, H. E. M., and Cools, R. (2013b). Serotonin and Aversive Pavlovian control of instrumental behavior in humans. *J. Neurosci.* 33, 18932–18939. doi: 10.1523/JNEUROSCI.2749-13.2013

Goodman, M., Carpenter, D., Tang, C. Y., Goldstein, K. E., Avedon, J., Fernandez, N., et al. (2014). Dialectical behavior therapy alters emotion regulation and amygdala activity in patients with borderline personality disorder. *J. Psychiatr. Res.* 57, 108–116. doi: 10.1016/j.jpsychires.2014.06.020

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., et al. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *J. Clin. Psychiatry* 69, 533–545.

Guitart-Masip, M., Duzel, E., Dolan, R., and Dayan, P. (2014). Action versus valence in decision making. *Trends Cogn. Sci.* 18, 194–202. doi: 10.1016/j.tics.2014.01.003

Gunderson, J. G. (2009). Borderline personality disorder: Ontogeny of a diagnosis. *Am. J. Psychiatry* 166, 530–539. doi: 10.1176/appi.ajp.2009.08121825

Gutteling, B. M., Montagne, B., Nijs, M., and van den Bosch, L. M. C. W. (2012). Dialectical behavior therapy: Is outpatient group psychotherapy an effective alternative to individual psychotherapy? Preliminary conclusions. *Compr. Psychiatry* 53, 1161–1168. doi: 10.1016/j.comppsych.2012.03.017

Hallquist, M. N., Hall, N. T., Schreiber, A. M., and Dombrovski, A. Y. (2018). Interpersonal dysfunction in borderline personality decision neuroscience perspective. *Curr. Opin. Psychol.* 21, 94–104. doi: 10.1016/j.copsyc.2017.09.011

Hazlett, E. A., Zhang, J., New, A. S., Zelmanova, Y., Goldstein, K. E., Haznedar, M. M., et al. (2012). Potentiated amygdala response to repeated emotional pictures in borderline personality disorder. *Biol. Psychiatry* 72, 448–456. doi: 10.1016/j.biopsych.2012.03.027

Hebart, M. N., and Gläscher, J. (2015). Serotonin and dopamine differentially affect appetitive and aversive general Pavlovian-to-instrumental transfer. *Psychopharmacologia* 232, 437–451. doi: 10.1007/s00213-014-3682-3

Heinz, A., Schlagenhauf, F., Beck, A., and Wackerhagen, C. (2016). Dimensional psychiatry: Mental disorders as dysfunctions of basic learning mechanisms. *J. Neural Transm.* 123, 809–821. doi: 10.1007/s00702-016-1561-2

Hinojosa-Aguayo, I., and González, F. (2020). Affect-driven impulsivity impairs human action control and selection, as measured through Pavlovian instrumental transfer and outcome devaluation. *Q. J. Exp. Psychol.* 73, 537–554. doi: 10.1177/1747021819883963

Hoertel, N., López, S., Wang, S., González-Pinto, A., Limosin, F., and Blanco, C. (2015). Generalizability of pharmacological and psychotherapy clinical trial results for borderline personality disorder to community samples. *Personal. Disord.* 6, 81–87. doi: 10.1037/per0000091

Huys, Q. J. M. (2018). Advancing clinical improvements for patients using the theory-driven and data-driven branches of computational psychiatry. *JAMA Psychiatry* 75, 225–226. doi: 10.1001/jamapsychiatry.2017.4246

Huys, Q. J. M., Browning, M., Paulus, M. P., and Frank, M. J. (2021). Advances in the computational understanding of mental illness. *Neuropsychopharmacology* 46, 3–19. doi: 10.1038/s41386-020-0746-4

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., et al. (2011). Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Comput. Biol.* 7:e1002028. doi: 10.1371/journal.pcbi.1002028.t002

Huys, Q. J. M., Gölzer, M., Friedel, E., Heinz, A., Cools, R., Dayan, P., et al. (2016). The specificity of Pavlovian regulation is associated with recovery from depression. *Psychol. Med.* 46, 1027–1035. doi: 10.1017/S0033291715002597

Iskric, A., and Barkley-Levenson, E. (2021). Neural changes in borderline personality disorder after dialectical behavior therapy–a review. *Front. Psychiatry* 12:772081. doi: 10.3389/fpsyt.2021.772081

Jacobson, N. S., and Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J. Consult. Clin. Psychol.* 59, 12–19.

Jocham, G., Klein, T. A., and Ullsperger, M. (2011). Dopamine-Mediated Reinforcement Learning Signals in the Striatum and Ventromedial Prefrontal Cortex Underlie Value-Based Choices. *J. Neurosci.* 31, 1606–1613. doi: 10.1523/JNEUROSCI.3904-10.2011

Jones, K. A., Menniti, F. S., and Sivarao, D. V. (2015). Translational psychiatry-light at the end of the tunnel. *Ann. N.Y. Acad. Sci.* 1344, 1–11. doi: 10.1111/nyas.12725

Krause-Utz, A., Winter, D., Niedtfeld, I., and Schmahl, C. (2014). The latest neuroimaging findings in borderline personality disorder. *Curr. Psychiatry Rep.* 16:438. doi: 10.1007/s11920-014-0438-z

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., et al. (1996). The reliability and validity of the outcome questionnaire. *Clin. Psychol. Psychother.* 3, 249–258.

Lana, F., and Fernández-San Martín, M. I. (2013). To what extent are specific psychotherapies for borderline personality disorders efficacious? A systematic review of published randomised controlled trials. *Actas Esp. Psiquiatr.* 41, 242–252.

Linehan, M. M. (1993). *Cognitive-behavioral Treatment of Borderline Personality disorder*. New York: NY: Guilford Press.

Ly, V., Cools, R., and Roelofs, K. (2014). Aversive disinhibition of behavior and striatal signaling in social avoidance. *Soc. Cogn. Affect. Neurosci.* 9, 1530–1536. doi: 10.1093/scan/nst145

Mânsson, K. N. T., Frick, A., Boraxbekk, C.-J., Marquand, A. F., Williams, S. C. R., Carlbring, P., et al. (2015). Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Transl. Psychiatry* 5:e530. doi: 10.1038/tp.2015.22

Minzenberg, M. J., Fan, J., New, A. S., and Tang, C. Y. (2007). Fronto-limbic dysfunction in response to facial emotion in borderline personality disorder: An event-related fMRI study. *Psychiatry Res.* 155, 231–243. doi: 10.1016/j.pscychresns.2007.03.006

Niedtfeld, I., Schmitt, R., Winter, D., Bohus, M., Schmahl, C., Sabine, C., et al. (2017). Pain-mediated affect regulation is reduced after dialectical behavior therapy in borderline personality disorder: A longitudinal fMRI study. *Soc. Cogn. Affect. Neurosci.* 12, 739–747. doi: 10.1093/scan/nsw183

Nitschke, J. B., Sarinopoulos, I., Oathes, D. J., Johnstone, T., Whalen, P. J., Davidson, R. J., et al. (2009). Anticipatory activation in the amygdala and anterior cingulate in generalized anxiety disorder and prediction of treatment response. *Am. J. Psychiatry* 166, 302–310. doi: 10.1176/appi.ajp.2008.07101682

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776. doi: 10.1016/j.conb.2004.10.016

Paulus, M. P., Huys, Q. J. M., and Maia, T. V. (2016). A roadmap for the development of applied computational psychiatry. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1, 386–392. doi: 10.1016/j.bpsc.2016.05.001

Perez, D. L., Vago, D. R., Pan, H., Root, J., Tuescher, O., Fuchs, B. H., et al. (2016). Frontolimbic neural circuit changes in emotional processing and inhibitory control associated with clinical improvement following transference-focused psychotherapy in borderline personality disorder. *Psychiatry Clin. Neurosci.* 70, 51–61. doi: 10.1111/pcn.12357

Pizzagalli, D. A. (2010). Frontocingulate dysfunction in depression: Toward biomarkers of treatment response. *Neuropsychopharmacology* 36, 183–206. doi: 10.1038/npp.2010.166

Pool, E., Brosch, T., Delplanque, S., and Sander, D. (2015). Stress increases cue-triggered "wanting" for sweet reward in humans. *J. Exp. Psychol. Anim. Learn. Cogn.* 41, 128–136. doi: 10.1037/xan0000052

Poser, B. A., Versluis, M. J., Hoogduin, J. M., and Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magn. Reson. Med.* 55, 1227–1235. doi: 10.1002/mrm.20900

Prevost, C., Liljeholm, M., Tyszka, J. M., and O'Doherty, J. P. (2012). Neural correlates of specific and general pavlovian-to-instrumental transfer within human Amygdalar Subregions: A high-resolution fMRI Study. *J. Neurosci.* 32, 8383–8390. doi: 10.1523/JNEUROSCI.6237-11.2012

Quail, S. L., Morris, R. W., and Balleine, B. W. (2016). Stress associated changes in Pavlovian-instrumental transfer in humans. *Q. J. Exp. Psychol.* 70, 675–685. doi: 10.1080/17470218.2016.1149198

Roiser, J. P., Elliott, R., and Sahakian, B. J. (2011). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology* 37, 117–136. doi: 10.1038/npp.2011.183

Salvador, R., Vega, D., Pascual, J. C., Marco, J., Canales-Rodríguez, E. J., Aguilar, S., et al. (2016). Converging medial frontal resting state and diffusion-based abnormalities in borderline personality disorder. *Biol. Psychiatry* 79, 107–116. doi: 10.1016/j.biopsych.2014.08.026

Sampedro, F., Farrés, C. C. I., Soler, J., Elices, M., Schmidt, C., Corripio, I., et al. (2021). Structural brain abnormalities in borderline personality disorder correlate with clinical severity and predict psychotherapy response. *Brain Imaging Behav.* 15, 2502–2512. doi: 10.1007/s11682-021-00451-6

Schmitgen, M. M., Niedtfeld, I., Schmitt, R., Mancke, F., Winter, D., Schmahl, C., et al. (2019). Individualized treatment response prediction of dialectical behavior therapy for borderline personality disorder using multimodal magnetic resonance imaging. *Brain Behav.* 9:e01384. doi: 10.1002/brb3.1384

Schmitt, R., Winter, D., Niedtfeld, I., Herpertz, S. C., and Schmahl, C. (2016). Effects of psychotherapy on neuronal correlates of reappraisal in female patients with borderline personality disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1, 548–557. doi: 10.1016/j.bpsc.2016.07.003

Schnell, K., and Herpertz, S. C. (2007). Effects of dialectic-behavioral-therapy on the neural correlates of affective hyperarousal in borderline personality disorder. *J. Psychiatr. Res.* 41, 837–847. doi: 10.1016/j.jpsychires.2006.08.011

Schulze, L., Schmahl, C., and Niedtfeld, I. (2016). Neural correlates of disturbed emotion processing in borderline personality disorder: A

multimodal meta-analysis. *Biol. Psychiatry* 79, 97–106. doi: 10.1016/j.biopsych.2015.03.027

Schulze, L., Schulze, A., Renneberg, B., Schmahl, C., and Niedtfeld, I. (2019). Neural correlates of affective disturbances: A comparative meta-analysis of negative affect processing in borderline personality disorder, major depressive disorder, and posttraumatic stress disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 220–232. doi: 10.1016/j.bpsc.2018.11.004

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., and Huettel, S. A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *J. Neurosci.* 30, 2490–2495. doi: 10.1523/JNEUROSCI.3319-09.2010

Soloff, P. H., Abraham, K., Ramaseshan, K., Burgess, A., and Diwadkar, V. A. (2017). Hyper-modulation of brain networks by the amygdala among women with Borderline Personality Disorder: Network signatures of affective interference during cognitive processing. *J. Psychiatr. Res.* 88, 56–63. doi: 10.1016/j.jpsychires.2016.12.016

Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., et al. (2016). Charting the landscape of priority problems in psychiatry, part 1: Classification and diagnosis. *Lancet Psychiatry* 3, 77–83. doi: 10.1016/S2215-0366(15)00361-2

Stoffers, J. M., Völlm, B. A., Rücker, G., Timmer, A., Huband, N., and Lieb, K. (2012). Psychological therapies for people with borderline personality disorder. *Cochrane Database Syst. Rev.* 2012:CD005652. doi: 10.1002/14651858.CD005652.pub2

Swart, J. C., Frobose, M. I., Cook, J. L., Geurts, D. E., Frank, M. J., Cools, R., et al. (2017). Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. *eLife* 6:e22169. doi: 10.7554/eLife.22169

Talmi, D., Seymour, B., Dayan, P., and Dolan, R. J. (2008). Human Pavlovian Instrumental Transfer. *J. Neurosci.* 28, 360–368. doi: 10.1523/JNEUROSCI.4028-07.2008

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, 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, 273–289. doi: 10.1006/nimg.2001.0978

Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/JNEUROSCI.0564-07.2007

van Timmeren, T., Quail, S. L., Balleine, B. W., Geurts, D. E. M., Goudriaan, A. E., and van Holst, R. J. (2020). Intact corticostriatal control of goal-directed action in Alcohol Use Disorder: A Pavlovian-to-instrumental transfer and outcome-devaluation study. *Sci. Rep.* 10:4949. doi: 10.1038/s41598-020-61892-5

Watson, P., Wiers, R. W., Hommel, B., and de Wit, S. (2014). Working for food you don't desire. Cues interfere with goal-directedfood-seeking. *Appetite* 79, 139–148. doi: 10.1016/j.appet.2014.04.005

Westlund Schreiner, M., Klimes-Dougan, B., Mueller, B. A., Nelson, K. J., Lim, K. O., and Cullen, K. R. (2019). Neurocircuitry associated with symptom dimensions at baseline and with change in borderline personality disorder. *Psychiatry Res. Neuroimaging* 290, 58–65. doi: 10.1016/j.pscychresns.2019.07.001

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage* 92, 381–397. doi: 10.1016/j.neuroimage.2014.01.060

Winter, D., Niedtfeld, I., Schmitt, R., Bohus, M., Schmahl, C., and Herpertz, S. C. (2017). Neural correlates of distraction in borderline personality disorder before and after dialectical behavior therapy. *Eur. Arch. Psychiatry Clin. Neurosci.* 267, 51–62. doi: 10.1007/s00406-016-0689-2

Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited–again. *Neuroimage* 2, 173–181. doi: 10.1006/nimg.1995.1023

Wunderlich, K., Rangel, A., and O'Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17199–17204. doi: 10.1073/pnas.0901077106

Wunsch, E.-M., Kliem, S., and Kröger, C. (2014). Population-based cost–offset estimation for the treatment of borderline personality disorder: Projected costs in a currently running, ideal health system. *Behav. Res. Ther.* 60, 1–7. doi: 10.1016/j.brat.2014.06.002