Current Biology

A causal role for the pedunculopontine nucleus in human instrumental learning

--Manuscript Draft--

Manuscript Number:	CURRENT-BIOLOGY-D-20-01052R2
Full Title:	A causal role for the pedunculopontine nucleus in human instrumental learning
Article Type:	Research Article
Corresponding Author:	Mathias Pessiglione INSERM PARIS, FRANCE
First Author:	Vasilisa Skvortsova, Ph.D.
Order of Authors:	Vasilisa Skvortsova, Ph.D.
	Mathias Pessiglione
	Stefano Palminteri
	Anne Buot
	Carine Karachi
	David Grabli
	Marie-Laure Welter
Abstract:	Summary A critical mechanism for maximizing reward is instrumental learning. In standard instrumental learning models, action values are updated on the basis of reward prediction errors (RPE), defined as the discrepancy between expectations and outcomes. A wealth of evidence across species and experimental techniques has established that RPE are signaled by midbrain dopamine neurons. However, the way dopamine neurons receive information about reward outcomes remains poorly understood. Recent animal studies suggest that the pedunculopontine nucleus (PPN), a small brainstem structure considered as a locomotor center, is sensitive to reward and sends excitatory projection to dopaminergic nuclei. Here, we examined the hypothesis that the PPN could contribute to reward learning in humans. To this aim, we leveraged on a clinical protocol that assessed the therapeutic impact of PPN deep brain stimulation (DBS) in three patients with Parkinson's disease. PPN local field potentials (LFP), recorded while patients performed an instrumental learning task, showed a specific response to reward outcomes in a low frequency (alpha-beta) band. Moreover, PPN DBS selectively improved learning from rewards but not from punishments, a pattern that is typically observed following dopaminergic treatment. Computational analyses indicated that the effect of PPN DBS on instrumental learning was best captured by an increase in subjective reward sensitivity. Taken together, these results support a causal role for PPN-mediated reward signals in human instrumental learning.

Summary:

A critical mechanism for maximizing reward is instrumental learning. In standard instrumental learning models, action values are updated on the basis of reward prediction errors (RPE), defined as the discrepancy between expectations and outcomes. A wealth of evidence across species and experimental techniques has established that RPE are signaled by midbrain dopamine neurons. However, the way dopamine neurons receive information about reward outcomes remains poorly understood. Recent animal studies suggest that the pedunculopontine nucleus (PPN), a small brainstem structure considered as a locomotor center, is sensitive to reward and sends excitatory projection to dopaminergic nuclei. Here, we examined the hypothesis that the PPN could contribute to reward learning in humans. To this aim, we leveraged on a clinical protocol that assessed the therapeutic impact of PPN deep brain stimulation (DBS) in three patients with Parkinson's disease. PPN local field potentials (LFP), recorded while patients performed an instrumental learning task, showed a specific response to reward outcomes in a low frequency (alpha-beta) band. Moreover, PPN DBS selectively improved learning from rewards but not from punishments, a pattern that is typically observed following dopaminergic treatment. Computational analyses indicated that the effect of PPN DBS on instrumental learning was best captured by an increase in subjective reward sensitivity. Taken together, these results support a causal role for PPN-mediated reward signals in human instrumental learning.

Keywords: reinforcement learning, reward prediction error, deep brain stimulation, local field potentials, peduncolopontine nucleus, dopamine, Parkinson’s disease

Introduction

A fundamental mechanism through which animals learn to optimize their actions is instrumental learning. According to this mechanism, action values are updated on the basis of reward prediction errors (RPE), defined as the discrepancy between obtained and expected outcomes[1]. Decades of single-cell recording studies in non-human primates have demonstrated that RPE are signaled by midbrain dopamine neurons[2–5]. The link between RPE and dopamine has also been documented in humans, using electrophysiological or hemodynamic recording of dopaminergic nuclei[6,7], and pharmacological interference with dopaminergic transmission[8,9].

Decisive advances on this question have come from recent experiments in rodents. Optogenetic techniques combined with electrophysiology in mice have shown that RPE signals in dopaminergic nuclei 1) are indeed emitted by dopamine neurons, 2) represent a similar subtractive function of outcome and expectation across neurons, and 3) are sufficient to trigger learning of associations between cues and outcomes[10–14]. However, less is known about the way dopamine neurons could possibly be informed about the components of RPE estimates – the representations of actual and expected rewards[15]. In particular, many investigations have focused on how information about reward, at the time of outcome, could be conveyed to dopamine neurons.

One promising candidate for an excitatory input to dopaminergic nuclei is the peduncolopontine nucleus (PPN)[16,17]. Traditionally, the PPN, located in the upper part of the lateral mesencephalon, is considered as a part of the mesencephalic locomotor area[18–20], together with the cuneiform nucleus (CN). More recently, the role of the PPN in reward learning, and its tight connection with dopaminergic nuclei, have come into light. Indeed, the PPN sends glutamatergic and cholinergic projections to the ventral tegmental area (VTA) and substantia nigra pars compacta (SNpc) that are susceptible to activate dopamine neurons[21–25]. These afferent projections from the PPN to dopamine neurons were found to carry information about reward[26–29]. Moreover, activation of PPN cholinergic and glutamatergic neurons contributes to the reinforcement of actions[30,31], which is impaired by PPN inactivation or lesion[32–35].

While the case for the PPN contributing to reward learning (via dopamine neurons) is well established in animals (mostly rodents), it has not been examined so far in humans. The reason is that the PPN is a tiny midbrain group of neurons difficult to locate, access, record

or manipulate with indirect, non-invasive techniques [20]. Yet this opportunity has been offered by clinical trials assessing the therapeutic effect of PPN deep brain stimulation (DBS) on the motor symptoms of Parkinson's disease (PD). By stimulating remaining neurons, the aim of DBS is to compensate for the partial PPN degeneration observed in PD[36,37]. The rationale for a beneficial impact on motor symptoms is two-fold: it might first activate remaining dopamine neurons, through the connections from PPN to VTA and SNpc, and second improve gait and postural disorders, due to its specific role in locomotion[38–41]. For the latter reason, PD patients recruited for PPN DBS trials suffer from the so-called axial symptoms (gait and postural disorders) that are poorly improved with dopatherapy or subthalamic nucleus stimulation, currently the standard target for DBS in PD[42–44]. The first clinical trials indeed reported some improvement of falls and freezing of gait in PD patients treated with PPN DBS[39,45,46].

Here, we leveraged on a clinical trial conducted in Paris[47], to record and stimulate the PPN of three PD patients performing instrumental learning tasks. Our working hypothesis was that the PPN might contribute to instrumental learning by providing reward-related information to dopamine neurons. Two predictions at least could be derived from this hypothesis: 1) PPN activity should signal reward outcomes and 2) PPN DBS should improve reward learning, in the same way as dopatherapy would do.

To test the first prediction, we recorded PPN local field potentials (LFP) while patients performed a first learning task, during the week following the implantation of DBS electrodes (Figure 1A, left). This first task was previously used in fMRI studies to distinguish between neural representations of choices (left vs. right hand movement) and their outcomes (reward vs. no reward)[48]. To test the second prediction, we examined choice behavior in a second learning task that patients performed several months after the surgery, once in the "on-DBS" and once in the "off-DBS" state (Figure 1A, right). This second task was previously used in pharmacological studies to specify the effect of dopaminergic drugs, as a selective improvement of reward learning, not punishment learning[9].

Results

PPN hemodynamic activity signals reward outcomes

To explore whether the PPN region signals reward-related events, before turning to LFP in PD patients, we reanalyzed an fMRI dataset collected in healthy volunteers performing learning task 1, as reported previously[48]. This task involved choosing between two cues with either a left-hand or right-hand button press, and learning the probabilistic associations with monetary reward from the observation of choice outcomes (Figure 2A and Figure S1AB). It was thus designed to disentangle between neural representations of choice (left vs. right cue) and outcome (reward vs. no reward receipt). It had previously been used to dissociate neural correlates of reward and movement in both young and aged volunteers, and to assess learning performance in several clinical populations, including PD patients[48,51,52].

Time series of fMRI activity were regressed against a GLM with the moment of choice split between two regressors (left vs. right response) and the moment of feedback split between four regressors (two responses, left vs. right, by two outcomes, reward vs. no reward). Individual contrast estimates were extracted within the bilateral anatomical PPN masks (Figure S1C) taken from a three-dimensional histological atlas of the basal ganglia[51,52]. Contrast estimates were analyzed using a 3-way repeated-measures ANOVA with outcome (reward vs. no reward), choice (left vs. right) and the side of PPN ROI (left vs. right) as three within-subject factors. Only the main effect of reward was significant ($F(1,19) = 12.7$, $p = 0.003$), not the main effect of choice side ($F(1,19) = 0.07$, $p = 0.79$) or ROI side $F(1,19) = 0.41$, $p = 0.53$). There was no detectable significant interaction involving the ROI side (all $p > 0.30$), suggesting that choice was similarly represented in the ipsi and contralateral PPN, and that reward similarly affected PPN region ipsi or contralateral to the response side.

We therefore pooled across the left and right PPN ROI for post-hoc analyses (Figure S1D). Consistently, we found that contrast estimates were significantly positive following reward outcomes ($t(19) = 3.13$, $p = 0.006$), but not after no reward outcomes ($t(19) = 0.88$, $p = 0.39$), with a significant difference between the two events ($t(19) = 3.58$, $p = 0.002$). Contrast estimates were also significantly positive at the time of choice, for both ipsilateral and contralateral responses (ipsilateral: $t(19) = 3.62$, $p = 0.0018$, contralateral: $t(19) = 3.33$, $p = 0.0035$), with no significant difference between the two ($t(19) = 0.254$, $p = 0.802$).

The results provided preliminary evidence that PPN activity discriminates between reward and no reward outcomes but not between low-level features such as spatial location on the screen (left vs. right cue) or motor response (left-hand versus right-hand movement). However, this result may not be very specific, given that many brain regions were also activated by reward in this task[48]. Moreover, the normalization to the MRI brain template is never accurate enough to ensure that the voxels covered by the anatomical mask really fall within PPN borders. To confirm these preliminary results, while overcoming the limitations inherent to fMRI, we performed similar analyses on LFP recorded in the PPN of PD patients during the same task.

PPN evoked potentials signal reward outcomes

PD patients performed learning task 1 while LFP were recorded from electrodes implanted in the PPN, during the week following surgery. Analysis of behavioral performance showed that the global pattern of responses was similar to previous observations (Figure 2B). Choice rate was 47.8/52.2, 41.3/58.7, 60.9/39.1, and 26.8/73.2% for the left/right responses to the four pairs of cues (respectively with 75/75, 25/25, 75/25, and 25/75% reward probability). This pattern indicates that in asymmetrical pairs, the cue with the higher probability of reward (75%) was more frequently selected on average (in 67.1% of responses). This average correct choice rate was higher than in age-matched healthy volunteers (62.0%, Figure 2B), but lower than in young healthy volunteers (77.9%, Figure S1B). When comparing directly the proportion of left/right responses between asymmetrical pairs (75/25 vs. 25/75%), the difference was significant at the group level ($p=0.032$, two-tailed t-test). However, at the individual level, the same test suggested significant learning in patients 2 and 3, but not in patient 1. Formal comparison between groups was precluded by the small sample sizes and the differences in testing conditions (PD patients were tested during LFP recording in the hospital room just after a major neurosurgery). Anyway, following on previous studies[50], our main analyses were focused on the response to reward outcomes, which does not depend on learning performance (contrary to the contrast between cues associated with low vs. high reward probability).

First, we tested whether LFP in the PPN would differ between left vs. right motor responses, and between reward vs. no reward outcomes, as we did in the fMRI data

analysis. LFP recordings were pooled across all contacts located within the PPN, left or right (Figure 1B). Fixed-effect analysis of time-frequency maps across PD patients indicated an increased activity around the time of choice (Figure 3A) in a low (alpha or beta) frequency band, which was also observed in each patient separately (Figure S2). However, there was no significant difference between ipsi and contralateral choice. On the contrary, a significant time-frequency cluster, again in a low-frequency band, was identified in the contrast between reward and no reward outcomes (Figure 3B). Inspection of individual time-frequency maps (Figure S3) revealed different patterns of activation in the different patients, with an overlap around the transition from alpha to beta band (10 to 20Hz).

Next, to better qualify this significant difference between reward and no reward outcomes, post-hoc tests were used to compare activity in this 10-20Hz band against zero. Results showed that the response to reward receipt, around 500ms following outcome onset, was significantly positive and survived a non-parametric cluster-wise correction for multiple comparisons (threshold at $p = 0.05$, two-tailed) both across all 15 contacts analyzed together (Figure 3C, left panel) and in each of the three patients (Figure 3C, right panel). The same test against zero, applied to trials ending with no reward outcomes, yielded no significant cluster of activity. We also checked that the difference between reward and no reward outcomes in the 10-20Hz band survived cluster-wise correction for multiple comparisons in each of the three patients.

Last, we performed additional regressions of PPN LFP activity against variants of a GLM containing factors susceptible to influence the response to outcomes. In the basic version used above, this GLM only contained an indicator regressor for reward vs. no reward outcomes. To examine whether the response to reward outcomes would depend on the choice made (left or right), we included in the GLM two additional regressors modeling the choice side (ipsi or contralateral to the recording contact) and the interaction between choice side and reward outcome. As with the other variants, the cluster signaling reward outcomes in time-frequency maps (Figure S4A) was virtually unchanged, while the interaction with choice side yielded no significant cluster. Thus, there was no evidence that the response to reward outcomes was dependent on the motor response.

Following a model-based approach, we added a regressor for reward expectation, to examine whether it would attenuate the response to outcomes, as it should do if the PPN

was signaling prediction errors. To model expected rewards, we fitted a basic Q-learning algorithm to behavioral choices (see Methods). This algorithm includes two functions: a learning function (delta rule), which updates cue values in proportion to outcome prediction errors and a choice function (softmax rule), which generates selection probabilities based on the difference between cue values. Reward expectation was modeled trial-by-trial as the expected value of the chosen option, generated using the fitted learning rate and the history of choices and outcomes. In time-frequency maps of regression estimates (Figure S4B), there was no significant cluster signaling reward expectation, while the cluster significantly signaling reward outcomes (vs. no reward) was similar to that obtained before (Figure 3B). Thus, only the outcome component of reward prediction errors (and not the expectation component) was reflected in PPN LFP activity.

Because the null result with reward expectation could have arisen from poor modeling of behavioral choices, we also tried a model-free approach. The prediction tested here is that if the PPN was signaling prediction errors, then responses to reward outcomes should have been diminished during the course of learning sessions, as these outcomes became more and more expected. We therefore included trial number and the interaction between trial number and reward outcome as additional regressors in the GLM. Again, time-frequency maps of regression estimates (Figure S4C) identified a significant cluster signaling reward outcomes very similar to that obtained before (Figure 3B), but no significant cluster related to the interaction with trial number. There was therefore no evidence that the response to reward outcomes was changed with learning. Altogether, these results suggest that the PPN is sensitive to reward outcomes in a rigid way that seems unaffected by the progressive building of expectations.

To recapitulate, LFP data analysis concurs with fMRI data analysis to show that PPN activity is sensitive to reward outcome in a learning task. However, this does not imply that PPN activity is causally involved in learning, i.e. in favoring choices that maximize reward outcome. To test this causal implication, we explored the effect of PPN DBS on choices in another instrumental learning task.

PPN DBS selectively enhances learning from reward outcome

The same patients performed a second learning task, which was previously[9,53] used to dissociate the neural correlates of reward- and punishment-based learning (Figure 4A). In this task, subjects select between abstract cues, by pressing a button for the top cue (“Go” response) or not pressing (“NoGo” response) for the bottom cue, and then learn contingencies from observing the outcome of their choice. Each session presents intermingled pairs of cues, associated with either 80/20% probability of monetary reward (1€) or 80/20% probability of monetary punishment (-1€). Thus, the task involves learning to maximize reward with some pairs, and to minimize punishment with others.

All patients performed three sessions of the task in the "on-DBS" state and three sessions in the "off-DBS" state (Figure 1A). The first session was considered as a practice session, following on previous analyses in PD patients[53]. A fixed-effect analysis of choice behavior was conducted across the six remaining sessions, using a two-way repeated-measures ANOVA with outcome valence (reward vs. punishment) and stimulation state (“on-DBS " vs. “off-DBS ”) as fixed factors. There were no main effects of outcome valence ($F(1,5) = 1.17$, $p = 0.33$) or stimulation ($F(1,5) = 3.03$, $p = 0.14$), but a significant valence by stimulation interaction: $F(1,5) = 7.28$, $p = 0.043$.

Post-hoc analyses indicated that reward learning was significantly better in the on-DBS than in the off-DBS state (78 ±7% vs. 56 ±9%, $t(5) = 2.86$, $p = 0.036$), while there was no significant difference in punishment learning (59 ±7% vs. 58 ±7%, $t(5) = 0.14$, $p = 0.89$). Also, there was no difference on global performance (across stimulation states) between reward and punishment learning ($t(5) = 1.08$, $p = 0.33$), as is usually observed in healthy volunteers[53]. This confirms that the two conditions were matched in difficulty, as the probabilistic contingencies were the same for reward and punishment outcomes.

We ran several permutation tests to verify that the observed difference in reward learning between on-DBS and off-DBS states was not due to chance (see Methods). Next, we computed the probability to observe by chance a difference (on-DBS - off-DBS) larger than the observed one: we obtained $p = 0.003$ for reward learning and $p = 0.401$ for punishment learning.

To check whether learning was affected by response type (Go vs. NoGo), we performed a three-way repeated-measures ANOVA with response type, learning condition and stimulation state as fixed factors on the data pooled across all three patients (12 sessions in

total). Only the condition by state interaction was significant ($F(1,5) = 10.48$, $p = 0.02$), all main effects and other interactions were not significant (all p-values > 0.14). These results confirm that there was no detectable bias in learning related to a potential motor difficulty in producing a Go response. Consistently, there was no significant difference in the rate of Go response between the on-DBS and off-DBS states, neither in the reward (on-DBS: 55.56 ±16.95%, off-DBS: 60.56 ± 17.18%, difference: $t = -0.72$, $p = 0.50$) or in the punishment condition (on-DBS: 43.33 ± 19.89%, off-DBS: 52.22 ±18.46%, difference $t = -1.23$, $p = 0.27$). Thus, we found no evidence for an impact of PPN DBS in the capacity to produce a motor response in the learning task.

Finally, we checked that the stimulation was affecting learning (i.e., choice improvement across trials) and not just the mean correct response rate. Linear regressions were performed to compare the slopes of cumulative obtained reward and cumulative avoided punishment (Figure 4B). There was a significant difference in slopes between off-DBS and on-DBS states for reward learning ($t(5) = 3.245$, $p = 0.023$) but not for punishment learning ($t(5) = 1.672$, $p = 0.155$), with significant learning condition (reward vs. punishment) by stimulation state (on-DBS vs. off-DBS) interaction: $t(5) = 3.011$, $p = 0.03$.

Examination of individual performance (Figure 4C) indicated that PPN DBS improved reward learning in all three patients, while the effect on punishment learning was inconsistent. However, these model-free analyses cannot specify which particular process was affected by PPN DBS, and whether it was the same across patients. To address this question, we fitted computational models to choice behavior.

PPN DBS increases the subjective sensitivity to reward outcome

To further explore the effect of PPN DBS on learning, we compared computational models in which the stimulation would affect different parameters. The null model is the standard Q-learning algorithm used to model the behavior in learning task 1, with a delta rule for learning and a softmax rule for choice[54]. This first model includes three parameters: the weights on reward and punishment prediction errors for value updating (learning rates α_R and α_P) and the weight on decision value for choice probability (inverse temperature β). All the other models are variants of this basic null model in which PPN DBS affects one target parameter specifically. Target parameters were multiplicative weights on reward or

punishment outcomes (K_R and K_P) in models 2 and 3, learning rates ($\alpha_R$and α_P) in models 4 and 5, and inverse temperature (β) in model 6.

Bayesian model selection identified model 2 as the most plausible in all three patients (with an exceedance probability of 0.95, see Table S3). The winning model was the same whether we used a fixed-effect model selection (assuming all patients implement the same model) or a random-effect model selection (assuming different patients can implement different models). In this winning model, each reward outcome has a weight of 1 in the off-DBS state, and a weight of K_R in the on-DBS state. Moreover, posterior estimates for the reward sensitivity parameter K_R were greater than 1 in all three patients (Figure 5A), suggesting that the subjective sensitivity to monetary reward was amplified by PPN DBS. Assuming that posterior distributions are gaussian, as is standard in Bayesian Variational Analysis[55], the probability of $K_R > 1$ was 0.980, 0.907 and 0.996 for patients P1, P2 and P3, respectively. We also verified that in model 3, the subjective punishment sensitivity K_P, capturing the effect of PPN DBS, was not significantly greater than 1 (Figure 5B). The mean posteriors for the other parameters of the winning model are provided in Table S4.

To assess the quality of model fitting in the different conditions, we computed Pearson's correlations across trials between observed and modeled correct choice rates (Figure S5). All correlations were significant (with $p < 0.0001$), with a similar explained variance (all r^2 between 0.45 and 0.65) for reward and punishment learning, whether in the on-DBS or off-DBS states. We also checked that the model could reproduce the critical qualitative observation: a selective effect of PPN DBS on reward learning. We simulated the K_R-only model using individual best fitting parameters and performed the same analysis on simulated choices as with observed choices (Figure 5B). The two-way repeated-measures ANOVA with learning condition (reward vs punishment) and stimulation state (on-DBS vs. off-DBS) revealed a significant condition by state interaction ($F(1,5) = 9.56$, $p = 0.03$). As in observed choices, this interaction in simulated choices was qualified as a significant improvement in reward learning ($t(5) = 3.39$, $p = 0.02$) but not punishment learning ($t(5) = 1.33$, $p = 0.24$). Thus, a selective increase in K_R was the most plausible account for the impact of PPN DBS on instrumental learning.

Discussion

In this study, we examined the hypothesis that the PPN may signal reward-related information and thereby contribute to reward learning in humans. Our two predictions were fulfilled: 1) PPN activity was responsive to reward outcome and 2) PPN DBS improved reward learning, specifically.

The neural response to reward outcome was assessed first using fMRI data collected in healthy participants and then iEEG data collected in PD patients performing the same learning task. Both recording techniques showed increased activity for reward but not for no-reward outcomes. The contrast between reward and no-reward outcomes in fMRI data was strongly significant but not specific to the PPN. The same contrast in a whole-brain analysis (see [50]) yielded activation in several regions such as the ventromedial prefrontal cortex and ventral striatum. At the statistical threshold used here (without correction for multiple comparisons), the brain response to reward was even more widespread. Given the size of the PPN region and the uncertainty in the normalization to standard anatomical template, we cannot be sure that the observed response was generated in the PPN.

The localization of recording electrodes in the native brain of PD patients offers better guarantee for accuracy. The response to reward outcome observed in LFP data was increased power in low-frequency (alpha-beta) bands. The timing was compatible with that of hemodynamic response, which is typically delayed by a few seconds. The frequency was low compared to the high-gamma activity that is classically considered as the LFP counterpart of cortical hemodynamic response[56–58], but common to LFP activity usually observed in the basal ganglia[59–61], including in the PPN of PD patients for motor signals[40,62]. It also overlaps with the low-frequency band (from 15 to 40 Hz) that is typically recommended and used for PPN DBS[39,45,63], which is rather low compared to standard high-frequency stimulation of the subthalamic nucleus[64,65].

Both fMRI and LFP activity remained insensitive to choices, meaning that they did not differ between left-hand and right-hand responses. Thus, the putative contribution of the PPN to instrumental learning would be to provide information about whether the action must be reinforced, not which action must be reinforced. Showing this dissociation was important because previous electrophysiological studies in rodents reported that PPN neurons encode both choice-related actions and reward outcomes[27,28]. The LFP

response recorded here is reminiscent of the reward signal conveyed by dopamine neurons, which is largely independent of the action to be reinforced, since it is equally observed during Pavlovian conditioning[66]. The independence from motor actions was also observed in the second learning task, where cues were chosen with either a Go or NoGo response. While PPN-DBS increased choice rate for the most rewarded cue, it did not increase the overall rate of Go responses. Thus, the contribution of the PPN to reward learning observed in our tasks seems independent from its role in motor processing that has been documented in previous studies[67,68].

One remaining issue is whether the PPN would also encode outcome expectation, i.e. the other component of RPE. We found no trace of reward expectation in LFP activity, which should manifest as reduced positive RPE following rewards, and even negative RPE in the absence of reward. However, it was difficult to test in our patients, because their learning curves were somewhat erratic, such that their expectations about reward outcomes were difficult to estimate. Such a poor learning performance is commonly observed in PD[69–73], and usually interpreted as a consequence of neuronal loss in dopaminergic nuclei, although it could also be attributed to PPN degeneration. If expectations remain low, outcomes remain surprising, such that it becomes difficult to distinguish between pure reward and RPE signals. Yet models of RPE computations in dopaminergic nuclei generally assume that expectation signals come from separate inputs and are subtracted from reward signals via GABAergic neurons[15]. In any case, what we can safely conclude is that PPN LFP signals the occurrence of reward outcomes. By this we do not mean that PPN LFP activity is proportional to reward value, which can only be assessed by varying reward magnitude, but simply that it selectively responds to reward and not to absence of reward.

To our knowledge, this is the first evidence obtained in humans for the implication of the PPN in reward processing. It is nonetheless consistent with a wealth of evidence in animals (rodents mostly) that PPN neurons signal reward outcomes and send this information to dopamine neurons[26–29]. The next question is how reward-related information reaches the PPN in the first place. There are plenty of possibilities, given the richness of afferent inputs to the PPN established in non-human primates[68,74–76], coming from both the basal ganglia (such as the globus pallidus and subthalamic nucleus) and cortical areas (such as the ventromedial prefrontal and anterior cingulate cortex). Another remaining question is whether the PPN reward signals that we observed here in the human brain are indeed

conveyed to dopamine neurons. Even if this is the case, the PPN may not be the only provider of reward-related information to dopaminergic nuclei, given the number of candidate inputs to VTA and SNpc found in rodents[77,78]. To assess whether PPN reward signals have a sufficient impact to causally modulate learning abilities, we turned to DBS.

Our results are in line with studies in rodents reporting that PPN stimulation enhances reinforcement[30,31]. In our task, the signature of PPN DBS was a selective improvement of reward learning, leaving punishment learning unaffected, hence excluding non-specific effects on attention or cognition. This signature has been previously observed following dopatherapy broadly speaking (mixing levodopa and dopamine receptor agonists) in both healthy participants and PD patients tested with various tasks, including the learning task used here[8,9,69–71]. The effect of PPN DBS on choice behavior was therefore compatible with an increase in dopamine release.

Moreover, computational analysis showed that the effect was best captured by a model in which PPN DBS increases the subjective sensitivity to reward outcomes. The opposite computational effect (decreased reward sensitivity) was found with both dopamine blockers and ventral striatum degeneration in humans, again using the same task[9,53]. Thus, the computational analysis strengthened the hypothesis that PPN DBS improves choice behavior by increasing dopamine release in the striatum. One difficulty, however, is to explain how a tonic intervention like PPN DBS (or dopatherapy, for that matter) can boost the impact of reward outcome, known to be mediated by phasic dopamine signals. A possible explanation is that increasing tonic dopamine helps the phasic signal, supposed to be reduced in PD, to pass a threshold above which it can achieve efficient reinforcement of cortical-striatal synapses[79]. A subtly different explanation is that PPN excitatory inputs do not increase tonic activity but amplify the response of dopamine neurons to reward signals, which may come from other afferent regions. On a more cautious note, the similarity in the behavioral effects of PPN DBS and dopatherapy does not prove that they share common targets at the neural level. Even if less parsimonious, the possibility remains that the two treatments exert their actions through independent neural mechanisms.

Our study has several limitations. An obvious limitation is the small number of patients who could be tested, due to an early ending of the clinical protocol, because of low benefit/risk tradeoff estimates in the first cases. Unfortunately, PPN DBS is still an experimental treatment, which has not reached the stage where it can be applied to large

cohorts of patients. The small sample size forced us to use fixed-effect analyses, which are common in non-human primate studies, but not in human clinical trials, for which random-effect analyses are standard. Another consequence is that we were unable to balance the order of stimulation states across patients. However, the two patients who exhibited a greater effect of PPN DBS were tested first in the On state and then in the Off state, which discards the possibility of reward learning improvement being confounded with practice effects. Another limitation is that the role of PPN in reward learning was tested in PD patients, who suffer from loss of not only dopamine neurons but also PPN neurons. Moreover, the techniques used here cannot tell whether PPN response to reward, or the impact of PPN DBS on reward learning, were mediated by glutamatergic or cholinergic neurons. Also, regarding DBS effects, there may be some concerns about the selectivity of the cylindrical DBS electrodes when targeting small structures such as the PPN[80]. Finally, we did not assess the interactions with dopatherapy, which would have required to double all testing sessions (to compare conditions with and without medication). One may speculate that the impact of PPN-DBS could be potentiated by dopaminergic treatment[26].

To conclude, we found evidence that PPN activity is sensitive to reward outcomes and that PPN stimulation boosts the subjective sensitivity to reward. These findings are consistent with the hypothesis raised from the literature in rodents that, in humans as well, the PPN might convey reward-related information to dopamine neurons. However, they are by no means a definitive proof that the hypothesis is correct: whether the reward signals observed in the PPN are actually used by dopamine neurons to compute RPE remains to be demonstrated. On a more clinical perspective, our findings suggest that the clinical impact of PPN DBS could go beyond the expected effects on gait and postural disorders. By boosting subjective reward sensitivity, PPN DBS could contribute to reduce the apathy that is commonly observed in PD, especially when dopatherapy is suppressed[65,81,82].

Acknowledgments

We thank Brian Lau for insightful comments and discussion; Amine El Helou and Thomas Andrillon for their help with the early stages of data analysis, and Sara Fernandez Vidal for her help with figure preparation.

V.S. was supported by Ecole des Neurosciences de Paris Ile-de-France (ENP); S.P. was supported by the ATIP-Avenir program. The study was funded by the program 'Investissements d'avenir' (ANR-10-IAIHU-06).

Author Contributions

S.P., D.G and M.P. conceptualized the study and designed the protocol; C.K. and M-L. W. performed the surgery and supervised clinical procedures; S.P. and A.B. collected behavioral and electrophysiology data; V.S., S.P. and M.P. analyzed the data and developed computational models; V.S. and M.P. wrote the initial draft; all authors contributed to editing the manuscript.

Declaration of interests

Carine Karachi reports having received lecture fees and research funding from Boston Scientific and Medtronic.

Figures Legends

Figure 1. General overview of the study.

A. Timeline of clinical and experimental events. PPN local field potentials were recorded using learning task 1, during the week following surgical implantation of DBS electrodes. Behavioral effects of PPN-DBS were tested several months later, using learning task 2, when patients came back to the hospital for clinical assessment. Learning task 2 was performed once in the off-DBS state and once in the on-DBS state, in a randomized order.

B. Location of recording sites. Electrodes implanted in the three tested patients (P1 – red, P2 – blue and P3 – yellow) are inserted in a 3D reconstruction of the peduncolopontine nucleus (PPN, transparent purple) and cuneiform nucleus (CN, transparent green). White asterisks indicate active contacts used for DBS.

See also Table S1 and Table S2.

Figure 2. Instrumental learning performance.

A. Example trial of learning task 1. Screenshots are shown from left to right, with durations in milliseconds. On every trial, subjects selected between left and right options represented by two visual cues, using their left-hand or right-hand index to press the corresponding button. The side of the selected cue (left in the example) was marked with a red pointer. Subjects could then observe the outcome of their choice (a 0.5€ reward or nothing) and update their estimates of cue-reward contingencies. Each session presented four different pairs of cues, associated with different combinations of reward probability (25/25, 75/75, 75/25, 25/75 %).

B. Behavioral performance of PD patients (N=3) during LFP recordings in hospital settings and aged healthy volunteers (N=8) tested in a previous study[49]. Histograms show the choice rate observed with the four pairs of cues associated to varying reward probability (light gray: symmetrical pairs, dark grey: asymmetrical pairs). Although patients tended to prefer responding with the right hand (to the two symmetrical pairs), the pattern of behavioral performance was qualitatively similar to that of controls, with higher choice rate

for more rewarded cues (within the two asymmetrical pairs). Error-bars are inter-session S.E.M.

See also Figure S1.

Figure 3. PPN potentials evoked by reward.

A. Time-frequency decomposition of response to choice (top panels) and outcome (bottom panels). Color code indicates power observed in each time-frequency bin of the map. Power was corrected for baseline measure (over a -500 to 0ms time window prior to fixation onset). Maps were averaged over all sessions and all available electrodes in all three patients. Only the contrast between reward and no-reward outcomes yielded significant differences, not the contrast between ipsi- and contra-lateral choice.

B. T-value map of the difference between reward and no reward outcomes. The only significant cluster ($p < 0.05$ after correction for multiple comparisons) was identified in a low-frequency band. Color code indicates the T-value for each time-frequency bin within the cluster. Dotted horizontal lines indicate the 10-20Hz band width explored in part D.

C. Time course of 10-20Hz activity following reward outcome onset, averaged over patients (left panel) or separately for each patient (right panel). Responses to no reward outcomes have been omitted in individual plots for the sake of visibility. Thick lines show time points at which activity is different from 0 ($p< 0.05$, after cluster-wise correction for multiple comparisons). Shaded area around the group mean represents inter-session S.E.M.

See also Figure S2-S4.

Figure 4. Behavioral effects of PPN DBS.

A. Example trials of learning task 2. Screenshots are shown from left to right, with durations in milliseconds. On every trial, subjects selected either the upper or lower cue, by pressing or not pressing a response button (Go or NoGo response). The selected cue was indicated with a red frame. Subjects could then observe the outcome of their choice (a 1€ reward, -1€ punishment, or nothing) and update their estimates of cue-outcome

contingencies. Each session presented novel pairs of cues, associated with either 80/20% reward probability (vs. nothing, as in top screenshots) or 80/20% punishment probabilities (vs. nothing, as in bottom screenshots).

B. Cumulative learning curves (left panel: cumulative reward obtained; right panel: cumulative punishment avoided) observed in the On (dark red and blue) and Off (light red and blue) PPN stimulation states. Dots represent real data fitted using linear regression and averaged across sessions and patients. Shaded areas are inter-session S.E.M. The slope of reward accumulation was specifically increased by PPN-DBS. $*p < 0.05$ (paired two-tailed t-tests), *n.s*. not significant.

C. Correct choice rates observed for reward (red) and punishment (blue) learning in the Off and On stimulation states. Histograms show group means and dots represent the three patients. The performance of healthy volunteers tested in a previous study[53] is shown as a reference point. Error bars represent inter-session S.EM. for patients 'data and inter-subject S.EM. for healthy volunteers. $*\ p < 0.05$ (paired two-tailed t-tests), *n.s*. not significant.

Figure 5. Computational account of PPN DBS effects.

A. Posterior estimates of subjective reward and punishment sensitivity parameters (K_R and K_P), estimated from models 2 and 3, where these parameters captured the effect of PPN DBS (relative to the Off state, where reward and punishment outcomes were assigned a reference weight of 1). Error bars represent the standard deviation of the individual posterior distribution. Reward sensitivity (K_R), more than punishment sensitivity (K_P), was increased way above 1 in the On state.

B. Average simulated performance for reward (red) and punishment (blue) learning conditions in "Off" and "On" stimulation states. Black dots are patients' data averaged across all "On" and "Off" sessions. Error bars represent inter-session SE for human data and simulations. $*p < 0.05$ (paired two-tailed t-tests), *n.s.* – not significant.

See also Figure S5 and Tables S3-S4.

STAR Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Mathias Pessiglione (mathias.pessiglione@gmail.com)

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The raw dataset supporting the current study has not been deposited in a public repository due to ethical regulations but synthetic dataset and Matlab codes are available on request to corresponding authors.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patients

Six patients with idiopathic Parkinson disease were recruited for a clinical trial aiming to assess the effects of PPN DBS on gait and balance disorders (trial #NCT020555261)[47]. As explained in the clinical report, two patients withdrew from the trial at an early stage due to surgery complications[47]. For technical reasons, data collection could not be completed in one patient, who was therefore not included in the final analysis, for which were retained the other 3 patients (2 females, mean age 60, range 46-70). Demographic and clinical characteristics are given in Table S1.

All patients gave their informed written consent to the participation in all parts of the study, for which they received no financial compensation. The study protocol was approved by local ethic committee (Comité de Protection des Personnes d'Ile-de-France, Paris 6).

METHOD DETAILS

Behavioral data collection

Patients performed two instrumental learning tasks, the first one (task 1) during LFP recording and the second one (task 2) 4 and 6 months after the surgery (Figure 1). Learning task 2 was assessed in two double-blind conditions: "on-DBS" and "off-DBS". Patients 1 and 3 performed task 2 in the On-Off order, whereas patient 2 did it in the reverse Off-On order, as well as excluded patient 4. Both tasks were presented using Cogent 2000 Matlab Toolbox (Wellcome Department of Imaging Neuroscience, London, UK)[83].

Task 1

Patients performed a version of the probabilistic instrumental learning task used in previous fMRI studies to dissociate the neural correlates of reward and movement[48–50][48]. Each session contained four pairs of visual cues from the Agathodaimon font. Each cue was associated with either 25 or 75% chance of getting 50 cents or nothing. The four pairs were associated with different combinations of reward probabilities: 25/25, 75/75, 25/75 and 75/25%. Task sessions were independent from each other and patients had to relearn new cue – outcome contingencies in every session. Each session contained 96 trials in total.

On every trial, patients had to make a choice between left and right cues presented simultaneously on the screen (Figure 2A). Patients had 3000ms to make their choice and had to keep pressing the button with her left or right index finger until the red pointer appeared on the screen below the selected cue indicating the choice for 500ms. Afterwards, a screen with feedback information (50 cents or 0 cents) appeared for 3000ms, before the onset of the next trial. If no response was made after 3000ms, the trial ended with a negative feedback and the next trial began. Patients were told that cues differed in probability of being rewarded and were encouraged to accumulate as much money as possible. However, they were not given any explicit information about the structure of the task and had to adjust their choices by trial and error. The difference between left- or right-hand choices in the asymmetrical pairs (75/25 and 25/75%) compared to symmetrical pairs (25/25 and 75/75%) indicated learning. Before the 3 test sessions, patients performed a short training session to familiarize themselves with the task.

Task 2

The same patients performed twice, (once On and once Off PPN DBS), two months apart, a task that involved learning from monetary gains and losses, used in previous fMRI and patient studies[9,53] (Figure 4A). All patients performed three sessions of the task in both "on-DBS" and "off-DBS" states, each session presenting novel cues to be learned. As in task 1, patients made binary choices between two visual cues, but this time each cue was associated with either a monetary gain (1€) or a monetary loss (-1€). In the reward condition, one cue was associated with 80% chance of winning 1€ and 20% of winning nothing, while the other cue had reverse contingencies. Symmetrically, in the punishment condition, one cue was associated with 80% chance of losing 1€ and 20% of losing nothing, and vice-versa for the other cue. A third "neutral" pair of cues did not result in any financial outcome and served as a motor control. Each session presented one pair of reward cues, one pair of punishment cues, and one neutral pair. The reward and punishment trials were thus intermixed within a session.

On every trial, patients had 4000ms to make a choice between the upper and lower cues: they pressed a button to select the upper one or did nothing to select the lower one (Figure 4A). The chosen cue was highlighted on the screen for 500ms and then the outcome was shown for 3000ms. Again, patients were told to maximize their payoffs, without any detail about how gain and loss conditions were distributed over cues.

Computational models

To get deeper insight about the effects of PPN-DBS on learning, we built up several variations of a basic Q-learning model. In all models, monetary gains and losses were coded as 1 and -1 and Q-values were initiated at 0.5 and -0.5 for reward and punishment learning conditions, respectively. Chosen Q-values at time *t* were updated proportionally to the prediction error, according to the Rescorla-Wagner learning rule:

$$(1)\ \ Q_{t+1} = Q_t + \alpha * (Outcome - Q_t),$$

where α is a learning rate (between 0 and 1) that was set separately for reward and punishment learning conditions. The choice was implemented using a softmax rule:

$$(2)\ P_{up} = \frac{1}{1+\exp(-\beta(Q_{up}-Q_{down}))},$$

where P_{up} is the probability of choosing the upper cue and β is an inverse temperature parameter that adjusts for the stochasticity of choice. This basic Q-learning model (model 1) included three free parameters that were independent from the stimulation. We next constructed several specifications of this model that made different predictions about the effects of PPN DBS on reward and punishment learning.

As we were interested in finding the parameter that best captures the difference in learning between on-DBS and off-DBS states, we focused on models where stimulation impacts only one parameter, and excluded models with combinatory effects on several parameters. First, we included a multiplicative sensitivity parameter that could modulate subjective perception of reward K_R (model 2) or punishment K_P (model 3) under PPN DBS. It was set to 1 in Off sessions but was allowed to vary between 0 and infinity in On sessions.

$$(3)\ Q_{t+1} = Q_t + \alpha_R * (\boldsymbol{K_R} * Outcome - Q_t), \text{ where}$$

$$\begin{cases} 0 < K_R < Inf \mid DBS = 1 \\ K_R = 1 \quad\quad\quad\ \mid DBS = 0 \end{cases}$$

These parameters capture potential differences in the sensitivity to reward or punishment outcomes and adjust both the slope and plateau of learning curves.

In the next two models, we assumed that stimulation primarily affected learning rate for either reward α_R(model 4) or punishment α_P (model 5) and thereby by adjusted the slope of learning curves without affecting the plateau.

$$(4)\ Q_{t+1} = Q_t + \boldsymbol{\alpha_R} * (Outcome - Q_t), \text{ where}$$

$$\begin{cases} \boldsymbol{\alpha_R} = \boldsymbol{\alpha_{R\,ON}} \mid DBS = 1 \\ \boldsymbol{\alpha_R} = \boldsymbol{\alpha_{R\,OFF}} \mid DBS = 0 \end{cases}$$

Finally, we included a model where the effect of the stimulation was independent from outcome valence but instead changed choice stochasticity by affecting the inverse temperature parameter β (model 6).

$$(5)\ P_{up} = \frac{1}{1+\exp(-\beta*(Q_{up}-Q_{down}))}, \text{ where}$$

$$\begin{cases} \beta = \beta_{ON} \mid DBS = 1 \\ \beta = \beta_{OFF} \mid DBS = 0 \end{cases}$$

All models except model 1 had four free parameters in total.

We inverted all 6 models, for ach patient separately, using the Variational Bayes approach under the Laplace approximation[55,84] implemented via a customary-built MatLab toolbox (available at http://mbb-team.github.io/VBA-toolbox/). This is an iterative method that approximates model evidence, which is difficult to track analytically, using variational free energy[85]. Model evidence represents a trade-off between accuracy (goodness of fit) and complexity (number of parameters). Model inversion was computed using the same priors for parameters shared across all models, and the same variance for parameters meant to capture the effect of PPN DBS. The most plausible effect of PPN DBS was selected using a random-effect model comparison, which assumes that different patients might implement different models [84,86]. A fixed-effect analysis (i.e., just summing model evidence over the three patients) yielded similar results.

To verify that the winning model was able to reproduce the observed effect of PPN DBS, we used it to simulate choices in 12 learning sessions, keeping the best-fitting individual parameters, and performed the same analysis as with observed choices. To assess the quality of fit, we also computed Pearson's correlations across trials, including all 6 On and 6 Off sessions between observed and simulated correct choice rate (Figure S5).

LFP data collection

Surgical procedure

Details of the surgical procedure and localization of the PPN target regions were previously reported[40,47]. Individual localization of the PPN target area was determined using both direct MRI navigation and a 3D histological atlas of the basal ganglia deformed to match the T1-weighted preoperative MRI[88]. The localization of the definitive DBS electrodes was performed using postoperative helicoidal CT scans registered to the preoperative T1-weighted MRI scans[52]. Local field potentials were recorded from the bilateral definitive DBS electrodes (model 3389, Medtronic Neurological division), with four cylindrical platinum-iridium contacts (1.27 mm in diameter and 1.5mm in length, 0.5 mm separation).

Signals were amplified, low-pass filtered at 250Hz and sampled at 512 Hz (Basis BE System, EB Neuro S.p.A).

PPN-DBS parameters

Contacts and parameters of the stimulation for each patient are reported in Table S2. Clinical effects of PPN DBS with and without L-dopa treatment have been described in details[47]. All patients were stimulated bilaterally and were withdrawn from the dopaminergic medication for at least 12 hours prior to the testing session.

fMRI data collection

To explore whether the PPN region encodes reward-related variables, we first analyzed the data collected in healthy controls (N = 20, 11F, age between 19 and 31) who performed the same instrumental learning task as PD patients (task 1) in the fMRI scanner. Data were preprocessed and analyzed using statistical parametric mapping (SPM8) software (Wellcome Trust Center for Neuroimaging, London, UK)[87]. Structural T1-weighted images were corregistered to the mean functional EPI, segmented and normalized to the standard anatomical template. Preprocessing of the EPI time series was identical to that reported in previous fMRI studies[48,50], including spatial realignment and normalization using the same transformation as structural images, except that we skipped the final spatial smoothing of the data to avoid blurring the BOLD signal between the PPN and its neighbors. The PPN region of interest (Figure S1A) was defined from a digital atlas of subcortical structures[51,52] mapped onto individual normalized brain scans.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were conducted using customary-built scripts and statistical toolbox in MATLAB (R2020a, Natick, Massachusetts: The MathWorks Inc.). All statistical details can be found in the results section and/or figure legends.

Behavioral data analysis

Task 1

Due to the small number of patients (N = 3), second-level statistical tests assessed fixed effects, with one data point per session, ignoring the differences between individuals. Two patients performed three sessions of the task. The last session in patient 1 was excluded from the analysis due to a large number of trials with no responses (> 30%). The remaining 8 sessions all had a proportion of missed trials below 15% (8.42 ±5.31% on average).

To assess performance in learning (Figure 2B), we compared the proportion of left/right responses between asymmetrical pairs (25/75 vs. 75/25).

Task 2

We defined a correct response as choosing the best cue in the reward learning condition and avoiding the worst cue in the punishment condition. Performance in the first session was much lower than in the two subsequent sessions for each of the three patients, as was observed in other clinical populations performing the same task[53]. We therefore considered this first session as a practice session and did not include it in the main behavioral data analysis that is based on 12 sessions in total (6 on-DBS and 6 off-DBS). Note that including this first session in the analysis would not change the overall pattern of choice behavior.

In order to assess the effects of learning condition (reward vs. punishment) and stimulation state (on-DBS vs. off-DBS), we ran a two-way repeated-measures ANOVA on session-wise mean correct choice rate, thus considering state and condition as fixed factors, and session as a random factor. The ANOVA was followed with planned pair-wise comparisons using two-tailed t-tests between performance in on-DBS vs. off-DBS states, separately for reward and punishment conditions. Additionally, to test for interaction with motor responses, we ran a three-way repeated-measures ANOVA on mean correct choice rate with learning condition (reward vs. punishment), stimulation state (on-DBS vs. off-DBS), and response type (Go vs. NoGo) as fixed factors, and session as a random factor.

Because the number of data points is low for parametric tests to be robust, we also performed a permutation test (N = 10000), to compute the probability that the observed results were due to chance level. On each permutation, the on-DBS/off-DBS session labels were randomly flipped and the differences between on-DBS and off-DBS learning performance were recomputed, separately for the reward and punishment conditions.

Significance was assessed as the proportion of permutations where the On-Off difference was positive and above the observed difference in actual performance.

In order to provide reference points for the learning performance observed in our three patients, we also report performance in young healthy volunteers (N = 20, 8F, mean age 43.6 ±2.8), as well as age-matched controls (N = 20, 8F, mean age 43.6 ±2.8) tested previously with the same task[48,49][53]. We did not make any statistical comparisons between groups because sample size was too small.

To quantify learning dynamics, we fitted a linear regression model to the trial-by-trial cumulated points won or lost, separately for reward and punishment learning conditions and for "on-DBS" and "off-DBS" states. On every trial, the cumulative point was increased by one when the outcome was a reward and decreased by one when it was a punishment. To show whether the balance was progressing over trials, we regressed these cumulative scores against trial number, separately for each of the 12 sessions. We next performed a two-way repeated-measures ANOVA with learning condition and stimulation states as fixed factors on the slopes (regression weights).

LFP data analysis

Preprocessing

Bipolar recordings were first computed between the adjacent contacts for each electrode by subtracting the signal of more ventral from more dorsal contact resulting in three (0-1, 1-2 and 2-3) contacts per side, with the 0-1 contact being the most ventral and 2-3 being the most dorsal. Notch filter was applied at 50 Hz to remove the line noise and data were band-pass filtered between 2 and 95 Hz. All sessions in each patient were manually examined for artifacts and trials with amplitudes above 500 mV or greater than 4SD of the amplitude distribution across sessions were removed. Contacts with artifacts in more than 70% of trials were excluded, leading to a total of 15 contacts available for analysis. The contacts included in the final dataset for each patient are reported in Table S2.

Time-frequency decomposition

Preprocessed data were z-scored separately for each bipolar recording, prior to the spectral decomposition, to obtain comparable values across contacts. Spatiotemporal frequency

maps were obtained using the multi-taper method implemented in Chronux Matlab library[88] http://chronux.org/. For each bipolar contact, power was calculated in 300ms sliding window with a 30ms step and 6 orthogonal tapers with a time bandwidth product equal to 5. Trial-by-trial spectrograms were normalized to the baseline fixation window (500ms) prior to the presentation of choice options and converted to decibels (dB).

Statistical analysis

Due to the small number of patients (N = 3), second-level statistical tests assessed fixed effects, with one data point per contact, ignoring the differences between individuals.

To test for the presence of reward-related signals in PPN LFP, we defined a window from -500ms to +2000ms around outcome display and split the time-frequency series between reward and no reward trials. At the first level, we computed the contrast between reward and no reward trials for each contact, resulting in 15 time-frequency maps. These contrast maps were then brought to a second-level analysis across contacts and tested against zero at every frequency and every time point.

To correct for multiple comparisons, i.e. testing over 47 frequencies (2:5:95 Hz) at 50 time points (-500:60:2500 ms), we used a cluster-based permutation test using the script routines implemented in the Fieldtrip signal processing toolbox for Matlab[89,90]. First, t-values were computed at every frequency and time point and the threshold for cluster selection was set to the 97.5 quintile of the t-distribution (two-tailed t-test with alpha-level of 0.05). Clusters were selected on the basis of their unsigned t-value, and constructed on the basis of temporal and spatial continuity, separately for positive and negative t-values. Cluster-level t-statistic was computed as the sum of t-values within the cluster. The permutation distribution was based on a maximum (unsigned) cluster-level statistic. Only clusters that had a Monte-Carlo p-value less than 0.025 (5% chance that the shuffled t-value exceeded the true t-value for a given cluster with a two-tailed test) were considered significant and are shown on Figure 3B. Importantly, this test informs about whether activity was different between reward and no reward conditions at each time-frequency point, but not about the directionality of the difference.

To further specify the observed difference in LFP activity, we extracted the (baseline-corrected) power time series within the significant frequency band, separately for reward and no-reward trials. The aim of this post-hoc analysis was to qualify the significant

difference as driven by increased or decreased activity following reward vs. no reward outcomes. For each trial, power time series were averaged across all contacts, either at the individual or group level (Figure 3C). To test for a response to reward outcomes, the power at each time point was compared to zero, using again a permutation tests to correct for multiple comparisons[89]. In each permutation, the power sign was flipped in half the trials, picked at random, and the t-value was recomputed based on the shuffled data. The procedure was repeated for a total of 10,000 permutations, providing the distribution of t-values under the null hypothesis (no response to reward). The test was deemed significant if the observed t-value exceeded that obtained by chance in more than 95% of permutations and for at least 8 consecutive time points.

To explore whether other variables such as reward expectation, trial number or choice side would affect LFP activity, we regressed trial-by-trial time-frequency maps against GLM that included these variables and their interaction with outcomes (coded 1 for reward and zero for no reward). Second-level statistical tests were then conducted as for the simple contrast between reward and no reward outcomes, employing the same permutation procedure to correct for multiple comparisons.

fMRI data analysis

At the subject level, we constructed a GLM with two regressors for button presses split between left- and right-hand choices and four regressors for outcome onsets split between left- and right-hand choices and between reward vs. no-reward outcomes. All six regressors were modeled as stick functions with duration set to zero and convolved with the canonical hemodynamic response. Regressors of no interest included the six motion parameters. We next extracted subject-by-subject the beta estimates at the moment of choice and outcome for all contrasts, which were averaged over all voxels within the bilateral masks delineating the PPN.

Choice-related response was analyzed using repeated-measures ANOVA with choice side (left vs. right) and PPN anatomical ROI side (left vs. right) as within-subject factors. Outcome-related response was analyzed using repeated-measures ANOVA with outcome (reward vs. no-reward), choice (left vs. right), and PPN anatomical ROI side (left vs. right) as within-subject factors.

References

1. Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction (Cambridge: MIT Press).

2. Schultz, W. (1997). A Neural Substrate of Prediction and Reward. Science (80-.). *275*, 1593–1599.

3. Bayer, H.M., and Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron *47*, 129–141.

4. Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. Nature *412*, 43–48.

5. Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T.K., Ueda, Y., Inokawa, H., Haruno, M., and Kimura, M. (2011). Dopamine neurons learn to encode the long-term value of multiple future rewards. Proc. Natl. Acad. Sci. U. S. A. *108*, 15462–7.

6. D'Ardenne, K., McClure, S.M., Nystrom, L.E., and Cohen, J.D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science *319*, 1264–7.

7. Zaghloul, K.A., Blanco, J.A., Weidemann, C.T., Mcgill, K., Jaggi, L., Baltuch, G.H., and Kahana, M.J. (2009). Human substantia nigra neurons encode unexpected financial rewards. Science (80-.). *323*, 1496–1499.

8. Frank, M.J., Seeberger, L.C., and O'Reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science *306*, 1940–3.

9. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature *442*, 1042–5.

10. Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., and Janak, P.H. (2013). A causal link between prediction errors, dopamine neurons and

learning. Nat. Neurosci. *16*, 1–10.

11. Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G.D., Bonci, A., de Lecea, L., and Deisseroth, K. (2009). Phasic Firing in Dopaminergic Neurons. Science (80-.). *324*, 1080–1084.

12. Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature *482*, 85–8.

13. Chang, C.Y., Esber, G.R., Marrero-Garcia, Y., Yau, H.-J., Bonci, A., and Schoenbaum, G. (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat. Neurosci. *19*, 111–116.

14. Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., and Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. Nature *525*, 243–246.

15. Watabe-Uchida, M., Eshel, N., and Uchida, N. (2017). Neural Circuitry of Reward Prediction Error. Annu. Rev. Neurosci. *40*, 373–394.

16. Mena-Segovia, J., Bolam, J.P., and Magill, P.J. (2004). Pedunculopontine nucleus and basal ganglia: distant relatives or part of the same family? Trends Neurosci. *27*, 585–588.

17. Kobayashi, Y., and Okada, K.-I. (2007). Reward prediction error computation in the pedunculopontine tegmental nucleus neurons. Ann. N. Y. Acad. Sci. *1104*, 310–23.

18. Olszewski, J., and Baxter, D. (1982). Cytoarchitecture of the Human Brainstem (Basel: S Karger AG).

19. Marín, O., Smeets, W.J.A.., and González, A. (1998). Evolution of the basal ganglia in tetrapods: a new perspective based on recent studies in amphibians. Trends Neurosci. *21*, 487–494.

20. Matsumura, M. (2005). The pedunculopontine tegmental nucleus and experimental parkinsonism: A review. J. Neurol. *252*, 5–12.

21. Lokwan, S.J.., Overton, P.., Berry, M.., and Clark, D. (1999). Stimulation of the pedunculopontine tegmental nucleus in the rat produces burst firing in A9 dopaminergic neurons. Neuroscience *92*, 245–254.

22. Floresco, S.B., West, A.R., Ash, B., Moore, H., and Grace, A.A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. Nat Neurosci *6*, 968–73.

23. Forster, G.L., and Blaha, C.D. (2003). Pedunculopontine tegmental stimulation evokes striatal dopamine efflux by activation of acetylcholine and glutamate receptors in the midbrain and pons of the rat. Eur. J. Neurosci. *17*, 751–762.

24. Pan, W., and Hyland, B.I. (2005). Pedunculopontine tegmental nucleus controls conditioned responses of midbrain dopamine neurons in behaving rats. J. Neurosci. *25*, 4725–32.

25. Dautan, D., Souza, A.S., Huerta-Ocampo, I., Valencia, M., Assous, M., Witten, I.B., Deisseroth, K., Tepper, J.M., Bolam, J.P., Gerdjikov, T. V, *et al.* (2016). Segregated cholinergic transmission modulates dopamine neurons integrated in distinct functional circuits. Nat. Neurosci. *19*, 1025–1033.

26. Okada, K., Toyama, K., Inoue, Y., Isa, T., and Kobayashi, Y. (2009). Different pedunculopontine tegmental neurons signal predicted and actual task rewards. J. Neurosci. *29*, 4858–70.

27. Norton, A.B.W., Jo, Y.S., Clark, E.W., Taylor, C.A., and Mizumori, S.J.Y. (2011). Independent neural coding of reward and movement by pedunculopontine tegmental nucleus neurons in freely navigating rats. Eur. J. Neurosci. *33*, 1885–96.

28. Thompson, J.A., and Felsen, G. (2013). Activity in mouse pedunculopontine tegmental nucleus reflects action and outcome in a decision-making task. J. Neurophysiol. *110*, 2817–2829.

29. Hong, S., and Hikosaka, O. (2014). Pedunculopontine tegmental nucleus neurons provide reward, sensorimotor, and alerting signals to midbrain dopamine neurons. Neuroscience *282C*, 139–155.

30. Yoo, J.H., Zell, V., Wu, J., Punta, C., Ramajayam, N., Shen, X., Faget, L.,

Lilascharoen, V., Lim, B.K., Hnasko, T.S., *et al.* (2017). Activation of Pedunculopontine Glutamate Neurons Is Reinforcing. J. Neurosci. *37*, 38–46.

31. Xiao, C., Cho, J.R., Zhou, C., Treweek, J.B., Chan, K., Mckinney, L., Yang, B., and Gradinaru, V. (2016). Cholinergic Mesopontine Signals Govern Locomotion and Reward through Dissociable Midbrain Pathways. Neuron *90*, 333–347.

32. Inglis, W.L., Olmstead, M.C., and Robbins, T.W. (2000). Pedunculopontine tegmental nucleus lesions impair stimulus--reward learning in autoshaping and conditioned reinforcement paradigms. Behav. Neurosci. *114*, 285–94.

33. Maclaren, D., Wilson, D.I.G., and Winn, P. (2013). Updating of action-outcome associations is prevented by inactivation of the posterior pedunculopontine tegmental nucleus. Neurobiol. Learn. Mem. *102*, 28–33.

34. Syed, A., Baker, P.M., and Ragozzino, M.E. (2016). Pedunculopontine tegmental nucleus lesions impair probabilistic reversal learning by reducing sensitivity to positive reward feedback. Neurobiol. Learn. Mem. *131*, 1–8.

35. Thompson, J.A., Costabile, J.D., and Felsen, G. (2016). Mesencephalic representations of recent experience influence decision making. Elife *5*.

36. Hirsch, E., Graybielt, A., Duyckaertst, C., and Javoy-Agid, F. (1987). Neuronal loss in the pedunculopontine tegmental nucleus in Parkinson disease and in progressive supranuclear palsy. Proc. Natl. Acad. Sci. *84*, 5976–5980.

37. Jenkinson, N., Nandi, D., Muthusamy, K., Ray, N.J., Gregory, R., Stein, J.F., and Aziz, T.Z. (2009). Anatomy, physiology, and pathophysiology of the pedunculopontine nucleus. Mov. Disord. *24*, 319–328.

38. Ballanger, B., Lozano, A.M., Moro, E., van Eimeren, T., Hamani, C., Chen, R., Cilia, R., Houle, S., Poon, Y.Y., Lang, A.E., *et al.* (2009). Cerebral blood flow changes induced by pedunculopontine nucleus stimulation in patients with advanced Parkinson's disease: A [15O] H2O PET study. Hum. Brain Mapp. *30*, 3901–3909.

39. Thevathasan, W., Pogosyan, A., Hyam, J. a, Jenkinson, N., Foltynie, T., Limousin, P., Bogdanovic, M., Zrinzo, L., Green, A.L., Aziz, T.Z., *et al.* (2012). Alpha

oscillations in the pedunculopontine nucleus correlate with gait performance in parkinsonism. Brain *135*, 148–60.

40. Lau, B., Welter, M.-L., Belaid, H., Fernandez Vidal, S., Bardinet, E., Grabli, D., and Karachi, C. (2015). The integrative role of the pedunculopontine nucleus in human gait. Brain *138*, 1284–96.

41. Pereira, E.A.C., Nandi, D., Jenkinson, N., Stein, J.F., Green, A.L., and Aziz, T.Z. (2011). Pedunculopontine stimulation from primate to patient. J. Neural Transm. *118*, 1453.

42. Karachi, C., Grabli, D., Bernard, F.A., Tandé, D., Wattiez, N., Belaid, H., Bardinet, E., Prigent, A., Nothacker, H., Hunot, S., *et al.* (2010). Cholinergic mesencephalic neurons are involved in gait and postural disorders in Parkinson disease. *120*.

43. Hamani, C., Moro, E., and Lozano, A.M. (2011). The pedunculopontine nucleus as a target for deep brain stimulation. J. Neural Transm. *118*, 1461–8.

44. Grabli, D., Karachi, C., Welter, M.-L., Lau, B., Hirsch, E.C., Vidailhet, M., and François, C. (2012). Normal and pathological gait: what we learn from Parkinson's disease. J. Neurol. Neurosurg. Psychiatry *83*, 979–85.

45. Ferraye, M.U., Debû, B., Fraix, V., Goetz, L., Ardouin, C., Yelnik, J., Henry-Lagrange, C., Seigneuret, E., Piallat, B., Krack, P., *et al.* (2010). Effects of pedunculopontine nucleus area stimulation on gait disorders in Parkinson's disease. Brain *133*, 205–14.

46. Goetz, L., Bhattacharjee, M., Ferraye, M.U., Fraix, V., Maineri, C., Nosko, D., Fenoy, A.J., Piallat, B., Torres, N., Krainik, A., *et al.* (2018). Deep Brain Stimulation of the Pedunculopontine Nucleus Area in Parkinson Disease: MRI-Based Anatomoclinical Correlations and Optimal Target. Neurosurgery *84*, 506–518.

47. Welter, M.-L., Demain, A., Ewenczyk, C., Czernecki, V., Lau, B., El Helou, A., Belaid, H., Yelnik, J., François, C., Bardinet, E., *et al.* (2015). PPNa-DBS for gait and balance disorders in Parkinson's disease: a double-blind, randomised study. J. Neurol.

48. Palminteri, S., Boraud, T., Lafargue, G., Dubois, B., and Pessiglione, M. (2009). Brain hemispheres selectively track the expected value of contralateral options. J. Neurosci. *29*, 13465–72.

49. Palminteri, S., Serra, G., Buot, A., Schmidt, L., Welter, M.-L., and Pessiglione, M. (2013). Hemispheric dissociation of reward processing in humans: Insights from deep brain stimulation. Cortex *49*, 2834–2844.

50. Worbe, Y., Palminteri, S., Hartmann, A., Vidailhet, M., Lehericy, S., and Pessiglione, M. (2011). Reinforcement Learning and Gilles de la Tourette Syndrome. Arch Gen Psychiatry *68*, 1257–1266.

51. Yelnik, J., Bardinet, E., Dormont, D., Malandain, G., Ourselin, S., Tandé, D., Karachi, C., Ayache, N., Cornu, P., and Agid, Y. (2007). A three-dimensional, histological and deformable atlas of the human basal ganglia. I. Atlas construction based on immunohistochemical and MRI data. Neuroimage *34*, 618–38.

52. Bardinet, E., Bhattacharjee, M., Dormont, D., Pidoux, B., Malandain, G., Schüpbach, M., Ayache, N., Cornu, P., Agid, Y., and Yelnik, J. (2009). A three-dimensional histological atlas of the human basal ganglia. II. Atlas deformation strategy and evaluation in deep brain stimulation for Parkinson disease. J. Neurosurg. *110*, 208–19.

53. Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., *et al.* (2012). Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. Neuron *76*, 998–1009.

54. Palminteri, S., Lefebvre, G., Kilford, E.J., and Blakemore, S.-J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. PLOS Comput. Biol. *13*, e1005684.

55. Daunizeau, J., Adam, V., and Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. PLoS Comput. Biol. *10*, e1003441.

56. Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150–

157.

57. Scheeringa, R., Fries, P., Petersson, K.-M., Oostenveld, R., Grothe, I., Norris, D.G., Hagoort, P., and Bastiaansen, M.C.M. (2011). Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. Neuron *69*, 572–83.

58. Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., *et al.* (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals. Nat. Neurosci. *23*, 664–675.

59. Brown, P., and Williams, D. (2005). Basal ganglia local field potential activity: Character and functional significance in the human. Clin. Neurophysiol. *116*, 2510–2519.

60. Brittain, J.-S., and Brown, P. (2014). Oscillations and the basal ganglia: Motor control and beyond. Neuroimage *85*, 637–647.

61. Khanna, P., and Carmena, J.M. (2015). Neural oscillations: beta band activity across motor networks. Curr. Opin. Neurobiol. *32*, 60–67.

62. Androulidakis, A.G., Mazzone, P., Litvak, V., Penny, W., Dileone, M., Doyle, L.M.F., Tisch, S., Di, V., and Brown, P. (2008). Oscillatory activity in the pedunculopontine area of patients with Parkinson's disease. Exp. Neurol. 211 *211*, 59–66.

63. Stefani, A., Lozano, A.M., Peppe, A., Stanzione, P., Galati, S., Tropepi, D., Pierantozzi, M., Brusa, L., Scarnati, E., and Mazzone, P. (2007). Bilateral deep brain stimulation of the pedunculopontine and subthalamic nuclei in severe Parkinson's disease. Brain *130*, 1596–607.

64. Frank, M.J., Samanta, J., Moustafa, A.A., and Sherman, S.J. (2007). Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science *318*, 1309–12.

65. Czernecki, V., Schüpbach, M., Yaici, S., Lévy, R., Bardinet, E., Yelnik, J., Dubois, B., and Agid, Y. (2008). Apathy following subthalamic stimulation in Parkinson

disease: A dopamine responsive symptom. Mov. Disord. *23*, 964–969.

66. Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. Nat. Rev. Neurosci. *17*, 183–195.

67. Gut, N.K., and Mena-Segovia, J. (2019). Dichotomy between motor and cognitive functions of midbrain cholinergic neurons. Neurobiol. Dis. *128*, 59–66.

68. Nowacki, A., Galati, S., Ai-Schlaeppi, J., Bassetti, C., Kaelin, A., and Pollo, C. (2019). Pedunculopontine nucleus: An integrative view with implications on Deep Brain Stimulation. Neurobiol. Dis. *128*, 75–85.

69. Palminteri, S., Lebreton, M., Worbe, Y., Grabli, D., Hartmann, A., and Pessiglione, M. (2009). Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. Proc. Natl. Acad. Sci. U. S. A. *106*, 19179–84.

70. Bódi, N., Kéri, S., Nagy, H., Moustafa, A., Myers, C.E., Daw, N., Dibó, G., Takáts, A., Bereczki, D., and Gluck, M. a (2009). Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain *132*, 2385–95.

71. Rutledge, R.B., Lazzaro, S.C., Lau, B., Myers, C.E., Gluck, M.A., and Glimcher, P.W. (2009). Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson's Patients in a Dynamic Foraging Task. J. Neurosci. *29*, 15104–15114.

72. Schmidt, L., Braun, E.K., Wager, T.D., and Shohamy, D. (2014). Mind matters: placebo enhances reward learning in Parkinson's disease. Nat. Neurosci. *17*, 1793–1797.

73. Skvortsova, V., Degos, B., Welter, M.-L., Vidailhet, M., and Pessiglione, M. (2017). A Selective Role for Dopamine in Learning to Maximize Reward But Not to Minimize Effort: Evidence from Patients with Parkinson's Disease. J. Neurosci. *37*, 6087–6097.

74. David Smith, A., and Paul Bolam, J. (1990). The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. Trends Neurosci. *13*, 259–265.

75. Aravamuthan, B.R., McNab, J.A., Miller, K.L., Rushworth, M., Jenkinson, N.,

Stein, J.F., and Aziz, T.Z. (2009). Cortical and subcortical connections within the pedunculopontine nucleus of the primate Macaca mulatta determined using probabilistic diffusion tractography. J. Clin. Neurosci. *16*, 413–20.

76. Mena-Segovia, J., and Bolam, J.P. (2017). Rethinking the Pedunculopontine Nucleus: From Cellular Organization to Function. Neuron *94*, 7–18.

77. Geisler, S., Derst, C., Veh, R.W., and Zahm, D.S. (2007). Glutamatergic afferents of the ventral tegmental area in the rat. J. Neurosci. *27*, 5730–5743.

78. Tian, J., Huang, R., Cohen, J.Y., Osakada, F., Kobak, D., Machens, C.K., Callaway, E.M., Uchida, N., and Watabe-Uchida, M. (2016). Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. Neuron *91*, 1374–1389.

79. Palminteri, S., and Pessiglione, M. (2017). Chapter 23 - Opponent Brain Systems for Reward and Punishment Learning: Causal Evidence From Drug and Lesion Studies in Humans. In, J.-C. Dreher and L. B. T.-D. N. Tremblay, eds. (San Diego: Academic Press), pp. 291–303.

80. Zitella, L.M., Mohsenian, K., Pahwa, M., Gloeckner, C., and Johnson, M.D. (2013). Computational modeling of pedunculopontine nucleus deep brain stimulation. J. Neural Eng. *10*, 045005.

81. LeBouc, R., Rigoux, L., Schmidt, L., Degos, B., Welter, M.-L., Vidailhet, M., Daunizeau, J., and Pessiglione, M. (2016). Computational dissection of dopamine motor and motivational functions in humans. J. Neurosci. *36*, 6623–6633.

82. Thobois, S., Ardouin, C., Lhommée, E., Klinger, H., Lagrange, C., Xie, J., Fraix, V., Coelho Braga, M.C., Hassani, R., Kistner, A., *et al.* (2010). Non-motor dopamine withdrawal syndrome after surgery for Parkinson's disease: predictors and underlying mesolimbic denervation. Brain *133*, 1111–27.

83. Wellcome Department of Imaging Neuroscience, London, U. Cogent 2000.

84. Rigoux, L., Stephan, K.E., Friston, K.J., and Daunizeau, J. (2014). Bayesian model selection for group studies - Revisited. Neuroimage *84*, 971–85.

85. Friston, K.J., and Stephan, K.E. (2009). Free energy and the brain. Synthese *159*,

1–39.

86. Penny, W.D. (2012). Comparing dynamic causal models using AIC, BIC and free energy. Neuroimage *59*, 319–30.

87. Wellcome Trust Center for Neuroimaging, London, U. Statistical Parametric Mapping Toolbox for fMRI data analysis.

88. Mitra, P., and Hemant, B. (2008). Observed Brain Dynamics (New York: Oxford University Press).

89. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–90.

90. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869.

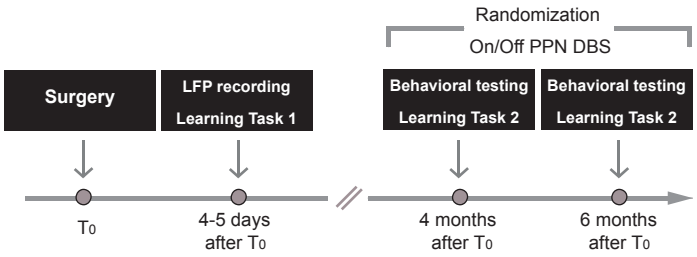
KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Bacterial and Virus Strains		
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Critical Commercial Assays		
Deposited Data		
Experimental Models: Cell Lines		
Experimental Models: Organisms/Strains		

Human patients with Parkinson's disease	N/A	N/A
Oligonucleotides		
Recombinant DNA		
Software and Algorithms		
Cogent 2000 Matlab Toolbox	[83]	http://www.vislab.ucl.ac.uk/cogent_2000.php
Fieldtrip	[90]	https://www.fieldtripttoolbox.org/
Chronux Matlab library	[88]	http://chronux.org/.
Statistical Parametric Mapping Toolbox for fMRI data analysis (SPM8)	[87]	https://www.fil.ion.ucl.ac.uk/spm/
Other		
Custom scripts for data analyses	The custom Matlab codes are available on request to corresponding authors.	N/A

Figure1

A


Randomization
On/Off PPN DBS
Surgery
LFP recording
Learning Task 1
Behavioral testing
Learning Task 2
Behavioral testing
Learning Task 2
T0
4-5 days
after T0
4 months
after T0
6 months
after T0

B

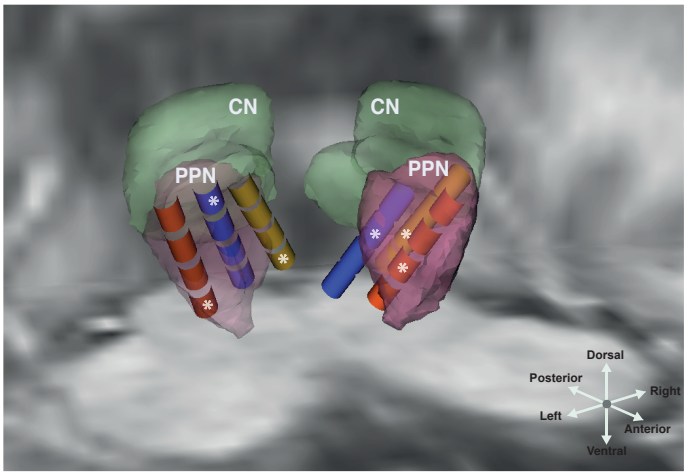
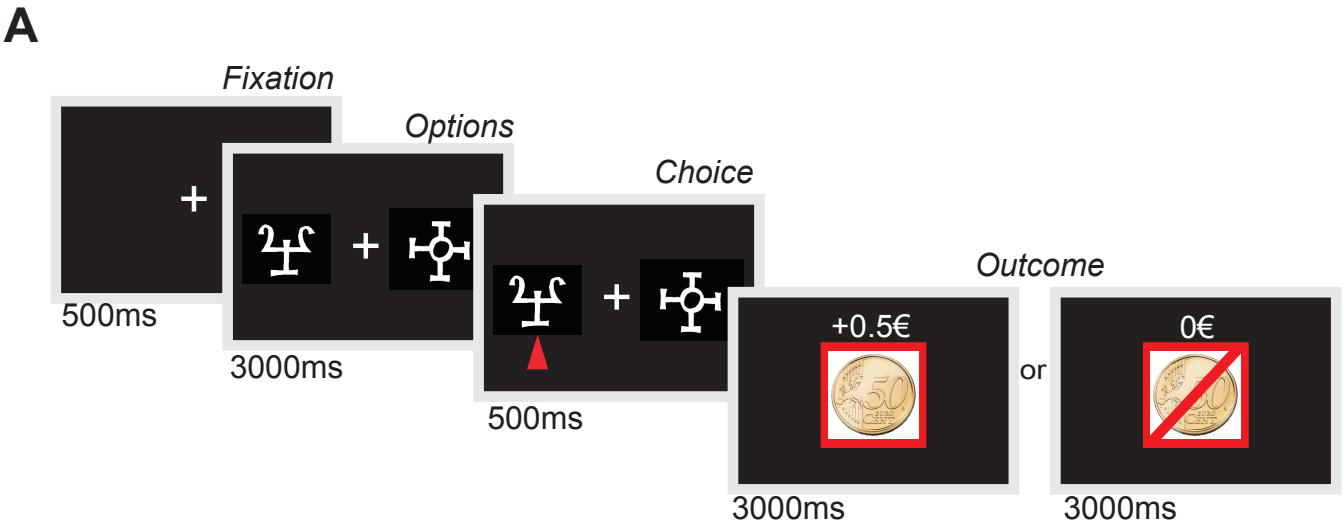

CN
CN
PPN
PPN
Dorsal
Posterior
Right
Left
Anterior
Ventral

Figure2


A
Fixation
Options
Choice
Outcome
500ms
3000ms
500ms
+0.5€
or
0€
3000ms
3000ms

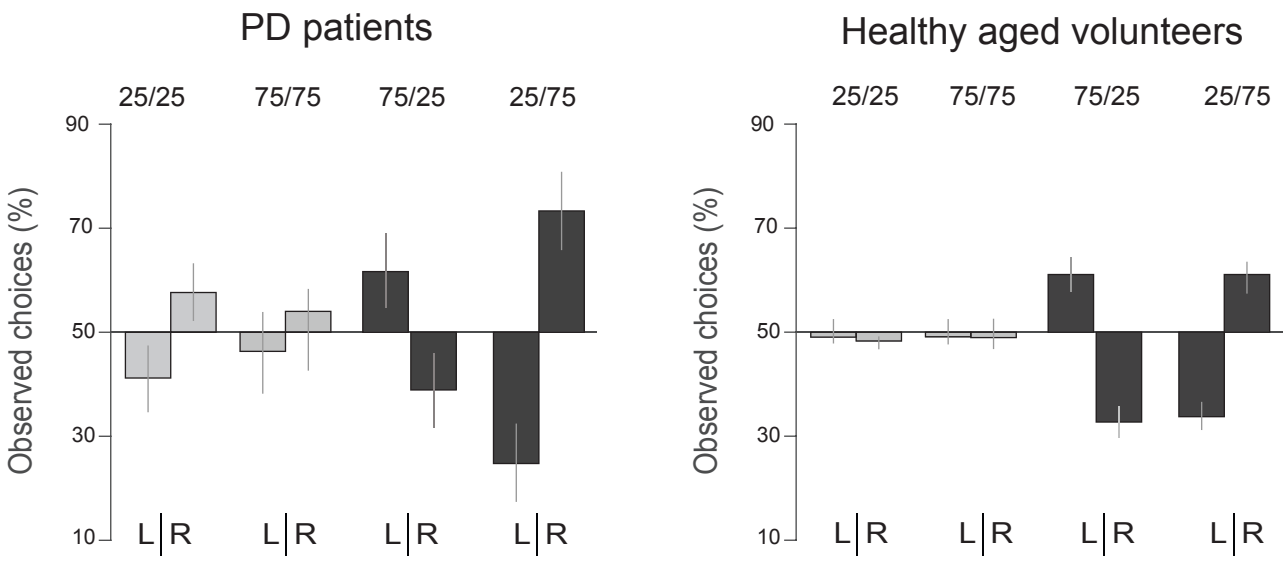
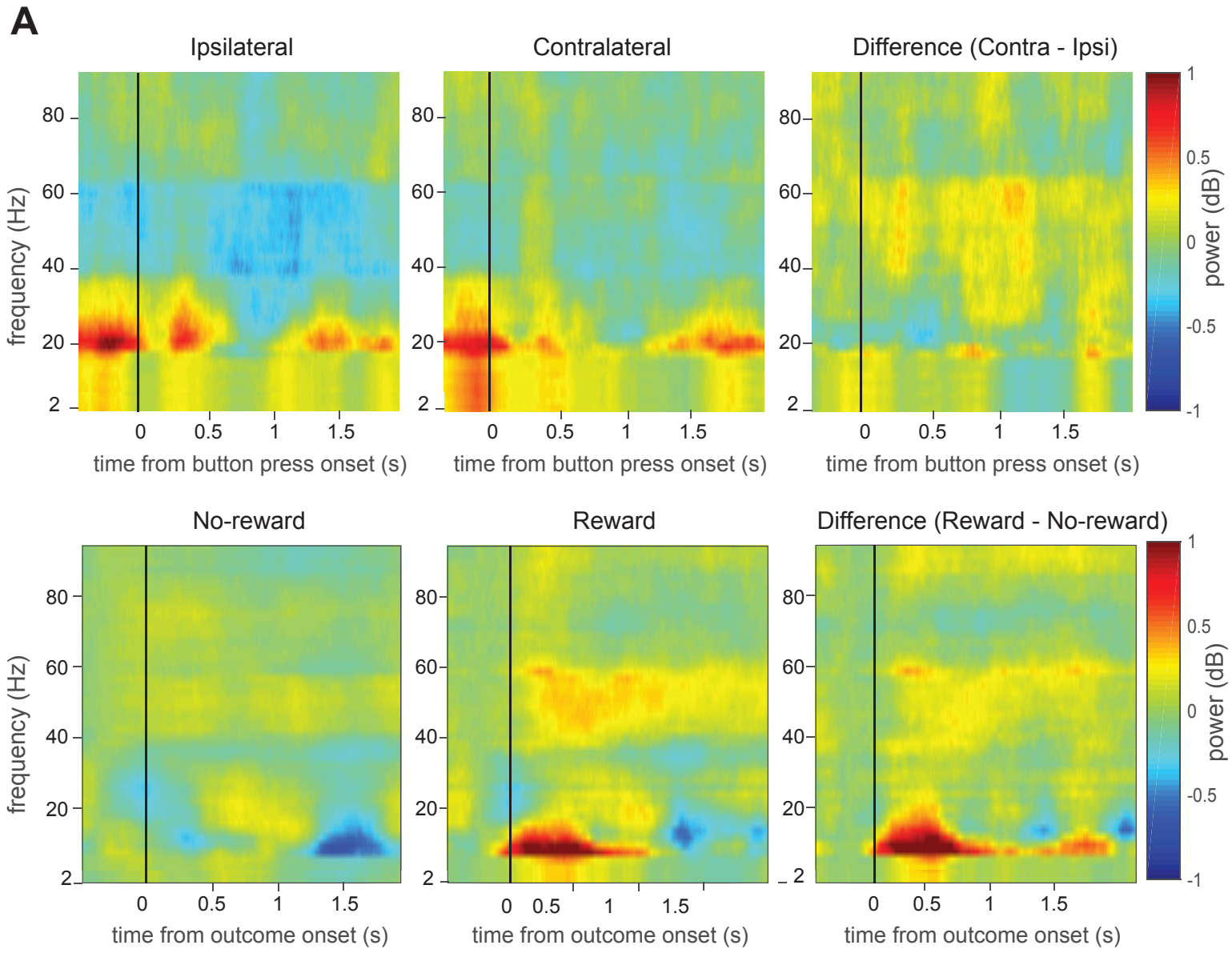
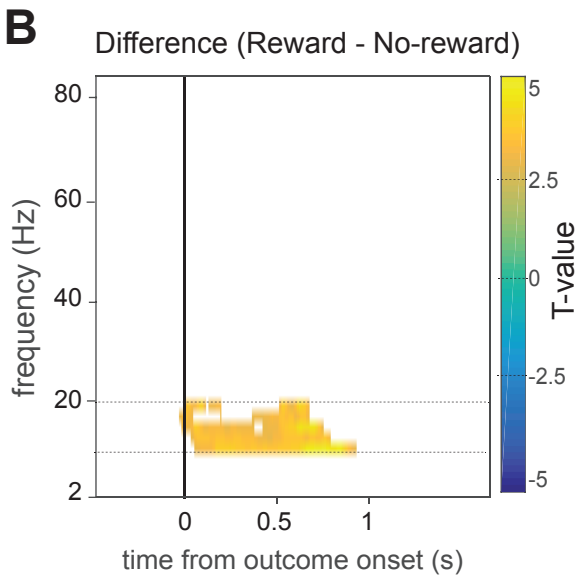

B
PD patients
Healthy aged volunteers
25/25
75/75
75/25
25/75
Observed choices (%)
90
70
50
30
10
L R

Figure3


A
Ipsilateral
Contralateral
Difference (Contra - Ipsi)
frequency (Hz)
time from button press onset (s)
power (dB)
No-reward
Reward
Difference (Reward - No-reward)
time from outcome onset (s)


B
Difference (Reward - No-reward)
frequency (Hz)
time from outcome onset (s)
T-value

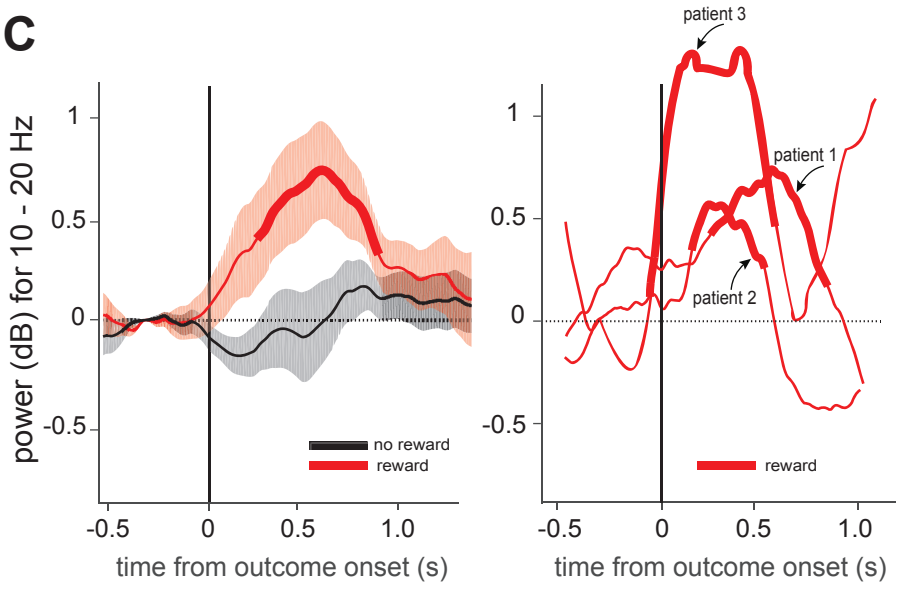

C
power (dB) for 10 - 20 Hz
time from outcome onset (s)
no reward
reward
patient 3
patient 1
patient 2

Figure4

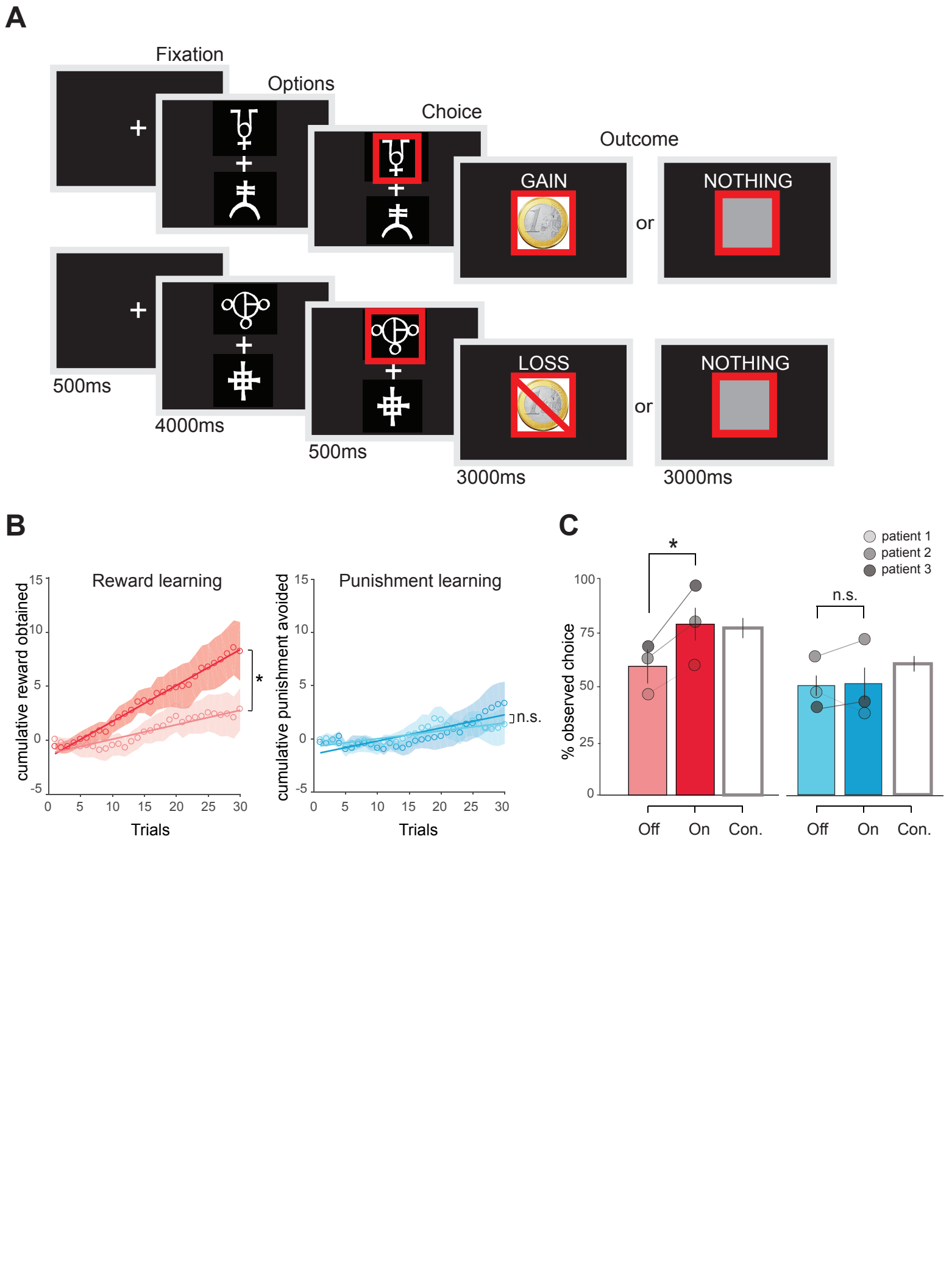
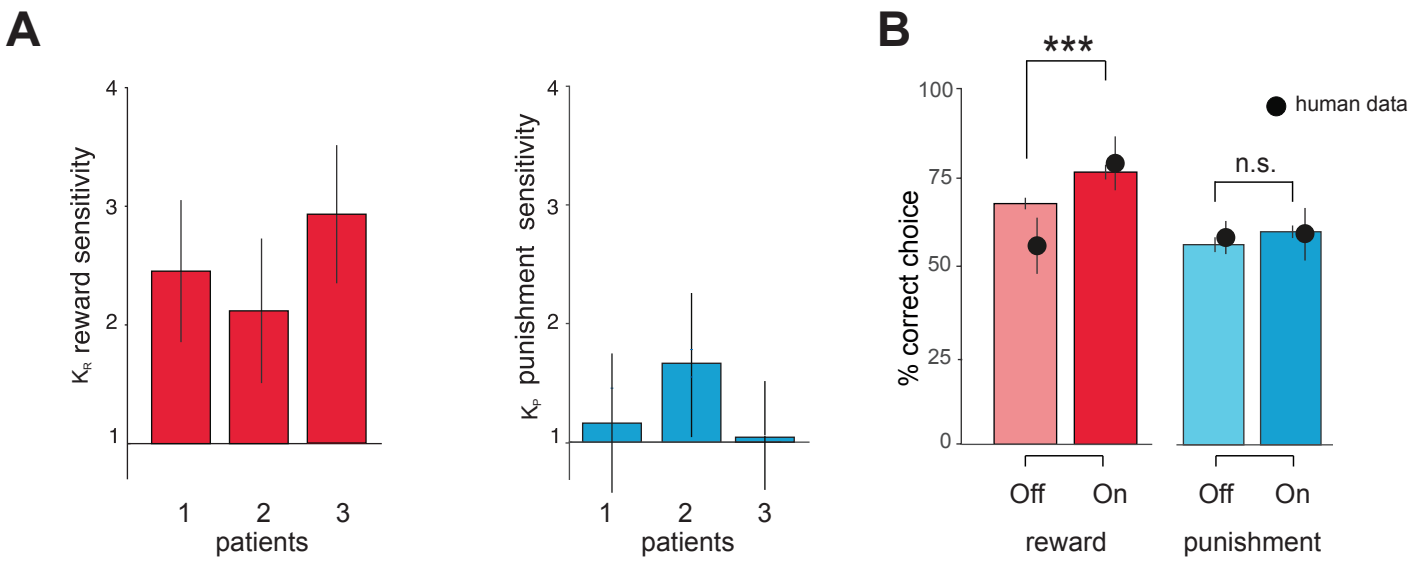

A
Fixation
Options
Choice
Outcome
GAIN
NOTHING
or
LOSS
NOTHING
or
500ms
4000ms
500ms
3000ms
3000ms
B
Reward learning
cumulative reward obtained
Trials
*
Punishment learning
cumulative punishment avoided
Trials
n.s.
C
patient 1
patient 2
patient 3
% observed choice
*
n.s.
Off
On
Con.
Off
On
Con.

Figure5


A
B
K_R reward sensitivity
K_P punishment sensitivity
patients
patients
1
2
3
4
***
n.s.
human data
% correct choice
100
75
50
25
0
Off
On
Off
On
reward
punishment

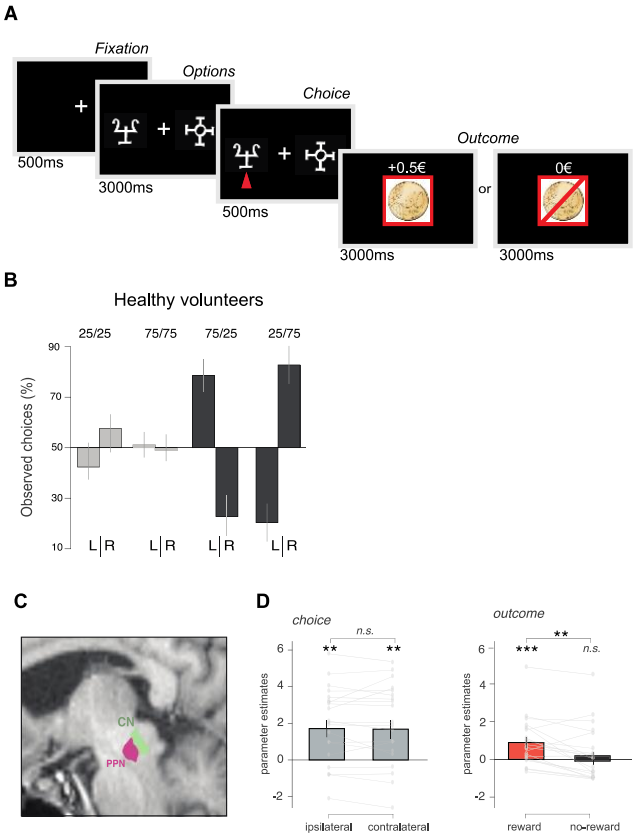


Figure S1. PPN hemodynamic response to reward. Related to Figure 2.

A. Example trial of learning task 1 (same as Figure 2). Screenshots are shown from left to right, with durations in milliseconds. On every trial, subjects selected between left and right options represented by two visual cues, using their left-hand or right-hand index to press the corresponding button. The side of the selected cue (left in the example) was marked with a red pointer. Subjects could then observe the outcome of their choice (a 0.5€ reward or nothing) and update their estimates of cue-reward contingencies. Each session presented four different pairs of cues, associated with different combinations of reward probability (25/25, 75/75, 75/25, 25/75 %).

B. Behavioral performance of young healthy volunteers (N=20) tested during fMRI data acquisition. Histograms show the choice rate observed with the four pairs of cues associated to varying reward probability (light gray: symmetrical pairs, dark grey: asymmetrical pairs). Even if the pattern of behavioral performance was similar to that of PD patients, young healthy volunteers were better at learning which cue was more rewarded (for the two asymmetrical pairs in dark grey). Error-bars are inter-subject S.E.M. The plot is reproduced with permission from[S1].

C. Sagittal view showing the localization of the PPN (purple) and adjacent CN (green) masks taken from an histological atlas of the basal ganglia[S2]. Masks are superimposed on the average T1 of healthy volunteers (N = 20) who performed learning task 1 in the MRI scanner, normalized to the MNI brain template.

D. Contrast estimates for ipsilateral vs. contralateral choice (left panel) and reward vs. no reward outcome (right panel). Regression estimates were extracted from bilateral PPN masks, contrasted at the subject level and then tested at the group level. Dots correspond to individual subject estimates. Hemodynamic activity was increased at the time of choice, but unaffected by response side. It was also increased at the time of outcome, specifically when reward was delivered. Error bars are inter-subject S.E.M. *** $p < 0.001$, ** $p < 0.01$, *n.s.* not significant.

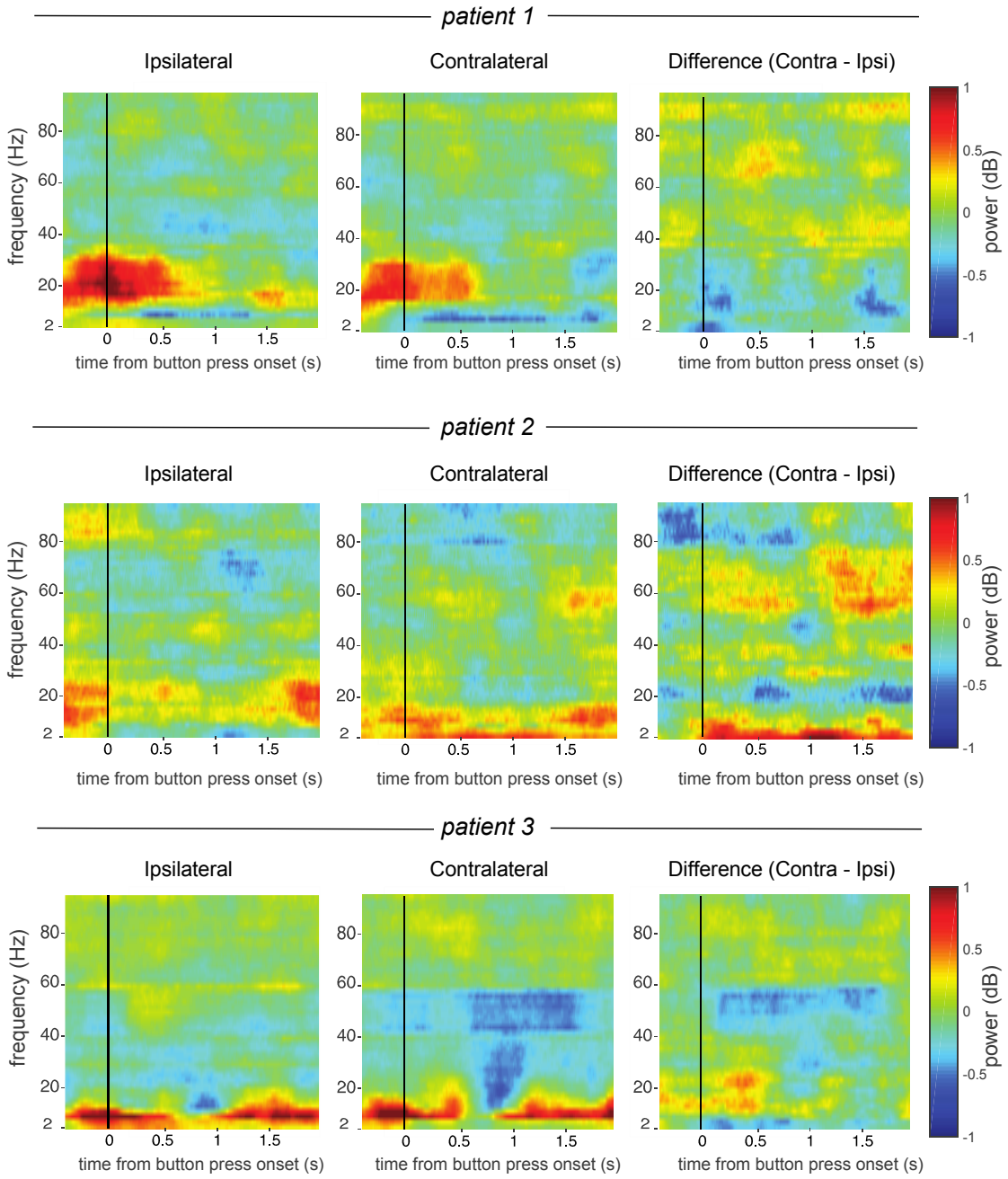


Figure S2. Individual time-frequency maps of choice-evoked activity. Related to Figure 3.

Color code indicates power observed in each time-frequency bin of the map. Power was corrected for baseline measure (over a -500 to 0ms time window prior to fixation onset). Although increased activity was observed in the alpha-beta band around choice onset in all three patients, there was no significant difference between ipsi- and contra-lateral responses.

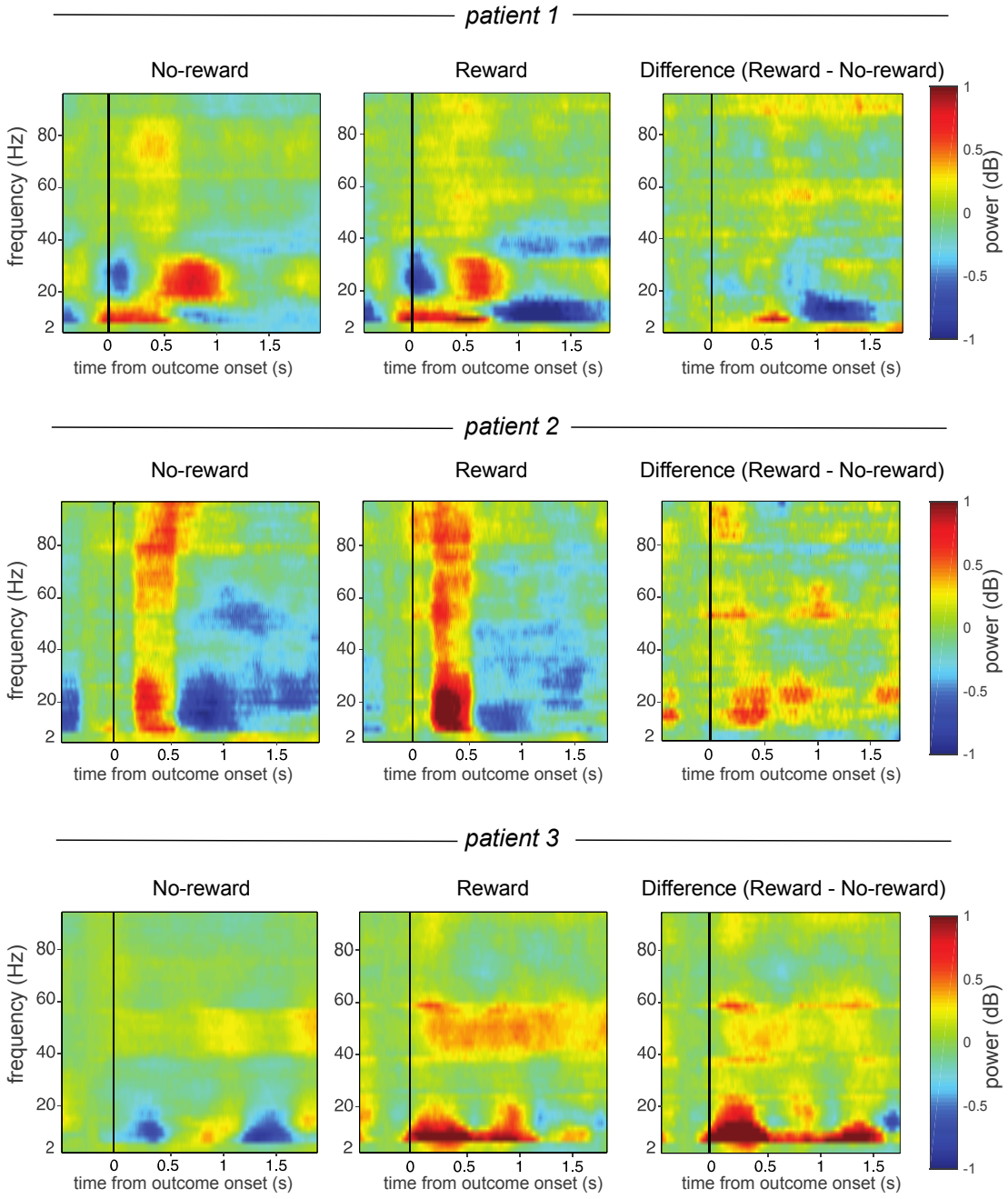


Figure S3. Individual time-frequency maps of outcome-evoked activity. Related to Figure 3.

Color code indicates power observed in each time-frequency bin of the map. Power was corrected for baseline measure (over a -500 to 0ms time window prior to fixation onset). Higher activity in the alpha-beta band following reward outcome was observed in all three patients.

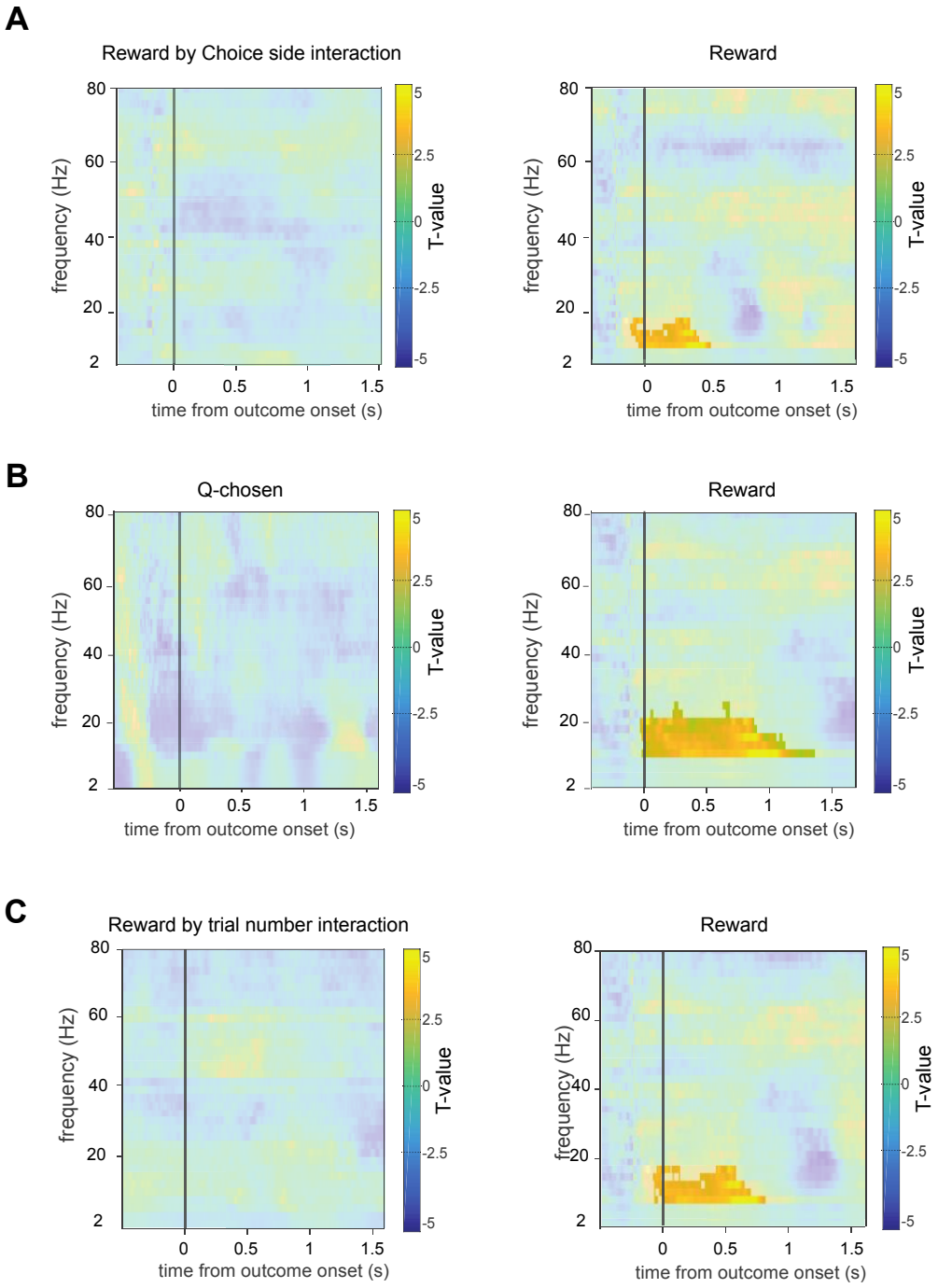


Figure S4. PPN potentials evoked by reward in alternative regression analyses. Related to Figure 3.

Time-frequency maps show T-values of the contrast between reward and no reward outcomes, in a GLM that also contained choice side (ipsi or contralateral to the recording contact) plus its interaction with reward outcome in A, reward expectation (chosen Q-value) in B, and trial number (within sessions) plus its interaction with reward outcome in C. Maps were averaged over all sessions and all available contacts in all three patients. Color code indicates the T-value obtained in each time-frequency bin of the map. Plain colors delineate the significant cluster ($p < 0.05$ after correction for multiple comparisons), whereas more transparent colors denote non-significant bins. None of the additional regressors yielded any significant activation, and the significant cluster signaling reward outcomes was similar in all cases.

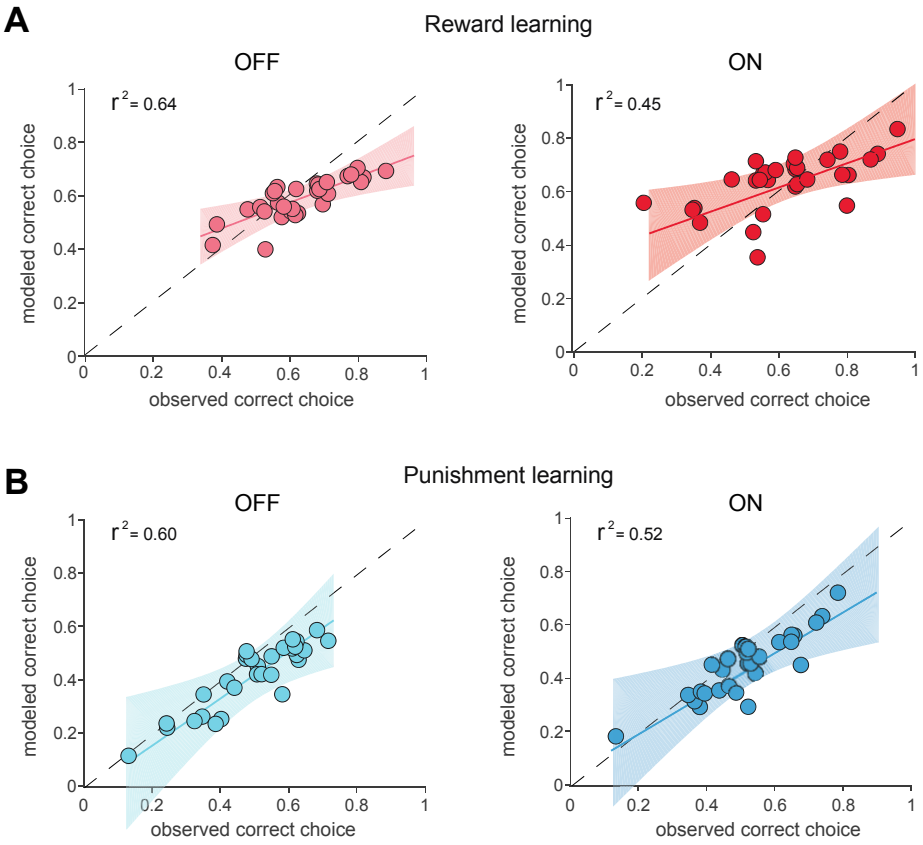


Figure S5. Quality of model fitting. Related to Figure 5.

Scatter plots show inter-trial Pearson's correlations between observed and modeled correct choice rates for reward and punishment learning in the On and Off stimulation states. PPN DBS effects were captured by the winning model with modulation of the subjective sensitivity to reward K_R. Solid lines are linear regression fits; shaded areas indicate 95% confidence intervals around linear regression estimates. Percentage of explained variance (r^2 estimates) is provided separately for every condition.

Patients	Gender,Age	Disease duration (years)	UPDRS score On/Off	L-dopa dosage (mg)	MMS score	MDRS score	WCST score
1	F, 70	18	16/45	570	25	140	15
2	F, 64	10	19/38	1050	25	141	12
3	M, 46	12	19/50	1300	27	139	20

Table S1. Demographic and clinical characteristics of patients. Related to Figure 1.

UPDRS - Unified Parkinson's disease rating scale, part III (range: 0 – 108; higher score indicates worse motor function; On and Off scores indicate motor performance with and without levodopa); medication dosage is expressed as levodopa equivalent; MMS - Mini-Mental State examination (range: 0 – 30); MDRS - Mattis Dementia Rating Scale (range 0 – 144); WCST - Wisconsin Card Sorting Test (range 0 – 20). For MMS, MDRS and WCST, higher score indicates better cognitive functioning.

Patients	Bipolar contacts used in LFP data analysis		Total number of contacts available	Contacts used for DBS		Coordinates in mm of the stimulated contacts (+ Left, - Right)			Stimulation parameters F(Hz) / P(ms) / A(V)	
	Left	Right		Left	Right	L	D	H	Left	Right
1	0-1 1-2 2-3	0-1 1-2 2-3	6	0-1 +	1-2 +	3.4 -5.6	8.7 7.8	-1.4 -1.6	40/60/ 3.1	40/60 3.1
2	1-2 2-3	1-2 2-3	4	3-2 +	2-3 +	3.2 -5.2	8.7 9.2	-2.5 -5.4	40/60/ 1.2	40/60/ 0.8
3	0-1 2-3	0-1 1-2 2-3	5	0-1 +	1-0 +	3.6 -2.2	6.3 7.8	-3.5 -4.6	20/30/ 1.3	20/60/ 2.4

Table S2. PPN electrodes and stimulation characteristics. Related to Figure 1.
LFP activity was computed as the difference between recordings from two adjacent electrodes, going from more ventral (contact 0) to more dorsal (contact 3). Lower indices for bipolar contacts indicate more ventral contact location. The coordinates of the stimulated contacts are given in millimeters from midline: L - laterality (– right side, + left side), D – ventro-dorsal distance from the floor of the fourth ventricle, H – rostro-caudal distance from a ponto-mesencephalic junction to the inferior colliculi caudal margin (– above this line, + below this line). Reproduced with permission from[S3].

Patients	Model 0	Model K_R	Model K_P	Model α_R	Model α_P	Model β
1	-166.52	-156.13	-167.10	-169.38	-168.58	-158.61
2	-178.16	-175.86	-177.92	-178.58	-178.60	-178.11
3	-104.92	-96.09	-108.01	-106.43	-107.05	-105.51
Summed	-449.53	-428.09	-453.03	-454.40	-454.24	-442.23
Expected Frequency	0.0417	0.7916	0.0417	0.0417	0.0417	0.0417
Exceedance Probability	0.0098	0.9478	0.0100	0.0085	0.0117	0.0122

Table S3. Results of Bayesian model comparison. Related to Figure 5.

The top four rows indicate approximated model evidence for each of the six models fitted to individual choices and summed across patients for the fixed-effect model selection (lower values indicate better fit). The two rows at the bottom indicate expected frequencies and exceedance probabilities obtained from random-effect model selection.

Patients	Reward learning rate	Punishment learning rate	Softmax inverse choice temperature	K_R reward sensitivity scaling parameter
1	0.22	0.26	1.23	2.43
2	0.47	0.60	3.57	2.09
3	0.70	0.61	0.39	2.89

Table S4. Fitted computational parameters. Related to Figure 5.

Lines show for the three patients the mean of each free parameter posterior distribution, obtained with the best-fitting model where PPN DBS only modulated reward sensitivity (K_R).

Supplemental References

S1. Palminteri, S., Boraud, T., Lafargue, G., Dubois, B., and Pessiglione, M. (2009). Brain hemispheres selectively track the expected value of contralateral options. J. Neurosci. *29*, 13465–72.

S2. Bardinet, E., Bhattacharjee, M., Dormont, D., Pidoux, B., Malandain, G., Schüpbach, M., Ayache, N., Cornu, P., Agid, Y., and Yelnik, J. (2009). A three-dimensional histological atlas of the human basal ganglia. II. Atlas deformation strategy and evaluation in deep brain stimulation for Parkinson disease. J. Neurosurg. *110*, 208–19.

S3. Welter, M.-L., Demain, A., Ewenczyk, C., Czernecki, V., Lau, B., El Helou, A., Belaid, H., Yelnik, J., François, C., Bardinet, E., *et al.* (2015). PPNa-DBS for gait and balance disorders in Parkinson's disease: a double-blind, randomised study. J. Neurol.