Data supplement for Yang et al., Effect of CBT on Biased Semantic Network in Panic Disorder: A Multicenter fMRI Study Using Semantic Priming. Am J Psychiatry (doi: 10.1176/appi.ajp.2019.19020202)

CONTENTS

S1. Theoretical background and hypotheses

According to the clinical observation of panic attacks triggered by agoraphobic situations and the bio-informational theory of Lang (1), we propose a biased semantic memory in patients with panic disorder. In their semantic network, panic-related concepts reveal negative valence (signified by the red color in Figure S1) and enhanced interconnection between them (e.g., elevator–dizziness; signified by the black double arrow in Figure S1). These characteristics of their panic-related semantic network are supposed to correlate with behavioral and neural alterations of PD. At the behavioral level, panic-related concepts, for example, panic trigger and panic symptoms word pairs should be rated with increased negative valence and enhanced relatedness. The processing of symptom words should be automatically facilitated when primed with trigger words because of their enhanced relatedness. At the neural level, the increased negative valence should activate the salience [anterior cingulate cortex (ACC) and insula] and defensive network (amygdala and thalamus; brain regions circled by red ellipses) and the priming-related facilitation correlated with activation suppression in semantic network caused by facilitated semantic access, retrieval, and selection [brain regions circled by black ellipses, e.g., the left inferior frontal gyrus (IFG) and temporal cortex]. Cognitive-behavioral therapy (CBT, first-line treatment for panic disorder) leads to clinical improvement and should reduce the behavioral and neural alterations (see Figure S1 the blue ellipse and arrows).

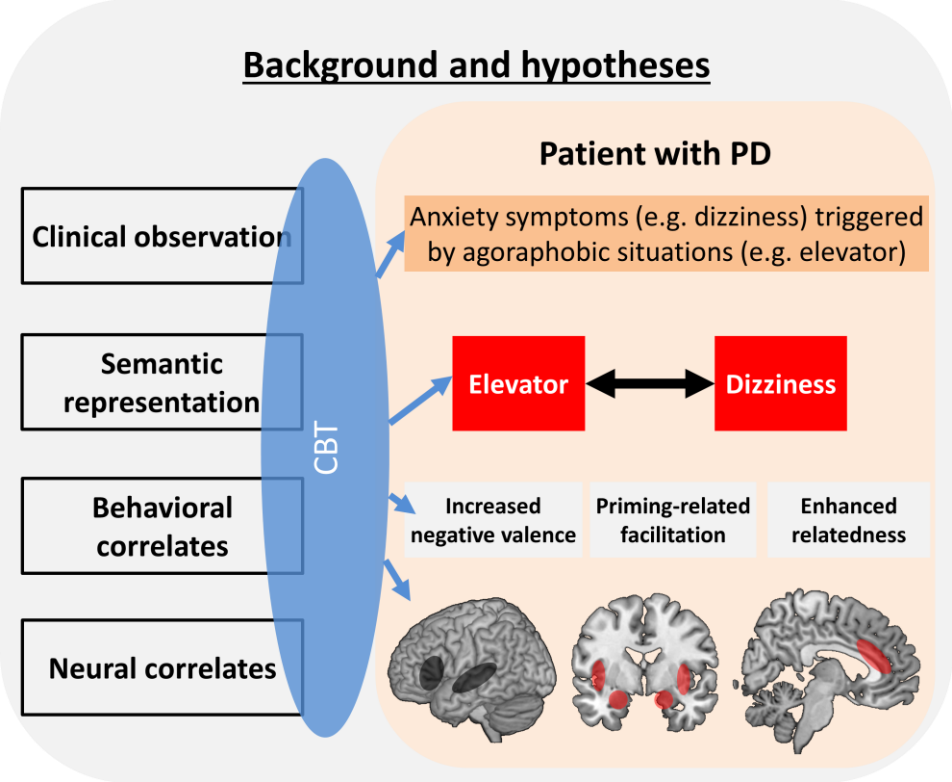


FIGURE S1: Schematic presentation of behavioral and neural correlates of panic-related semantic network

S2. Inclusion and exclusion criteria for patients and clinical assessments

S2.1 Patients: inclusion and exclusion criteria

Inclusion criteria for all patients were as follows: (A) a current primary diagnosis of panic disorder with or without agoraphobia (PD/PDA; evidenced by the Composite International Diagnostic Interview [CAPI-WHO-CIDI; DIAX-CIDI version]); (B) a score ≥ 3 on the Clinical Global Impressions Scale (CGI); (C) an age of 18–65 years.

Exclusion criteria were (A) comorbid DSM-IV-TR psychotic or bipolar I disorder; (B) current alcohol dependence/current abuse or dependence on benzodiazepine and other psychoactive substances; (C) current suicidal intent; (D) borderline personality disorder; (E) concurrent ongoing psychopharmacological treatment for PD/PDA or another mental disorder; (F) antidepressant or anxiolytic pharmacotherapy. The HCs were free of current or past medical, neurological, or psychiatric illness. Exclusion criteria for both groups were a cardiac pacemaker, ferromagnetic metal implants, tattoos, or permanent make up with ferromagnetic colors.

S2.2 Clinical Assessments

Assessors: Assessors took part in a three-day training course. This included conducting clinical interviews, knowledge of operational procedures, and training in the web-based study database. The assessors were subsequently certified wherein proper administration of the instruments was tested. Biweekly conference calls were held to maintain consistent strategies and address questions. Patients directly entered self-administered clinical assessments into an internet-based computer interface. Patients were trained by their study therapist in the use of the computer program. All data were linked with the corresponding login password so that every change of the database was time-stamped and could be tracked. The database was saved at a central data coordinating center (study coordination center; KKS at the Technische Universität Dresden) that also ensured data security. The database was checked regularly, and the time of entry was compared against the scheduled entry time. Therapists and clinical directors of each

center received regular feedback about the quality and timeliness of data for each of their patients.

A maximum of two weeks before the fMRI measurements, clinical assessments were performed on each patient. The following clinical interviews and questionnaires were used in the assessments:

Composite International Diagnostic Interview (CIDI): The standardized computer-administered personal CIDI is administered by expert interviewers and systematically assesses all DSM–IV disorders. The diagnoses derived by the CIDI have been demonstrated to be reliable and valid (2).

Hamilton Anxiety Rating Scale (HAM-A/SIGH-A): The SIGH-A is a clinician-rated interview that measures a broad range of anxiety symptoms, with scores that range between 0 and 56. The SIGH-A was utilized in this study. SIGH-A scores have been documented with high interrater reliability and test-retest reliability (3).

Clinical Global Impression (CGI): The CGI is a clinician-rated scale that measures the overall severity of a disorder, with scores that range between 1 (no symptoms) and 7 (among the most severely ill patients). To maximize reliability, we instructed interviewers to consider panic symptoms, anxiety, anticipatory anxiety, avoidance, and functional level before making the global rating. The rating thus represents a global assessment of the patient's severity and functioning concerning presenting symptoms (4).

Body Sensations Questionnaire (BSQ): The BSQ (5) is a questionnaire with 17 items measuring fear of the body sensations associated with high arousal and panic, such as rapid heartbeat. Items are rated on 1- to 5-point scales ranging from "not frightened or worried by this sensation" to "extremely frightened by this sensation." The BSQ has a high internal consistency (e.g., alpha = 0.88) and acceptable test-retest reliability (r = 0.67).

Agoraphobic Cognitions Questionnaire (ACQ): The ACQ (5) is a self-report questionnaire with 14 items measuring maladaptive thoughts about the possible consequences of panic (e.g.,

dizziness). Items are rated on 1- to 5-point scales ranging from "thought never occurs when I'm anxious" to "thought always occurs when I'm anxious." The BSQ has a high internal consistency (e.g., alpha = 0.80) as well as a high test-retest reliability (r = 0.86).

Both BSQ and ACQ were administrated only in the study centers providing CBT (Study center Greifswald, Marburg, and Münster). Patients from study center Berlin and Dresden (where fMRI-scans were performed only at baseline) were not measured with these two questionnaires.

Mobility Inventory (MI): The MI is a self-report questionnaire that measures the degree of agoraphobic avoidance across 27 situations, each of which is rated with respect to being in that situation alone or accompanied by another person (6). The mean score, as well as the Alone and Accompanied subscale (range 1–5), is reported with the 7-day version of the Mobility Inventory. The MI-7 is identical to the original MI except that respondents are instructed to report only on the previous seven days. Scores of the MI are highly reliable and sensitive to change.

Panic and Agoraphobia Scale (PAS): The PAS is a self-report questionnaire that measures the severity of panic attacks, avoidance, anticipatory anxiety, disability, and worries about health. Scores on the PAS have been demonstrated to have good reliability and sensitivity to change (7).

Anxiety Sensitivity Index (ASI): The ASI is a 16-item self-report measuring subjects beliefs about potential harmful consequences of anxiety-related symptoms. Each item is rated on a five-point scale from 0 (very little) to 4 (very much) (8)**.**

Beck Depression Inventory-II (BDI-II): The BDI-II is a 21-item self-report questionnaire measuring severity of symptoms associated with depression. Symptoms are rated on a four-point scale within the time frame of the past seven days (9).

Brief Symptom Inventory (BSI): The BSI is a self-report symptom scale designed to measure levels of psychopathology (10). The BSI is a shortened form of the SCL-90-R Symptom Checklist and included 53 items and nine subscales *Somatization, Obsessive-compulsive, Interpersonal*

Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. Participants were asked to rate their degree (from 0 “not-at-all” to 4 “extremely”) of suffering from each problem during the past week.

S2.3 Primary outcome clinical variables

Primary outcome parameters targeted domains of global anxiety and panic/agoraphobic-specific symptomatology. They were assessed at baseline, an intermediate assessment after the fourth session, posttreatment, and 6-month follow-up. Specific exposure-related questions were assessed after every session (for a more extensive description of the assessments and methods of this randomized controlled trial and the research network, see 11). Primary outcome variables included expert clinical judgment, measured with (1) the SIGH-A and (2) the CGI, (3) the panic disorder and agoraphobia severity measured by questionnaire PAS, (4) number of panic attacks during the last week (respective item of the PAS), and (5) agoraphobic avoidance as measured by the MI.

Since the current fMRI study deals with the general effects of CBT on the neural correlates of PD, and not the specific effects of successful psychotherapy, we did not compare therapy responders with nonresponders. However, we investigated the correlations between the clinical improvements and the changes in neural activation. Therefore, we did not apply any criterion for therapy response. The criteria for therapy-responders will be reported in our upcoming clinical research paper.

S2.4 Neuropsychological tests

To compare the neuropsychological performance of PD and HC, Trail Making Test (TMT) and Digit Span Forward and Backward were administered to measure the participants’ speed of processing, mental flexibility, executive functions and working memory (12,13). Regensburg Word Fluency Test (RWT) was used to measure their phonologic and semantic word fluency (14).

S2.5 Procedure of recruitment and measurement of healthy controls

Healthy control subjects were recruited by public advertisement or selected from pools of healthy controls, who have already participated in fMRI studies and agreed to participate in further studies. After the primary telephone screening, only healthy control subjects without current or past mental health problem, family history of mental health problem, history of neurological disorders, and contraindication for MRI measurement were invited for further measurements. They were assessed with CIDI, HAM-A/SIGH-A, ASI, BDI, BSI, NEO five-factor inventory (NEO-FFI), and all neuropsychological tests. However, about 10% of the healthy control subjects (mostly from only one center) did not undergo clinical assessment due to a management problem.

S3. Manualized, exposure-based cognitive behavioral therapy for PD and PDA

The highly standardized and controlled CBT protocol was adopted from Gloster et al. (2011), which were demonstrated as effective (15). This CBT protocol consisted of psychoeducation and an individualized behavioral analysis of the patient's symptoms and coping behavior, providing the treatment rationale of exposure and implementing exposure exercises.

For patients with PD only, CBT was administered in 6 twice-weekly sessions. Exposure sessions included only interoceptive exposure. Two variants of therapy were applied, which only differ in their sequence of standard interoceptive exposure exercises.

For patients with PD and comorbid agoraphobia (PDA), CBT was administered in 12 twice-weekly sessions. Additionally, to the above therapeutic components for patients with PD only, patients with PDA received therapist guided in-situ exposure to patients' individual agoraphobic situations. Again, two variants of therapy were applied, which only differ in the specific implementation of in-situ exposure exercises (additional provocation of bodily sensations or not).

Because the variants of CBT were identical in content, structure, and doses of exposure and demonstrated significant symptom reduction after CBT according to preliminary analyses, groups for variants of treatment were collapsed in the current study.

Since both patients with PD and patients with PDA have benefited from CBT and revealed significant clinical improvements, the two groups were collapsed in the current study to maximize the sample size.

Therapists: Before the beginning of the study, all therapists (either licensed psychotherapists or psychology graduates currently on CBT-training) received a three-day training workshop. The treatment was explained and practiced using roleplay. Thereafter, therapists were evaluated during a recorded role-play of selected parts of the treatment to ensure competence and adherence. Those who passed (N=39) this test were eligible to treat patients in the study. Weekly supervision and videotaping of all sessions, except those sessions that included

exposure in vivo, was implemented to maintain therapy integrity and identify violations of the protocol. Eighty-four randomly selected videotapes in the overall clinical trial (including patients not participating in the fMRI task) were evaluated in terms of treatment integrity. Two independent raters who were not involved in study recruitment or patient treatment were trained, and rated therapists' adherence and competence on a developed scale (4-point Likert scale ranging from 0=not adherent at all to 3=total adherence). Mean treatment adherence was rated between good and very good (M=2.64, SD=.31) with excellent inter-rater reliability (ICC=.93).

S4. Study centers, technical parameters for fMRI data acquisition, and measures of quality control for fMRI measurement

fMRI measurements were conducted in five German centers [Berlin (Center 1), Dresden (Center 2), Greifswald (Center 3), Marburg (Center 4), and Münster (Center 5)], which were participating centers for the national research initiative Panic-Net (funded by the German Federal Ministry of Education and Research, BMBF). These centers have a long-standing tradition of collaborative multi-center fMRI studies, (e.g., 16–19). All fMRI data were acquired using 3T scanners. The following scanners were used: a 3T Philips Achieva scanner (Philips Medical Systems, Best, Netherlands) in Münster; a 3T Siemens Trio scanner (Siemens AG, Erlangen, Germany) in Dresden and Marburg; a 3T Siemens Verio scanner (Siemens AG, Erlangen, Germany) in Greifswald; and a 3T General Electric Healthcare scanners (General Electric Healthcare, Milwaukee, Wisconsin) in Berlin. A standard operating procedure was created to guide the experimenter and standardize the measurements in each center. Before the experiment was conducted in each center, a coordinator carefully evaluated the equipment and experimental procedure.

fMRI data acquisition: fMRI brain images were acquired using 3T MRI scanners in each center. A total of 435 transversal functional images (echo-planar images, 64 × 64 matrix; 30 slices ascending; field of view [FoV] = 230 mm; repetition time [TR] = 2,000 ms; echo time [TE] = 30 ms; flip angle = 90; slice thickness = 3.8 mm; voxel resolution = 3.6 × 3.6 mm) that covered the whole brain and were positioned parallel to the intercommissural line (anterior commissure—posterior commissure) were recorded. Routine quality control measures were performed to ensure a high standard of fMRI data acquisition and data quality.

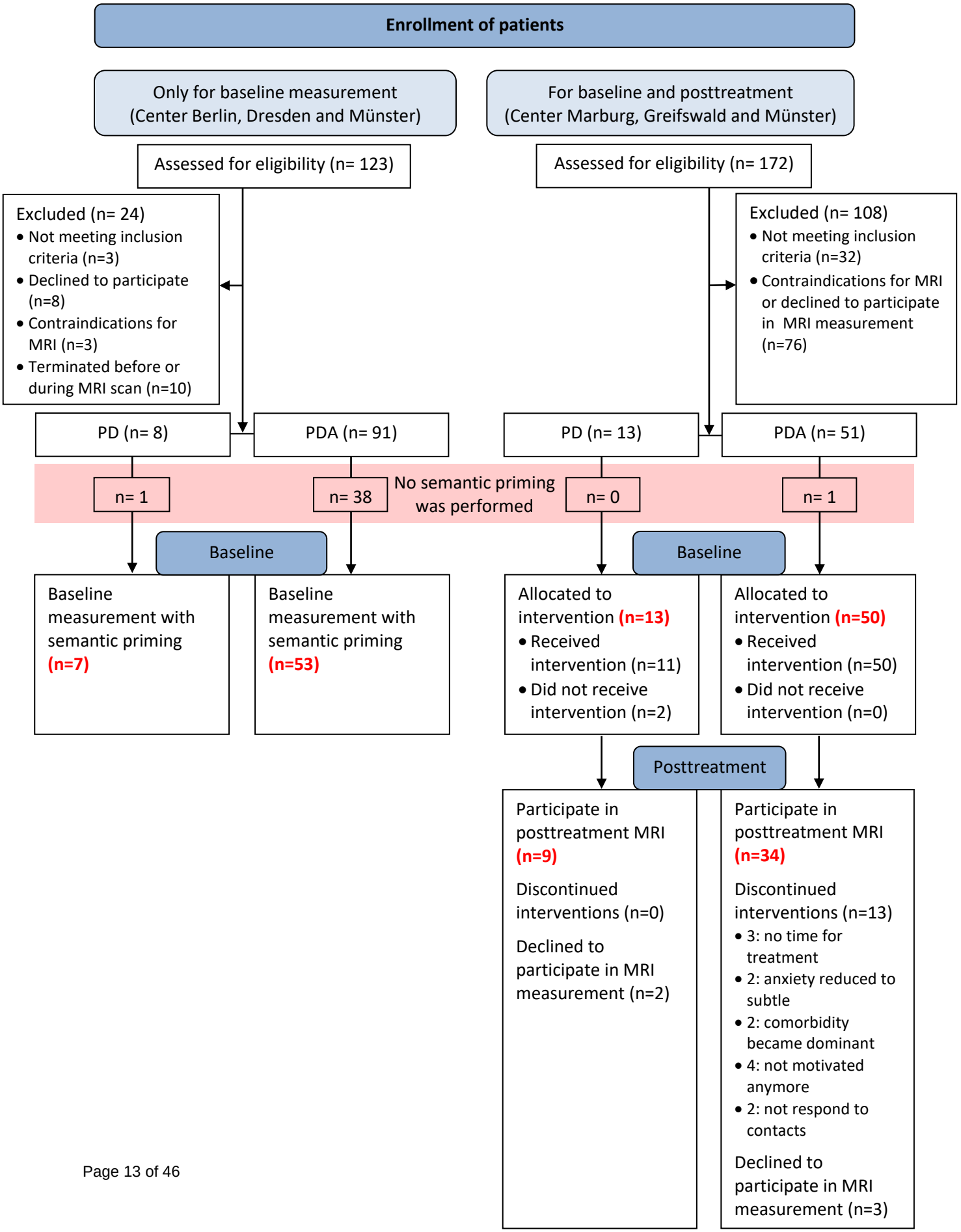
As a further quality control, we performed a phantom measurement after each session to determine the stability of the scanner by quantifying the signal fluctuation (percentage signal change, PSC; see (20). The PSC of the experimental fMRI data and head movement were also checked before the data analysis. PD and HC revealed no significant differences in PSC parameters ($P = 0.23$).

Procedure: Besides the semantic priming paradigm, two other tasks embedded in the Panic-Net study were also performed by the subjects during the fMRI measurement. The semantic priming paradigm and another paradigm on interoceptive processing were always performed as the first or second paradigms, which was counterbalanced in the scan order. Including a structural scan at the end of the experiment, the total scan duration was about 50 minutes.

Patients with PD were measured before and after CBT. The two measurements have an average time interval of 12.31±5.38 weeks. The HCs were also measured twice with an average time interval of 12.04±6.25 weeks. The two groups revealed no significant differences in the time interval (t=0.22, p=.83).

S5. CONSORT Panic-Net II Flow Diagram and Attritions

FIGURE S2: Patient flow chart in Panic-Net II



In the Panic-Net-II study, three study centers participated in the clinical trial and treated the patients with PD or PDA. These patients were measured twice (at baseline and posttreatment). Two study centers only recruited patients with PD and PDA for baseline measurement (see Figure S2). The study center Münster also recruited an additional three patients for baseline measurement only after the termination of the clinical trial.

In the clinical trial, 51 PDA and 13 PD patients, who underwent baseline fMRI measures with semantic priming task, were assigned to treatments. However, 34 PDA and 9 PD patients participated in the posttreatment fMRI measurement. The reasons for dropouts are the following (see Figure S2): one patient did not enter the treatment (dropout before treatment); one patient started to take medication (dropout before treatment); three patients could not manage to take the treatment regularly (new job, limited time, etc.); two patients experienced only subtle anxiety after some sessions and were therefore not motivated for further sessions; two patients terminated the treatment because their comorbidity became dominant; four patients were not motivated for treatment anymore; two patients did not respond to contacts; five patients withdrew their consent to participate in the fMRI postmeasurement (however, they completed the treatment).

In total, 125 Patients and 152 HC were recruited in the five centers. Forty-three patients and 52 HC were measured twice (pre-/post-CBT or waiting period, respectively). Two patients with fMRI measurements were excluded afterward because they did not fulfill the DSM-IV diagnostic criterion for PD (one had only panic attacks; one was diagnosed with a social anxiety disorder). Further quality control steps were applied for the behavioral and fMRI data from the remaining 123 PD and 152 HC.

S6. Quality control for behavioral and fMRI data and resulting data quality

Quality control steps were applied for the behavioral and fMRI data from the 123 PD and 152 HC. First, we excluded five PD and two HC, since they had more than 25% response error during the lexical decision task. The behavioral data, including response time and error rate of 118 PD and 150 HC, were qualified for the further analyses of priming effects. A subsample (n=78) of the 150 HC were used to investigate the neural correlates of anxiety sensitivity. Results were published by Yang et al. (21) in the Journal "*Social, Cognitive and Affective Neuroscience*." Among these participants, 113 PD and 139 HC performed the rating task after the fMRI scan, which were included for the analyses of effects in relatedness and valence. Second, we excluded fMRI data of all the left-handed participants, since the left-handed participants had a higher incidence of atypical language lateralization (22). This step resulted in an exclusion of five PD and two HC, who reported a left dominance in the Edinburgh Handedness Inventory. Finally, quality control was applied to functional MRI data to exclude the datasets with serious flaws, which could affect/mislead our fMRI results based on the group. Together, 10 MRI data sets of the control group (5%) and 12 data sets of the patient group (7.7%) were excluded from the total sample due to insufficient fMRI data quality identified by the following hierarchical four steps procedure.

The MRI data quality assurance started with **(step 1)** a visual inspection of the neurological abnormalities according to T1 structural MRI image (cysts, tumor, lesion, neurodegeneration) and raw fMRI image (drop out of signal in the brain regions). Next, **(step 2)** head movement was set at 1.5 voxel sizes to exclude data sets (maximal movement > 5.4 mm). **Step 3** assessed variability in single-subject whole-brain functional volumes, determined using the Artifact Recognition Toolbox (http://www.nitrc.org/projects/artifact_detect). Individual whole-brain fMRI volumes meeting at least one of two criteria were flagged and regressed out when determining task-specific effects: 1) significant mean-volume signal intensity variation (i.e., within volume mean signal greater or less than 4 standard deviations of mean signal of all volumes in time series), and 2) individual volumes where scan-to-scan movement exceeded 2 mm translation or 2° rotation in any direction. Participants with 5% or more flagged volumes

were excluded from the analysis. In the final quality assurance **(step 4)**, we excluded the fMRI data sets with insufficient overall activity and insufficient activity in visual (during stimulus presentation) and premotor cortices (during button press). Since our participants were presented with semantic stimuli quite frequently, and they were asked to respond to the stimuli with their left hand, we expected strong neural activation as well as BOLD fMRI responses within the whole brain, the visual cortex, and the right premotor cortices. We conducted a quality control using the results from the first-level analysis for each participant. In the model of first-level analyses, the hemodynamic response triggered by the target words in all six conditions (including N–N, T–N, N–S, T–S, N–P, and T–P) was modeled with a canonical haemodynamic response function. Realignment parameters were included as regressors of no interest to account for movement artifacts. A high-pass filter with a cutoff period of 128 seconds was applied. The parameter estimate (ß) and t-statistic image of each condition were calculated for each subject. To explore the general task-related neural activation at an individual level, we contrasted all active conditions to baseline using the first level results. The threshold was set at $p<0.005$, uncorrected, which is quite liberal for such basic visual-motor effects. The average number of activated voxels for all participants was 42,769. However, there were fewer participants (n=8, less than 3%) revealing less than 2000 activated voxels (Figure S3 A), which were excluded from our fMRI sample. Furthermore, we performed a second-level group analysis by entering the parameter estimates for all six conditions (N–N, T–N, N–S, T–S, N–P and T–P) as within-subject variable into a flexible factorial analysis using SPM8. The two groups (PD & HC) were not separated in this analysis. We checked the global activation during the active task conditions using an effect of interest (EOI) analysis. We identified one peak in the right premotor cortex (F=311.0) and very strong activation in the visual cortex (F=202.3). We extracted the beta values from the sphere with a radius of 8 mm around the peak voxel in the right premotor cortex and the V1. We excluded the 2% (n=8) of the dataset with the least activation in the two brain regions (Figure S3 B). Since excluded data sets from this step have overlap with the exclusion from the previous steps, this step has resulted in the exclusion of only 7 data sets (n=4 for patients and n=3 for HC).

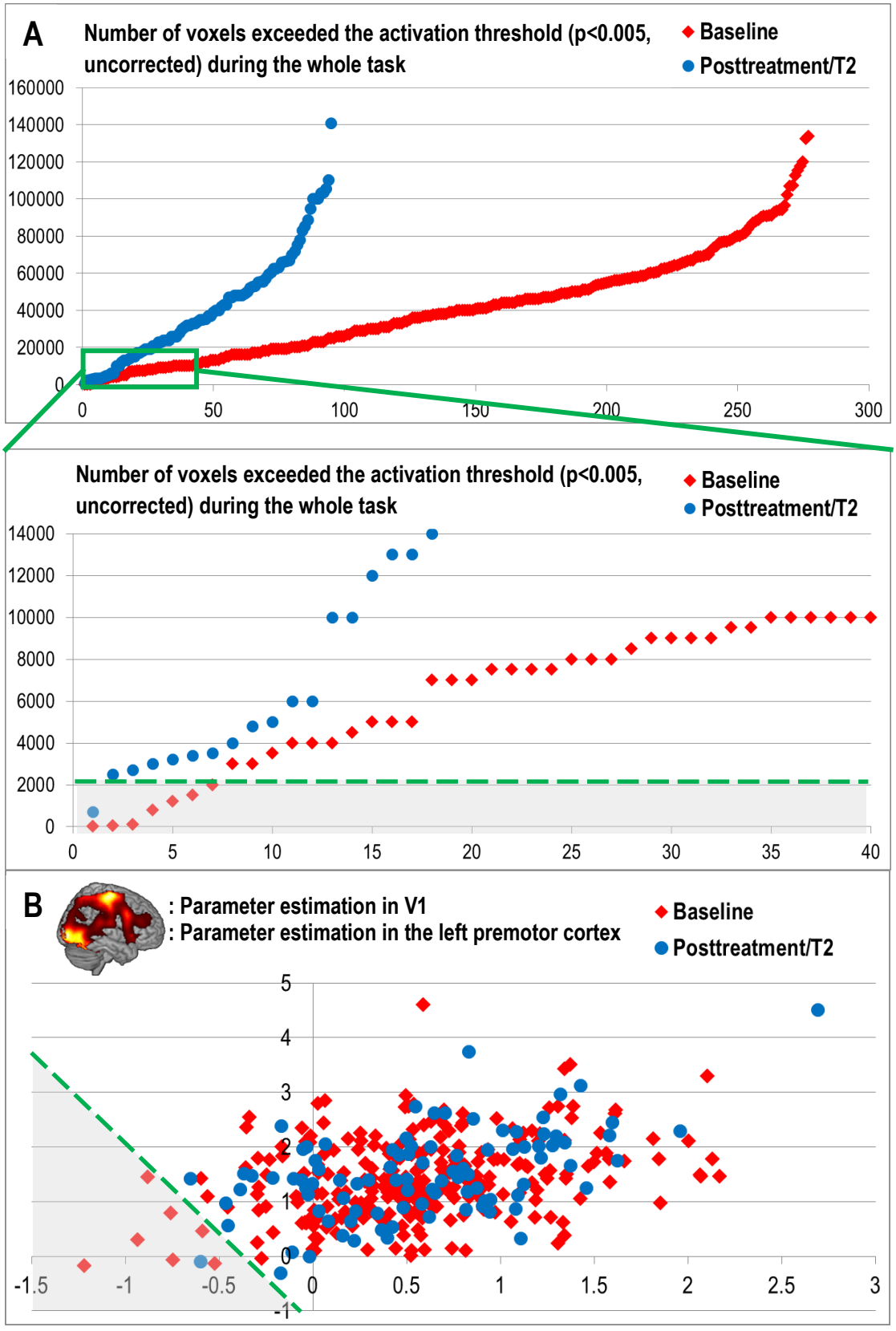


FIGURE S3: fMRI data quality checks according to the activation at the individual level. A: 2% of the data sets, which have shown the least voxels exceeded the activation threshold of p<0.005 during the active conditions in the lexical decision task. B: 2% of the data sets, which were identified as having insufficient activation both in the V1 and the premotor cortex during the active conditions.

The entire quality control procedure is illustrated in Figure S4. The number of data sets, which were excluded because of their flawed data quality, is also presented in this figure, separately for patients and healthy control subjects.

After the exclusion of data sets with insufficient data quality, a strict matching with regard to gender, age, study center, set, and pseudo-randomized trials' order of the semantic priming paradigm was performed with the remaining 103 data sets of patients and 139 data sets of the control group at baseline leading to a final sample of 103 patients and 103 matched controls. Likewise, 39 patients with both qualified baseline and posttreatment fMRI data were matched with 39 HC with qualified baseline and T2 fMRI data. The group distribution in the final sample with regard to gender, age, education, center, tobacco use, sets of semantic material, and trails' order are shown in Table S1, which presents social demographic and psychological characteristics of patients with PD or PDA and healthy control subjects in the subsample for fMRI analyses. PD and HC could not be matched perfectly at posttreatment/T2 regarding gender. Study center could also not be completely matched due to the uneven exclusion of flawed data across centers.

After our quality assurance steps and matching procedure, we still observed a significant difference in the years of education (X^2=14.21, df=6, p=0.03) and tobacco use (X^2=9.30, df=3, p=0.03) between our four groups. Although patients had significantly lower education than HC, however, their word fluency was not significantly different from each other (F<0.62, Table S1). We controlled for the influence of tobacco use and years of education by adding them in the second level analysis model as covariant. The other variables, including age, gender, set, trails' order, and PSC, are all comparable in the two groups (P>0.24, Table S1).

Since we have applied a relatively liberal exclusion criterion for head motion (> 5.4mm) to maximize the included patients, we compared the final sample of PD and HC to ensure that there were no significant group differences and the neural activations are not related to the head motion. In our final sample, the patients showed comparable head motion compared to HC at both time points (see Table S1; F=0.77; p=0.51). Furthermore, the head motion was not correlated to neural activation in brain regions showing group differences (e.g. ACC: r_{PD}=0.1; r_{HC}=0.05; left superior temporal gyrus: r_{PD}=0.07; r_{HC}=0.04).

FIGURE S4: Flow chart for data quality control steps and samples for analyses

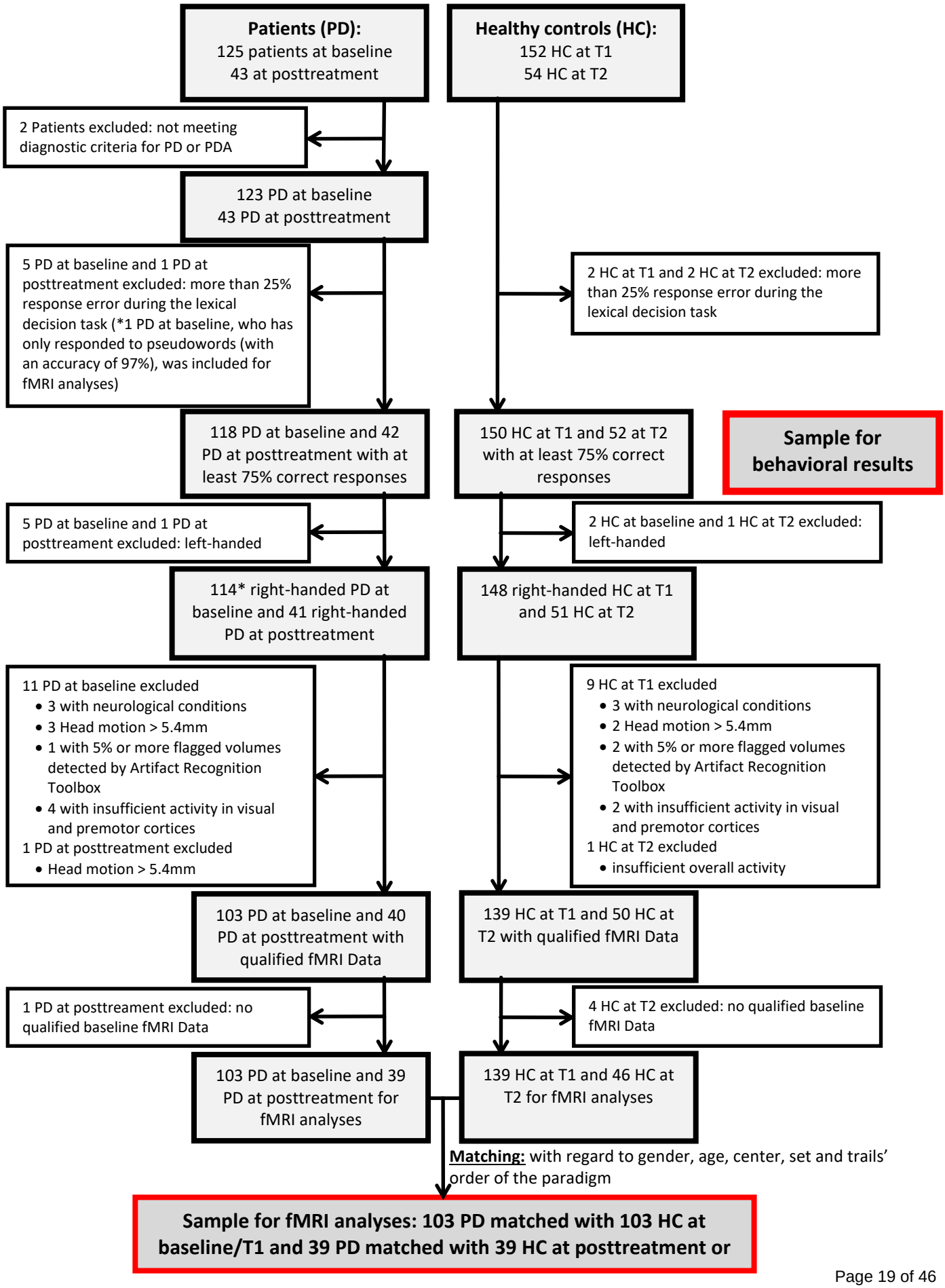


TABLE S1: Sociodemographic and clinical characteristics of patients with panic disorder (PD) and healthy control subjects (HC) in fMRI subsample

	PD		HC		
	Baseline (*n* = 103)	Post-treatment (n = 39)	T1 (*n* = 103)	T2 (*n* = 39)	*Tests*
Age in years	31.84±9.95	32.03±10.52	31.83±10.46	32.72±10.66	F=0.08
Female gender	59 (57%)	22 (56%)	59 (57%)	23 (59%)	X^2=0.06
Tobacco use	49 (44%)	24 (57%)	36 (33%)	14 (37%)	X^2=9.30*
Years of education					X^2=14.21*
≤ 8	5	1	2	1	
9 – 11	34	11	15	6	
≥ 12	64	27	86	32	
Study Center					X^2=54.12***
Center 1	29	0	25	0	
Center 2	18	0	20	0	
Center 3	25	19	21	16	
Center 4	20	13	23	14	
Center 5	11	7	14	9	
Set 1 or 2	49/54	17/22	52/51	17/22	X^2=0.85
Trials' order A or B	46/57	18/21	46/57	20/19	X^2=0.58
Head motion	1.39±0.85	1.38±0.94	1.22±0.87	1.29±0.89	F=0.77
Digit span forward	7.59±1.92	7.79±1.82	8.09±2.10	7.97±2.16	F=1.12
Digit span backward	7.26±1.97	7.36±2.07	7.79±2.02	7.74±1.93	F=1.41
TMT-A	27.07±8.86	27.03±10.17	25.43±9.48	24.50±7.78	F=1.10
TMT-B	57.72±21.36	56.85±14.95	56.20±21.42	55.59±18.74	F=0.15
RWT-P	10.15±3.67	10.90±3.82	10.65±3.68	10.13±3.65	F=0.62
RWT-K	12.89±4.07	13.13±4.19	13.26±3.62	12.79±3.02	F=0.23
CGI	4.93±0.83	3.46±1.15			t=7.00***
SIGH-A	18.61±8.80	15.28±8.16			t=1.96*
PAS	22.24±8.63	11.94±7.54			t=6.20***
MI alone	2.41±0.93	1.52±0.55			t=6.72***
BSQ†	48.58±10.35	36.40±11.31			t=5.25***
ACQ†	2.23±0.58	1.74±0.48			t=4.18***
ASI	33.07±12.00	23.06±12.85	9.79±5.98	9.87±7.51	F=97.25***
BDI-II	12.29±7.36	11.17±8.44	2.21±2.61	1.61±2.15	F=57.12***
BSI	54.11±31.60	33.46±27.39	8.26±8.50	9.26±10.51	F=66.85***

M: Mean; SD: Standard deviation; TMT: Trail Making Test; RWT: Regensburg Word Fluency test; ACQ: Agoraphobic Cognition Questionnaire; ASI: Anxiety Sensitivity Index; BDI-II: Beck Depression Inventory-II; BSI: Brief Symptom Inventory; BSQ: Body Sensations Questionnaire; CGI: Clinical Global Impression; MI: Mobility Inventory; PAS: Panic and Agoraphobia-Scale; SIGH-A: Structured Interview Guide for the Hamilton Anxiety Scale; *: $p < 0.05$; ***: $p < 0.001$; ES: Effect size in terms of Cohen's *d;* †: patients from study center 1 and center 2 have not provided data for these questionnaires.

S7. Clinical characteristics of PD subsamples, and response to CBT

In the current study, there are three different subsamples. They have different sample sizes: 1. Patients with PD at baseline (n=118, 55 of them were measured only one time, and they were not entered into the treatment); 2. Patients who entered treatments (n=63, however, 20 of them did not participate in the fMRI posttreatment measure; "Dropout," one patient made too many mistakes in the fMRI task at posttreatment); 3. The subsample of patients who finished their treatment and were measured twice (baseline and posttreatment, n=42).

To illustrate the comparability of the subsamples in their clinical characteristics, we have conducted ANOVA-tests for every clinical measurement at baseline. No significant differences were detected (P>0.21), indicating great similarity of subsamples in our study (for clinical characteristics of all three subsample, see Table S2). This allows us to compare subsample at postmeasurement (n=42) with the whole PD sample (n=118) in their clinical, behavioral, and fMRI effects.

TABLE S2: Baseline clinical characteristics of the whole PD sample and the subsamples

Clinical measurements	Whole PD sample (n=118)	PD treatment subsample (n=63)	PD subsample with postmeasurement (n=42)	*F*
CGI	4.84±0.88	4.84±0.92	4.74±0.73	0.24
SIGH-A	18.27±8.72	20.44±9.14	20.40±9.04	1.58
PAS	21.74±8.74	22.61±9.50	21.76±9.81	0.24
MI alone	2.36±0.94	2.41±0.90	2.38±0.90	0.05
BSQ†	47.59±10.31	47.59±10.31	47.21±10.10	0.08
ACQ†	2.20±0.59	2.20±0.59	2.17±0.59	0.04
ASI	32.32±12.08	34.11±12.56	34.76±13.40	0.83
BDI-II	12.04±7.16	13.38±7.56	14.02±7.71	1.28
BSI	53.72±31.74	58.19±34.13	55.98±31.59	0.39

ACQ: Agoraphobic Cognition Questionnaire; ASI: Anxiety Sensitivity Index; BDI-II: Beck Depression Inventory-II; BSI: Brief Symptom Inventory; BSQ: Body Sensations Questionnaire; CGI: Clinical Global Impression; MI: Mobility Inventory; PAS: Panic and Agoraphobia-Scale; SIGH-A: Structured Interview Guide for the Hamilton Anxiety Scale; †: patients from study center Berlin and Dresden have not provided data for these questionnaires. Therefore, they have the same mean value as the PD subsample entered treatments.

Clinical improvements were calculated for subsamples of patients with PD or PDA in the clinical trials (see Table S3). Responders to CBT were defined as follows: SIGH-A with at least 50% reduction from baseline; CGI score of "mild," "borderline," or "no" disability; no panic attacks in the previous week; MI no more than 1.5; and PAS total 8 or lower. Comparisons of the subgroups with PDA did not reveal significant group differences at baseline and posttreatment. The clinical improvement and response rate in three subsamples of PDA were also not significant.

TABLE S3: Clinical improvement and response after CBT in three subsamples of patients in clinical trial

	Patients with fMRI and clinical baseline and posttreatment[1]				**Patients with baseline fMRI and posttreatment clinical measurements[2]**				**Patients with only baseline fMRI and clinical measurements (dropouts)**				**Tests (Only PDA)**
	PD (n=9)		**PDA (n=31)**		**PD (n=2)**		**PDA (n=5)**		**PD (n=2)**		**PDA (n=14)**		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
CGI													
Baseline	4.11	0.33	4.94	0.73	5.00	0.00	5.60	1.14	3.00	2.83	5.07	0.73	F=1.61
Posttreatment	2.88	1.13	3.55	1.06	3.00	0.00	3.20	1.79					F=0.38
Improvement	1.23	0.89	1.39	1.12	2	0.00	2.4	0.89					F=3.70†
Response rate	75%		42%		100%		80%						X^2=2.50
SIGH-A													
Baseline	15.67	7.16	22.90	8.32	24.50	17.68	20.60	12.03	11.50	16.26	18.57	7.51	F=1.28
Posttreatment	11.88	6.51	15.77	8.13	13.50	3.54	13.40	8.91					F=0.36
Improvement	3.79	3.49	7.13	7.39	11.00	14.14	7.20	6.65					F=0
Response rate	13%		32%		50%		20%						X^2=0.30
PAS													
Baseline	11.33	4.85	24.94	8.75	26.00	8.49	27.60	10.90	16.00	16.97	23.93	6.56	F=0.35
Posttreatment	6.88	5.36	12.06	7.76	7.50	0.71	11.20	8.41					F=0.05
Improvement	4.45	3.60	12.87	7.77	18.50	7.78	16.40	9.94					F=0.83
Response rate	75%		32%		100%		40%						X^2=0.12
MI alone													
Baseline	1.36	0.37	2.71	0.77	1.69	0.55	2.68	1.26	1.76	1.07	2.55	0.76	F=0.19
Posttreatment	1.17	0.15	1.56	0.57	1.11	0.05	1.96	0.84					F=1.80
Improvement	0.19	0.30	1.15	0.75	0.57	0.60	0.72	1.02					F=1.27
Response rate			58%				40%						X^2=0.57

1: Three PDA patients with fMRI posttreatment measure failed to give their clinical measurements after treatment; 2:This subsample included two patients, who dropped out from the treatment, however, participated in the clinical postmeasurement. Improvement: Baseline minus Posttreatment. Responders of CBT were defined as follows: CGI score of "mild," "borderline," or "no" disability; SIGH-A at least 50% reduction from baseline; PAS total 8 or lower; and MI no more than 1.5. CGI: Clinical Global Impression; MI: Mobility Inventory (response rate for PD was not calculated, since most of patients with PD showed a very low score on MI at baseline); PAS: Panic and Agoraphobia-Scale; SIGH-A: Structured Interview Guide for the Hamilton Anxiety Scale. Statistical tests were applied only to groups of patients with PDA, since patients with PD were too few (n=2). †: p=0.063.

S8. Study stimuli, fMRI-design, and a pilot study with PD

The semantic priming paradigm has been described in detail elsewhere (21) and is based on our previous work (23–25). Six conditions were constructed in this paradigm: N–N (neutral-related); T–N (unrelated); N–S (unrelated); T–S (fear-related); and two pseudoword conditions, N–P and T–P (Table S4). The neutral–related word pairs were selected from an existing data set that had been developed and validated previously (23,24). The panic-trigger/symptom word pairs were taken from an earlier study (25). The stimuli (i.e., primes, targets, and word pairs) were matched between conditions according to their lexical frequency (26), word length, syllables, and concreteness (27). Thus, no significant differences between the conditions ($P_{uncorrected}>0.07$) were observed in the final stimulus sets (see Supplementary Material of (21). Pseudowords were pronounceable German words constructed by changing one or two consonants in real words.

TABLE S4: Experimental design of the semantic priming paradigm and examples of stimuli

Prime \ Target	Neutral (N)	Panic Symptom (S)	Pseudoword (P)
Neutral (N)	**N–N** *cup – pot* *television –entertainment*	**N–S** *cup – suffocate* *television – faint*	**N–P** *cup – salkom* *television - holsch*
Panic Trigger (T)	**T–N** *supermarket – pot* *forest – entertainment*	**T–S** *supermarket – suffocate* *forest – faint*	**T–P** *supermarket – tuneu* *forest – faussak*

In our rapid event-related fMRI design, small blocks of stimuli containing one to three trials from the same condition were constructed, with a small intertrial interval (ITI) of 1.5–2.5 s ($M_{ITIsmall}$ = 2 s). Longer ITIs (jittered 3–5 s; $M_{ITIlong}$ = 4 s) were placed randomly between the small blocks so that the overlapping hemodynamic response functions (HRF) could be deconvolved.

The fear trigger–symptom (T–S) word pairs used in this study were evaluated in a pilot study by ten patients with PD and agoraphobia (PDA) and 20 healthy control subjects (HC). Results are already published as supplementary material for the article by Yang et al., 2016 (21). They were

asked to rate the valence (from -3, “highly negative,” to 3, “highly positive”) and relatedness (from 1, “unrelated,” to 7, “highly related”) of the word pairs. For both sets of T–S word pairs, patients with PDA rated T–S word pairs as more unpleasant than healthy subjects (Set 1: M_{PDA}=-1.24, SD=0.46; M_{HC}=-0.58, SD=0.26, p<0.001; Set 2: M_{PDA}=-1.14, SD=0.45; M_{HC}=-0.61, SD=0.21, p<0.001) and reported higher relatedness between the fear-trigger/fear-symptom word pairs (Set 1: M_{PDA}=4.14, SD=1.00; M_{HC}=1.85, SD=0.52, p<0.001; Set 2: M_{PDA}=3.79, SD=1.02; M_{HC}=1.87, SD=0.62, p<0.001). Additionally, both sets of stimuli revealed high equivalence for PDA.

After the fMRI scan, all subjects were invited to rate the valence and relatedness of the word pairs with real words as targets. Out of the 118 patients with PD, 113 completed the ratings after the fMRI experiment at the baseline. The perceived relatedness of panic trigger-symptom word pairs by patients with PD was significantly (r=.33, p<.01) correlated with their agoraphobic avoidance behavior measured with the mobility inventory (Figure S5).

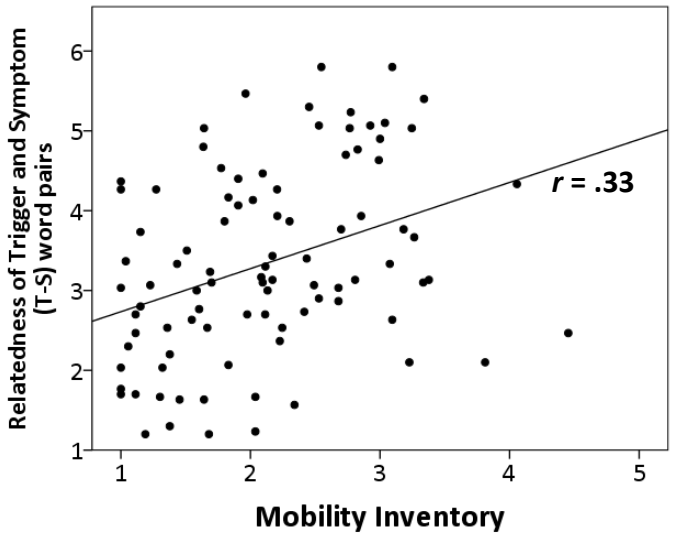


FIGURE S5: Correlation between the score in mobility inventory and the perceived relatedness of panic trigger-symptom word pairs in patients with panic disorder before treatment

S9. Calculation of the fMRI results' significance threshold in cluster level corrected for whole-brain multiple comparisons

The following conditions were considered: 1) Semantic priming has proved to have a very small effect size (d=0.10 to 0.26; for meta-analyses see (28,29); 2) a generally high arousal of patients in the MRI scanner could reduce the difference of neural activation between experimental and control conditions; 3) clinical subgroups and genetic variance in patients could have led to high variability in our results; thus, we have decided to report individual voxel activity significant for $P<0.05$. However, in reporting the fMRI results, the correction for whole-brain multiple comparisons is crucial (30). We decided to report activation clusters extent threshold corrected for multiple comparisons to $P<0.05$, as specified via Monte Carlo simulations (for a detailed description of this approach see 31,32). The cluster extent threshold procedure relies on the fact that given spurious activity or noise (voxel-wise type-I error), the probability of observing increasingly large (spatially contiguous) clusters of activity systematically decreases. To calculate the cluster threshold, we used the Matlab script "cluster_threshold_beta.m" written by Slotnick, S.D. (downloaded from the webpage www2.bc.edu/sd-slotnick/scripts.htm). The parameters used for this calculation are the following: image matrix=64×64×30; slices =30; dim_xy=3.59; dim_z=4.18; FWHM=13.3; dim_resampled=2; p_corrected=0.05; p_voxel=0.05; iterations=1000. The simulation associated with the aforementioned individual voxel and corrected p-values resulted in a cluster extent threshold of 122 contiguous resampled voxels (976 mm^3), an extent threshold that was enforced for all contrasts. However, considering the recent controversial discussion about the eligibility of cluster threshold in fMRI study (33,34), our significant neural effects should also be viewed with caution.

S10. Baseline comparisons of PD with HC in rating and behavioral results

For the investigation of baseline psychopathological alterations in PD, we conducted three separate repeated-measures analyses of variance (ANOVA): with relatedness and valence rating, and response time (RT) as dependent variables; target (panic-symptom and neutral) and prime (panic-trigger and neutral) as the within-subject variables; and the group (PD vs. HC) as the between-subject factor only for baseline data. To correct the p-value for multiple comparisons during our post hoc t-tests of behavioral effects, a Bonferroni correction was adopted. Since we conducted 24 tests [3 variables (relatedness, valence, and RT) × (2 groups × 2 priming effects (neutral and fear) + group difference in 4 conditions)], according to Bonferroni correction for multiple comparisons, p=0.05/24, namely 0.002, was set as the significance threshold for these post hoc tests.

Relatedness at baseline. An overview of behavioral effects is presented in Table 2 and Figure S6. We observed significant main effects for target ($F(1,250)=2087.17$, $p<0.001$) and prime ($F(1,250)=3309.10$, $p<0.001$), predominately driven by the very high relatedness rating of N–N. Furthermore, we obtained significant interactions for group × target ($F(1,250)=64.39$, $p<0.001$); group × prime ($F(1,250)=38.29$, $p<0.001$); and group × target × prime ($F(1,250)=40.47$, $p<0.001$). Post hoc tests indicated higher perceived relatedness for N–N compared to T–N (PD: $t=66.99$, df=112, $p<0.001$; HC: $t=72.39$, df=138, $p<0.001$) and for T–S compared to N–S (PD: $t=17.64$, df=112, $p<0.001$; HC: $t=14.92$, df=138, $p<0.001$) in both groups. Importantly, compared to HC, patients reported higher relatedness for the panic-related T–S condition only ($t=7.13$, df=250, $p<0.001$).

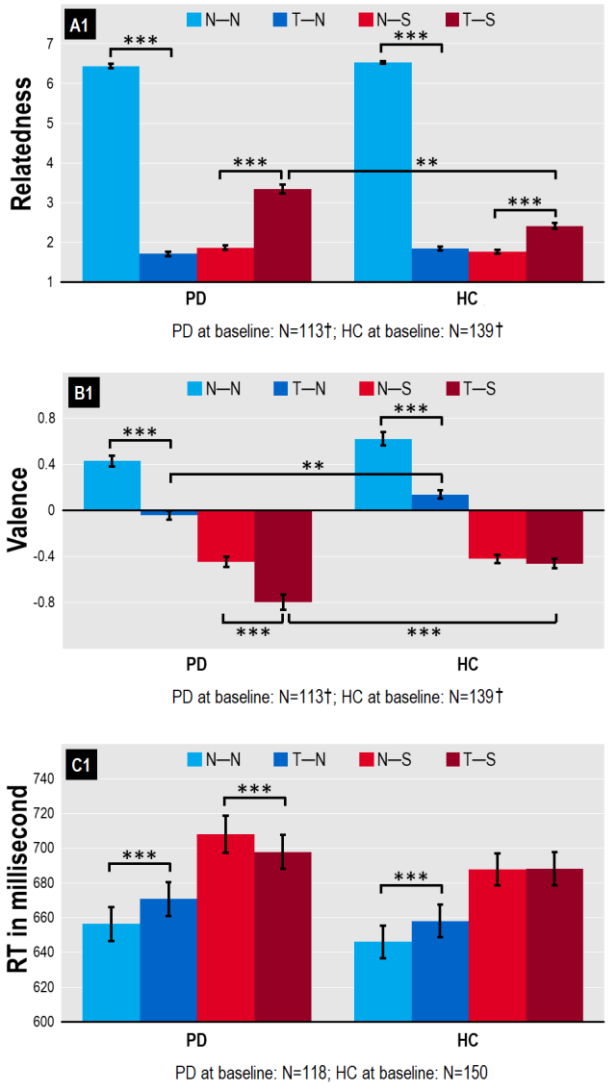


FIGURE S6: *Results of relatedness and valence ratings, as well as reaction time, during the lexical decision task at baseline/T1. From top to bottom of this figure, results for relatedness (A1), valence (B1), and response time (C1) for each condition, including N–N (neutral-neutral; related), T–N (trigger-neutral; unrelated), N–S (neutral-symptom; unrelated), and T–S (trigger-symptom; panic-related), are presented. PD: patients with panic disorder; HC: healthy control subjects; †: smaller n for PD and HC when some of the participants failed to give their rating after the semantic priming task; **: p<0.01; ***: p<0.001.*

Valence at baseline. We found significant main effects for target ($F(1,250)=439.76$, $p<0.001$) and prime ($F(1,250)=290.53$, $p<0.001$), indicating that panic-symptom target words and panic-trigger prime words were more unpleasant than neutral words for both groups. Among the interaction effects, we found significant target × prime ($F(1,250)=51.27$, $p<0.001$); group × prime ($F(1,250)=14.05$, $p<0.001$); and group × target × prime ($F(1,250)=16.78$, $p<0.001$) interactions. Post hoc tests revealed that both patients and HC perceived T–N as more negative than N–N (PD: $t=11.28$, df=112, $p<0.001$; HC: $t=10.87$, df=138, $p<0.001$); however, only PD patients perceived higher negative valence for T–S compared to N–S (PD: $t=7.53$, df=112, $p<0.001$; HC: $t=1.88$, df=138, p=0.06). Group comparisons revealed that T–N and T–S conditions

were rated as more negative by patients, in contrast to HC (T–N: t=3.26, df=250, p=0.001; T–S: t=4.47, df=250, p<0.001).

Errors at baseline. During the lexical decision task, both the patients and HC made a few errors (see Table 2), and there were no group differences in terms of accuracy.

RT at baseline. A repeated-measures ANOVA of RTs yielded significant main effects for target ($F(1,266)=492.04$, $p<0.001$) and prime ($F(1,266)=4.54$, $p<0.05$). The significant target × prime interaction ($F(1,266)=44.16$, $p<0.001$) suggested priming effects in both groups. The significant group × target × prime interaction ($F(1,266)=5.37$, $p<0.05$) indicated group differences in semantic priming effects. Post hoc tests revealed a reduction in response time when neutral targets were preceded with related neutral primes (N–N<T–N) in both groups (PD: t=3.90, df=117, p<0.001; HC: t=4.58, df=149, p<0.001). Most importantly, the preceding panic-triggers led to shorter RT for panic-symptom targets only in PD (t=3.92, df=117, p<0.001), but not in HC (t=0.41, df=149, p=0.68). Post hoc tests revealed no group differences in RTs for any conditions.

S11. Rating and behavioral results in the subsamples

S11.1 fMRI subsample at baseline (103 pairs of matched PD and HC)

The fMRI subsample affirmed most of the significant effects in the greater sample for behavioral results. Only the interaction of group × prime ($F(1,198)=3.05$, $p=0.08$) and group × target × prime ($F(1,198)=2.39$, $p=0.12$) for RT did not survive the significance tests. The insignificant effects in the sample for behavioral results remained insignificant in the fMRI subsample. All the insignificant post hoc tests in the sample for behavioral results also remained insignificant in the fMRI subsample.

TABLE S5: Relatedness rating, valence rating, response time and percentage of error response in the lexical decision task for fMRI-subsample (n=103) at baseline

Group	Condition	Relatedness	SD	Valence	SD	RT (in ms)	SD (in ms)	Priming	Error (in %)	SD (in %)
PD	N–N	6.41	0.59	0.41	0.50	659.4	105.9	13.1***	1.6	3.2
	T–N	1.68	0.56	-0.06	0.42	672.5	106.9		2.1	3.6
	N–S	1.84	0.61	-0.44	0.49	711.6	118.9	11.1***	3.1	3.6
	T–S	3.35	1.17	-0.79	0.71	700.5	107.2		3.1	5.6
HC	N–N	6.51	0.42	0.58	0.70	642.1	122.4	13.7***	1.6	2.9
	T–N	1.83	0.56	0.09	0.51	655.8	122.6		1.9	2.8
	N–S	1.74	0.63	-0.42	0.46	685.4	121.4	-0.2	3.1	3.8
	T–S	2.44	0.95	-0.45	0.52	685.6	124.6		3.3	3.6

For relatedness and valence, the sample sizes of PD and HC are 96 and 92, respectively. For RT and percentage of error response, both PD and HC have a sample size of 103. Priming refers to semantic priming effects: (T–N > N–N) or (N–S > T–S) in RT; RT (in ms): reaction time in millisecond; SD: standard deviation; PD: patients with panic disorder; HC: healthy control subjects; ***: $p < 0.001$

S11.2 Treatment effects on rating and response time in fMRI subsample (103 pairs of matched PD and HC at baseline vs. 39 pairs at posttreatment)

Comparisons of T–S vs. N–S for the relatedness, valence rating, and RT in the fMRI subsample at each measuring point (baseline and posttreatment/T2) are presented in Table S6. Again, the fMRI subsample affirmed all significant effects in the greater sample for behavioral results

(results of the whole sample see the main text). The insignificant effects in the sample for behavioral results remained insignificant in the fMRI subsample. All the insignificant post hoc tests in the sample for behavioral results also remained insignificant in the fMRI subsample.

TABLE S6: Comparisons of T–S vs. N–S in the whole sample and fMRI-subsample at baseline and posttreatment

	Relatedness(T–S>N–S)		Valence(T–S<N–S)		RT(T–S<N–S)	
	Baseline/T1	Posttreatment/T2	Baseline/T1	Posttreatment/T2	Baseline/T1	Posttreatment/T2
Whole PD sample						
PD	1.47	0.94	0.35	0.17	12.54	4.12
HC	0.66	0.69	0.04	0.05	1.04	6.82
fMRI-subsample						
PD	1.49	0.92	0.36	0.17	13.12	-0.18
HC	0.68	0.74	0.05	0.06	2.37	8.94

PD: patients with panic disorder; HC: healthy control subjects; RT (in ms): reaction time in millisecond

S12. Neural correlates of panic priming effects in PD at baseline and posttreatment and their differences compared with healthy controls

In this section, we report the neural correlates of panic priming effects in PD at baseline and posttreatment and their differences compared to HC as additional supporting evidence. An overview is presented in Figure S7. The significant clusters of panic priming in the patient group are reported in S11.1 and Table S7. Group comparisons (PD vs. HC) of panic priming are presented in S12.2 and Table S8.

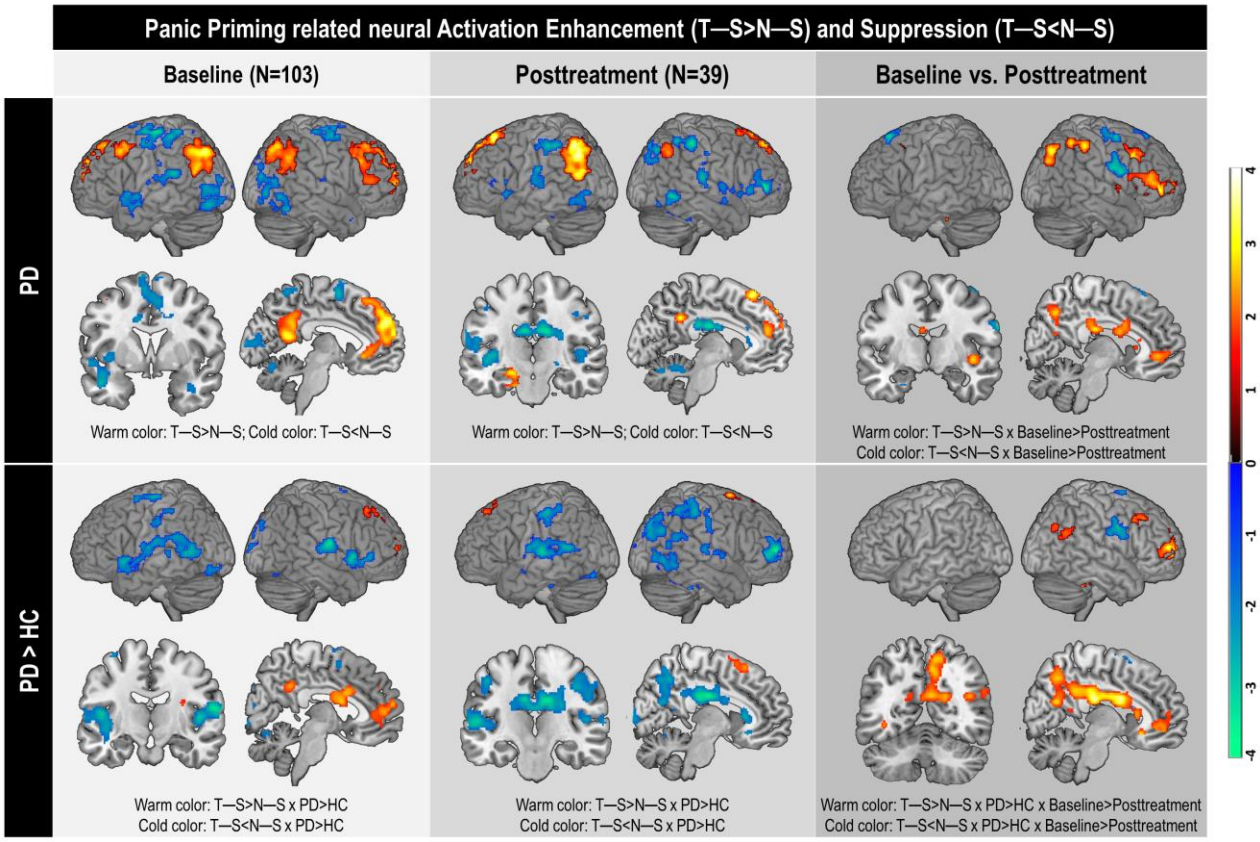


FIGURE S7: Neural correlates of panic priming in patients with PD (n=103) and their group differences compared with healthy control subjects (n=103) at both baseline and posttreatment. Panic-priming effects (T–S>N–S or T–S<N–S) are shown separately in PD alone or PD vs. HC for baseline and posttreatment measurements. The right row shows baseline vs. posttreatment to illustrate the CBT effects. The warm color shows the panic-priming–related activation enhancement (T–S>N–S); the cold color shows the panic-priming–related activation suppression (T–S<N–S). The colored bars display the level of t-values. PD: patients with panic disorder; HC: healthy control subjects; N–S (neutral-symptom word pairs); T–S (trigger-symptom; panic-related).

S11.1 Neural correlates of panic priming effects in PD at baseline and posttreatment

PD showed panic priming-induced neural activation suppression (T–S<N–S) and enhancement (T–S>N–S) in overlapping brain regions at baseline and posttreatment. These activations, as well as the group × time interaction, are presented in Table S7. They are listed in rows next to each other for easier comparison. Since the biggest cluster in each contrast is huge and spreads into multiple cortices. The largest cluster has 5284 voxels, extending into multiple anatomical brain areas. Therefore, we used the Automated Anatomical Labeling (AAL) toolbox of SPM to name the brain regions with active voxels. The peak voxel, as well as the number of active voxels in a given brain region, are reported. The sub-clusters/brain regions with less than 60 voxels were not reported, except the amygdala.

Panic-priming effect in PD. (1) Activation enhancement (T–S>N–S): For PD patients at both baseline and posttreatment, we found activation within the left angular gyrus/inferior parietal lobule (IPL), dorsal medial prefrontal cortex (dmPFC)/superior frontal gyrus (SFG) and middle cingulate cortex (MCC). In addition, patients at baseline showed greater activation in the middle frontal gyrus (MFG), ACC, posterior cingulate cortex (PCC)/precuneus, and right IPL. (2) Priming-related activation suppression (T–S<N–S): Patients showed suppression at both time points in the bilateral superior temporal gyri (STG), visual cortices, and insula. However, PD at posttreatment demonstrated additional suppression in the MCC, right STG, MFG/IFG, insula, and amygdala.

Treatment effect on panic-priming in PD. Following CBT, activation for T–S>N–S significantly decreased in the ACC, PCC/precuneus, right MFG, and IFG. At posttreatment compared to baseline, patients showed greater T–S<N–S (activation suppression) in the MCC, right MFG, and right insula/hippocampus. The reduction of priming-related activation suppression by CBT was limited to small clusters in the sensorimotor cortex.

TABLE S7: Neural correlates of panic priming effects in patients with panic disorder (PD)

		PD at baseline			PD at posttreatment			Baseline>Posttreatment		
Anatomical region	**BA**	**MNI coordinates**	**t-value**	**no. Voxels**	**MNI coordinates**	**t-value**	**no. Voxels**	**MNI coordinates**	**t-value**	**no. Voxels**
Panic priming-related activation enhancement (T–S > N–S)										
mPFC	8,9,32	-8,52,28	3.64	1731	-12,28,58	3.70	504	8,46,-2	2.59	130
Precuneus	7,30	-4,-54,16	3.58	939				8,-62,42	3.01	246
Left angular gyrus	7,40	-44,-72,44	3.53	778	-52,-62,38	3.84	768			
Left SFG	8,10	-16,60,14	3.51	719	-18,28,56	4.01	492			
Left IPL	7,40	-36,-74,44	3.50	264	-46,-58,50	3.73	308			
MCC	24,31	8,-50,34	3.02	254	-8,-48,34	3.02	153	12,-22,24	3.59	71
Right SFG	10	22,58,10	2.97	688				24,58,4	3.16	122
Right angular gyrus	7	36,-68,50	2.93	710	44,-60,40	2.46	134	32,-70,46	3.02	126
PCC	23	-4,-48,24	2.90	299						
Left MOG	7,19	-38,-74,40	2.87	206				32,-72,44	3.27	189
Left MFG	8	-28,20,48	2.86	671						
ACC	32	0,40,2	2.80	548				-2,24,-8	3.35	310
Right MFG	8,9	40,22,34	2.68	679				34,18,40	3.21	325
Right IPL	40	48,-54,40	2.42	184						
Right IFG	46	50,34,12	2.24	144				42,36,12	2.84	336
Left MTG	39				-58,-64,18	3.37	209			
Left hippocampus					-20,-20,-12	3.20	339			
Right MFG	46							38,44,2	3.49	283
Right IPL/supramarginal gyrus	40							48,-36,52	3.08	437
Right insula								42,-8,-4	2.78	143
Right putamen/insula								26,10,-14	2.72	191
Right hippocampus								12,-30,-10	2.51	144
Cuneus	31							16,-70,24	2.39	218
Panic priming-related activation suppression (T–S < N–S)										
Left precentral gyrus	6	-40,-18,66	3.58	178						
Left STG	13,22	-42,-6,-14	3.52	307	-42,-14,0	2.96	87			
Right ITG	37	46,-54,-22	3.36	216	50,-56,-6	3.17	210			
Left fusiform gyrus	19,37	-22,-70,-16	3.21	177	-38,-44,-18	2.88	211			
Right fusiform gyrus	37	46,-54,-20	3.21	118						
Right STG/insula	13	40,-20,-6	3.14	237	44,-8,-6	3.58	186			
Left STG	22	-56,-40,20	2.94	212	-56,-26,2	2.34	112			
Left postcentral gyrus/ SPL	5	-26,-44,68	2.93	340						
Right precentral gyrus/ SFG	6	40,-4,64	2.92	159						
Left insula		-38,8,0	2.89	240	-24,38,-12	3.34	493			
Left cerebellum		-18,-68,-18	2.89	256	-26,-64,-26	3.65	735			
SMA	6	-6,0,62	2.86	451				-12,30,62	3.29	193
Left temporal pole	21	-44,4,-22	2.71	213						
Right MOG	19	38,-86,18	2.70	306						
Left IOG	18	-44,-78,-14	2.70	245						
Calcarine gyrus	17,31	4,-96,-4	2.59	283	15,-72,26	3.30	607			
Right SOG	7,18	20,-84,18	2.57	292	32,-72,44	3.09	358			
Left postcentral gyrus	1,2	-44,-24,34	2.56	72	-66,-18,24	2.92	117			
Left MOG	19	-26,-84,4	2.51	131						
Right fusiform gyrus	19	34,-66,-14	2.39	95	35,-70,-14	2.66	117			
Precuneus	7,31	-6,-54,62	2.36	152	22,-62,22	3.13	575			
Right insula					38,44,4	4.24	690			
Right hippocampus					20,-10,12	3.73	109			
Right IPL/postcentral gyrus	40				50,-34,52	3.54	293			

Right cerebellum					22,-46,-26	3.21	416			
MCC	23				6,-26,30	3.15	175			
Left ITG	37				-40,-44,-18	3.01	293			
Left IPL	2				-45,-29,48	2.91	321			
		PD at baseline			**PD at posttreatment**			**Baseline>Posttreatment**		
Anatomical region	**BA**	**MNI coordinates**	**t-value**	**no. Voxels**	**MNI coordinates**	**t-value**	**no. Voxels**	**MNI coordinates**	**t-value**	**no. Voxels**
TABLE S7 continued										
Right STG/supramarginal gyrus	40				52,-22,14	2.87	312			
Left postcentral gyrus	2				-48,-30,48	2.84	252			
Right MTG	21				62,-52,-8	2.83	64			
Right IFG	10				40,42,0	2.76	225			
Right parahippocampal	19				32,-34,-10	2.68	98			
Right amygdala					22,-6,-12	2.61	16			
Left putamen					-22,16,-6	2.60	133			
Left IFG	13				-34,18,10	2.46	342			
Right putamen					26,14,-6	2.40	259			
Right postcentral gyrus								64,4,24	3.17	134
Right precentral gyrus								42,-2,62	3.13	140
Hippocampus								-20,-22,-14	2.50	132

Significance level: uncorrected $p < 0.05$, cluster with at least 122 voxels; Brain regions with active voxels were identified by the Automated Anatomical Labeling (AAL) toolbox of SPM. Those brain regions were not presented, where less than 60 voxels were activated (except for amygdala). PD: patients with panic disorder; HC: healthy control subjects; L: Left; R: Right; ACC; Anterior cingulate cortex; IFG: Inferior frontal gyrus; IPL: Inferior parietal lobule; ITG: Inferior temporal gyrus; MCC: Middle cingulate cortex; MFG: Middle frontal gyrus; mPFC: Medial prefrontal cortex; MTG: Middle temporal gyrus; PCC; Posterior cingulate cortex; SFG: Superior frontal gyrus; SMA: Supplementary motor cortex; SPL: Superior parietal lobule; STG: Superior temporal gyrus.

S12.2 Neural correlates of group differences and CBT effects in panic priming

We found group differences (PD>HC) in panic priming-induced neural activation suppression (T–S<N–S) and enhancement (T–S>N–S) in overlapping brain regions at baseline/T1 and posttreatment/T2. These activations are presented in Table S8.

Group differences in panic-priming effect at baseline are reported in the main text (See results). At posttreatment, the group difference of priming-induced activation enhancement was limited only in a small cluster of the supplementary motor area. Accordingly, PD showed stronger priming-related activation suppression (T–S<N–S) in bilateral tempo-parietal cortices as well as in the right prefrontal cortex, MCC, and precuneus compared to HC.

Group × time interaction in panic-priming effect. Following CBT, activation for T–S>N–S significantly decreased in the ACC, MCC, PCC/precuneus, and right SFG/MFG in PD vs. HC. For priming-related activation suppression (T–S<N–S), changes were limited to small clusters in the sensorimotor cortex.

TABLE S8: Neural correlates of panic priming effects in patients with panic disorder (PD) compared to healthy control

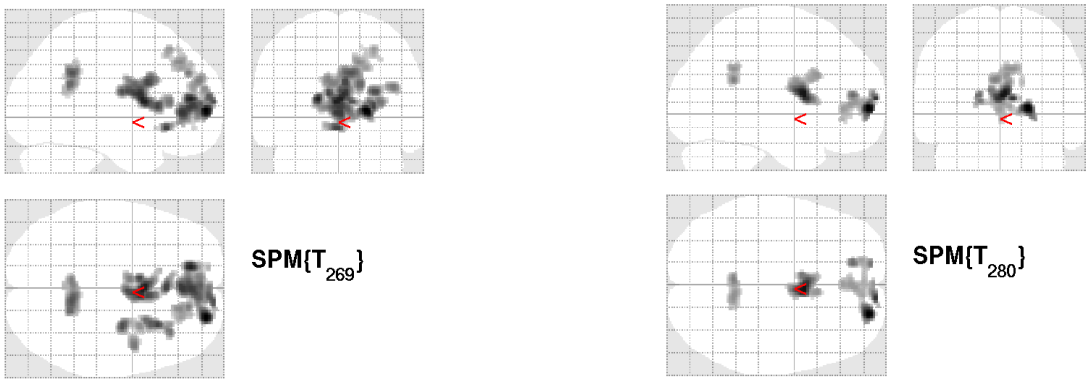
		PD>HC at baseline			PD>HC at posttreatment			PD>HC × Baseline>Posttreatment		
Anatomical region	**BA**	**MNI coordinates**	**t-value**	**no. Voxels**	**MNI coordinates**	**t-value**	**no. Voxels**	**MNI coordinates**	**t-value**	**no. Voxels**
Panic priming-related activation enhancement (T–S > N–S)										
Right SFG	10	22,58,6	3.04	131				24,58,4	3.82	343
ACC	32	0,40,4	2.67	459				-6,26,-8	3.50	473
Right Putamen/insula		28,20,8	2.58	78				26,10,10	2.93	102
ACC/MCC	24	6,14,30	2.57	101						
mPFC	10,32	-14,46,14	2.50	554				2,42,-4	2.57	164
MCC	23,31	10,-48,34	2.49	130				10,-24,30	3.50	530
Precuneus	7,23	16,-50,24	2.35	93				8,-62,44	2.86	1017
Right SFG/mPFC	9	14,44,34	2.26	304						
PCC	31	-6,-48,26	2.23	183				8,-42,30	2.70	171
SMA					14,6,70	2,87	294			
Right MFG	10							26,56,2	3.37	225
Right Cerebellum								32,-40,-32	3.04	172
cuneus	23							18,-58,20	2.66	277
Right MFG	8							36,22,42	2.54	220
Left Cerebellum								-20,-40,-28	2.42	131
Right angular gyrus	39							40,-60,22	2.38	143
Panic priming-related activation suppression (T–S < N–S)										
Right rolandic operculum		60,-12,16	3.57	287	48,-22,12	2.63	74			
Right IFG	47	56,14,-2	3.31	139						
Left STG/supramrginal gyrus	13,22	-48,-8,-8	3.24	782	-42,-12,0	3.32	896			
Right Insula		46,-10,6	3.18	498	46,-10,0	2.26	85			
Left rolandic operculum		-46,-10,8	3.17	199						
Left Insula		-44,-8,-4	3.10	469	-40,-12,2	3.55	195			
Right STG	22,42	62,-14,10	3.02	337	46,-8,-8	2.39	361			
Left MTG	22	-62,-58,6	2.94	426	-56,-30,4	2.90	260			
Right Cerebellum		13,-80,-18	2.83	171	32,-40,-32	3.32	416			
Calcarine/cuneus	17,18	6,-96,-4	2.75	649	6,-96,6	2.69	210			
Left postcentral gyrus	2,40	-44,-32,48	2.70	238	-44,-24,48	3.07	276			
Left postcentral gyrus	3,40	-44,-32,48	2.70	99	-64,-14,16	2.56	180			
Left precentral gyrus	4	-38,-20,66	2.64	97						
Left IPL	2,40	-32,-40,48	2.52	116	-46,-26,48	2.99	193			
Left fusiform/lingual gyrus	19	-22,-80,-18	2.44	155						
SMA	6	4,4,59	2.23	163				12,6,72	2,48	168
Left temporal pole	38	-48,16,-10	2.17	90						
Right SFG	10				34,58,10	3.82	217			
Right MFG	46				36,54,10	3.52	455			
MCC	23				6,-26,30	3.46	506			
ACC	32				-6,24,-8	3.15	130			
Cuneus/Calcarine	31				24,-64,22	3.05	586			
Precuneus	31				22,-64,22	3.00	765			
Left Cerebellum					-20,-40,-28	2.90	654			
SOG/MOG	31				26,-64,22	2.83	658			
Right IPL	40				50,-34,52	2.79	321			
Right postcentral gyrus	2				50,-32,52	2.71	509			
Left putamen/insula					-32,10,-10	2.65	180			
Right insula					30,14,-18	2.64	165			
Right MTG	21				43,-33,-2	2.47	141			
Right ITG	37				60,-56,-6	2.36	123			
Right angular gyrus	40				54,-54,24	2.30	131			
Right precentral/ postcentral gyrus	6							62,0,34	3.06	135

Significance level: uncorrected $p < 0.05$, cluster with at least 122 voxels; Brain regions with active voxels were identified by the Automated Anatomical Labeling (AAL) toolbox of SPM. Brain regions were not presented, where less than 60 voxels were activated (except for amygdala). PD: patients with panic disorder; HC: healthy control subjects; ACC; Anterior cingulate cortex; IFG: Inferior frontal gyrus; IPL: Inferior parietal lobule; ITG: Inferior temporal gyrus; MCC: Middle cingulate cortex; MFG: Middle frontal gyrus; mPFC: Medial prefrontal cortex; MTG: Middle temporal gyrus; PCC; Posterior cingulate cortex; SFG: Superior frontal gyrus; SMA: Supplementary motor cortex; SPL: Superior parietal lobule; STG: Superior temporal gyrus.

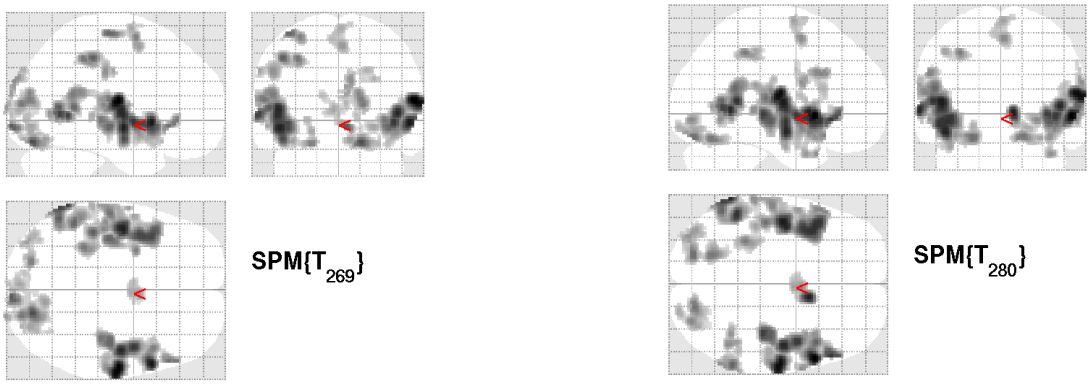
S13. Comparison of the fMRI results in analyses with and without covariates

Analysis with covariates **Analysis without covariates**

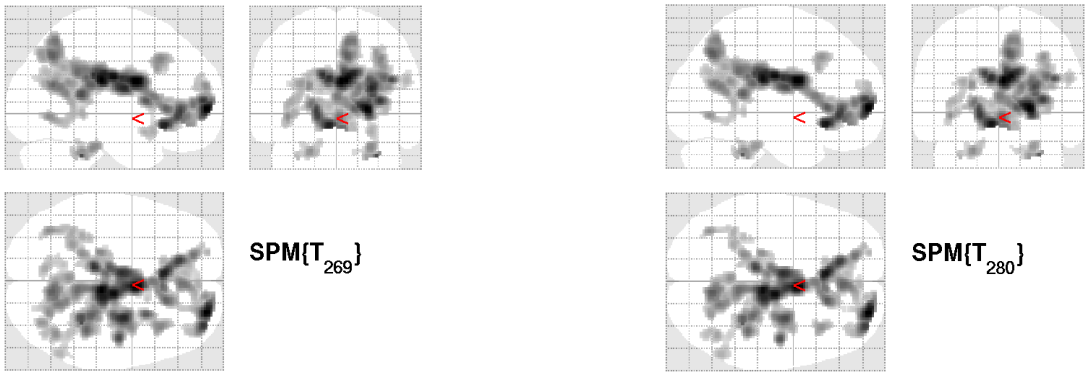
Baseline/T1: T–S>N–S × PD>HC



Baseline/T1: T–S<N–S × PD>HC



T–S>N–S × PD>HC × Baseline/T1>Posttreatment/T2



As the results show, the analysis without covariates does reveal slightly weaker effects in the three major contrasts in our study. However, the effect pattern is quite similar, indicating a very small probability of significant biasing effects by including covariates. Therefore, we chose to report the results from the analysis with covariates in our main text.

S14. Correlations between behavioral/neural changes and clinical improvements

To explore the correlations between the CBT induced clinical improvements and behavioral/neural changes, we calculated the clinical improvements measured by disorder relevant questionnaires (baseline>posttreatment for PAS, measurement of panic attack and agoraphobia; for MI, measurement of avoidance behavior and for SIGH-A, measurement of general anxiety symptoms), the behavioral baseline>posttreatment in Relatedness (T–S>N–S), Valence (T–S<N–S), RT (T–S<N–S) and baseline>posttreatment in neural activation ACC (T–S>N–S), and PCC (T–S>N–S). As proof of specificity, we also explored the correlations between the clinical improvements and the changes in panic priming-related neural suppression in bilateral STG [baseline>posttreatment in left STG (T–S<N–S) and right STG (T–S<N–S)]. Correlation coefficients are shown in Table S9. Clinical improvements were correlated positively with the reduction of relatedness rating and neural activation for T–S>N–S in PCC and in ACC from baseline to posttreatment. However, the change of panic priming-related neural suppression (baseline>posttreatment × T–S<N–S) in the left and right temporal cortices was not associated with clinical improvements.

TABLE S9: Correlations between CBT induced behavioral and neural changes and clinical improvements in patients with panic disorder

Clinical improvement (baseline>posttreatment) in:	Reduction of comparison (T-S vs. N-S) from baseline to posttreatment						
				Neural enhancement[1] (T–S>N–S) in:		Neural suppression[1] (T–S<N–S) in:	
	Relatedness T–S>N–S	Valence T–S<N–S	RT T–S<N–S	ACC‡	PCC	left STG	right STG
PAS	**0.39***	0.21	-0.28	0.21	**0.34***	0.11	0.07
MI	**0.32***	0.13	-0.26	**0.46***	**0.28†**	0.17	0.09
SIGH-A	0.06	0.21	-0.21	0.05	-0.10	0.04	0.06

[1]: parameter estimates of T-S>N-S or T-S<N-S in brain regions were extracted using the VOI function of SPM8 for each patient; RT: reaction time in millisecond; MI: Mobility Inventory; PAS: Panic and Agoraphobia-Scale; SIGH-A: Structured Interview Guide for the Hamilton Anxiety Scale; ACC‡: anterior cingulate cortex (MNI coordinator of the peak voxel: -6/26/-8, cluster size = 537 voxels) with exclusion of one outlier; PCC: posterior cingulate cortex (MNI coordinator of the peak voxel: 8/-44/32, cluster size = 272 voxels); STG: superior temporal gyrus (MNI coordinator of the peak voxel: left: -46/-8/-6, cluster size = 2266 voxels; right: 60/-12/16, cluster size = 1600 voxels); *: $p < 0.05$; †: $p < 0.10$.

S15. Summary of results and their implications

The baseline alterations were found at the behavioral and neural levels. All three behavioral hypotheses (increased negative valence, enhanced relatedness, and priming-related facilitation) were verified with significant comparisons between the patients with PD and HC. On the neural level, PD (compared to HC) showed enhanced activation in ACC and suppressed activation in the bilateral superior temporal gyrus/insula for processing of symptom words primed with trigger words (vs. neutral words). The CBT-related reductions of alterations are signified by the ↓ in Figure S8. First, we found reduced clinical symptoms in patients with PD after CBT. On the behavioral level, the rating of negative valence, as well as the rating of relatedness, was reduced after CBT. On the neural level, the only reduction of ACC activation was found following CBT. Our results suggest a reduction of negative valence (signified by the red color) and the emotional association of the panic-related concepts (signified by the red part of the double arrow) since the activation suppression in semantic-related brain regions was not normalized after CBT. The significant correlation between clinical improvement and neural activation change suggests a probable mediational role of ACC functioning as a mechanism of action of CBT.

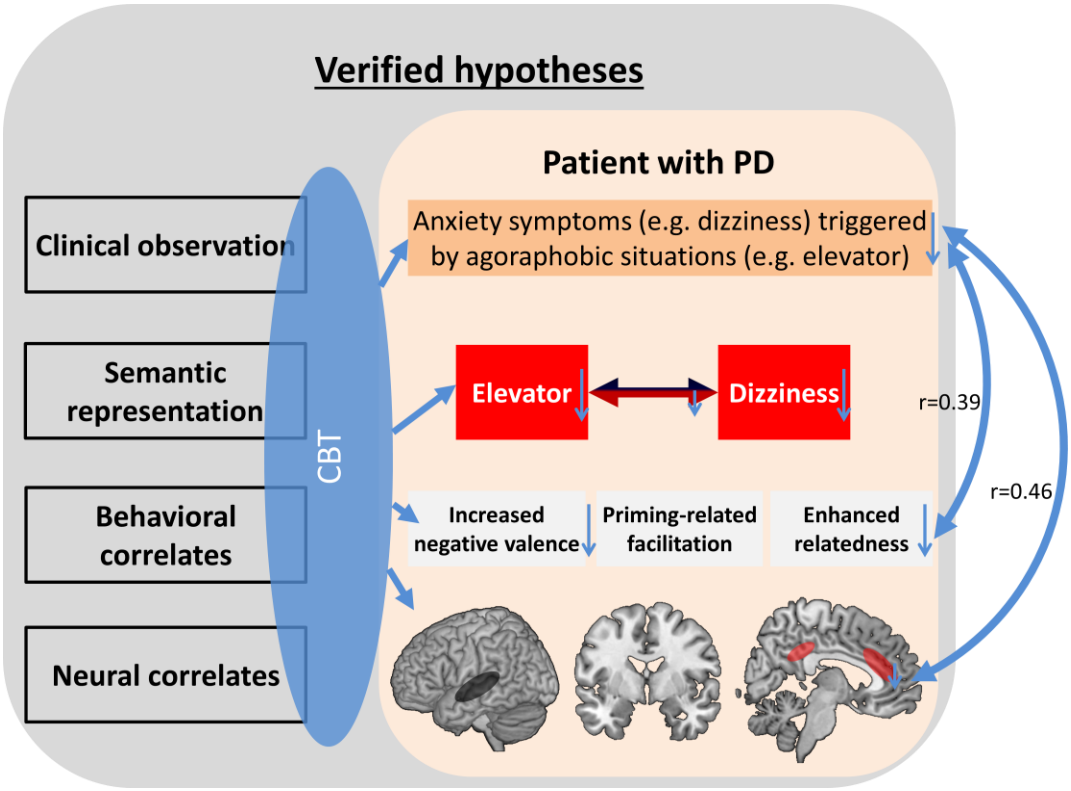


FIGURE S8: Summary of results on the schematic presentation of the biased semantic network in PD. ↓ symbolized the reduction of each characteristic. Double arrows illustrated the correlations between the reduction in panic and agoraphobic symptoms and the reduction in rating and neural activation.

S16. Comparisons of PDA with PD in terms of behavioral and neural effects

To maximize the sample size in our analyses, we included not only the patients with panic disorder and agoraphobia but also patients with panic disorder only (without agoraphobia, about 20% of the whole patient sample: 21 of 118 at baseline and 9 of 42 at posttreatment). We analyzed the clinical, behavioral, and fMRI data with Generalized Estimating Equations (GEE) analyses (using IBM SPSS statistics) because it allows dropouts.

Clinically, we found no group × time interaction for general anxiety symptoms (CGI: X^2=0.57, p=0.45, SIGH-A: F=2.74, p=0.10, and BDI: X^2=1.43, p=0.23) and panic related symptoms (BSQ: X^2=0.02, p=0.89, ACQ: X^2=2.68, p=0.10, and ASI: X^2=0.36, p=0.55). However, we found significant group × time interaction for agoraphobic symptoms (PAS: PDA at baseline=25.87±8.99, PDA at posttreatment=12.04±8.05, PD at baseline=14.00±8.25, PD at posttreatment=7.00±4.74; X^2=16.30, p<0.001 and MI: PDA at baseline=2.75±0.85, PDA at posttreatment=1.58±0.63, PD at baseline=1.46±0.39, PD at posttreatment=1.15±0.14, X^2=37.79, p<0.001) indicating greater reduction in symptoms in PDA group. These results are in line with our hypothesis that the additional exposure in agoraphobic situations of the 12 (compared to the 6) session CBT specifically reduces the agoraphobic symptoms in PDA.

Since the experimental materials contain agoraphobic situations as prime words, patients with PDA and PD could exhibit differences in their responses to panic priming effects. We used post hoc comparisons between PDA and PD to estimate their differences in behavioral and neural effects comparing T–S vs. N–S. First, we calculated the corresponding effects (T–S vs. N–S) of relatedness and valence rating, response time, and neural activation. For neural effects, we focused on the brain regions revealing baseline group (PD vs. HC) differences and group × time effects, including the ACC, PCC, and bilateral STG. We extracted parameter estimates (eigenvariates, extracted using the VOI function of SPM8) of clusters in ACC and PCC for contrast T–S>N–S or of clusters centered in bilateral STG for contrast T–S<N–S for each patient. The descriptive results are shown in Table S10. As an overview, positive therapeutic effects were found in both patients with PD and PDA. However, the effects in PD were generally weaker than in PDA. We used GEE with these behavioral and neural effects as dependent

variables and diagnoses (PD vs. PDA) and time (baseline vs. posttreatment) as independent variables. The diagnoses and diagnoses × time effects are of interest. Test results are presented in Table S11.

Significant effects of diagnoses were only found in the neural enhancement in PCC in processing T–S>N–S (X^2=5.13, p=0.02; PD showed generally higher activation for T–S (vs. N–S) compared to PDA). Additionally, there is a meaningful difference between PD and PDA in relatedness rating of T–S>N–S (PDA at baseline=1.55; PD at baseline=1.00; PDA at posttreatment=0.96; PD at posttreatment=0.83; X^2=3.93, p=.05), indicating less relatedness perceived by patients with PD only. For Valence T–S<N–S, patients with PDA showed general higher negative valence rating than PD (PDA at baseline=0.39; PD at baseline=0.16; PDA at posttreatment=0.18; PD at posttreatment=0.14, X^2=6.18, p=.01) and patients with PDA compared to PD showed greater reduction after CBT (diagnoses × time: X^2=3.71, p=.05). All diagnoses × time interactions were non-significant (X^2 values<2.78, p>0.10). Our results suggest a relatively high similarity in PD and PDA regarding their behavioral effects and neural activation patterns and support our fusion of the two groups in the analyses.

TABLE S10: Behavioral and neural effects of T–S vs. N–S in PD and PDA at baseline and posttreatment

Groups	Relatedness T–S>N–S	Valence T–S<N–S	RT T–S<N–S	Neural enhancement[1] (T–S>N–S) in:		Neural suppression[1] (T–S<N–S) in:	
				ACC	PCC	left STG	right STG
PD Baseline	1.00	0.16	7.11	0.31	0.54	0.19	0.26
PD Posttreatment	0.83	0.14	-7.34	0.12	0.49	0.12	0.09
Baseline>Posttreatment	0.18	0.02	14.45	0.19	0.05	0.07	0.17
PDA Baseline	1.55	0.39	13.59	0.15	0.17	0.14	0.12
PDA Posttreatment	0.96	0.18	3.85	-0.18	-0.06	0.21	0.24
Baseline>Posttreatment	0.59	0.21	9.74	0.33	0.23	-0.07	-0.12

[1]: parameter estimates of T-S>N-S in brain regions (clusters) were extracted using the VOI function of SPM8 for each patient; RT: reaction time in milliseconds; ACC: anterior cingulate cortex (MNI coordinator of the peak voxel: -6/26/-8, cluster size = 537 voxels); PCC: posterior cingulate cortex (MNI coordinator of the peak voxel: 8/-44/32, cluster size = 272 voxels); STG: superior temporal gyrus (MNI coordinator of the peak voxel: left: -46/-8/-6, cluster size = 2266 voxels; right: 60/-12/16, cluster size = 1600 voxels); PD: panic disorder; PDA: panic disorder with agoraphobia.

TABLE S11: Wald X^2-values of ANOVAs comparing PD and PDA in behavioral and neural effects of T–S vs. N–S

Depended variables	Relatedness T–S>N–S	Valence T–S<N–S	RT T–S<N–S	Neural enhancement[1] (T–S>N–S) in:		Neural suppression[1] (T–S<N–S) in:	
				ACC	PCC	left STG	right STG
Diagnoses	3.93*	6.18*	2.20	1.26	5.13*	0.02	0.01
Diagnoses × Time	2.78	3.71*	1.35	0.12	0.35	0.26	0.77

[1]: parameter estimates of T-S>N-S in brain regions (clusters) were extracted using the VOI function of SPM8 for each patient; RT: reaction time in milliseconds; ACC: anterior cingulate cortex (MNI coordinator of the peak voxel: -6/26/-8, cluster size = 537 voxels); PCC: posterior cingulate cortex (MNI coordinator of the peak voxel: 8/-44/32, cluster size = 272 voxels); STG: superior temporal gyrus (MNI coordinator of the peak voxel: left: -46/-8/-6, cluster size = 2266 voxels; right: 60/-12/16, cluster size = 1600 voxels); PD: panic disorder; PDA: panic disorder with agoraphobia; diagnoses: PD vs. PDA; time: baseline vs. posttreatment; *: p≤0.05.

References

1. Lang PJ. A Bio-Informational Theory of Emotional Imagery. Vol. 16, Psychophysiology. 1979. p. 495–512.

2. Wittchen H-UU. Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res [Internet]. 1994;28(1):57–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8064641

3. Katherine Shear M, Bilt J Vander, Rucci P, Endicott J, Lydiard B, Otto MW, et al. Reliability and validity of a Structured Interview Guide for the Hamilton Anxiety Rating Scale (SIGH-A). Depress Anxiety. 2001;13(4):166–78.

4. Guy W. Clinical global impression scale. ECDEU Assess Man Psychopharmacol Vol DHEW Publ No ADM 76. 1976;

5. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of far in agoraphobics: The Bodily Sensations Questionnaire and the Agoraphobic Cognitions Questionnaire. Journal of Consulting and. Clin Psychol. 1984;32(6):1090−1097.

6. Chambless DL, Caputo GC, Jasin SE, Gracely EJ, Williams C. The Mobility Inventory for Agoraphobia. Behav Res Ther. 1985;

7. Bandelow B. Panic and Agoraphobia Scale (PAS). Seattle, Toronto, Bern, Göttingen: Hogrefe & Huber Publishers; 1999.

8. Reiss S, Peterson RA, Gursky DM, McNally RJ. Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. Behav Res Ther. 1986;

9. Beck AT, Steer RA, Brown GK. Manual for the Beck depression inventory-II. San Antonio, TX Psychol Corp. 1996;

10. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. Psychol Med [Internet]. 1983 Aug 9;13(03):595. Available from: http://www.journals.cambridge.org/abstract_S0033291700048017

11. Gloster AT, Wittchen HU, Einsle F, Höfler M, Lang T, Helbig-Lang S, et al. Mechanism of action in CBT (MAC): methods of a multi-center randomized controlled trial in 369 patients with panic disorder and agoraphobia. Eur Arch Psychiatry Clin Neurosci [Internet]. 2009 Nov 30;259(S2):155–66. Available from: http://link.springer.com/10.1007/s00406-009-0065-6

12. Wechsler D. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). San Antonio. 2008;

13. Tombaugh TN. Trail Making Test A and B: Normative data stratified by age and education. Arch Clin Neuropsychol. 2004;

14. Aschenbrenner S, Tucha O, Lange KW. RWT: Regensburg word fluency test. Göttingen: Hogrefe; 2001.

15. Gloster AT, Wittchen H-U, Einsle F, Lang T, Helbig-Lang S, Fydrich T, et al. Psychological treatment

for panic disorder with agoraphobia: A randomized controlled trial to examine the role of therapist-guided exposure in situ in CBT. J Consult Clin Psychol [Internet]. 2011;79(3):406–20. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0023584

16. Kircher T, Arolt V, Jansen A, Pyka M, Reinhardt I, Kellermann T, et al. Effect of Cognitive-Behavioral Therapy on Neural Correlates of Fear Conditioning in Panic Disorder. Biol Psychiatry [Internet]. 2013 Jan;73(1):93–101. Available from: http://dx.doi.org/10.1016/j.biopsych.2012.07.026

17. Lueken U, Straube B, Konrad C, Wittchen H-U, Ströhle A, Wittmann A, et al. Neural Substrates of Treatment Response to Cognitive-Behavioral Therapy in Panic Disorder With Agoraphobia. Am J Psychiatry [Internet]. 2013;170(11):1345–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23982225

18. Straube B, Lueken U, Jansen A, Konrad C, Gloster AT, Gerlach AL, et al. Neural correlates of procedural variants in cognitive-behavioral therapy: A randomized, controlled multicenter fMRI study. Psychother Psychosom. 2014;83(4):222–33.

19. Reif A, Richter J, Straube B, Höfler M, Lueken U, Gloster AT, et al. MAOA and mechanisms of panic disorder revisited: From bench to molecular psychotherapy. Mol Psychiatry. 2014;19(1):122–8.

20. Stöcker T, Schneider F, Klein M, Habel U, Kellermann T, Zilles K, et al. Automated quality assurance routines for fMRI data applied to a multicenter study. Hum Brain Mapp. 2005;25(2):237–46.

21. Yang Y, Lueken U, Wittmann A, Holtz K, Kleint NI, Herrmann MJ, et al. Neural correlates of individual differences in anxiety sensitivity: an fMRI study using semantic priming. Soc Cogn Affect Neurosci [Internet]. 2016 Aug;11(8):1245–54. Available from: https://academic.oup.com/scan/article-lookup/doi/10.1093/scan/nsw024

22. Knecht S. Handedness and hemispheric language dominance in healthy humans. Brain [Internet]. 2000 Dec 1;123(12):2512–8. Available from: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/123.12.2512

23. Sass K, Habel U, Kellermann T, Mathiak K, Gauggel S, Kircher T. The influence of positive and negative emotional associations on semantic processing in depression: An fMRI study. Hum Brain Mapp. 2014;35(2):471–82.

24. Sass K, Habel U, Sachs O, Huber W, Gauggel S, Kircher T. The influence of emotional associations on the neural correlates of semantic priming. Hum Brain Mapp. 2012;33(3):676–94.

25. Sass K, Kircher T, Gauggel S, Habel U. An fMRI-study on semantic priming of panic-related information in depression without comorbid anxiety. Psychiatry Res - Neuroimaging [Internet]. 2014;222(1–2):37–42. Available from: http://dx.doi.org/10.1016/j.pscychresns.2014.02.008

26. Baayen RH, Piepenbrock R, Van Rijn H. The CELEX lexical database (CD-ROM). Philadelphia Linguistics Data Consortium University of Pennsylvania. 1993.

27. Wilson M. MRC psycholinguistic database: Machine-usable dictionary, version 2.00. Behav Res Methods, Instruments, Comput. 1988;

28. Lucas M. Semantic priming without association: a meta-analytic review. Psychon Bull Rev [Internet]. 2000;7(4):618–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11206202

29. Laver GD, Burke DM. Why do semantic priming effects increase in old age? A meta-analysis. Vol. 8, Psychology and Aging. 1993. p. 34–43.

30. Poldrack RA, Fletcher PC, Henson RN, Worsley KJ, Brett M, Nichols TE. Guidelines for reporting an fMRI study. Neuroimage. 2008;40(2):409–14.

31. Slotnick SD, Schacter DL. A sensory signature that distinguishes true from false memories. Nat Neurosci. 2004;7(6):664–72.

32. Slotnick SD, Moo LR, Segal JB, Hart J. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cogn Brain Res. 2003;17(1):75–82.

33. Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci. 2016;113(28):7900–5.

34. Slotnick SD. Cluster success: fMRI inferences for spatial extent have acceptable false-positive rates. Cogn Neurosci [Internet]. 2017;8(3):150–5. Available from: https://doi.org/10.1080/17588928.2017.1319350