Arousal, exploration and the locus coeruleus-norepinephrine system

Jepma, M.

Chapter 4

Neurocognitive function in dopamine-β-hydroxylase deficiency

This chapter is published as: Jepma, M., Deinum, J., Asplund, C.L., Rombouts, S.A., Tamsma, J.T., Tjeerdema, N., Spapé, M.M., Garland, E.M., Robertson, D., Lenders, J.W., & Nieuwenhuis, S. (in press). Neurocognitive function in dopamine-β-hydroxylase deficiency. *Neuropsychopharmacology*

Abstract

Dopamine-β-hydroxylase (DβH) deficiency is a rare genetic syndrome characterized by the complete absence of norepinephrine in the peripheral and the central nervous system. DβH-deficient patients suffer from several physical symptoms, which can be treated successfully with L-threo-3,4-dihydroxyphenylserine, a synthetic precursor of norepinephrine. Informal clinical observations suggest that DβH-deficient patients do not have obvious cognitive impairments, even when they are not medicated, which is remarkable given the important role of norepinephrine in normal neurocognitive function. The present study provided the first systematic investigation of neurocognitive function in human DβH deficiency. We tested five DβH-deficient patients and ten matched healthy control participants on a comprehensive cognitive task battery, and examined their pupil dynamics, brain structure and the P3 component of the electroencephalogram. All participants were tested twice; the patients were tested once ON and once OFF medication. Magnetic resonance imaging scans of the brain revealed that the patients had a smaller total brain volume than the control group, which is in line with the recent hypothesis that norepinephrine has a neurotrophic effect. In addition, the patients showed an abnormally small or absent task-evoked pupil dilation. However, we found no substantial differences in cognitive performance or P3 amplitude between the patients and the control participants, with the exception of a temporal-attention deficit in the patients OFF medication. The largely spared neurocognitive function in DβH-deficient patients suggests that other neuromodulators have taken over the function of norepinephrine in the brains of these patients.

Introduction

The locus coeruleus–norepinephrine (LC–NE) system is one of the major neuromodulatory systems in the brain. For a long time, investigators have associated this system with basic functions such as arousal and the sleep-wake cycle (Aston-Jones et al., 1984; Jouvet, 1969), and with various neuropsychiatric disorders such as depression and attention-deficit hyperactivity disorder (Ressler and Nemeroff, 2001; Siever and Davis, 1985). In addition, recent studies have shown that the LC-NE system is involved in more specific cognitive functions, such as memory, attention, perception, and decision making (Aston-Jones and Cohen, 2005; Robbins, 1997; Sara, 2009). These findings suggest that norepinephrine (NE) is essential for normal cognitive function in humans.

DβH deficiency is a rare genetic syndrome that is characterized by the congenital absence of the enzyme dopamine-β-hydroxylase (DβH), which is responsible for the conversion of dopamine (DA) to NE (Man in't Veld et al., 1987a; Robertson et al., 1986). As a result, DβH deficiency is characterized by a complete lack of NE and epinephrine in both the central and the peripheral nervous system (Man in't Veld et al., 1987a). There are currently approximately 15 patients with DβH deficiency known worldwide. These patients suffer from several physical symptoms, including severe orthostatic hypotension, fatigue and impaired exercise tolerance (Robertson and Garland, 2010). The only effective treatment of DβH deficiency involves administration of the drug L-threo-3,4-dihydroxyphenylserine (DOPS, droxidopa), which is converted directly into NE via L-aromatic-amino-acid decarboxylase, thereby bypassing DβH (Biaggioni and Robertson, 1987; Goldstein, 2006; Man in 't Veld et al., 1987b). Studies in rats and mice have shown that DOPS crosses the blood-brain barrier, and activates the production of NE in the central nervous system as well as the peripheral nervous system (Ishikawa et al., 1987; Kato et al., 1987a,b; Semba et al., 1985; Thomas et al., 1998). Treatment with DOPS results in a dramatic relief of physical symptoms and a substantial improvement of the quality of life of DβH-deficient patients.

The biochemical features, autonomic physiology and physical symptoms associated with human DβH deficiency have already been described in several studies (e.g., Mathias et al., 1990; Robertson et al., 1991; Thompson et al., 1995; Timmers et al., 2004). In addition, a post-mortem microscopic examination of the brain of one DβH-deficiency patient has revealed no histological abnormalities and no evidence for neuronal loss (Cheshire et al., 2006). However, to date there have been no systematic studies on cognitive and brain function in DβH deficiency. Informal clinical observations suggest that even before starting treatment, DβH-deficient patients do not have obvious cognitive impairments, which is striking given the large amount of evidence that NE plays an important role in normal cognitive function (Sara, 2009). This suggests that more carefully controlled laboratory tests may reveal subtle neurocognitive deficits in DβH-deficient patients that have remained unnoticed in informal observations.

The present study provides the first systematic evaluation of neurocognitive function in DβH deficiency. We tested 5 patients with DβH deficiency on a battery of cognitive tasks that have been proposed to depend on normal noradrenergic function, including an emotional working-

memory task (Chamberlain et al., 2006; Oei et al., 2010) and a temporal-attention task (attentional-blink task; De Martino et al., 2007; Nieuwenhuis et al., 2005a; Warren et al., 2009), expecting that these tasks would reveal possible abnormalities in the DβH-deficient patients. In addition, we examined task-evoked changes in pupil diameter, and recorded the electroencephalogram (EEG) during a target-detection task to examine event-related potential (ERP) correlates of noradrenergic activity (Liu et al., 2009; Nieuwenhuis et al., 2005b; Pineda et al., 1989). To assess whether potential abnormalities in performance were restricted to NE-mediated tasks, we also tested the patients on a spatial-attention task that does not probe noradrenergic function (Greenwood et al., 2005; Nieuwenhuis et al., 2007). Finally, we acquired an MRI scan of the patients' brain to assess possible abnormalities in brain volume and structure. We tested the patients once ON and once OFF DOPS medication, and compared their results with those of a matched healthy control group.

Materials and methods

Participants

We tested five DβH-deficient patients (two Dutch, two American, and one Canadian) and ten healthy controls (all Dutch). The two American patients were brothers, and the other patients were unrelated (see Supplementary Table 1 for the patients' demographic and clinical details). The genetic mutations in the DBH gene have been identified for all patients. Patient 1 is homozygous for the IVS1 +2T>C mutation, a mutation of the 5' splice site in the first intron which leads to abnormal splicing and hence a dysfunctional protein. Patient 2 is homozygous for a missense mutation in 764G>T (C255F; Deinum et al., 2004). Patients 3 and 4 are heterozygous for both the IVS1 +2T>C mutation and the 991G>A (D331N) missense mutation. Patient 5 is homozygous for two missense mutations in 259G>A (V87M) and 991G>A (D331N). Patient 5 also has a rare mosaic deletion at chromosome 11p13 [46,XX,del(11)(p12p14)/46,XX] which is unrelated to her DβH-deficiency (Erez et al., 2010).

The patient and control group were matched for age, sex and IQ (Table 1). We used the Vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS III, Wechsler, 1997) and the Raven's Standard Progressive Matrices test (SPM; Raven et al., 1988) to estimate IQ. The Dutch patients and their controls were matched for educational level as well. Given the different educational systems in the US and the Netherlands it was not possible to match the American patients and their Dutch control participants in terms of educational level; hence we matched for estimated IQ instead of educational level. Participants gave written informed consent before participation, and the study was approved by the medical ethics committee of the Leiden University Medical Center and the Institutional Review Board of Vanderbilt University.

Table 1. Demographic details of the control group and the patient group (means ± standard deviations)

	Control group (N = 10)	Patient group (N = 5)
Age (years)	24.6 ± 11.0	24.4 ± 10.0
Sex (proportion female)	6/10	3/5
Interval between test sessions (days)	7.5 ±3.2	7.6 ± 2.7
Scaled WAIS-III vocabulary score	8.6 ± 2.3	11.4 ± 3.4
Raven's SPM score	44.5 ± 6.9	45.6 ± 4.6
Estimated IQ (based on SPM score)	106.5 ± 11.1	107.2 ± 8.6

Notes: WAIS = Wechsler Adult Intelligence Scale, highest possible scaled vocabulary score = 19; SPM = standard progressive matrices, highest possible score = 66.

General procedure

All participants were tested twice on the same cognitive-task battery, with an intervening period of six to thirteen days. The patient and control groups had similar intervening periods (Table 1). Two patients were tested ON medication on the first test day and OFF medication on the second test day, and the other three patients were tested in the opposite order. Two of these patients had never been on DOPS medication before and started taking medication at least two days before the second test day. The other patients stopped taking their daily medication four to thirteen days before the OFF-medication test day and stayed off medication up to and including this day. Preceding and during the ON-medication test day, the patients took their DOPS medication as usual (see Supplementary Table 1 for the patients' demographic and clinical details).

The task battery included five cognitive tasks, described below and, in more detail, in Appendix I. At the beginning and end of each test day, participants completed the Positive Affect and Negative Affect Schedule (PANAS; Watson et al., 1988; translated into Dutch by Peeters et al., 1996). To measure catecholamine levels, we collected blood and 24-hour urine samples from the patients, prior to each test session (Table 1). Blood samples were taken after fifteen minutes of supine rest. We also collected blood samples from most control participants. Since we expected no differences in catecholamine levels between the two sessions for the control participants, their blood samples were collected only once. Finally, on one of the test days a structural T1-weighted MRI brain scan was acquired (see Appendix I for details of acquisition and analysis).

Emotional working-memory task

NE plays an important role in emotional memory (i.e., Chamberlain et al., 2006). The well-known phenomenon that emotional events are memorized better than neutral events (e.g., Cahill and McGaugh, 1998), for example, is associated with β-adrenergic-dependent modulations of amygdala-hippocampus interactions (Strange et al., 2003; Strange and Dolan, 2004). In addition, emotional distractor stimuli impair working-memory performance to a higher degree than neutral distractor stimuli (e.g., Buchner et al., 2004; Dolcos and McCarthy, 2006; Oei et al., 2009, 2010), an effect that is reduced by administration of the β-adrenergic antagonist propranolol (Oei et al.,

2010). We examined the effects of emotional and neutral distractor stimuli on performance in the working-memory task used by Oei et al. (2009, 2010).

Each trial of this task started with the presentation of either one or four letters (the target set), which had to be held in memory for later recognition. The target set was followed by a 1,500-ms delay period during which either a neutral picture or a negatively arousing picture was presented. After this, four letters (the probe set) were presented and participants had to indicate, as quickly and accurately as possible, whether or not the probe set contained a letter from the preceding target set.

Attentional-blink task

The attentional-blink paradigm is the most commonly used paradigm for investigating attentional selection in the temporal domain (for a review see Martens and Wyble, 2010). The attentional blink refers to a deficit in processing the second of two target stimuli that are presented in close temporal succession. This deficit is most severe when the second target is presented within 200-400 ms after the first target (Raymond et al., 1992), and is thought to result from competition between the two target stimuli for limited attentional resources (Shapiro et al., 1997). When the two targets are presented within approximately 200 ms, performance is often spared (e.g., Hommel and Akyürek, 2005), a phenomenon termed “lag-1 sparing”.

The temporal dynamics of the LC-NE system suggest that the LC-NE system mediates attentional selection in the temporal domain (Cohen et al., 2004; Dayan and Yu, 2006; Usher et al., 1999). LC neurons exhibit a phasic increase in activity shortly following task-relevant or otherwise motivationally significant stimuli (Aston-Jones et al., 2000). The resulting transient release of NE in cortical areas temporarily increases the responsivity of these areas to their input, which selectively facilitates the processing of the eliciting stimulus (Berridge and Waterhouse, 2003; Servan-Schreiber et al., 1990). Phasic increases in LC activity are followed by a brief refractory period during which LC-NE-mediated facilitation of information processing is temporarily unavailable (e.g., Aghajanian et al., 1977). These temporal dynamics of the LC-NE system suggest that the attentional blink may be mediated by the LC-NE system (Nieuwenhuis et al., 2005a; Warren et al., 2009). Consistent with this idea, β-adrenergic blockade impaired detection of the second target in an attentional-blink task (De Martino et al., 2007).

On each trial of this task, participants viewed a rapid serial visual presentation (RSVP) stream consisting of 2 target stimuli (T1 and T2; digits) and multiple distractor stimuli (letters), presented for about 100 ms each. The temporal distance between T1 and T2 was 1, 2, 3 or 7 items. Following each stream, participants were asked to report T1 and T2.

Visual-search task

This task examined attentional selection in the spatial domain. The spatially-nonspecific pattern of LC projections to the cortex suggests that the LC-NE system does not mediate spatial attention (Cohen et al., 2004; Greenwood et al., 2005; Nieuwenhuis et al., 2007). This task was

included to assess whether possible performance abnormalities of the DβH-deficient patients were restricted to NE-mediated tasks. On each trial of this task, participants searched for a target stimulus (a red vertical bar) among a variable number of distractor stimuli (green vertical bars and red horizontal bars) in a visual-search array, and indicated as quickly as possible whether the target stimulus was present or absent.

Oddball tasks combined with EEG measurement

We examined the P3, a prominent component of the scalp-recorded event-related brain potential. The P3 component is a broad, positive, large-amplitude potential which peaks between 300 and 400 ms following presentation of stimuli in any sensory modality (Sutton et al., 1965), and is largest over central-parietal midline electrodes. The amplitude of the P3 is strongly affected by the subjective probability and motivational significance of the eliciting stimulus: P3 amplitude increases with decreasing probability and with increasing motivational significance of the eliciting stimulus. In contrast, with the exception of tone intensity (Roth et al., 1984), P3 amplitude is relatively insensitive to physical stimulus properties. Several lines of evidence suggest that the P3 reflects the phasic response of the LC-NE system to the outcome of stimulus evaluation and decision making, and the consequent effects of the noradrenergic potentiation of information processing (reviewed in Nieuwenhuis et al., 2005b; see also Liu et al., 2009; Pineda et al., 1989).

The most common paradigm for studying the P3 is the oddball task, in which infrequent target stimuli are embedded in a series of frequently presented non-target stimuli (standards), and participants have to respond to each target stimulus but not to the standard stimuli. We measured participants' EEG while they performed visual and auditory versions of the oddball task, and assessed the P3 elicited by target stimuli.

Pitch-discrimination task combined with pupillometry

We examined participants' pupil diameter during performance of a pitch-discrimination task. Although the luminance level is the most important determinant of pupil diameter, there are also small but reliable changes in pupil diameter related to cognitive processing (Beatty and Wagoner, 1978; Kahneman, 1973). A large number of studies have shown that task processing is accompanied by a rapid increase in pupil diameter, and that the size of this pupil dilation reflects the information-processing load (e.g., Hess and Polt, 1964).

Several studies have reported that DβH-deficient patients have small pupils, but a normal pupillary light reflex and accommodation response (Biaggioni et al., 1990; Man in 't Veld et al., 1987a; Robertson et al., 1986). In addition, one study reported a prolonged redilation time following the light reflex in a sibling pair with DβH deficiency (Smith and Smith, 1999). The light reflex and accommodation response both produce pupil constrictions, which are subserved by the iris sphincter muscles. These muscles are innervated by cholinergic input from the parasympathetic nervous system. In contrast, pupil dilation is controlled by the iris dilator muscles which are activated

primarily via noradrenergic innervation of α-1 adrenoceptors (Hoffman and Taylor, 2001). This suggests that task-evoked pupil dilations in DβH-deficient patients might be abnormal.

On each trial of this task, a sequence of two tones was presented, and participants had to indicate whether the second tone was higher or lower in pitch than the first. We analyzed participants' baseline pupil diameter and their pupil dilation in response to the second tone.

Results

The control participants' behavioral, EEG and pupil data were analyzed by means of repeated-measures ANOVAs, with session (session 1 vs. session 2) and the independent task variables as within-subject factors. We tested whether the critical measures/effects in each patient OFF medication deviated from those in the control group using a modified *t*-test developed specifically to compare individual patients with a small control group (Crawford and Howell, 1998). In addition, we examined the effects of medication on the patients' scores, using the regression-based method developed by Crawford and Garthwaite (2006; see Appendix I for details of these analyses).

We focus our description of the results on the critical measures/effects of each task. The full factorial analyses of the data, the PANAS (i.e., subjective state) data, and results of the individual participants are reported in the Appendix II.

Catecholamine concentrations

Table 2 shows the average plasma and urine NE and DA concentrations in the patient group ON and OFF medication, and the plasma concentrations in the control group (see Supplementary Table 2 for the data from the individual patients). When OFF medication, two of the patients (patients 3 and 4) had plasma NE concentrations that were significantly lower than that in the control group [ps (1-tailed) < 0.03; Crawford and Howell's (1998) modified t-test] and the other patients had undetectable plasma NE concentrations. The apparent extremely low residual plasma NE concentration in patients 3 and 4 were likely due to technical artifacts, since plasma concentrations of the NE metabolite dihydroxyphenylglycol (DHPG) were extremely low in these patients when they were OFF medication. DHPG concentrations in patients 3 and 4 OFF medication were lower than 0.03 nmol/l, which is less than 1% of normal. As expected, all patients' plasma and urine NE concentrations were higher when ON compared to OFF medication, and this effect was especially pronounced for the urine concentrations. For the ON-medication session, the plasma NE concentrations of patient 1 and 5 did not differ significantly from the control group [p(1-tailed) = 0.09 and 0.08, respectively], but the plasma NE concentrations of patient 3 and 4 were still lower than that in the control group [p(1-tailed) = 0.048 and 0.049, respectively].

When OFF medication, all patients had higher plasma DA concentrations than the control group (all *p*s < 0.001). Although most patients' plasma DA concentrations were lower when ON compared to OFF medication, the ON medication concentration was still larger than that in the

control group for all but one patient. The medication effects on the urine DA concentrations were less consistent; patients 1 and 2 had higher urine DA concentrations when ON medication, whereas patients 3, 4 and 5 showed the opposite effect.

Table 2. Plasma and urine catecholamine concentrations in the control group and the patient group OFF and ON medication (means ± standard deviations).

	Healthy controls[+]	Patients OFF	Patients ON
Plasma NE	1.46 ± 0.45	0.10 ± 0.12	0.57 ± 0.13
Urine NE	-	5.50 ± 5.40	9682 ± 4839
Plasma DA	0.06 ± 0.02	1.28 ± 1.43	0.40 ± 0.40
Urine DA	-	1271 ± 903	793 ±379

Notes: [+] plasma concentrations were determined for 6 control participants; OFF = off medication; ON = on DOPS medication; all concentrations are in nmol/l; see Supplementary Table 2 for the catecholamine concentrations of the individual patients and missing data.

Emotional working-memory performance

The critical measure in this task was the interfering effect of emotional relative to neutral distractors on reaction time (RT). As expected, the control participants responded more slowly on trials with emotional compared to neutral distractors [$F(1, 7) = 14.7$, $p = 0.006$]. In addition, consistent with previous studies (Oei et al., 2009, 2010), distractor type interacted with target presence [$F(1, 7) = 16.3$, $p = 0.005$], indicating that the emotional-interference effect on RT was significant on target-present trials [$F(1, 7) = 43.9$, $p < 0.001$; effect range = 80 - 299 ms] but not on target-absent trials [$F(1, 7) = 0.75$, $p = 0.42$].

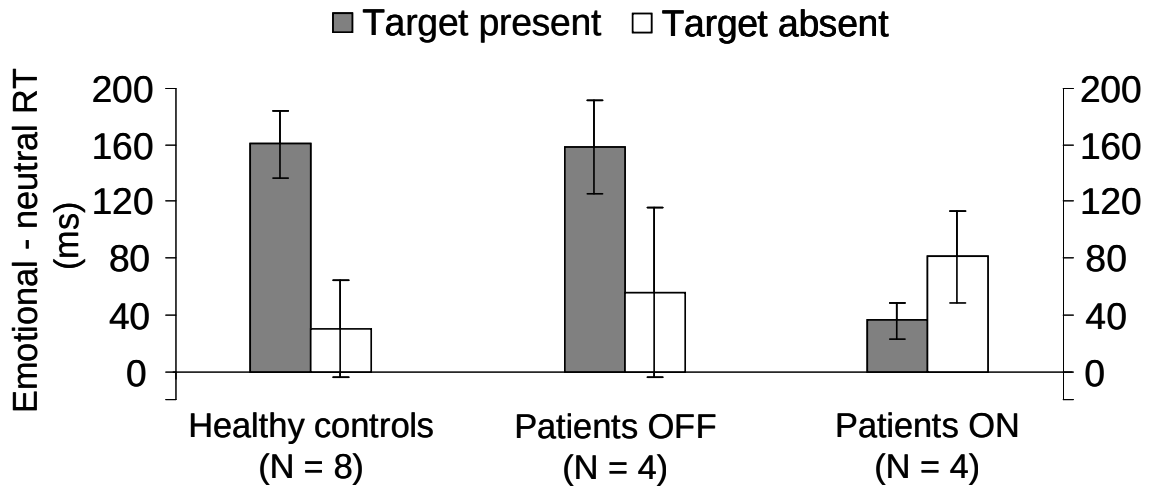


Figure 1. Average emotional-interference effect (i.e., RT on trials with emotional relative to neutral distractors) for the control group and the patient group OFF and ON medication, as a function of target presence (error bars are standard errors of the means). Because session did not interact with distractor type or target presence in the control group, the results from the control group are averaged across the two sessions.

Figure 1 shows the average increase in correct RT on trials with emotional relative to neutral distractors as a function of target presence, in the control group and in the patient group OFF and ON medication. When OFF medication, all patients showed an emotion-related slowing of responses on target-present trials that did not differ from the effect in the control group (effect range = 72 - 226 ms; all *t*s(7) < 0.8; *p*s > 0.24; Table 4; see Supplementary Figure 2 for the individual effects). In addition, all patients showed a smaller emotional interference effect when they were ON compared to OFF medication, but this medication effect did not differ significantly from the control group's practice effect in any of the patients (all *p*s > 0.08; Table 4). The normal emotional-interference effect in the patients OFF medication, and the finding that this interference effect was less pronounced when the patients were ON medication are both remarkable given the evidence that emotional-interference effects are normally mediated by NE.

The full factorial analysis of the effects of target presence, working-memory load, distractor type and session on correct RT and accuracy in the control group is reported in Appendix II and in Supplementary Figure 1.

Attentional-blink performance

Figure 2 shows the average T1 accuracy (upper panels) and T2 accuracy (lower panels; contingent on correct T1 identification) in the control group and the patient group, as a function of lag (1, 2, 3 or 7) and session. The T2 accuracy curves show a pattern that is characteristic of attentional blink data: lag-1 sparing, followed by a drop in performance for lags 2 and 3 (i.e., the attentional blink), and a recovery of performance at lag 7. This pattern was expressed in a significant effect of lag in the control group [$F(3, 27) = 12.1$, $p = 0.001$].

The critical measure in this task is the size of the attentional blink, which we defined as the decrease in T2 identification accuracy at lags 2 and 3, relative to lag 7 (Maclean and Arnell, 2010). When OFF medication, the patient group showed a larger attentional blink than the control group (average = 33.5% vs. 16.7%), but the difference from the control group only approached significance in patient 1 (Table 4; see Supplementary Figure 3 for the individual T2 accuracy curves). In addition, the patients showed a smaller attentional blink when they were ON compared to OFF medication: for three of the four patients tested on this task, the effect of medication on attentional-blink size was significantly larger than the practice effect in the control group (*p*s < 0.05; Table 4). The fourth patient also showed a marked increase in T2 accuracy when ON compared to OFF medication, but this did not result in a significant effect on attentional-blink size because the enhancing effect of medication was present at lags 2, 3 and 7. Together, these findings suggest that T2 identification accuracy during the attentional blink was impaired in the patients OFF medication, and that this impairment was restored by the DOPS medication.

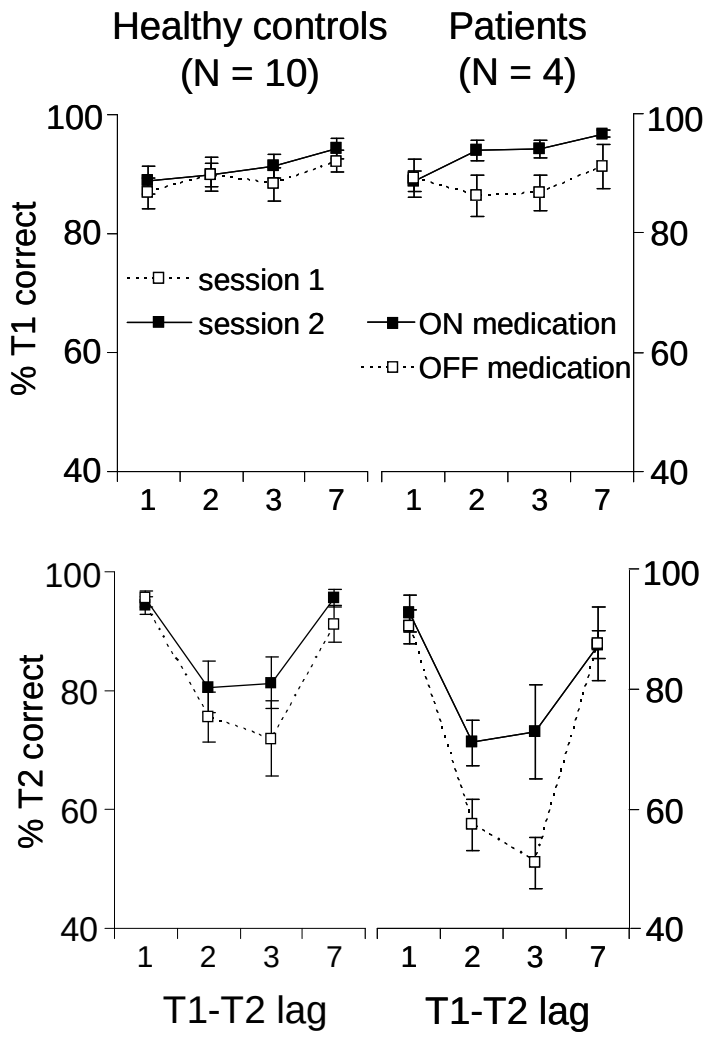


Figure 2. Average T1 and T2 identification accuracy in the attentional-blink task for the control group and the patient group, as a function of lag and session (error bars are standard errors of the means). Trials on which T1 and T2 were accurately identified but in the wrong order were treated as correct. As is usual, T2 accuracy is reported contingent on accurate identification of T1.

Visual-search performance

The critical measure in this task was the effect of set size (i.e., the total number of items in the search display) on RT. As expected, RT in the control group showed an increasing trend with set size [$F(2, 18) = 29.7$, $p < 0.001$], and set-size effects were larger for target-absent than target-present trials [$F(2, 18) = 7.8$, $p = 0.004$]. The variation in set size allowed us to derive the function relating RT to set size. The slope of this function measures the cost for adding additional items to the display and is often interpreted as "search efficiency," with steeper slopes indicating slower, less efficient search.

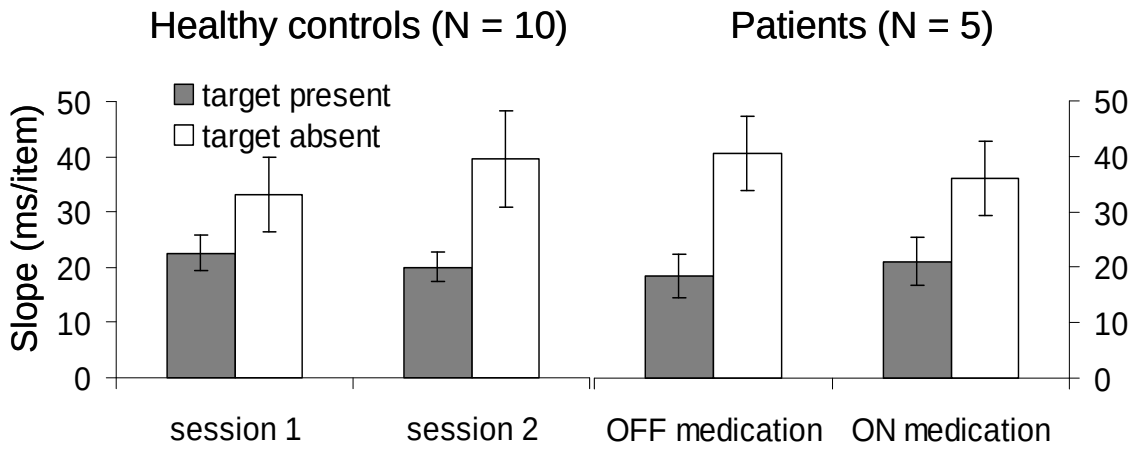


Figure 3. Average visual-search slopes for the control group and the patient group, as a function of target presence and session (error bars are standard errors of the means).

Figure 3 shows the average slopes for the control group and the patient group, as a function of target presence and session. The average slopes in the patient group were very similar to those in the control group, both ON and OFF medication. In the OFF-medication session, none of the patients' slopes deviated significantly from the control group (all *t*s(9) < 1.2; *p*s > 0.13; Table 4; see Supplementary Figure 5 for the individual slopes). In addition, the effects of medication did not differ significantly from the control group's practice effect in any of the patients (all *p*s > 0.11; Table 4). These results indicate that the patients had normal visual search efficiency, both ON and OFF medication.

The full factorial analysis of the effects of target presence, set size and session in the control group is reported in Appendix II and in Supplementary Figure 4.

The P3 component of the electroencephalogram

P3 amplitudes were maximal at electrode Pz in both the control group and the patient group; hence we focused our analyses on this electrode position. Figure 4 shows the grand average waveforms for standard and target stimuli in the visual and auditory oddball task, for the control group and the patient group ON and OFF medication. As expected, P3s were much larger for target stimuli than for standard stimuli. Figure 5 shows the P3 amplitudes of the individual participants.

When OFF medication, patient 5 showed a significantly smaller P3 amplitude than the control group in both the auditory and the visual oddball task, and patient 4 showed a significantly smaller P3 amplitude than the control group in the visual oddball task only (Table 4). For the other patients, P3 amplitude did not differ significantly from the control group. The effect of medication on P3 amplitude did not differ significantly from the control group's test-retest effect in any of the patients (all *p*s > 0.19; Table 4). These findings suggest that some but not all patients showed a P3 that was smaller than the P3 in the normal population, independently of whether they were ON or OFF medication.

The analyses of target-detection performance (RT and accuracy) are reported in Appendix II and in Supplementary Figure 6.

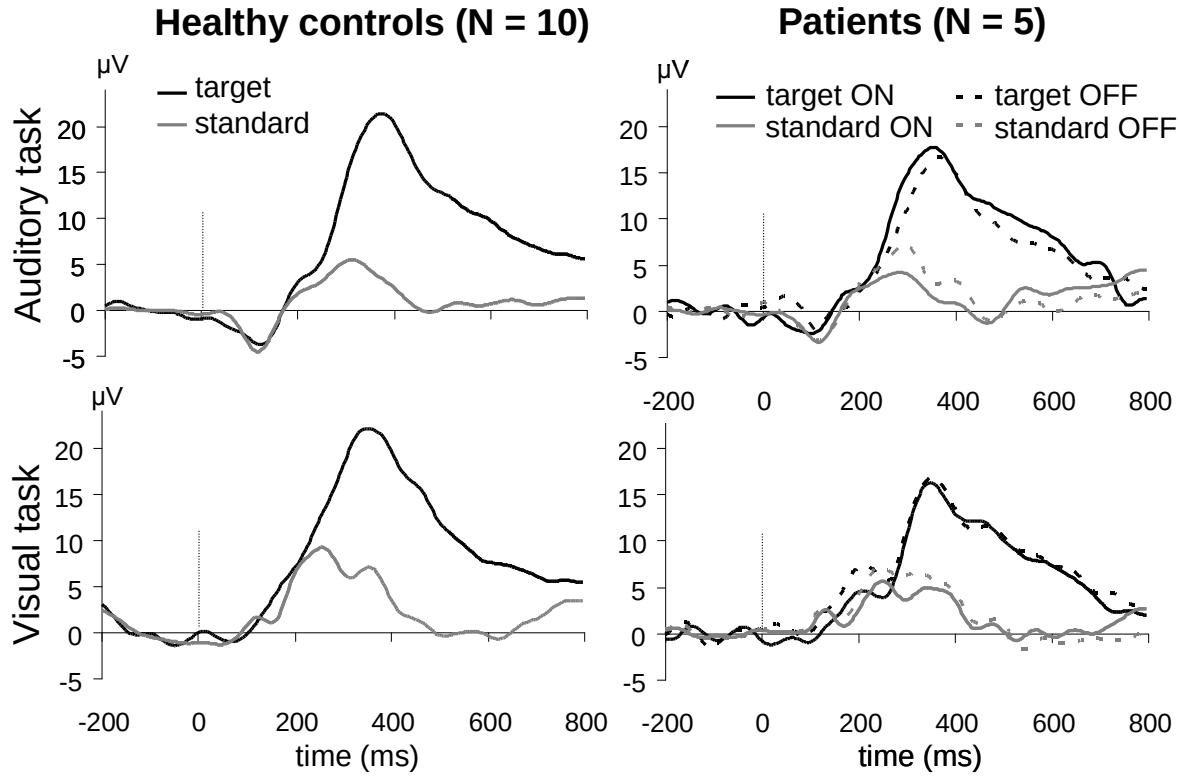


Figure 4. Grand-average waveforms for electrode Pz for the control group and the patient group, time-locked to the onset of the target and standard stimuli, in the auditory and visual oddball tasks. Because P3 amplitude in the control group did not differ across sessions [$F(1, 9) = 0.1$, $p = 0.72$], the data for the control participants are averaged across the two sessions.

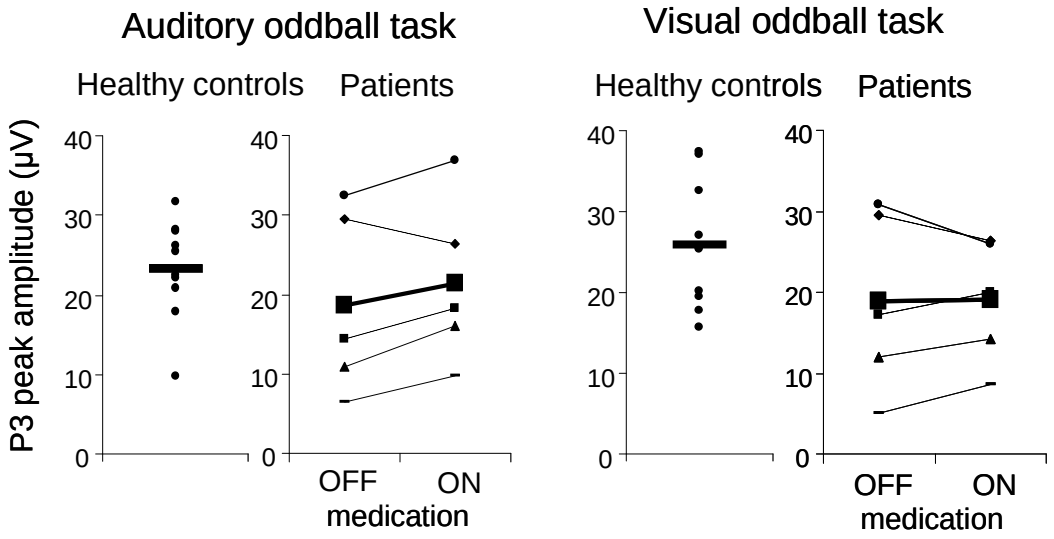


Figure 5. P3 amplitudes for the control participants and the patients in the auditory and visual oddball task. The bold lines indicate the grand average amplitudes, and the thinner lines and points indicate the amplitudes of each individual participant. Because there was no effect of session in the control group, the data for the control participants are averaged across the two sessions.

Pupil diameter during the pitch-discrimination task

The average baseline pupil diameter in the control group was 3.86 mm (SD = 0.56), and did not differ across the two sessions [$t(7) = 0.31$, $p = 0.77$]. When OFF medication, patient 2 had significantly smaller pupils than the control group. Patient 5 had significantly larger pupils than the control group which was due to a genetic deficit unrelated to her DβH deficiency (Erez et al., 2010). The other patients' baseline pupil diameter did not differ significantly from the control group (Table 4; see Supplementary Table 1 for each patient's baseline pupil diameter). Remarkably, patient 4 had significantly smaller pupils when he was ON compared to OFF medication. For the other patients, there was no significant effect of medication on baseline pupil diameter (Table 4).

We next assessed the magnitude of the task-evoked pupil dilations. As expected, all control participants showed a substantial pupil dilation following the comparison tone (average pupil dilation = 0.16 mm; SD = 0.04). Pupil dilation in the control group was not significantly affected by session [$F(1, 7) = 2.3$, $p = 0.17$] or tone-discrimination difficulty [$F(3, 21) = 2.4$, $p = 0.09$]. Figure 6 shows the time course of the grand-average pupil dilation following the comparison tone, for the control group and the patient group ON and OFF medication. When OFF medication, all but one patient showed significantly smaller task-evoked pupil dilations than the control group (see Supplementary Table 1 for each patient's average pupil dilation). Remarkably, patient 4 showed a significantly smaller pupil dilation when ON compared to OFF medication. The pupil dilation of patient 3 was also significantly affected by medication, but this result must be interpreted with caution because this patient's pupil dilations were negative in both sessions. For the other patients, there was no significant effect of medication on the task-evoked pupil dilation (Table 4).

The analyses of tone-discrimination performance (RT and accuracy) are reported in Appendix II and in Supplementary Figure 7.

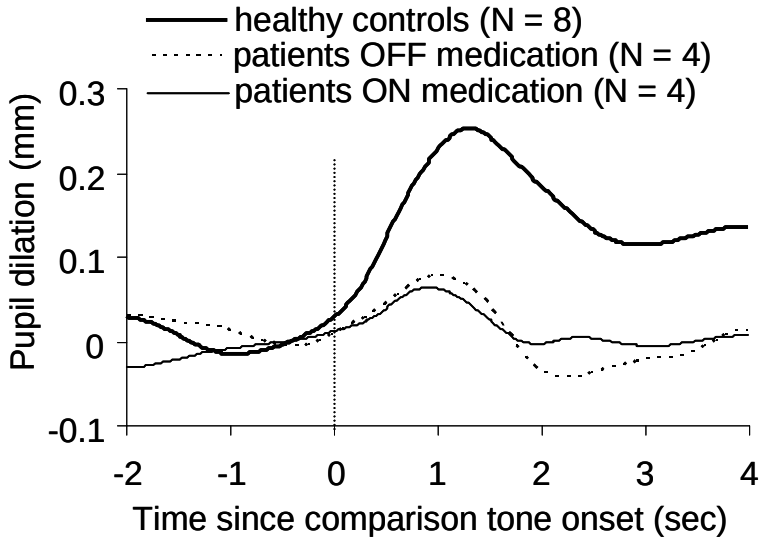


Figure 6. Time course of the grand-average pupil dilations in response to the comparison tone, for the control group and the patient group ON and OFF medication

Brain structure

Table 3 shows the average total brain volumes and the percentages of grey matter, white matter and cerebrospinal fluid (CSF) in the patient group and the control group, separately for the male and female participants. Four of the five patients had a smaller total brain volume than the control group. However, the proportions of grey matter, white matter and CSF did not differ from the control group in any of the patients (Table 4; see Supplementary Table 1 for the data of the individual patients).

The voxel-based morphometry analysis (Appendix I) revealed no significant topographic differences in grey matter volume between the patient group and the control group. The TCFE-corrected *p*-values for both the controls > patients contrast and the patients > controls contrast were larger than 0.34 in all voxels, suggesting that there were no trends for a group difference in grey matter distribution in any brain region. Together, these results suggest that most of the patients had an overall smaller brain than the control group, but that this difference was not confined to a specific tissue type or brain region.

Table 3. Whole-brain volume and percentage of grey matter, white matter and cerebrospinal fluid for the control group and the patient group, separately for the male and female participants (means ± standard deviations).

	Control group		Patient group	
	Men (N = 4)	Women (N = 5)	Men (N = 2)	Women (N = 3)
Brain volume (dm^3)	1.72 ± 0.05	1.46 ± 0.05	1.51 ± 0.06	1.30 ± 0.11
% grey matter	47.2 ± 0.7	44.1 ± 1.8	47.0 ± 1.5	44.7 ± 3.2
% white matter	38.7 ± 1.2	39.3 ± 2.0	38.5 ± 0.5	38.9 ± 2.2
% CSF	14.1 ± 1.7	16.5 ± 2.4	14.5 ± 1.0	16.4 ± 1.4

Notes: we did not collect MRI data from one female control participant; CSF = cerebrospinal fluid

Table 4. For each critical effect/measure, the *p* value reflecting the significance of the difference between each patient's OFF medication score and the average score of the control group (Crawford and Howell, 1998), and the *p* value indicating the significance of the deviation of each patient's medication effect from the control group's practice effect (Crawford and Garthwaite, 2006). *p* values < 0.05, which indicate that the estimated percentage of the normal population that would show a more extreme effect is smaller than 5%, are bold-faced.

	patient				
Patient OFF medication vs. control group	1	2	3	4	5
Emotional-interference effect on RT in target-present trials	-	0.44	0.25	0.29	0.41
Attentional-blink size	0.051	-	0.19	0.10	0.38
Visual search efficiency in target-present trials	0.36	0.15	0.14	0.15	0.41
Visual search efficiency in target-absent trials	0.41	0.23	0.36	0.24	0.50
P3 amplitude auditory oddball task	0.10	0.34	0.16	0.09	**0.04**
P3 amplitude visual oddball task	0.27	0.32	0.13	**0.048**	**0.01**
Baseline pupil diameter	-	**0.03**	0.45	0.22	**0.002**$^{+}$
Pupil dilation response	-	**0.03**	**0.003**	0.21	**0.001**
Brain volume (dm^3)	0.29	**0.006**	**0.03**	**0.01**	**0.02**
% grey matter	0.46	0.16	0.16	0.10	0.053
% white matter	0.33	0.39	0.33	0.46	0.12
% cerebrospinal fluid	0.39	0.30	0.45	0.30	0.35
Patient's medication effect vs. control group's practice effect	1	2	3	4	5
Emotional-interference effect on RT in target-present trials	-	0.19	0.09	0.18	0.19
Attentional-blink size	**0.045**	-	**0.003**	**0.049**	0.24
Visual search efficiency in target-present trials	0.45	0.28	0.28	0.28	0.39
Visual search efficiency in target-absent trials	0.22	0.11	0.18	0.50	0.16
P3 amplitude auditory oddball task	0.19	0.44	0.41	0.38	0.19
P3 amplitude visual oddball task	0.21	0.27	0.29	0.36	0.35
Baseline pupil diameter	-	0.25	0.20	**0.003**	0.21
Pupil dilation response	-	0.34	**0.03**	**0.04**	0.08

- = no data were collected; $^{+}$ this patient had significantly larger pupils than the control group, which was due to a genetic defect unrelated to DBH deficiency: a deletion on the short arm of chromosome 11 (Erez et al., 2010)

Discussion

The present study was the first systematic investigation of neurocognitive function in DβH deficiency. We tested five DβH-deficient patients and a matched healthy control group on a comprehensive cognitive task battery. In addition, we examined whether the patients differed from the control group with regard to the P3 component of the electroencephalogram, pupil dynamics and brain structure.

The patients' performance on most cognitive tasks did not differ substantially from the healthy control group, irrespective of whether they were ON or OFF DOPS medication. More specifically, the patients showed normal visual-search efficiency, tone-discrimination performance and target-detection performance, and a normal emotional-interference effect. In addition, we found an intact P3 component in most patients. Since DOPS medication effectively ameliorates DβH-deficient patients' orthostatic hypotension, medication-related changes in blood pressure and consequent effects on fatigue and affective state are important factors to take into account when comparing the patients' performance ON versus OFF medication. However, it is unlikely that these factors were responsible for the lack of medication effects on cognitive performance, for the following reasons. First, potential effects of fatigue or other physical symptoms on task performance would predict impaired performance when patients were OFF relative to ON medication, which was not found in most tasks. Second, the patients reported no substantial differences in affective state between the two sessions (Supplementary Table 3). Third, the critical measures in our cognitive tasks were difference scores (i.e. differences between task conditions), hence general medication-related effects on performance would cancel out in these difference scores.

The only cognitive function that was affected in the patients OFF medication was attentional selection in the temporal domain, as reflected by an increased attentional blink (i.e., impairment in processing the second of two target stimuli that are presented in close temporal succession). The attentional blink has not only been associated with NE (De Martino et al., 2007; Nieuwenhuis et al., 2005a; Warren et al., 2009), but also with dopamine (DA; Colzato et al., 2008); Colzato et al. have provided indirect evidence that higher DA levels are associated with a smaller attentional blink. Because DβH-deficient patients do not convert DA to NE, they are not only characterized by a lack of NE but also by increased DA levels (Man in 't Veld, 1987a), and DOPS medication both increases NE levels and reduces the excessive DA levels (Man in 't Veld et al., 1987b; Thomas et al., 1998). Thus, based on the patients' DA levels, it would be predicted that the patients OFF medication would show a smaller attentional blink than the healthy control group, and that the patients would show a smaller attentional blink OFF medication than ON medication. Since the opposite effects were found, this strongly suggests that the increased attentional blink in the patients OFF medication was due to the absence of NE rather than the excess of DA.

The largely spared neurocognitive function in the DβH-deficient patients is remarkable given the large body of evidence suggesting that the LC-NE system plays an important role in many

aspects of neurocognitive function (for recent reviews see Robbins and Arnsten, 2009; Sara, 2009). For example, individual differences in noradrenergic genotype in the normal population are predictive of performance on cognitive tasks measuring attention (Greene et al., 2009) and working memory (Parasuraman et al., 2005), and have been related to vulnerability to several psychiatric disorders (e.g., Cubells and Zabetian, 2004; Roman et al., 2002). In addition, DβH-knockout mice that lack NE due to a targeted disruption of the DβH gene show several behavioral deficits, including impairments in active-avoidance learning (Thomas and Palmiter, 1997a), memory retrieval (Murchison et al., 2004), and maternal and social behavior (Marino et al., 2005; Thomas and Palmiter, 1997b). Finally, pharmacological, neurophysiological, and lesion studies in animals suggest that the LC-NE system plays a crucial role in regulating the optimization of behavioral performance (e.g., Aston-Jones and Cohen, 2005; Bouret and Sara, 2005). It must be noted, however, that our task battery did not address all aspects of cognitive function. For example, we did not assess higher-level cognitive functions such as executive control and exploratory behavior. Therefore, our results leave open the possibility that the patients have subtle cognitive deficits that were not revealed by our task battery. In addition, although our data clearly indicate that there were no substantial abnormalities in the patients' performance on our test battery, it cannot be excluded that there were some subtle differences which failed to reach significance due to a lack of power of our experimental design.

Although the patients' relatively normal performance on our cognitive task battery is striking, it is consistent with informal clinical observations that DβH-deficient patients do not have obvious cognitive impairments or psychiatric disorders. Indeed, the absence of mental problems in most DβH-deficient patients that have been encountered so far has intrigued investigators in the areas of depression and schizophrenia (Cubells and Zabetian, 2004). It is especially remarkable that the patients OFF medication did not show impaired performance on cognitive tasks that are normally mediated by the LC-NE system (e.g., the emotional working-memory task), and showed a relatively intact P3 component, which is thought to reflect the noradrenergic potentiation of information processing (Liu et al., 2009; Nieuwenhuis et al., 2005b; Pineda et al., 1989). These findings suggest that alternative neural mechanisms and/or neuromodulatory systems compensate for the absence of NE in DβH-deficient patients. Previous findings that DβH-deficient patients have a relatively normal sleep pattern (Tulen et al., 1990; 1991), although the sleep-wake cycle is normally mediated by the LC-NE system (Hobson et al., 1986; Jouvet, 1969), are consistent with this idea.

Since DβH is responsible for the conversion of DA to NE, it is thought that DA rather than NE is stored and released by noradrenergic neurons in DβH-deficient patients. Indeed, plasma DA levels in DBH-deficient patients respond to various physiological and pharmacological manipulations that normally affect plasma NE levels (Man in 't Veld, 1987a; Robertson et al., 1986), although it remains to be determined whether this also applies to DA levels in the central nervous system. Thus, a possible explanation for the spared neurocognitive function in DβH deficiency is that DA has, to some extent, taken over the function of NE in the brains of DβH-

deficient patients. Obviously, a functional replacement of NE by DA would require the presence of postsynaptic receptors with DA affinity in noradrenergic synapses. Studies in mice suggest that some α2-adrenergic receptor subtypes have a comparable affinity for DA and NE (Zhang et al., 1999), whereas α1- and β-adrenegic receptors have a much lower affinity for DA than for NE (Zhang et al., 2004). However, since the congenital absence of NE may have altered the affinity of adrenergic receptors, it is unknown whether the same receptor characteristics apply to DβH-deficient patients. Another possible explanation for a functional replacement of NE by DA is that DβH-deficient patients have an increased density of postsynaptic DA receptors on noradrenergic synapses. A recent positron emission tomography (PET) study in mice suggests that DβH knockout mice have a normal density of D2 dopamine receptors in the high-affinity state (Skinbjerg et al., 2010), which does not support this hypothesis. However, since results from DβH-knockout mice might not be generalizable to human DβH-deficient patients, the assessment of DA receptor densities in human DβH-deficient patients, for example using PET scanning, remains an important objective for future studies.

It is interesting to note that the first study that used gene targeting to produce DβH-deficient mice found that the majority of DβH-deficient embryos died in mid-gestation and only 5% reached adulthood (Thomas et al., 1995). To prevent embryonic lethality, subsequent studies using DβH-knockout mice have supplied the embryos with adrenergic agonists (isoproterenol and phenylephrine) and DOPS via the maternal drinking water, such that NE is present in the DβH-knockout mice until birth. Thomas et al.'s (1995) results suggest that the human DβH-deficient patients may represent the minority of DβH-deficiency cases that have survived this condition. If this is true, an interesting speculation is that these patients were able to survive because they happened to have optimal dopaminergic or noradrenergic genotypes to compensate for the absence of NE. Future studies might assess this possibility by examining whether the frequency of occurrence of specific alleles of dopaminergic and noradrenergic genes (e.g., the COMT, DAT, and the dopamine and noradrenergic receptor genes) in DβH-deficient patients deviates from those in the normal population.

In contrast to the generally normal neurocognitive function in the DβH-deficient patients, we did find clear abnormalities in their task-evoked pupil dilation response. The task-evoked pupil dilation was very small or absent in most of the patients, which might be due to a decreased noradrenergic innervation of the iris dilator muscle. However, it is also possible that the abnormal pupil dynamics in some of the patients resulted from ocular abnormalities unrelated to their DBH deficiency; this might explain why the pupil-dilation response was not restored by DOPS medication. Importantly, the patients' small or absent task-evoked pupil dilations did not reflect a decreased processing of the task-related stimuli, since their performance on the tone-discrimination task during which their pupils were measured was not impaired.

The patient group also differed from the control group with regard to total brain volume: all but one patient had a significantly smaller brain volume than the control group, but the relative proportions of grey matter, white matter and cerebrospinal fluid, and the distribution of grey matter

volume across the brain did not deviate from those in the control group. The smaller brain volume in most DBH-deficient patients is in line with recent findings suggesting that NE has a neurotrophic effect on cortical neurons (e.g., Counts and Mufson, 2010; Kalinin et al., 2007; Madrigal et al., 2007, 2009). Apparently, the patients' decreased brain volume did not result in cognitive impairments; this suggests that although the patients have a smaller number of neurons, their neurons are intact and make proper connections.

To conclude, our findings suggest that neurocognitive function in human DβH-deficient patients is largely spared, even when they are OFF medication, but that their total brain volume is smaller than that of the normal population. The normal neurocognitive function in DβH-deficient patients is striking given the important role of NE in normal cognition, but corroborates informal clinical observations that most patients do not have obvious cognitive impairments. Our findings suggest that DβH-deficient patients have developed alternative mechanisms to compensate for the absence of NE in the brain, possibly through a functional replacement of NE by DA; the nature of these compensatory mechanisms remains to be explored by future studies.

Appendix I: 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° x 1.2° 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 ± SE: valence 2.4 ± 0.8, arousal 6.6 ± 0.4), the other half were emotionally neutral (M ± SE: valence 5.1 ± 0.6, arousal 3.3 ± 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° x 1.2° 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 0.5×0.5°, presented for 1s 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 0.9×0.9°. 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 0.9×0.9° 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 6×6 matrix (8.7° horizontally × 9.6° 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 (0.7×1.3°) 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 x 1.7°) 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 μV/2 ms] and slow drifts [200 μV/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 = 8º, 140 slices). The voxel size was 0.88 x 0.88 x 1.2 mm^3.

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 (Willemsen 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 gluthatione (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.

Appendix II: 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 DβH deficiency				diabetic	irregular and large pupils+
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^
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#	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^3)	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; + due to a genetic defect unrelated to DβH deficiency: a deletion on the short arm of chromosome 11; * these two patients had never been on DOPS medication before the study; ^ this patient had consumed 7 doses of 300 mg before she was tested ON medication, and was feeling normal at that time; # 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

	Plasma NE		Urine NE		Plasma DA		Urine DA	
patient	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 (ps > 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 DβH-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 ± standard deviations)

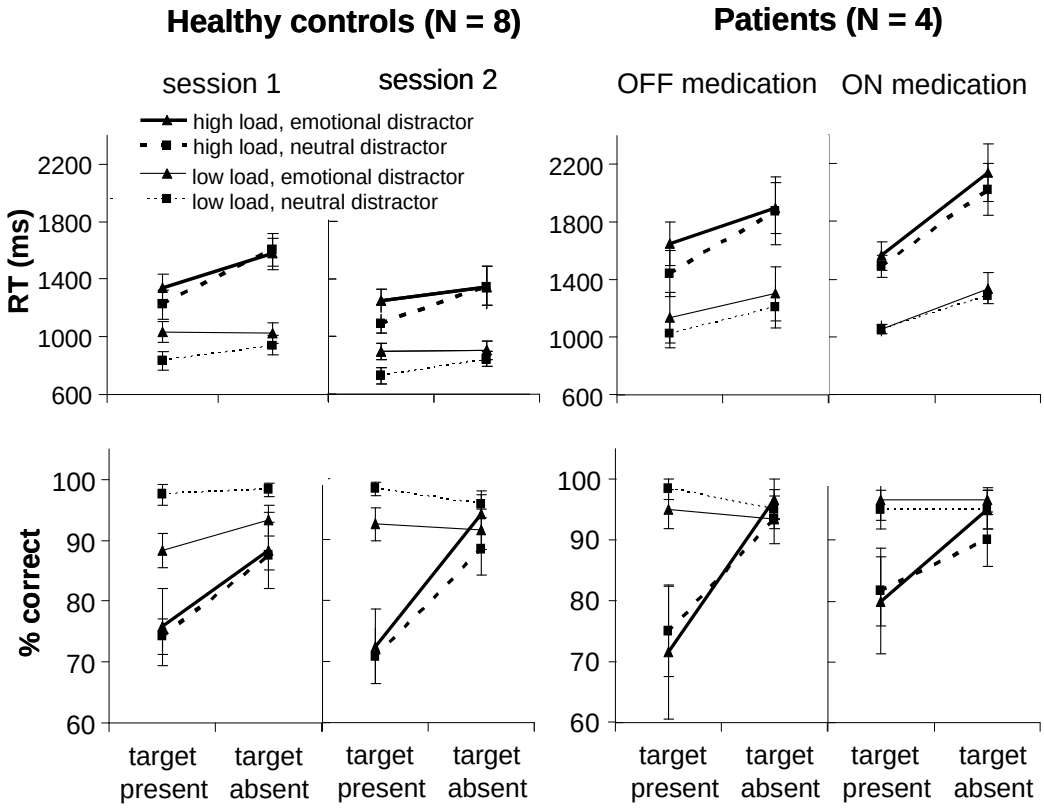
		Control group (N = 10)		Patient group (N = 5)	
		Session 1	Session 2	ON	OFF
Positive affect	beginning	3.0 ± 0.3	2.7 ± 0.4	2.7 ± 0.6	2.5 ± 0.3
	end	2.8 ± 0.5	2.6 ± 0.5	2.4 ± 0.6	2.5 ± 0.5
Negative affect	beginning	1.2 ± 0.3	1.1 ± 0.2	1.2 ± 0.3	1.4 ± 0.6
	end	1.2 ± 0.2	1.1 ± 0.2	1.2 ± 0.2	1.1 ± 0.2

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

Emotional working-memory performance

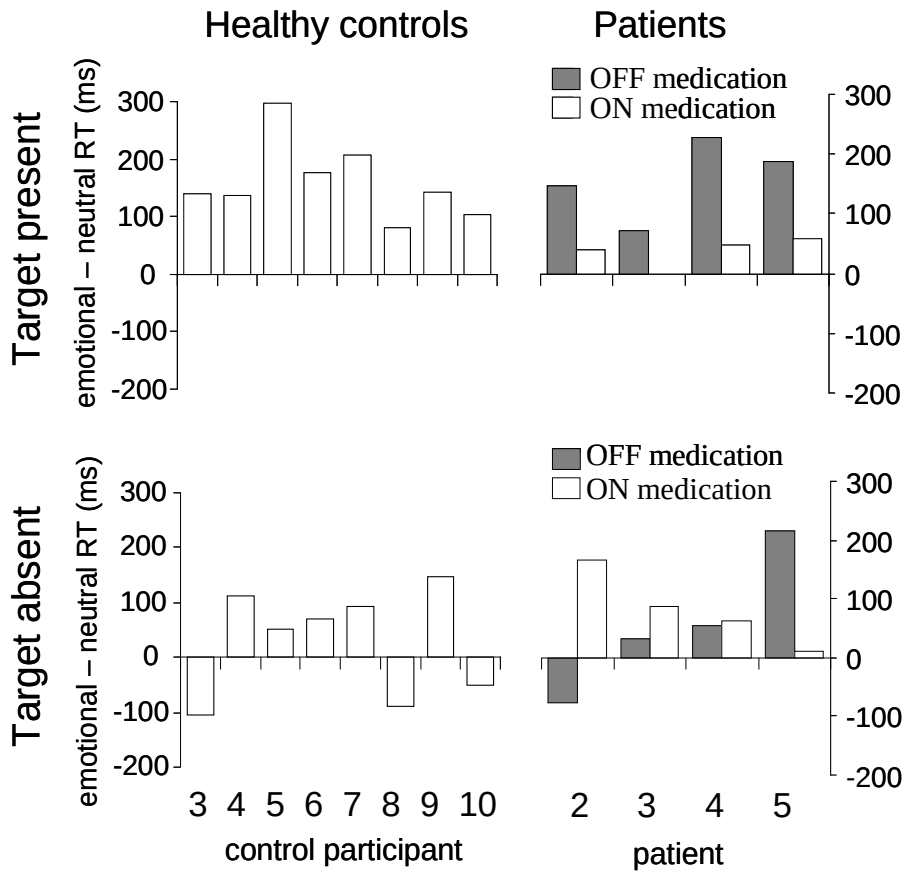
A repeated-measures ANOVA on RT in the control group yielded significant main effects of working-memory load [$F(1, 7) = 65.7$, $p < 0.001$], target presence [$F(1, 7) = 8.9$, $p = 0.02$] and distractor type [$F(1, 7) = 14.7$, $p = 0.006$]. There was a main effect of session as well [$F(1, 7) =$

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.

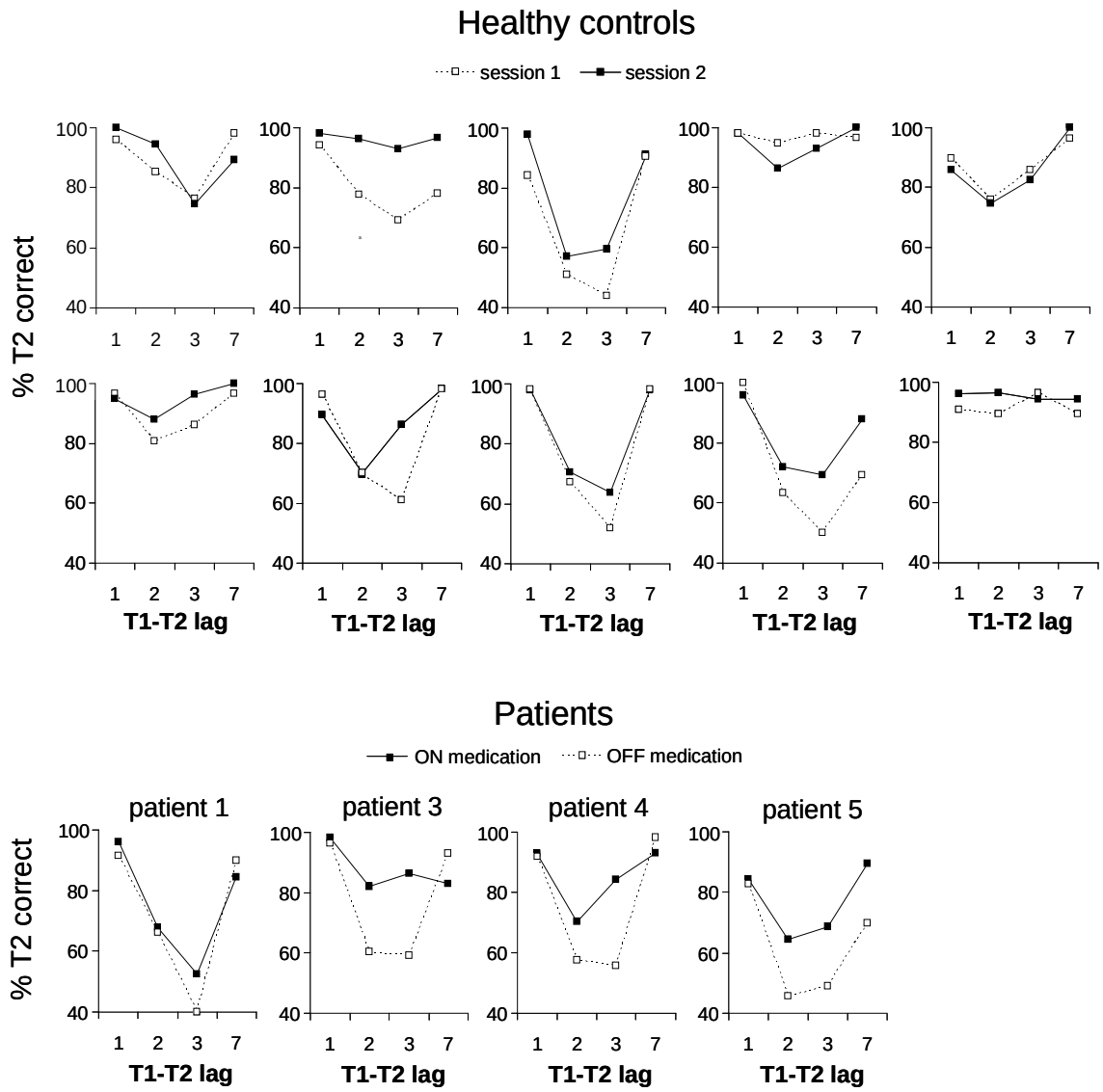


Supplementary Figure 2. Each individual participant's emotional-interference effect on RT as a function of target presence. Because session did not interact with distractor type or target presence in the control group, the data from the control participants are averaged across sessions.

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 ps > 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 [$F(3, 27) = 2.1$, $p = 0.13$].

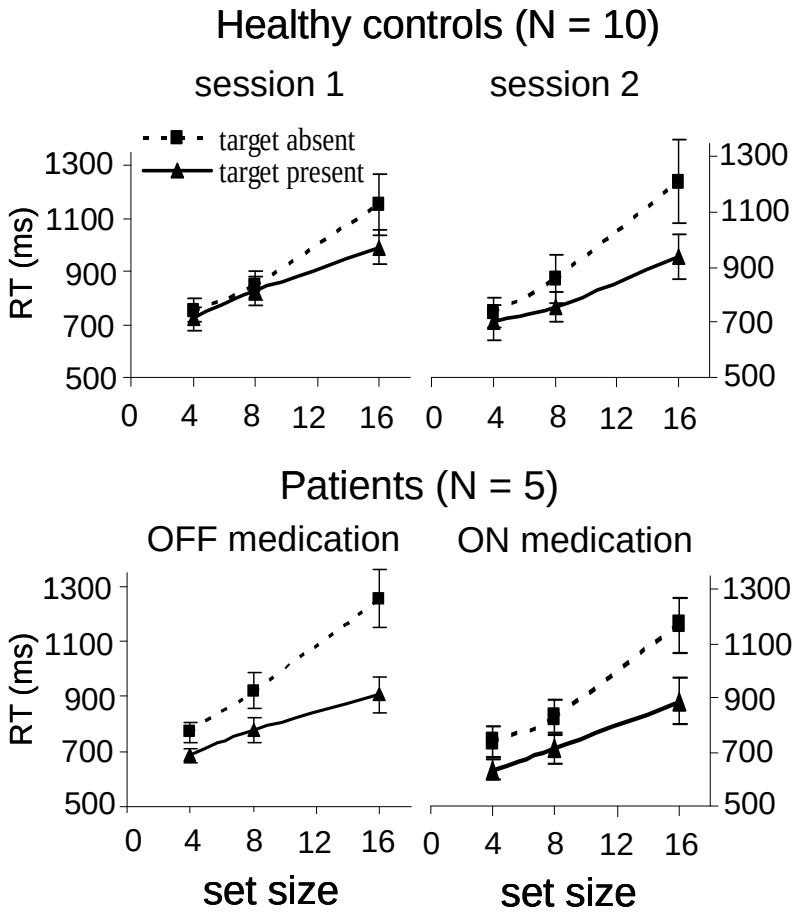


Supplementary Figure 3. Each individual participant's T2 accuracy as a function of lag and session.

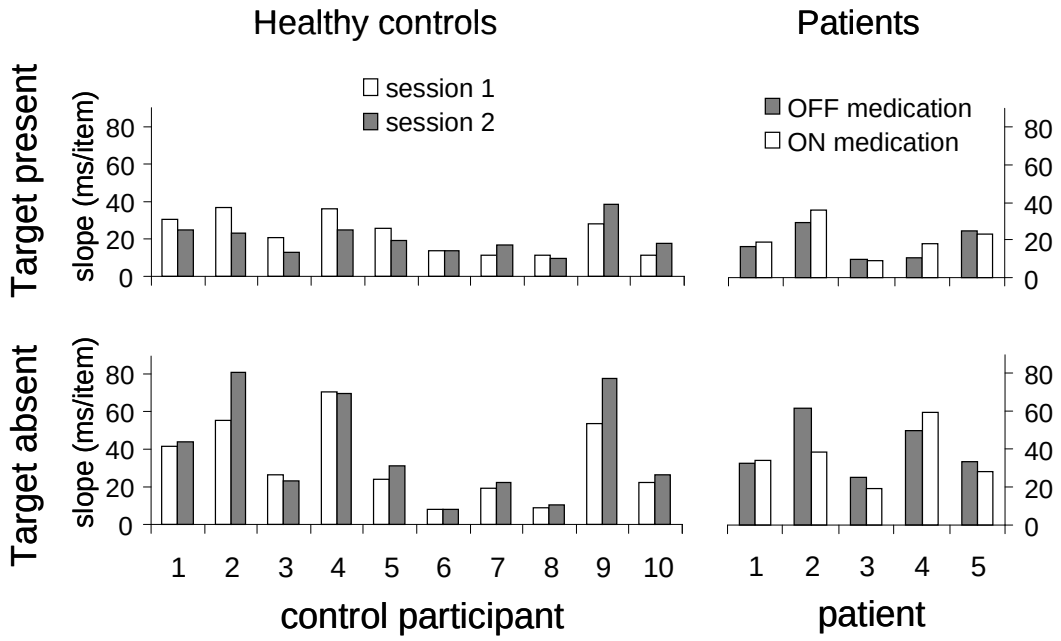
Visual-search performance

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.

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.

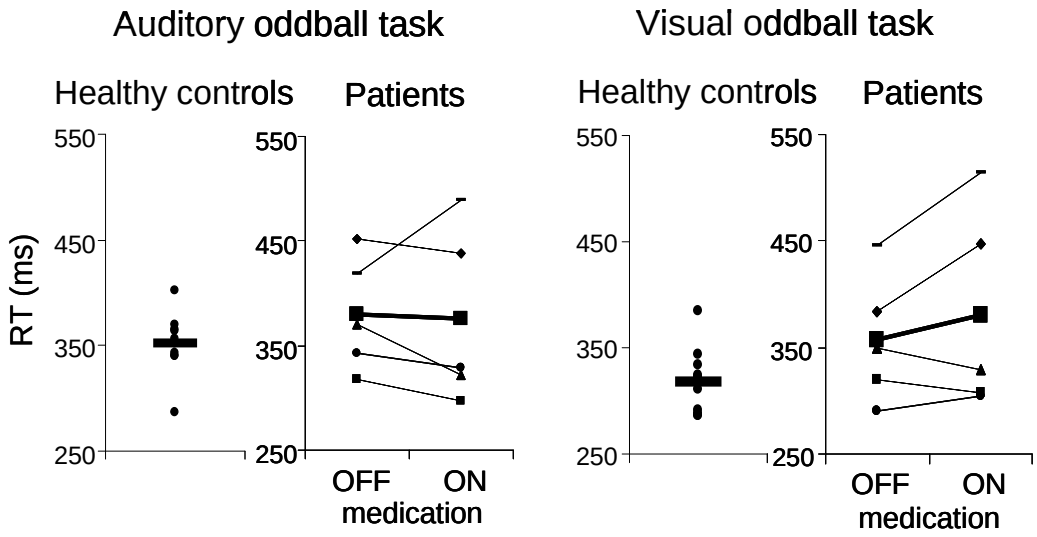


Supplementary Figure 4. Average correct RT for the control group and the patient group in the visual-search task as a function of target presence, set size and session (error bars are standard errors of the mean).



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

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.

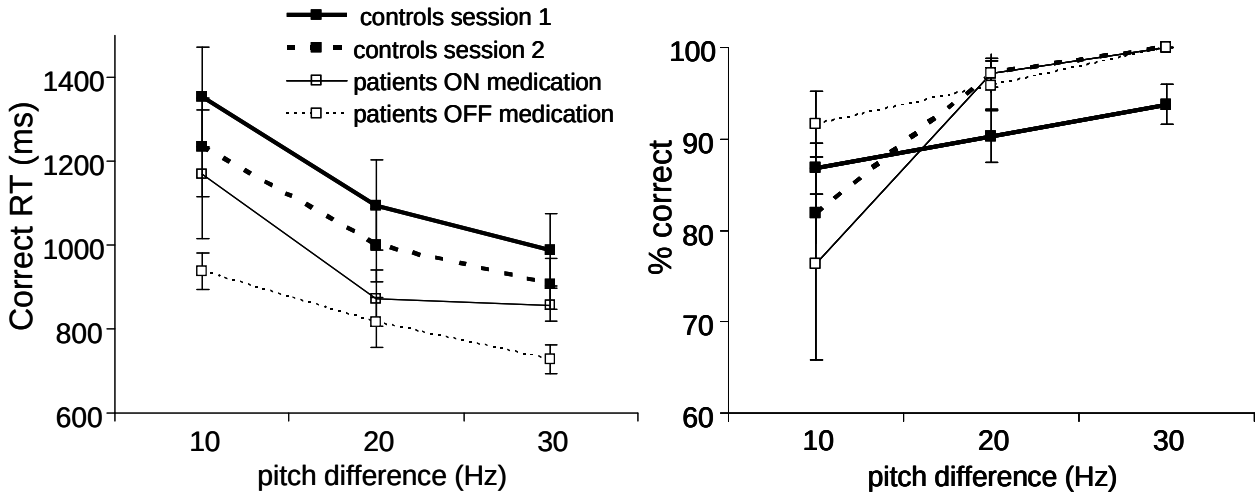
Target-detection RT in the control group did not differ across sessions [$F(1, 9) = 0.25$, $p = 0.63$], but was shorter in the visual task than in the auditory task [$F(1, 9) = 10.6$, $p = 0.01$].

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)