

## Supplementary Methods

### **Emotional working-memory task**

We used a modified Sternberg item-recognition task (Oei et al., 2009; Oei et al., 2010; Sternberg, 1966). Each trial started with a fixation cross presented for 1 s in the center of the screen. Following the fixation cross, either one or four capital letters (the target set,  $1.5^\circ \times 1.2^\circ$  per letter) appeared on the screen for 1 s. The target set had to be held in memory during the following 1.5-s delay period. During this delay period, a picture was presented. Pictures were selected from the International Affective Picture System (IAPS; Lang, Bradley, and Cuthbert, 2005). Half of the pictures were negatively arousing ( $M \pm SE$ : valence  $2.4 \pm 0.8$ , arousal  $6.6 \pm 0.4$ ), the other half were emotionally neutral ( $M \pm SE$ : valence  $5.1 \pm 0.6$ , arousal  $3.3 \pm 0.7$ ), as rated on a 1-9 point scale (Lang, Bradley, & Cuthbert, 2005). Following the delay period, four capital letters were presented (the probe set,  $1.5^\circ \times 1.2^\circ$  per letter). On half of the trials the probe set contained one letter from the target set, and on the other half of the trials the probe set did not contain a letter from the target set. The probe set was followed by an intertrial interval of 2 s.

The participants' task was to indicate whether one of the probe letters had been part of the last target set or not, by pressing the 'z' or the 'm' key. The key assignment was balanced across participants. Participants were instructed to respond as fast and as accurately as possible. The probe set stayed on the screen until the participant made a response. If the participant did not respond within 3 s, the trial ended automatically and a 'TOO SLOW' message appeared on the screen. Prior to the start of the experimental session, participants viewed on-screen instructions and were given 8 practice trials. The experimental session consisted of 15 repetitions of the factorial combination of working-memory load (1 or 4 target letters), distractor type (neutral or negative picture) and target presence (target present or target absent). The task lasted approximately 18 minutes.

### **Attentional-blink task**

*Stimuli.* Stimuli were presented in black against a light grey background. Each trial started with a fixation cross measuring  $0.5 \times 0.5^\circ$ , presented for 1 s in the center of the screen. Subsequently, the fixation cross was replaced by a rapid serial visual presentation (RSVP) stream of 19 uppercase letters and 2 digits, each measuring approximately  $0.9 \times 0.9^\circ$ . Each letter was randomly drawn (without replacement) from the alphabet and presented for 74 ms, followed by a 24-ms blank interval. "I," "O," "Q," and "S" were left out as they resemble digits too much. The two digits (T1 and T2) were randomly drawn without replacement from the set 2 to 9. T1 was presented 10 to 13

temporal positions from the beginning of the stream. The temporal distance between T1 and T2 was either one, two, three or seven items, corresponding to lags of 98, 196, 294, and 686 ms.

*Procedure.* The participant's task was to identify both T1 and T2 by typing the digits in order on a standard keyboard after the end of the RSVP stream. Participants were instructed to guess whenever they failed to identify a digit. The two keyboard entries were followed by the presentation of a feedback stimulus for 150 ms (e.g., '+, -' to indicate that T1 was correct and T2 was incorrect). After a 1-s blank screen, the next trial started. Each participant started with 12 practice trials, three for each lag, randomly intermixed. This was followed by six blocks of 40 trials each with each block containing ten repetitions of each lag.

### Visual-search task

*Stimuli.* Each trial started with a white fixation cross measuring  $0.9 \times 0.9^\circ$  against a dark background, presented for 500 ms in the center of the screen. Subsequently, the fixation cross was replaced by a search display, which consisted of 4, 8, or 16 items that were randomly plotted in the cells of an imaginary  $6 \times 6$  matrix ( $8.7^\circ$  horizontally  $\times 9.6^\circ$  vertically) with some random jitter within the cells. On half of the trials, the target, a vertical red bar, was present in the array. On the other half of the trials, the target was absent. The distractors were vertical green bars and horizontal red bars. Thus, the target was defined by a specific conjunction of features (color and orientation).

*Procedure.* On each trial, the participant's task was to report whether or not the target ( $0.7 \times 1.3^\circ$ ) was present by giving a response with their left or right index finger using the 'z' and 'm' keys on the computer keyboard. The keyboard entry was immediately followed by a 1,000-ms blank screen after which the next trial started. Participants performed two blocks of 96 trials each, with each block containing 16 repetitions of the factorial combination of set size (4, 8, or 16) and trial type (target present or absent) presented in random order. Prior to the start of the experimental session, participants viewed on-screen instructions and were given 12 practice trials. The task instructions encouraged participants to respond as quickly as possible while minimizing the number of errors. Performance feedback was provided at the end of each block.

### Oddball tasks and EEG measurement

*Oddball tasks.* Participants performed a visual and an auditory oddball task. In the visual oddball task, a series of black crosses and circles ( $1.7 \times 1.7^\circ$ ) was presented on a light grey background. Each stimulus was presented for 250 ms and the interval between two successive stimuli was 2500 ms. In the auditory oddball task, a series of 1000-Hz and 2000-Hz tones (75 dB) was presented. Each tone lasted 150 ms and the interval between two successive tones was 2100 ms

Participants were instructed to make speeded key-press responses with the dominant hand to target stimuli (circles/2000-Hz tones, 20% of the trials) but not to non-target stimuli (crosses/1000-Hz tones, 80% of the trials). Each task consisted of 30 target trials and 120 non-target trials.

*EEG recording.* For the Dutch patients and all control participants, EEG activity was recorded from 24 Ag/AgCl scalp electrodes (Fp1, AFz, Fz, F3, F7, FCz, Cz, C3, T7, CPz, Pz, P3, P7, POz, O1, Oz, O2, P8, P4, C4, T8, F8, F4, Fp2). In addition, two electrodes were placed at the left and right mastoid. We measured the horizontal and vertical electro-oculogram (EOG) using bipolar recordings from electrodes placed approximately 1 cm lateral of the outer canthi of the two eyes and from electrodes placed approximately 1 cm above and below the participant's left eye. For the American patients, EEG, EOG and mastoid activity was recorded from a high-density array of 128 Ag/AgCl electrodes embedded in soft sponges (Geodesic Sensor Net, EGI, Inc., Eugene, OR, USA),

*Signal processing and data analyses.* For the Dutch patients and the control participants, the signal was DC amplified and digitized with a BioSemi ActiveTwo system (BioSemi B.V., Amsterdam, The Netherlands) at a sampling rate of 256 Hz. For the American patients, the signal was DC amplified and digitized with a Net Amps 200 amplifier at a sampling rate of 250 Hz, using Net Station 4.3 software (EGI, Inc., Eugene, OR, USA). Each active electrode was referenced offline to the average of the left and right mastoids. EEG and EOG were high-pass filtered at 0.1 Hz. We extracted single-trial epochs for a period from 200 ms before until 800 ms after stimulus onset. Ocular and eyeblink artifacts were corrected using the method of Gratton, Coles, and Donchin (1983) as implemented in Brain Vision Analyzer. Epochs with other artifacts (spike artifacts [50  $\mu$ V/2 ms] and slow drifts [200  $\mu$ V/200 ms]) were also discarded. Then, for each participant, task and stimulus type (target/standard), averaged waveforms aligned to a 200-ms prestimulus baseline were generated. The P3 amplitude was defined as the most positive peak in the 200–600-ms time window after the stimulus. We focused our analyses on the electrode position at which the P3 amplitude in response to target stimuli was largest.

### Pitch-discrimination task and pupillometry

*Stimuli and procedure.* Participants performed an auditory pitch-discrimination task (Gilzenrat et al., 2010; Kahneman and Beatty, 1967) while their pupils were continuously measured. They were seated in front of a computer monitor displaying a blank medium gray field, and were instructed to hold gaze within a central fixation square delineated by a thin black border subtending 10° of visual angle. Participants were presented sequences of two sinusoidal tones (72

(dB, 250 ms), and were instructed to indicate whether the second of the two tones was higher or lower in pitch than the first. Each trial began with an 850-Hz reference tone. This tone was followed 3 s later by the comparison tone, which ranged from 820 Hz to 880 Hz in steps of 10 Hz. Participants were instructed to respond as quickly and accurately as possible upon hearing the comparison tone. All participants pressed a left key if the second tone was lower and a right key if the second tone was higher than the first tone. Four seconds after the comparison tone, participants received a 250-ms feedback sound that informed them of their accuracy. The feedback was followed by a variable intertrial interval, chosen randomly between 4 and 8 s. Prior to the start of the experimental session, participants viewed on-screen instructions and were given a short block of practice trials at easiest discriminability to familiarize them with the task.

Participants performed two blocks of 36 trials, in counterbalanced order, with each block lasting approximately 10 minutes. In total, participants received 18 trials in which the reference tone and comparison tone were of equal pitch (i.e., impossible-discrimination trials), and they always received negative feedback on these trials. On the other trials, the comparison tone was selected randomly without replacement from the set [820, 830, 840, 860, 870, and 880 Hz], such that participants encountered all of these comparison tones nine times.

The experiment was conducted at a slightly dimmed illumination level. For the Dutch patients and the control participants, the left and right pupil diameters were recorded at 60 Hz using a Tobii T120 eye tracker, which is integrated into a 17-inch TFT monitor (Tobii Technology, Stockholm, Sweden). For the American patients, the left pupil diameter was recorded at 120 Hz using an Applied Systems Laboratory EYE-TRAC 6000 system (ASL, Bedford, MA, USA). These patients used a chinrest and headrest that positioned them 38.5 cm from a Sony Trinitron Multiscan E540 computer monitor.

*Pupil analysis.* Pupil data were processed and analyzed using Brain Vision Analyzer (Brain products, Gilching, Germany). Artifacts and blinks were removed using a linear-interpolation algorithm. We assessed the baseline pupil diameter prior to trial onset, and the pupil dilation following the comparison tone. To determine baseline pupil diameter, we averaged the pupil data during the two seconds immediately preceding the reference tone. The pupil dilation evoked by the comparison tone was measured as the average deviation from the baseline in the 3 s following onset of the comparison tone.

## **Positive Affect Negative Affect Scale (PANAS)**

At the beginning and the end of each test session, participants completed the PANAS (Watson, Clark, & Tellegen, 1988; translated in Dutch by Peeters, Ponds, & Boon-Vermeeren, 1999), which consists of 10 negative and 10 positive mood terms. For each term, participants indicated to what extent they currently felt that way, using a 5-point response scale with values from 1-very slightly or not at all, to 5-extremely.

## **MRI**

*Acquisition.* All MRI scans were obtained on a 3-T Philips Achieva MRI scanner, using a three-dimensional T1-weighted gradient echo sequence (TR = 9.8 ms; TE = 4.6 ms; flip angle = 8°, 140 slices). The voxel size was 0.88 x 0.88 x 1.2 mm<sup>3</sup>.

### *Analysis.*

The structural images were brain extracted (Smith, 2002), and the resulting images were segmented into grey matter, white matter and cerebrospinal fluid (CSF; Zhang 2001). We determined each participant's total brain volume, as well as the proportions of grey matter, white matter and CSF, and assessed whether these measures in each patient differed from those in the control group, using the modified independent-samples *t*-test developed by Crawford and Howell (1998). Because there are sex differences in total brain volume and in the proportion of grey matter (e.g., Cosgrove et al., 2007), we compared the female patients to the female control participants and the male patients to the male control participants.

To assess the presence of regionally specific differences in grey matter density between the patient group and the control group, we performed a voxel-based morphometry-style analysis (Ashburner and Friston, 2000, Good et al., 2001) using FSL tools (FMRIB's Software Library; Smith et al., 2004). The grey-matter partial volume images were aligned to MNI152 standard space using affine registration (Jenkinson and Smith 2001, Jenkinson et al., 2002), and the resulting images of the five patients and five matched healthy controls were averaged to create a study-specific grey-matter template. The native grey matter images were then non-linearly re-registered to this grey-matter template, modulated, and smoothed with an isotropic Gaussian kernel with a sigma of 2 mm. We used permutation-based non-parametric inference within the framework of the general-linear model, to assess whether there were brain regions with a significantly lower grey matter density in the patient group than in the control group, and vice versa (5000 permutations). We used threshold-free cluster enhancement (TFCE), a new method for finding significant clusters in MRI data without having to define an initial cluster-forming threshold (Smith and Nichols,

2009). Statistical maps were thresholded at  $p < 0.05$ , corrected for multiple comparisons across space.

### Catecholamine measurement

Plasma catecholamine concentrations were measured using high performance liquid pressure (HPLC) analysis. For the Dutch participants, fluorometric detection was used (Willemse et al., 1995). Within- and between-run coefficients of variation for plasma norepinephrine were 4.1% and 6.1% at a level of 1.76 nmol/l, respectively, and the analytical detection limit for norepinephrine was 0.002 nmol/l. For the American patients, electrochemical detection was used (Holmes et al., 1994). The coefficient of variation for plasma norepinephrine was 4.5% at a level of 1.51 nmol/l, and the analytical detection limit for norepinephrine was 0.024 nmol/l. Catecholamines were collected in ice-chilled 10 ml Vacutainer tubes (Becton-Dickenson Co., Franklin Lakes, NJ) containing 0.2 ml of a solution of EGTA (0.25 mol/L) and glutathione (0.20 mol/l).

### Statistical analyses

We compared the results of the patients OFF medication to those of the control participants using a modified independent-samples *t*-test developed specifically to compare an individual patient with a small control group (Crawford and Howell, 1998). This method maintains the Type I error rate (false positives) at the specified (5%) level regardless of the size of the control sample (Crawford and Garthwaite, 2005). The *p* value obtained by this method indicates whether the patient's score is significantly different from the control group, and also provides an unbiased point estimate of the abnormality of the patient's score; that is, it reflects the estimated proportion of the control population that would obtain a more extreme score (Crawford and Garthwaite, 2006a). We used this method to test whether the critical measures/effects in each patient OFF medication deviated from those in the control group, using a statistical threshold of  $p < 0.05$  (1-tailed). To control for potential practice effects, we compared the results of the patients that were tested OFF medication on the first study day with the control group's results on the first study day, and the results of the patients that were tested OFF medication on the second study day with the control group's results on the second study day.

We next examined the effects of medication on the patients' scores, using a regression-based method developed by Crawford and Garthwaite (2006b). The control participants' data were used to generate regression equations that predicted their scores in the second session from their scores in the first session, and vice versa (i.e., the practice effect). These regression equations were then used to predict each patient's ON-medication score from their OFF-medication score, and it was tested

whether there was a significant difference between the predicted and observed ON-medication scores. Like Crawford and Howell's (1998) modified *t*-test, this method controls the Type 1 error rate even when the size of the control sample is small. The *p* value obtained by this method provides an estimate of the abnormality of the difference between each patient's predicted and observed ON-medication scores, which reflects the estimated proportion of the control population that would show a larger difference. We used this method to test for each patient's critical measures/effects whether the effect of medication was significantly larger than the practice effect in the control group, using a statistical threshold of  $p < 0.05$  (1-tailed). For the patients that were tested ON medication on the first study day, the predicted ON-medication scores were based on the regression equation in which the control participants' scores on the second study day predicted their scores on the first study day. For the patients that were tested ON medication on the second study day, the opposite regression equation was used.

## Supplementary Results

**Supplementary Table 1.** demographic and clinical characteristics of each patient

patient	1	2	3	4	5
Age (years)	25	41	15	22	19
Sex	Female	female	male	male	female
Nationality	Dutch	Dutch	American	American	Canadian
Scaled WAIS-III vocabulary score	8	8	13	16	12
Raven's SPM score	43	45	41	53	46
Estimated IQ (based on SPM score; Peck, 1970)	98	112	100	119	107
Abnormalities unrelated to D $\beta$ H deficiency				diabetic	irregular and large pupils <sup>+</sup>
Order of ON and OFF medication test days	ON-OFF	ON-OFF	OFF-ON	OFF-ON	OFF-ON
Period on medication before study participation	1.5 years	6 years	-*	-*	4 years
Period on medication before ON-medication test day	1.5 years	6 years	2.5 days	3.5 days	2 days <sup>^</sup>
Period off medication before OFF-medication test day	13 days	7 days	whole life	whole life	4 days
Interval between test sessions (days)	13	7	6	6	6
Missing data (tasks)	EWM, pupil <sup>#</sup>	AB			
Baseline pupil diameter OFF medication (mm)	-	2.35	3.79	4.34	6.29
Baseline pupil diameter ON medication (mm)	-	2.39	3.46	2.83	6.70
Task-evoked pupil dilation OFF medication (mm)	-	0.064	-0.048	0.129	-0.088
Task-evoked pupil dilation ON medication (mm)	-	0.099	-0.053	0.060	-0.017
Brain volume (dm <sup>3</sup> )	1.42	1.20	1.55	1.46	1.28
% grey matter	43.9	41.9	48.1	45.9	48.2
% white matter	40.4	40.0	38.1	38.9	36.4
% cerebrospinal fluid	15.7	18.0	13.8	15.2	15.4

Notes: patients 3 and 4 are brothers, the other patients are unrelated; AB = attentional-blink task; EWM = emotional-working memory task; <sup>+</sup> due to a genetic defect unrelated to D $\beta$ H deficiency: a deletion on the short arm of chromosome 11; \* these two patients had never been on DOPS medication before the study; <sup>^</sup> this patient had consumed 7 doses of 300 mg before she was tested ON medication, and was feeling normal at that time; <sup>#</sup> emotional-working memory and pupillometry data were not collected for the two matched control participants of patient 1 either.

**Supplementary Table 2.** Each patient's plasma and urine catecholamine concentrations for the ON and OFF medication sessions

patient	Plasma NE		Urine NE		Plasma DA		Urine DA	
	ON	OFF	ON	OFF	ON	OFF	ON	OFF
1	0.72	ND	4682	ND	1.08	2.95	1163	669
2	-	-	4212	ND	0.39	2.73	187	21
3	0.46	0.22	14390	11	0.20	0.20	914	1670
4	0.47	0.17	12536	11	0.27	0.25	695	2242
5	0.64	ND	12588	6	0.06	0.25	1005	1757

Notes: all values are in nmol/l; ON = on DOPS medication, OFF = off medication; ND = not detectable; patient 2's plasma NE concentrations were unmeasurable due to interfering peaks in the chromatogram.

### Subjective state

Supplementary Table 3 shows the average positive and negative affect (PANAS) scores in the control group and the patient group at the beginning and end of each test session. The control participants' positive affect score was higher in the first than in the second session [ $F(1, 9) = 9.6, p = 0.01$ ], and higher at the beginning than at the end of the test sessions [ $F(1, 9) = 6.6, p = 0.03$ ]. The control participants' negative affect scores were very low, and were not significantly affected by session or point in time ( $p > 0.18$ ). The patient group reported an overall slightly lower positive affect than the control group. The patient group's negative affect score was identical to that of the control group, except for a somewhat higher negative affect at the beginning of the OFF-medication session. This suggests that medication status did not have substantial effects on the patients' affective state, which is surprising given previous findings that social anxiety and mood symptoms were diminished by L-DOPS treatment in two D $\beta$ H-deficient siblings (Critchley et al., 2000).

**Supplementary Table 3.** Positive and negative affect scores at the beginning and end of each test session in the control group and the patient group (means  $\pm$  standard deviations)

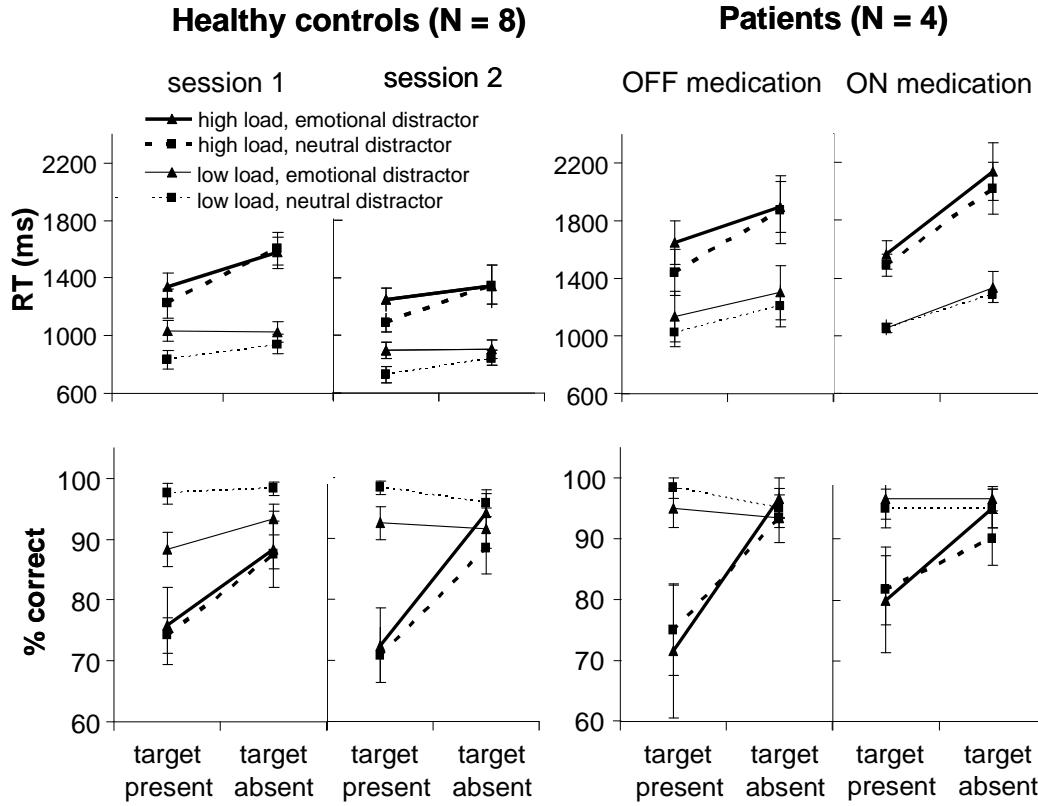
		Control group (N = 10)		Patient group (N = 5)	
		Session 1	Session 2	ON	OFF
Positive affect	beginning	$3.0 \pm 0.3$	$2.7 \pm 0.4$	$2.7 \pm 0.6$	$2.5 \pm 0.3$
	end	$2.8 \pm 0.5$	$2.6 \pm 0.5$	$2.4 \pm 0.6$	$2.5 \pm 0.5$
Negative affect	beginning	$1.2 \pm 0.3$	$1.1 \pm 0.2$	$1.2 \pm 0.3$	$1.4 \pm 0.6$
	end	$1.2 \pm 0.2$	$1.1 \pm 0.2$	$1.2 \pm 0.2$	$1.1 \pm 0.2$

Note: range of both scales = 1-5; ON = on DOPS medication, OFF = off medication

## Emotional working-memory performance

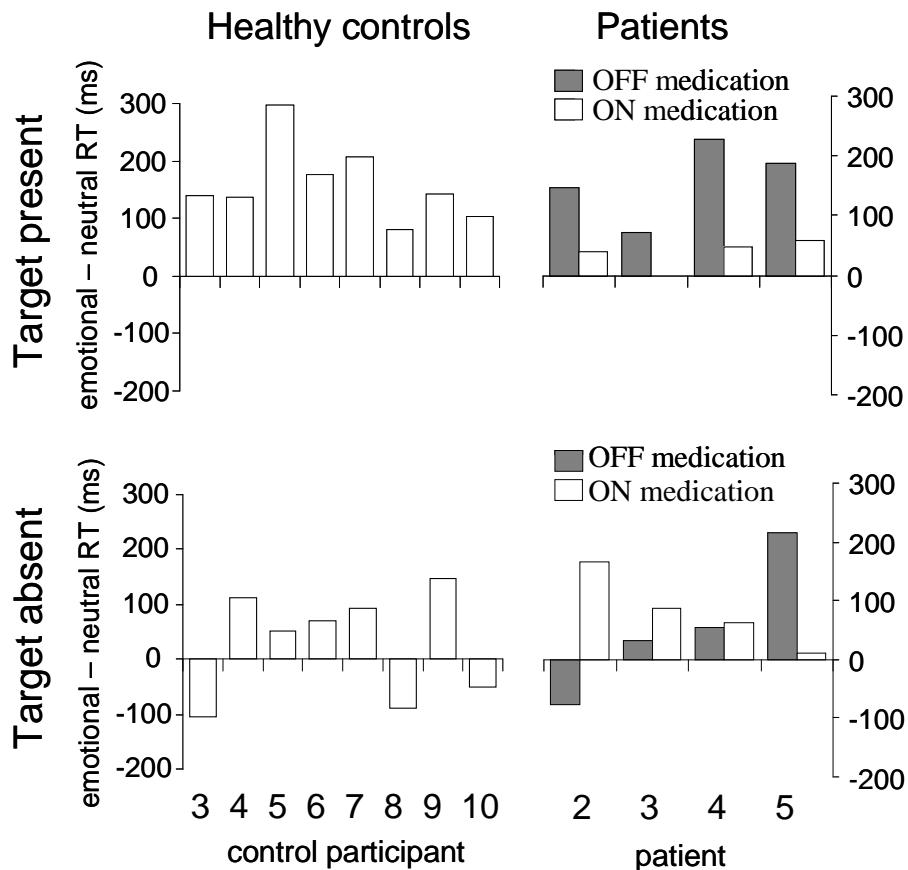
A repeated-measures ANOVA on RT in the control group yielded significant main effects of working-memory load [ $F(1, 7) = 65.7, p < 0.001$ ], target presence [ $F(1, 7) = 8.9, p = 0.02$ ] and distractor type [ $F(1, 7) = 14.7, p = 0.006$ ]. There was a main effect of session as well [ $F(1, 7) = 21.5, p = 0.002$ ]. In addition, distractor type interacted with target presence [ $F(1, 7) = 16.3, p = 0.005$ ], indicating that the interfering effect of emotional distractors on RT was larger on target-present than on target-absent trials. Finally, there was an interaction between session and working-memory load [ $F(1, 7) = 11.9, p = 0.01$ ], indicating that the effect of working memory load on RT was larger in session 1 than in session 2.

**Supplementary Figure 1.** Average correct RT and accuracy for the control group and the patient group in the emotional working-memory task, as a function of target presence, working-memory load, distractor type and session (error bars are standard errors of the mean).



As expected, accuracy in the control group was significantly affected by working-memory load [ $F(1, 7) = 30.8, p = 0.001$ ] and target presence [ $F(1, 7) = 26.9, p = 0.001$ ]. In addition, there was an interaction between working-memory load and target presence [ $F(1, 7) = 11.6, p = 0.01$ ]. There were no main effects of session ( $p = 0.93$ ) or distractor type ( $p = 0.22$ ) on accuracy.

**Supplementary Figure 2.** Each individual participant's emotional-interference effect on RT as a function of target presence. Because session did not interact with distractor type or target presence in the control group, the data from the control participants are averaged across sessions.

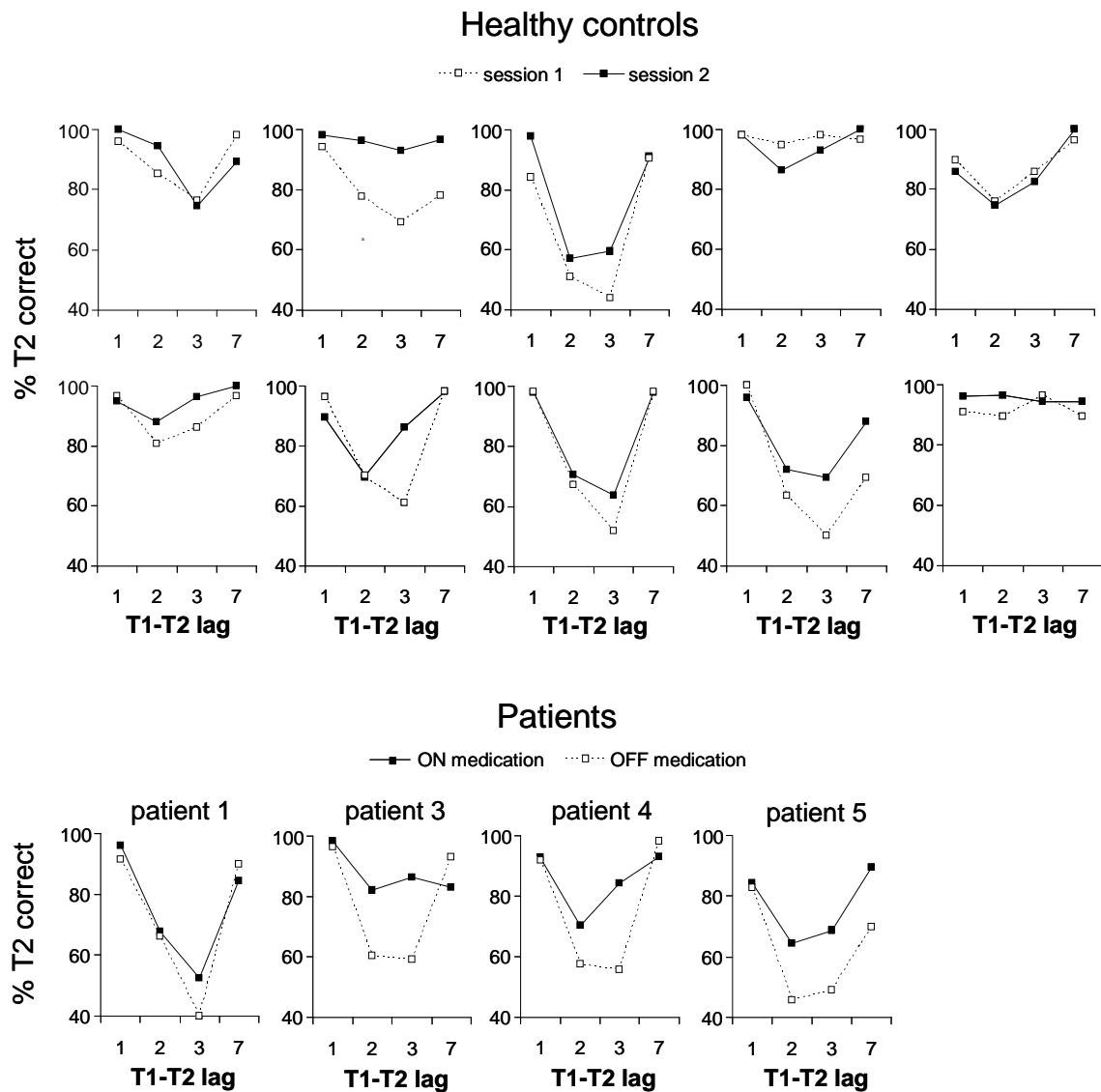


### Attentional-blink performance

T1-identification accuracy in the control group showed an increasing trend with lag [ $F(3, 27) = 4.1, p = 0.01$ ], but did not differ between the two sessions [ $F(1, 9) = 2.2, p = 0.17$ ]. Although most patients showed a numerically lower T1 identification accuracy when OFF compared to ON medication, the medication effect on T1 accuracy was not significantly different from the control group's test-retest effect in any of the patients (all  $p > 0.09$ , according to Crawford & Garthwaite's regression-based method)

T2-identification accuracy in the control group was better in session 2 than in session 1 [ $F(1, 9) = 7.7, p = 0.02$ ]. In addition, T2-identification accuracy was affected by lag [ $F(3, 27) = 12.1, p = 0.001$ ], reflecting the characteristic shape of the attentional-blink curve. There was no significant interaction between session and lag [ $F(3, 27) = 2.1, p = 0.13$ ].

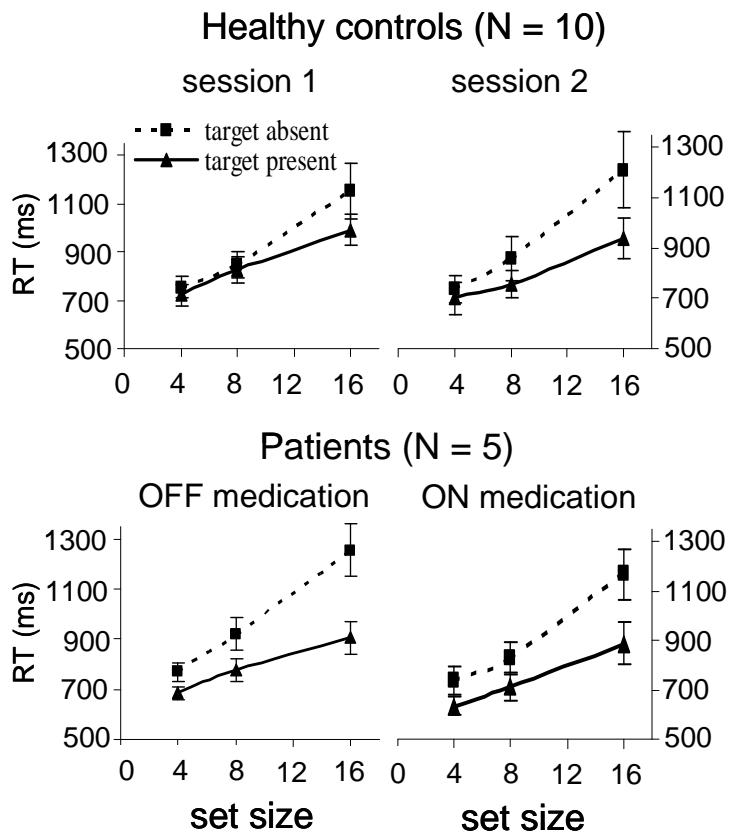
**Supplementary Figure 3.** Each individual participant's T2 accuracy as a function of lag and session.



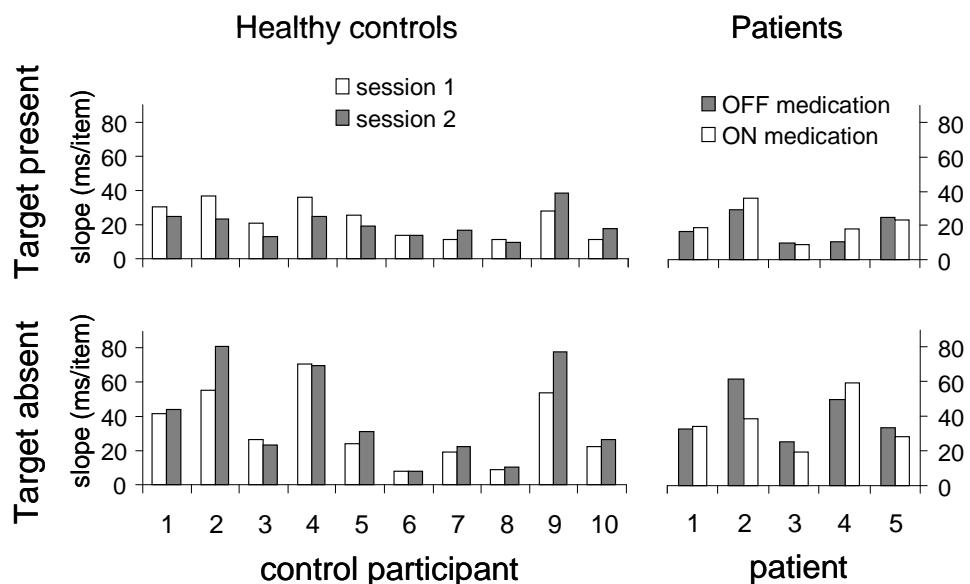
### Visual-search performance

As expected, RT in the control group was significantly affected by set size [ $F(2, 18) = 29.7, p < 0.001$ ] and target presence [ $F(1, 9) = 6.6, p = 0.03$ ]. In addition, there was a significant interaction of these two variables [ $F(2, 18) = 7.8, p = 0.004$ ], indicating that set-size effects were larger for target-absent trials. There was no main effect of session ( $p = 0.64$ ), but session interacted with target presence [ $F(1, 9) = 10.6, p = 0.01$ ], indicating that the effect of target presence was larger in session 2. There was also a three-way interaction between session, target presence and set size [ $F(2, 18) = 3.9, p = 0.04$ ], indicating that the interaction between target presence and set size was more pronounced in session 2.

**Supplementary Figure 4.** Average correct RT for the control group and the patient group in the visual-search task as a function of target presence, set size and session (error bars are standard errors of the mean).



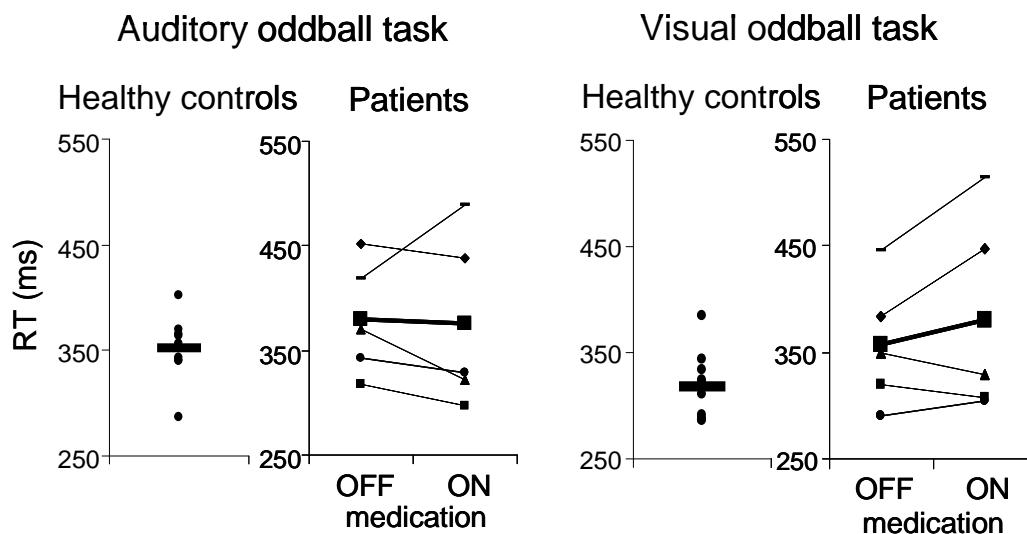
**Supplementary Figure 5.** Slopes in the visual-search task for each individual participants, as a function of target presence and session.



Error rates were rather low (average 4.6% in the control group, 2.9% in the patient group ON medication and 2.2% in the patient group OFF medication), and were not affected by set size in the control group ( $p = 0.41$ ), indicating that the increasing RT with set size was not due to a speed-accuracy trade-off.

### Target-detection performance in the oddball tasks

**Supplementary Figure 6.** Correct RT for the control participants and the patients in the auditory and visual oddball task. The bold lines indicate the average values across participants, and the thinner lines and points indicate the RTs of each individual participant. Because there was no effect of session in the control group, the data for the control participants are averaged across the two sessions.



Target-detection RT in the control group did not differ across sessions [ $F(1, 9) = 0.25, p = 0.63$ ], but was shorter in the visual task than in the auditory task [ $F(1, 9) = 10.6, p = 0.01$ ].

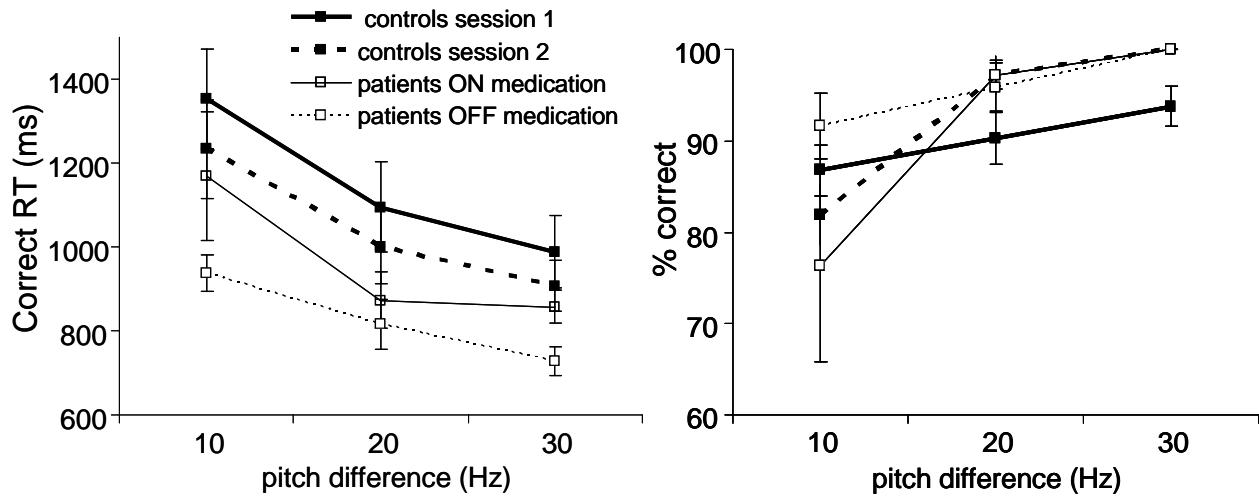
Target-detection accuracy was very high (mean accuracy > 98% in both groups and sessions), and did not differ across sessions [ $F(1, 9) = 1.6, p = 0.24$ ] or task [ $F(1, 9) = 2.0, p = 0.19$ ] in the control group.

### Pitch-discrimination performance in the tone-discrimination task

The control group's RT increased as a function of tone-discrimination difficulty [ $F(2, 14) = 37.4, p < 0.001$ ]. There was no significant main effect of session on RT ( $p = 0.08$ ), and no interaction between session and difficulty ( $p = 0.9$ ).

The control group's accuracy decreased with increasing difficulty [ $F(2, 14) = 14.1, p < 0.001$ ]. There was no significant main effect of session on accuracy ( $p = 0.06$ ), and no interaction between session and difficulty ( $p = 0.08$ ).

**Supplementary Figure 7.** Average correct RT and percentage of correct responses in the tone-discrimination task for the control group and the patient group, as a function of the pitch difference between the two tones (i.e., pitch-discrimination difficulty) and session (error bars indicate standard errors of the mean)



## Supplementary References

- Ashburner J, Friston KJ. Voxel-based morphometry--the methods. [Review]. *Neuroimage* 2000; 11: 805-21.
- Cosgrove KP, Mazure CM, Staley JK. Evolving knowledge of sex differences in brain structure, function, and chemistry. [Review]. *Biol Psychiatry* 2007; 62: 847-55.
- Crawford JR, Garthwaite PH. Testing for suspected impairments and dissociations in single-case studies in neuropsychology: Evaluation of alternatives using Monte Carlo simulations and revised tests for dissociations. *Neuropsychology* 2005; 19: 318-31.
- Crawford JR, Garthwaite PH. Methods of testing for a deficit in single case studies: Evaluation of statistical power by Monte Carlo simulation. *Cognitive Neuropsychology* 2006a; 23: 877-904.
- Crawford JR, Garthwaite PH. Comparing patients' predicted test scores from a regression equation with their obtained scores: a significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology* 2006b; 20: 259-71.
- Crawford JR, Howell DC. Comparing an individual's test score against norms derived from small samples. *Clin Neuropsychol* 1998; 12: 482-86.
- Critchley H, Mathias CJ, Dolan RJ. Congenital dopamine B-hydroxylase deficiency and schizotypy. *Schizophr Res* 2000; 41: 254.
- Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci* 2010; 10: 252-69.
- Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage*. 2001; 14: 21-36.
- Gratton G, Coles MG, Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol* 1983; 55: 468-84.
- Holmes C, Eisenhofer G, Goldstein DS. Improved assay for plasma dihydroxyphenylacetic acid and other catechols using high-performance liquid chromatography with electrochemical detection. *J Chromatogr B Biomed Appl* 1994; 653: 131-8.
- Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* 2001; 5: 143-56.
- Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 2002; 17: 825-41.
- Kahneman D, Beatty J. Pupillary responses in a pitch-discrimination task. *Percept Psychophys* 1967; 2: 101-5.
- Lang PJ, Bradley MM, Cuthbert BN. International Affective Picture System (IAPS): Digitized photographs, instruction manual, and affective ratings (Tech. Rep. No. A-6). Gainesville: University of Florida, Center for Research in Psychophysiology. 2005.
- Oei NY, Tollenaar MS, Elzinga BM, Spinhoven P. Propranolol reduces emotional distraction in working memory: a partial mediating role of propranolol-induced cortisol increases? *Neurobiol Learn Mem* 2010; 93: 388-95.
- Oei NY, Tollenaar MS, Spinhoven P, Elzinga BM. Hydrocortisone reduces emotional distracter interference in working memory. *Psychoneuroendocrinology* 2009; 34: 1284-93.
- Peck OF. The conversation of Progressive Matrices and Mill Hill Vocabulary raw scores into deviation IQs. *Journal of Clinical Psychology* 1970; 26: 67-70.
- Peeters FPML, Ponds RWHM, Vermeeren MTG. Affectivity and self-report of depression and anxiety (Dutch). *Tijdschr Psych* 1996; 38: 240-50.
- Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002; 17:143-55.
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 2004; 23: S208-19.

- Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 2009; 44: 83-98.
- Sternberg S. High-speed scanning in human memory. *Science* 1966; 153: 652-4.
- Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 1988; 54: 1063-70.
- Willemse JJ, Ross HA, Jacobs MC, Lenders JW, Thien T, Swinkels LM, Benraad TJ. Highly sensitive and specific HPLC with fluorometric detection for determination of plasma epinephrine and norepinephrine applied to kinetic studies in humans. *Clin Chem* 1995; 41: 1455-60.
- Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 2001; 20: 45-57.