Stimulation of the vagus nerve reduces learning in a go/no-go reinforcement learning task

Anne Kühnel[1*], Vanessa Teckentrup[2], Monja P. Neuser[2],
Quentin J. M. Huys[?], Caroline Burrasch[2], Martin Walter[2-5], & Nils B. Kroemer[2*]

[1] Max Planck Institute of Psychiatry and International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany

[2] Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany

[3] Otto-von-Guericke University Magdeburg, Department of Psychiatry and Psychotherapy, Germany

[4] Clinical Affective Neuroimaging Laboratory, Magdeburg, Germany

[5] Leibniz Institute for Neurobiology, Magdeburg, Germany

Corresponding authors*

Anne Kühnel, anne_kuehnel@psych.mpg.de
Dr. Nils B. Kroemer, nils.kroemer@uni-tuebingen.de
Calwerstr. 14, 72076 Tübingen, Germany

Abstract

When facing decisions to approach rewards or to avoid punishments, we often figuratively go with our gut. While the impact of metabolic state such as hunger on motivation is well documented, the role of vagal feedback signals originating from the gut in adjusting instrumental actions is still largely elusive. Consequently, we investigated the effect of non-invasive transcutaneous vagus nerve stimulation (tVNS) vs. sham (randomized cross-over design) on approach and avoidance behavior using an established go/no-go reinforcement learning paradigm (Guitart-Masip et al., 2012) in 39 healthy, overnight-fasted participants. First, mixed-effects logistic regression analysis of choice accuracy showed that tVNS acutely impaired learning, p = .045, regardless of the required action or valence of the reward. Second, in line with mixed-effects results, computational reinforcement learning models showed that tVNS acutely reduced the learning rate *(*$\Delta\alpha$ = -0.092, p_{boot} = .002) and these changes were more pronounced for trials incurring punishment *(*$\Delta\alpha_{Pun}$ = -0.081, p_{boot} = .012 vs. $\Delta\alpha_{Rew}$ = -0.031, p = .22). However, tVNS had no effect on go biases, pavlovian response biases or response time indicating that changes in performance were not driven by changes in action execution, but speed of contingency learning. To conclude, our results highlight a novel role of vagal afferent input in modulating reinforcement learning by tuning the learning rate according to homeostatic needs.

Introduction

To survive, organisms must procure energy by approaching options that pay off while avoiding costly options, potentially incurring punishments. Fundamental learning mechanisms have evolved to support this vital optimization of instrumental actions [1–4]. One key challenge is to balance short-term and long-term goals of reward-related behavior. For example, being patient in light of temptation to receive bigger returns in the future is often beneficial to maximize long-term outcomes. However, forfeiting immediate rewards during a hungry state [5] in the same manner may put an individual at risk of starvation [1]. Likewise, increasing the thriftiness of actions may help to promptly improve the energy balance even though one might fail to explore an option that could be much better in the long run [6]. Although it seems imperative to adjust value-based decisions and learning according to homeostatic needs, little is known about the neurobiological mechanisms subserving such adaptations in humans to date. One plausible candidate for modulatory input onto circuits involved in reward learning would be a caloric feedback signal [7] originating from the gut.

Signals about metabolic and homeostatic state are largely transmitted via the vagus nerve which connects peripheral organs such as the gut with the brain. Vagal afferents terminate in the nucleus tractus solitarii, NTS, [8] a hub further relaying metabolic information to the mid- and forebrain [8,9] including to dopaminergic neurons in the substantia nigra. Within that pathway, vagal afferents have been shown to modulate dopaminergic [10,11], but also noradrenergic [12], and cholinergic signaling [13]. Accordingly, endogenous stimulation of the gut with nutrients evokes dopamine responses in the dorsal striatum tracking caloric value [14,15]. These dopamine signals are critical for appetitive conditioned learning [10,16,17] as well as motivated behavior [18,19]. Similarly, vagal afferent signalling regulates food intake [16] and stimulation of the vagus nerve has been associated with reduced food intake and decreased weight gain [20]. More broadly, episodic and spatial memory function [21] as well as cognitive flexibility [22] and mood [23] are also influenced by vagal signaling. Collectively, these results suggest that vagal signals conveying the metabolic state influence a wide variety of behaviors including appetitive learning via alterations in dopaminergic signaling.

Until recently, research in humans has been limited by the invasive nature of cervical vagus nerve stimulation (VNS). Still, VNS has been related to a broad range of behavioral effects such as enhancing cognitive functioning [24] and memory retention [25,26]. Additionally, VNS has repeatedly been shown to reduce depressive and anxiety symptoms [25]. Lately, non-invasive transcutaneous VNS (tVNS) has become feasible, opening new avenues for research and treatment. It is commonly applied via the ear targeting the auricular branch of the vagus nerve, which has been shown to affect projections to the NTS in preclinical studies [27]. Comparably, studies using tVNS with concurrent fMRI have revealed enhanced activity in the NTS and other interconnected brain regions including the dopaminergic midbrain and nucleus accumbens [28,29]. Moreover, tVNS had similar positive effects on depressive symptoms [30,31], memory retention [32,33], and cognitive performance [34,35] as implanted cervical VNS. Despite the recent progress, detailed understanding of the link between vagally mediated metabolic signaling and reward-related alterations, which could explain anti-depressant effects, is lacking.

Previous studies investigating the effects of (t)VNS mainly focused on processes predominantly associated with the noradrenergic system such as cognitive control, fear learning and extinction. Nonetheless, since vagal signals also modulate dopaminergic transmission, reward learning and motivation should also be influenced by vagal stimulation. While response vigor has been linked to tonic dopamine concentration [36] potentially reflecting average reward rates, learning via reward prediction errors (RPE) has been linked to phasic dopamine signals [37,38]. The magnitude of RPEs and subsequent learning is in turn influenced by tonic dopamine levels [39,40]. Perhaps counterintuitively, high dopamine tone may reduce the constraint of actions imposed by previous rewards [40] as phasic signals are then proportionally smaller (value theory, [41]) or choices rely less on learned values (thrift theory, [2]). Moreover, learning is differentially supported by the dopamine circuitry depending on the specific task [42], with distinct but interacting circuits underlying learning from rewards or punishment. Consequently, vagal feedback may lead to alterations in reward learning, response vigor, and action selection mediated by dopaminergic signaling and mapping the effects of tVNS onto these motivational facets would shed new light on the endogenous modulation of reward seeking.

In the present crossover study, we therefore applied tVNS (vs. sham) to mimic metabolic signaling via vagal afferents and tested its effects on reward learning, which may be mediated by changes in dopamine levels. Reward learning was probed with a valence dependent go/no-go learning paradigm established by Guitart-Masip et al. [42] delineating instrumental action learning and pavlovian control. In line with the value and thrift theories of dopamine, we hypothesized that participants' performance during tVNS would be impaired, as increased tonic dopamine levels reduce the impact of phasic RPEs. We then used computational reward learning models to investigate specific valence or action dependent changes in performance. In a previous study, Guitart-Masip et al. [43] reported a decrease in pavlovian bias by increased dopamine levels (after L-DOPA administration) thereby improving

Commented [1]: Has this actually been an a-priori hypothesis or a result of the exploration?

performance in incongruent action-valence combinations while reducing performance in congruent action-valence combinations. Hence, we expected tVNS to lead to a similar attenuation of the pavlovian influence on instrumental learning. In addition, we explored potential effects of tVNS on greater go response rates or faster response time, which would be indicative of a heightened response vigor.

Methods

Participants

In total, 44 individuals participated in the study. They were physically and mentally healthy, German speaking, and right-handed, as determined by a telephone interview (24 female; M_{age}= 25.5 years ± 3.7; M_{BMI}= 23.0 ± 3.0; 17.93 - 30.9 kg/m^2). For the current analysis, five participants had to be excluded (n=4: did not complete both sessions of the task, n=1: did not make any go response). The final sample included 39 participants (23 female, M_{age}= 25.5 years ± 4.0; M_{BMI}= 23.0 ± 3.0). The institutional review boards of the University of Tübingen approved the study and we obtained informed consent from all participants prior to taking part in the experiment.

Experimental procedure

Participants were required to fast overnight (i.e., >8h hours prior to the visit) for both experimental sessions. Sessions were conducted in a randomized, single-blind manner as the experimenter was not blind to the stimulation condition (for information on the device, see SI). Nevertheless, participants were close to chance in guessing the correct condition (60%; p_{binomial} = .049) suggesting that blinding was effective. Sessions started between 7.00 am and 10:15 am and lasted about 2.5h in total. After participants arrived for the first session, they provided written informed consent. Next, we collected anthropometric and state-related information (see SI) before the tVNS electrode was placed on the left ear to target the auricular branch of the vagus nerve. In line with the stimulation procedure by Frangos et al. [28], the electrode was located at the left cymba concha for tVNS and at the left earlobe for sham stimulation. Stripes of surgical tape served to secure the electrode in place. Individual stimulation strength was assessed for every session separately using pain VAS ratings ("How intensely do you feel pain induced by the stimulation?” ranging from 0 (“no sensation”) to 10 (“strongest sensation imaginable”). Stimulation was initiated at an amplitude of 0.1 mA and increased by the experimenter by 0.1-0.2 mA at a time. Participants rated the sensation after every increment until ratings settled around the value 5 (corresponding to “mild prickling”). Then, the stimulation continued throughout the task block according to the default stimulation protocol of the device (i.e., alternating blocks of stimulation on and off for 30s each). Within this block, participants completed a food-cue reactivity task (~20 min), an effort allocation task (~40 min), and the instrumental learning task (~15 min).

After completing state-related questions, participants received their monetary rewards according to task performance (M_{paid} = 5.79€ ± 2.41; 0.70-10.06) and compensation (either monetary as a 32€ fixed amount or partial course credit). Both visits followed the same standardized protocol.

Paradigm

We hypothesized that tVNS would affect reinforcement learning via changes in dopaminergic neurotransmission. However, increases in dopaminergic transmission are not universally translated into increases in performance as expected changes critically depend on the nature of the task. Due to the well-known characteristics of the dopaminergic circuit [44,45], we sought to disentangle effects of the stimulation on action- or valence-dependent learning. To this end, we used a previously established go/no-go reinforcement learning task [46–48]. In this task, participants learn state-action contingencies and receive rewards or punishments. Each trial consisted of three stages (Figure 1). First, participants saw a fractal cue (state) out of a set of four different fractals per session. These fractals were initially randomized to one of the four possible combinations of the go × win two-factorial design of the task.

Second, participants had to do a target detection task by pressing a button (go) or withholding their response (no-go). Third, they saw the outcome of the state-action combination, which was either a win (5 cents), omission (no win/punishment, 0 cents), or a loss (-5 cents). Using trial and error, participants had to learn which action following each fractal was best in terms of maximizing wins or minimizing losses.

To facilitate reinforcement learning, the outcomes were presented probabilistically. Thus, participants had 80% chances to win after correct state-action sequences, 20% chances to win after incorrect sequences for rewarded trials as well as 80% chances to avoid losses after correct and 20% chances to avoid losses after incorrect sequences for punished trials. Participants were instructed about the probabilistic nature of the task and that either go or no-go responses could be the correct response for a given fractal. There was no change in the contingencies over time. To ensure that participants understood the task, they were queried before starting the experiment. In total, the task included 240 trials (60 trials per condition) and took about 15 min to be completed.

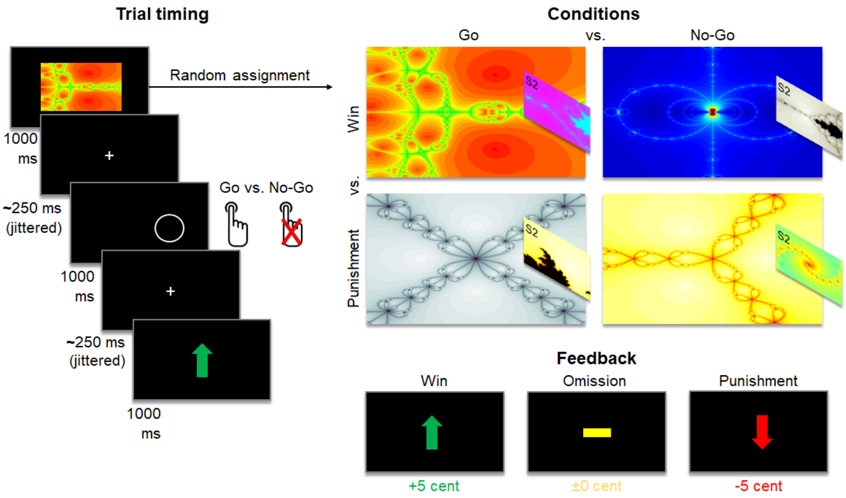

Figure 1: Schematic summary of the go/no-go reinforcement learning task. To maximize the total payoff, participants had to learn which action (go vs. no-go) during the target detection stage following a given fractal was associated with the best possible outcome (i.e., receiving reward or avoiding impending punishment). These contingencies were randomly assigned and had to be learned by trial and error. S2 = Session 2

Data analysis

Behavioral data

Full mixed-effects analysis of the go/no-go reinforcement learning task

To estimate the effects of tVNS on choice accuracy, we defined a full mixed-effects analysis as implemented in HLM 7 [49]. Effects of the conditions were modeled by predicting if a given choice (Bernoulli distribution) was correct based on the regressors go (dummy coded), win (dummy coded), and the interaction term go × win in a generalized linear model. To assess tVNS effects, the model included terms for the stimulation condition (dummy coded, 0 = sham, 1 = tVNS) and interactions of the stimulation term with the condition regressors (i.e., stimulation × go, stimulation × win, stimulation × go × win). Furthermore, we included a trial regressor capturing improvements in accuracy across trials, which was transformed by the natural logarithm and mean centered. At the participant level, we calculated two models that included random effects for all intercepts and slopes: model 1 controlled only for order whereas model 2 additionally controlled for gender, and BMI. We also tested an additional interaction term stimulation x ln(trial), but found that the coefficient estimate was highly correlated with the stimulation main effect. Thus, we excluded this term to avoid redundancy. All other random effects were complementary and showed significant between-subject variance (p < .001). Analogous to using expectation maximization in the computational model, we obtained empirical Bayes estimates, which take the distribution at the group level into account as individual estimates of the stimulation effects.

Reinforcement learning model

To dissociate which facet of instrumental action learning was altered by tVNS, we fit reinforcement-learning models to participant's behavior starting with the winning model detailed in Guitart-Masip et al. [42,46] as standard model. Here, participants learn stimulus (s) specific action (a) values (Q) that are updated at each trial t according to the Rescorla-Wagner rule as follows:

$$Q_t(s_t, a_t) = Q_{t-1}(s_t, a_t) + \alpha(\rho r_t - Q_{t-1}(s_t, a_t)),$$

with learning rate alpha ($\alpha \in [0,1]$), reward sensitivity rho, a positive free parameter quantifying the individual importance of reward and obtained rewards r_t coded as -1 in case of punishment, 1 in case of reward and 0 if participants received neither reward nor punishment. Further, agents learn action-independent values (V) of each state updated after the same rule indicating if a stimulus is associated with punishments or rewards.

$$V(s_t) \; = V_{t-1}(s_t) + \alpha(\rho r_t \; - \; V_{t-1}(s_t)),$$

Action values (Q) and stimulus values (V) at each trial are used to compute action weights as follows:

$$W_t(a,s) = \; \{Q_{t}(a,s) \; + b \; + \; \pi V_t(s), \;\; a = go \; Q_t(a,s), \qquad else$$

Where *b* is a free parameter that reflects a constant bias to choose the go option. The influence of pavlovian tendencies (e.g. increased go behavior in potentially rewarding situations and avoidance in aversive situations) is parameterized by π, a positive free parameter. The pavlovian parameter inhibits the go tendency in conditions that are associated with punishments and thus have negative learned state-values (V), while it increases go tendencies in conditions associated with reward and positive state-values. Consequently, this leads to impaired learning in incongruent (e.g. go-to-avoid punishment) trials.

The action at each trial is selected based on action probabilities that are estimated by passing action weights (W) through a softmax function (Eq. x) and adding a noise parameter $(lapse, \xi \; \in [0,1])$ modulating the influence of learned expectations on subsequent decisions.

$$p(s_t) = \left[\frac{exp\; exp\; (W(s_t))}{\sum_a \quad exp\;(W(a'|s_t}\right](1 - \xi) + \frac{\xi}{2}$$

Subsequently, we fit three further models to disentangle possible effects depending on reward valence by estimating either learning rate, learning rate and

Deleted: ted

reward sensitivity, or learning rate, reward sensitivity and pavlovian bias for reward and punishment conditions separately.

Model fitting and selection

Models were fit using hierarchical expectation maximization (EM) as described by Huys et al. [50]. In this approach, individual parameters as well as the underlying group distribution parameters are estimated iteratively. The current group distributions were used as priors to estimate individual level parameters using Laplace approximation in the E-step. Consequently, in the M-step, group-level distributions were updated based on the new individual parameter estimates and their uncertainty. Repeated sessions were treated as independent measurements and one underlying distribution was fit over all participants and measurements. Reward sensitivity and pavlovian bias parameters were log transformed and learning rate and noise parameters were transformed using the inverse sigmoid function to ensure theoretical parameter constraints.

Model fit was assessed using group-level integrated BIC (iBIC, [50]) where model fit and model complexity across all measurements are taken into account. As better group-level fit may be driven by large improvements in few participants, we additionally used likelihood ratio tests to determine the best fitting model for each participant. The winning model not only had to be parsimonious, but parameter estimates had to be stable in order to reliably quantify stimulation effects. We therefore tested stability of parameter estimates over 10 EM initializations and subsequently used mean parameters for further analysis. Furthermore, recovery of observed behavior based on simulations with estimated parameters was assessed.

Deleted: ted

Deleted: ted

Deleted: ,

Deleted: ,

Statistical analysis and software

We assessed all tVNS effects using a statistical significance threshold of $p < .05$ (two-tailed) and corrected for multiple comparisons across the five parameters in the standard computational model analysis. We also planned to correct across the condition-specific interaction terms in the mixed-effects model, but they did not reach uncorrected significance. To account for non-normal distributions of parameters from the computational model, differences in parameter estimates between the sham and tVNS condition were tested using bootstrapping (n = 1000 resamples). We performed all data analyses with Matlab v2016a (computational model) or HLM v7 (mixed-effects models) and data visualisation with R v5.0.1 and R Deducer [51].

Commented [2]: Maybe add how you corrected for multiple comparisons

Results

tVNS reduces choice accuracy across conditions

We first analyzed the performance of participants by estimating effects of reward valence, required action, and stimulation on accuracy in a full mixed-effects model. In line with previous studies, overall accuracy was higher in conditions requiring a go response (t = 5.93, p < .001), whereas reward valence only influenced accuracy in interaction with the required action (valence: t = 0.83, p = .412, valence × action: t = 7.198, p < .001). In other words, participants performed worse in the *go-punishment* and *no-go-win* conditions in which pavlovian biases (e.g. approach reward and avoid punishment) and required behavior were incongruent.

Next, we assessed main and interaction effects of tVNS vs. sham stimulation on choice accuracy. Across conditions, tVNS reduced accuracy (t = -2.08, p = .045; model uncorrected for BMI and sex: t = -1.98, p = .055). However, we observed no interaction effects with action (*t* = -0.46, p = .646) or valence (*t* = 0.78, p = .44). Nevertheless, order of stimulation (sham/tVNS first) did influence stimulation effects (t = -3.60, p < .001) with stronger impairments in overall performance if tVNS was applied first. Consequently, we controlled for stimulation order in the analyses. Notably, acute tVNS-induced reduction in performance did not lead to deficits in the second session with higher day-to-day improvements in the group that received tVNS first, t = 2.05, p = .048.

Deleted: ,

Deleted: ,

Deleted: ,

Commented [3]: It's a learnign task; the test-retest are tricky; especially when conditions change. I wonder whether the first test should be cross-sectional comparison between sham and tVNS at the first session. Do you see evidence for your hypotheses at this point?

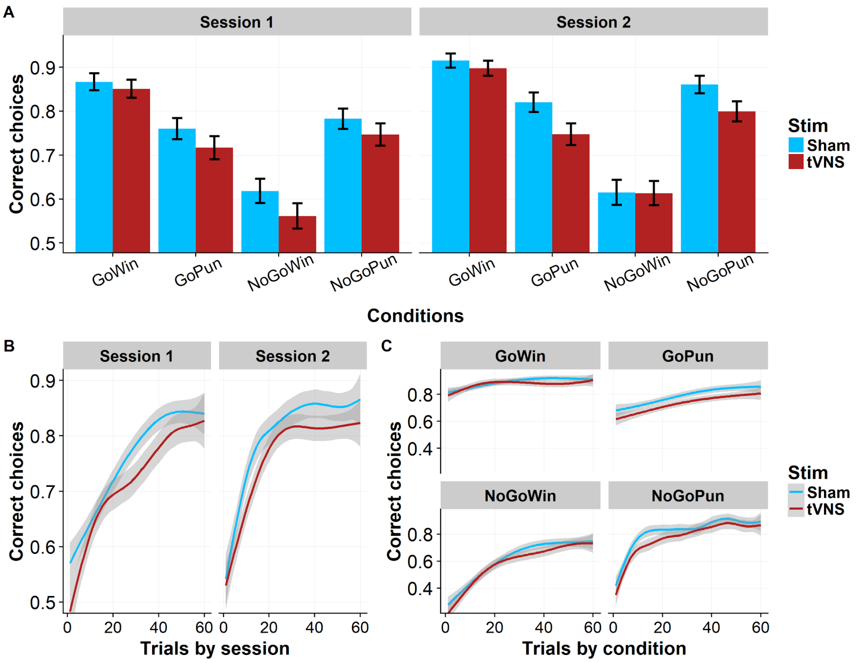

Figure 2: Choice accuracy is reduced in the tVNS condition compared to sham stimulation. A: Mean choice accuracy for tVNS and sham stimulation in each session and condition. Error bars depict 95% confidence intervals. B: Choice accuracy for tVNS and sham stimulation over trials separated by session indicate stronger tVNS-induced reduction of choice accuracy in session 1. C: Choice accuracy for tVNS and sham stimulation over trials separated by condition do not suggest action- or valence-specific effects of tVNS.

Commented [4]: I generally want to see the real data and simulations on top of each other. Only then can I really consider whether the model actually captures the data; and if so how well it does; and where it fails. Looking at these curves, I do wonder whether in session 1 you do see more of a Pavlovian effect, cross=sectionally; and whether the learning itself is faster, but the asymptote lower, though seeing the individual plots does suggest a reduction of learning rate in at least some circumstances.

tVNS reduces the learning rate in a computational model of behavior

To further characterize which learning processes were affected by tVNS leading to impaired performance, we then fitted a computational reward-learning model (Guitart-Masip et al. 2012) using an expectation maximization algorithm to regularize parameter estimates. We estimated five parameters controlling choice behavior over time: learning rate, reward sensitivity, go bias, pavlovian bias, and noisiness of choices for each session and calculated differences between tVNS and sham sessions. Average data generated based on individually estimated parameters corresponded well with observed data (Figure S?) and parameter estimates were sufficiently stable (see SI).

Impaired performance during tVNS was mainly reflected in a reduced learning rate alpha (Δα = -0.092, p = .009, p_{boot} = .002, corrected for stimulation order: t = -2.741 p = .009). Additionally, participant's choices in the tVNS condition were 'noisier' and less dependent on learned action values (Δξ = 0.035, p = .086, p_{boot} = .05), although only nominally significant before correction for multiple testing. Stimulation effects on performance were also recovered in the simulated data based on individual parameter estimates (Figure 3C-D, t = , p=?).

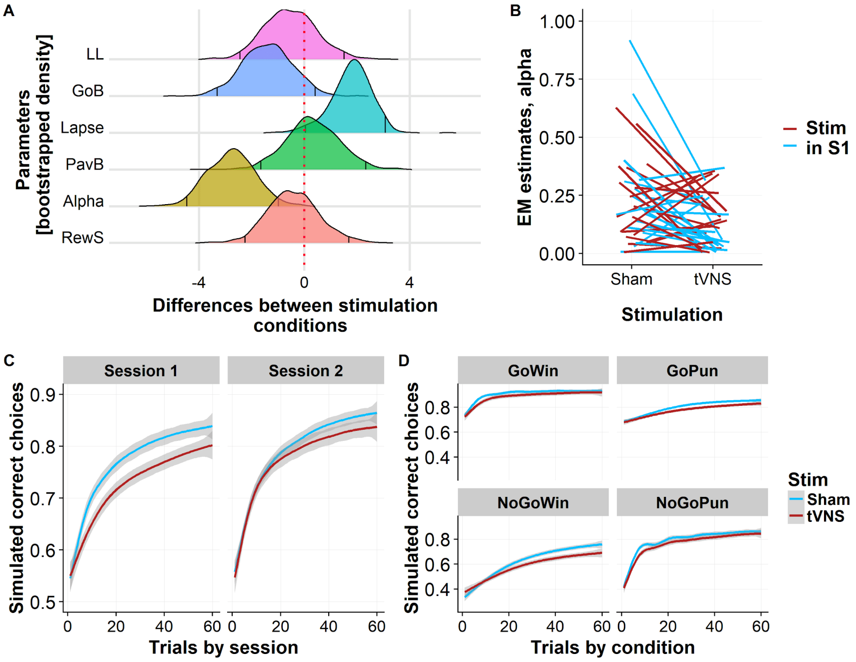

Figure 3: Reduced choice accuracy in the tVNS condition is driven by a reduced learning rate, α, and increased choice stochasticity in the 5-parameter computational model. A: Bootstrapped density plots of the differences in individual parameter estimates between tVNS and sham stimulation. Lines indicate 95% confidence intervals. B: Individual changes in the learning rate indicate mainly a reduction of high baseline learning rates after tVNS. Stimulation in session 1: blue = sham, red = tVNS C: Choices simulated from individual parameter estimates recover participants' choice patterns and stimulation effects for sessions. D: Recovered choice patterns indicate no difference in stimulation effects depending on valence or action. LL = Log-Likelihood, GoB = Go-bias, PavB = Pavlovian Bias (π), RewS = Reward sensitivity (ρ)

Commented [5]: I'd also plot the bar graphs to show that/if the model captures the overall effects in the two groups equally well.

Commented [6]: Is the legend cut in Fig 3B?

Valence-specific effects of tVNS may be captured by modeling separate parameters for rewards and punishments. Therefore, we also built an extended 6-parameter model assuming separate learning rates. The 6-parameter model provided a better model fit on the group level (ΔiBIC = 263), although on an individual level, model fit was only significantly improved for 27 out of 78 runs. Nonetheless, stability of individual parameter estimates was sufficient (SD = 0.0028 - 0.5043). Subsequent estimation of tVNS effects revealed that the slower learning rate during tVNS stimulation was predominantly driven by a decrease of the learning rate in the punishment condition ($\Delta\alpha_{pun}$ = -0.081, p = .019 , p_{boot} = .012, corrected e2for order: t = -2.516, p = .016) while decreases of alpha in reward conditions were less pronounced and non-significant ($\Delta\alpha_{rew}$ = -0.031, p = .219 p_{boot} = .21, corrected for order: t = -1.244, p = .211). However, the interaction between stimulation × valence for the learning rate was not significant (F(1,37) = 1.975, p = .168), indicating only weak specificity of the tVNS effect on punishment learning. In contrast to the 5-parameter model, tVNS did not affect choice stochasticity in the extended model (Δξ = -0.0031, p = .863, p_{boot} = .941). Again, stimulation effects on performance were recovered in the averaged simulated data (Figure 4B-C, t = -1.984 p = .055).

Commented [7]: Again I think it's important to demonstrate by comparison with the actual data how this does improve fit, so maybe again add the mean data lines to the.figure and show the bars.

Commented [8]: Unklar wie der range zustande kommt. Ist das über Parameter?

Deleted: ,

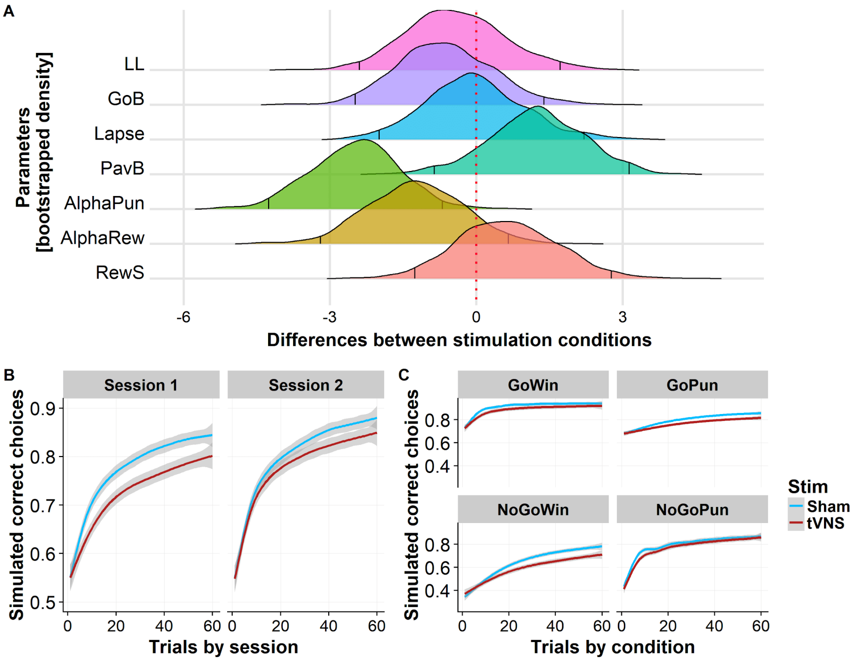

Figure 4: Reductions in learning rate are driven by slowed learning in punishment conditions. A: Bootstrapped density plots of the differences in individual parameter estimates between tVNS and sham stimulation. Lines indicate 95% confidence intervals. B: Choices simulated from individual parameter estimates recover participants' choice patterns and stimulation effects for sessions. C: Recovered choice patterns indicate no difference in stimulation effects depending on valence or action. LL = Log-Likelihood, GoB = Go-bias, PavB = Pavlovian Bias (π), RewS = Reward sensitivity (ρ)

We then explored more complex models by additionally separating reward sensitivity and/or pavlovian bias for reward and punishment as previously described [52]. However, across multiple iterations, individual estimates became increasingly volatile (SD =) precluding their use to reliably estimate within-subject stimulation effects (for details, see SI).

Commented [9]: Wieviele Iterationen braucht es denn um da einen sinnvollen Wert angeben zu können

Commented [10]: Vielleicht sollten wir doch einfach 10 reseeds machen, dann ist es einheitlich.

tVNS effects depend on baseline characteristics

Behavioral effects of dopamine increases are known to depend on baseline dopamine levels and one well-established indicator of baseline dopamine tone is BMI [53,54]. Indeed, tVNS effects on general accuracy (*t* = 1.987, *p* = .055), as well as learning rate(*t* = 2.351, *p* = .024) depended partly on participants' BMI. More specifically, tVNS reduced the speed of acquisition more strongly in participants with a low (healthy) BMI, who have a higher dopamine tone in the striatum compared to overweight participants. Furthermore, tVNS effects on the learning rate were strongly dependent on the intercept (i.e., average learning rate across both sessions) as shown by a strong correlation(*r* = -0.886, *p* < .001) between the tVNS main effect and the intercept in a mixed-effects model accounting for regression to the mean (Figure 3b).

Deleted: ,

Deleted: ,

Deleted: ,

Deleted: ,

Deleted: ,

tVNS does not affect response time

Lastly, we estimated effects of tVNS on response time as an indicator of alterations in response vigor. However, no significant changes in reaction time were observed(*t* = 0.826, *p* = .414 (Figure S.1)), further corroborating that impaired performance was mediated by slowed learning and not altered action selection.

Deleted: indication

Deleted: ,

Discussion

The vagus nerve rapidly transmits metabolic state signals to the brain and thereby modulates the dopamine system. Here, we investigated changes in instrumental reinforcement learning, which is critically dependent on dopamine, after emulating vagal feedback signals using tVNS. Importantly, we found that tVNS reduced overall accuracy of choices driven by a slowed acquisition of action contingencies, predominantly for punishments. In contrast, action- or valence-specific biases were unaffected by tVNS. In line with the value hypothesis of dopamine [40,41], the observed attenuation of the learning rate may be explained by an increase in dopamine tone leading to a lower signal-to-noise ratio of phasic dopamine signals [40,41]. Thus, using the novel non-invasive stimulation of the vagus nerve, our results provide evidence that metabolic feedback signals may alter reward learning by tuning the speed of acquisition according to homeostatic need.

Vagal feedback signals evoked by tVNS acutely impaired choice accuracy in valenced go-/no-go learning. This general impairment of performance in the tVNS condition was mirrored by lower learning rates. This is in agreement with the value theory of dopamine and further studies showing that the impact of phasic RPE signals on actions depends on dopamine tone [40,41]. In short, increased dopamine tone leads to a comparably smaller signal-to-noise ratio if phasic signals are unaffected. Thereby, learning from phasic signals evoked by action outcomes might be slowed. Accordingly, reduced learning after L-DOPA administration has been reported in patients with Parkinson's disease [55,56] as well as healthy participants [57]. Similarly, reduced learning rates are in line with the dopamine overdose hypothesis [58], especially as stimulation effects were dependent on indicators of baseline dopamine tone. For instance, tVNS markedly reduced learning rates when they were high overall, indicative of a low dopamine tone independent of the stimulation. Moreover, tVNS reduced learning rates more strongly in healthy weight participants, who were previously shown to have a medium dopamine tone in the striatum that is relatively higher compared to overweight and obese individuals [53,54]. Taken together, this indicates that tVNS-induced increases in tonic dopamine led to reduced learning rates predominantly in participants with high (or “ideal”) baseline dopamine function.

In addition to a slower learning rate, impaired performance in the task may also be caused by an increase in choice stochasticity as predicted by the thrift hypothesis of dopamine [2]. Here, increased tonic dopamine would indicate heightened average reward and energy availability [36], consequently leading to more exploration reflected in an uncoupling of learned value and choice [59]. In agreement with this account of dopamine functioning, we did find that tVNS was associated with an increase in decision noise [60,61]. However, the increase was not significant after correction for multiple testing and not consistent across models suggesting limited effects at best. These discrepancies may partly be explained by different parameterizations of the computational models. Whereas choice stochasticity is often captured with the temperature parameter in the softmax function, this parameter is separated in the common task model proposed by Guitart-Masip et al. to differentiate reward sensitivity from actual decision noise. Comparably, Guitart Masip et al. [43] did not report changes in the noise parameter in the same task using the same parametrization after L-DOPA administration. Collectively, these results suggest that tVNS primarily affects action contingency learning and not solely noise in value-based decisions.

In contrast to our hypothesis, tVNS did neither affect response-specific biases such as go or pavlovian biases nor response times in any condition. In previous studies, pharmacologically-induced increases in tonic dopamine modulated pavlovian [43] or motivational biases [62] and differentially affected learning from rewards versus punishments [63,64]. Whereas we also observed that punishment learning was more affected than reward learning, there was no significant interaction between valence-dependent learning rates and tVNS effects. Thus, tVNS-induced effects were generally independent of valence or the required action, which is not in line with previous pharmacological interventions increasing dopaminergic transmission. One possibility is that tVNS affects multiple transmitter systems and their interplay may therefore lead to different behavioral alterations. However, many dopaminergic drugs such as L-DOPA also act on other transmission systems [65] suggesting that this is an insufficient explanation. Another possibility is that modulatory effects of tVNS are more confined within the motivational circuit compared to systemic drug administration. For example, it is conceivable that tVNS could alter the balance between fast reinforcement learning, primarily linked to the amygdala, and slow reinforcement learning, primarily linked to the striatum. It has been shown that chronic tVNS increases functional connectivity between the amygdala and the prefrontal cortex in depressed patients [31] whereas repetitive VNS acutely reduces amygdala-evoked responses in the prefrontal cortex of rats [66]. Thus, future research may help to resolve these questions by detailing corresponding alterations in motivational circuits as our study leads to testable predictions about shifting the balance more towards slow striatal reinforcement learning [67]. To conclude, tVNS appears to reduce the speed of contingency learning without altering action-related processing, but more research is needed to establish differences in these processes between endogenous versus exogenous modulations of dopaminergic transmission.

While behavioral effects of tVNS including reduced accuracy and learning rates can be explained by a modulation of the dopaminergic system, the study has several limitations. First, the vagus nerve projects to multiple brain areas and tVNS is not specific, affecting various transmitter systems. Most prominently, tVNS is also associated with heightened noradrenergic signaling mediated by projections to the locus coeruleus. Increased noradrenaline signaling during tVNS has mainly been associated with improved memory performance mediated by increased arousal and attention [33]. Nonetheless, phasic noradrenaline (NA) signals have also been shown to track unsigned prediction errors or, in other words, surprise [68,69]. Surprise signals are critical for learning and, accordingly, treatment with an NA reuptake inhibitor was associated with comparable baseline-dependent changes in learning rate as reported here [70]. Since action contingencies are fixed throughout the task, it is not possible to clearly dissociate dopaminergic and noradrenergic processes acting via signed (reward) or unsigned (surprise) prediction errors, respectively. Moreover, as dopamine is the precursor of noradrenaline, future studies disentangling both systems are necessary. Second, the within-subject cross-over design offers increased statistical power to detect stimulation effects, especially considering baseline dependence. Nonetheless, repeated completion of the task may have affected performance and therefore modulated stimulation effects. We accounted for order effects in the statistical analyses, but replication in independent groups would be preferable.

To summarize we showed that vagal signals impair choice accuracy by acutely reducing learning speed in a reinforcement learning task. These findings indicate that vagal afferents modulate dopamine signaling and, in accordance with the value theory of dopamine function, slower acquisition may be due to a reduced signal-to-noise ratio of evoked phasic dopamine. We conclude that how much we learn from rewards and losses may therefore depend on the metabolic state signaled by the vagus nerve. Thereby, rapid learning which actions in a given state lead to future reward or punishment could be facilitated during a hungry state compared to a less deprived state. Critically, this behavioral flexibility with respect to the current metabolic state seems to be less pronounced in overweight participants with a lower baseline dopamine tone, which is in line with the reported reduced sensitivity to peripheral metabolic feedback [71]. Furthermore, reported anti-depressant effects of tVNS may partly rely on reduced learning, especially from punishments, as this may compensate for the reported increased punishment sensitivity in depressed patients [52,72]. More broadly, reduced learning and dependence on learned contingencies may also offer the possibility to prevent over-reliance on learned action-outcome combinations and encourage exploring changes in behavior. In turn, this could lead to greater behavioral flexibility that may be advantageous in many environments.

Acknowledgement

We thank xyz. The study was supported by the University of Tübingen, Faculty of Medicine fortune grant #2453-0-0. CB, VT, & NBK received salary support from the University of Tübingen, Faculty of Medicine fortune grant #2453-0-0. MPN received salary support from the Else Kröner-Fresenius Stiftung, grant 2017_A67

Author contributions

NBK was responsible for the study concept and design. VT implemented the task. CB & MPN collected data under supervision by MW & NBK. AK, QJMH, & NBK conceived the method including statistical and computational models. AK & NBK processed the data and performed the data analysis. AK & NBK wrote the manuscript. All authors contributed to the interpretation of findings, provided critical revision of the manuscript for important intellectual content and approved the final version for publication.

Financial disclosure

The authors declare no competing financial interests.

References

[1] Keramati M, Gutkin B. Homeostatic reinforcement learning for integrating reward collection and physiological stability. Elife 2014;3. doi:10.7554/eLife.04811.

[2] Beeler JA. Thorndike's Law 2.0: Dopamine and the Regulation of Thrift. Front Neurosci 2012;6. doi:10.3389/fnins.2012.00116.

[3] Korn CW, Bach DR. Maintaining homeostasis by decision-making. PLoS Comput Biol 2015;11:e1004301.

[4] Kroemer NB, Small DM. Fuel not fun: Reinterpreting attenuated brain responses to reward in obesity. Physiol Behav 2016;162:37–45.

[5] Skrynka J, Vincent BT. Hunger increases delay discounting of food and non-food rewards 2017. doi:10.31234/osf.io/qgp54.

[6] Beeler JA, Frazier CRM, Zhuang X. Putting desire on a budget: dopamine and energy expenditure, reconciling reward and resources. Front Integr Neurosci 2012;6:49.

[7] Veldhuizen MG, Babbs RK, Patel B, Fobbs W, Kroemer NB, Garcia E, et al. Integration of Sweet Taste and Metabolism Determines Carbohydrate Reward. Curr Biol 2017;27:2476–85.e6.

[8] de Lartigue G. Role of the vagus nerve in the development and treatment of diet-induced obesity. J Physiol 2016;594:5791–815.

[9] Grill HJ, Hayes MR. Hindbrain neurons as an essential hub in the neuroanatomically distributed control of energy balance. Cell Metab 2012;16:296–309.

[10] Tellez LA, Medina S, Han W, Ferreira JG, Licona-Limón P, Ren X, et al. A gut lipid messenger links excess dietary fat to dopamine deficiency. Science 2013;341:800–2.

[11] Han W, Tellez LA, Perkins MH, Perez IO, Qu T, Ferreira J, et al. A Neural Circuit for Gut-Induced Reward. Cell 2018. doi:10.1016/j.cell.2018.08.049.

[12] Roosevelt RW, Smith DC, Clough RW, Jensen RA, Browning RA. Increased extracellular concentrations of norepinephrine in cortex and hippocampus following vagus nerve stimulation in the rat. Brain Res 2006;1119:124–32.

[13] Hulsey DR, Hays SA, Khodaparast N, Ruiz A, Das P, Rennaker RL 2nd, et al. Reorganization of Motor Cortex by Vagus Nerve Stimulation Requires Cholinergic Innervation. Brain Stimul 2016;9:174–81.

[14] Ferreira JG, Tellez LA, Ren X, Yeckel CW, de Araujo IE. Regulation of fat intake in the absence of flavour signalling. J Physiol 2012;590:953–72.

[15] de Araujo IE, Ferreira JG, Tellez LA, Ren X, Yeckel CW. The gut-brain dopamine axis: a regulatory system for caloric intake. Physiol Behav 2012;106:394–9.

[16] de Lartigue G, Ronveaux CC, Raybould HE. Deletion of leptin signaling in vagal afferent neurons results in hyperphagia and obesity. Mol Metab 2014;3:595–607.

[17] Davis JF, Tracy AL, Schurdak JD, Tschöp MH, Lipton JW, Clegg DJ, et al. Exposure to elevated levels of dietary fat attenuates psychostimulant reward and mesolimbic dopamine turnover in the rat. Behav Neurosci 2008;122:1257–63.

[18] Palmiter RD. Dopamine signaling in the dorsal striatum is essential for motivated behaviors: lessons from dopamine-deficient mice. Ann N Y Acad Sci 2008;1129:35–46.

[19] Palmiter RD. Is dopamine a physiologically relevant mediator of feeding behavior? Trends Neurosci 2007;30:375–81.

[20] Val-Laillet D, Biraben A, Randuineau G, Malbert CH. Chronic vagus nerve stimulation decreased weight gain, food consumption and sweet craving in adult obese minipigs. Appetite 2010;55:245–52.

[21] Suarez AN, Hsu TM, Liu CM, Noble EE, Cortella AM, Nakamoto EM, et al. Gut vagal sensory signaling regulates hippocampus function through multi-order pathways. Nat Commun 2018;9. doi:10.1038/s41467-018-04639-1.

[22] Klarer M, Weber-Stadlbauer U, Arnold M, Langhans W, Meyer U. Cognitive effects of subdiaphragmatic vagal deafferentation in rats. Neurobiol Learn Mem 2017;142:190–9.

[23] Klarer M, Arnold M, Günther L, Winter C, Langhans W, Meyer U. Gut vagal afferents differentially modulate innate anxiety and learned fear. J Neurosci 2014;34:7067–76.

[24] Wang G-J, Yang J, Volkow ND, Telang F, Ma Y, Zhu W, et al. Gastric stimulation in obese subjects activates the hippocampus and other regions involved in brain reward circuitry. Proc Natl Acad Sci U S A 2006;103:15641–5.
[25] Howland RH. Vagus Nerve Stimulation. Curr Behav Neurosci Rep 2014;1:64–73.
[26] Ghacibeh GA, Shenker JI, Shenal B, Uthman BM, Heilman KM. The influence of vagus nerve stimulation on memory. Cogn Behav Neurol 2006;19:119–22.
[27] He W, Jing X-H, Zhu B, Zhu X-L, Li L, Bai W-Z, et al. The auriculo-vagal afferent pathway and its role in seizure suppression in rats. BMC Neurosci 2013;14:85.
[28] Frangos E, Ellrich J, Komisaruk BR. Non-invasive Access to the Vagus Nerve Central Projections via Electrical Stimulation of the External Ear: fMRI Evidence in Humans. Brain Stimul 2015;8:624–36.
[29] Kraus T, Hösl K, Kiess O, Schanze A, Kornhuber J, Forster C. BOLD fMRI deactivation of limbic and temporal brain structures and mood enhancing effect by transcutaneous vagus nerve stimulation. J Neural Transm 2007;114:1485–93.
[30] Fang J, Rong P, Hong Y, Fan Y, Liu J, Wang H, et al. Transcutaneous Vagus Nerve Stimulation Modulates Default Mode Network in Major Depressive Disorder. Biol Psychiatry 2016;79:266–73.
[31] Liu J, Fang J, Wang Z, Rong P, Hong Y, Fan Y, et al. Transcutaneous vagus nerve stimulation modulates amygdala functional connectivity in patients with depression. J Affect Disord 2016;205:319–26.
[32] Jacobs HIL, Riphagen JM, Razat CM, Wiese S, Sack AT. Transcutaneous vagus nerve stimulation boosts associative memory in older individuals. Neurobiol Aging 2015;36:1860–7.
[33] Burger AM, Verkuil B, Van Diest I, Van der Does W, Thayer JF, Brosschot JF. The effects of transcutaneous vagus nerve stimulation on conditioned fear extinction in humans. Neurobiol Learn Mem 2016;132:49–56.
[34] Steenbergen L, Sellaro R, Stock A-K, Verkuil B, Beste C, Colzato LS. Transcutaneous vagus nerve stimulation (tVNS) enhances response selection during action cascading processes. Eur Neuropsychopharmacol 2015;25:773–8.
[35] Sellaro R, van Leusden JWR, Tona K-D, Verkuil B, Nieuwenhuis S, Colzato LS. Transcutaneous Vagus Nerve Stimulation Enhances Post-error Slowing. J Cogn Neurosci 2015;27:2126–32.
[36] Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 2007;191:507–20.
[37] Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997;275:1593–9.
[38] Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 2013;16:966–73.
[39] Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 2006;442:1042–5.
[40] Kroemer NB, Lee Y, Pooseh S, Eppinger B, Goschke T, Smolka MN. L-DOPA reduces model-free control of behavior by attenuating the transfer of value to action 2016. doi:10.1101/086116.
[41] Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, et al. Mesolimbic dopamine signals the value of work. Nat Neurosci 2016;19:117–26.
[42] Guitart-Masip M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. Go and no-go learning in reward and punishment: interactions between affect and effect. Neuroimage 2012;62:154–66.
[43] Guitart-Masip M, Economides M, Huys QJM, Frank MJ, Chowdhury R, Duzel E, et al. Differential, but not opponent, effects of L -DOPA and citalopram on action learning with reward and punishment. Psychopharmacology 2014;231:955–66.
[44] Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 2004;306:1940–3.

[45] Cox SML, Frank MJ, Larcher K, Fellows LK, Clark CA, Leyton M, et al. Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. Neuroimage 2015;109:95–101.
[46] Guitart-Masip M, Economides M, Huys QJM, Frank MJ, Chowdhury R, Duzel E, et al. Differential, but not opponent, effects of L -DOPA and citalopram on action learning with reward and punishment. Psychopharmacology 2014;231:955–66.
[47] Guitart-Masip M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. Go and no-go learning in reward and punishment: interactions between affect and effect. Neuroimage 2012;62:154–66.
[48] Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ. Modeling Avoidance in Mood and Anxiety Disorders Using Reinforcement Learning. Biol Psychiatry 2017;82:532–9.
[49] Raudenbush SW, Bryk AS, Cheong YF, Congdon (Jr. R, Du Toit M. HLM 7: Hierarchical Linear and Nonlinear Modeling. 2011.
[50] Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. PLoS Comput Biol 2011;7:e1002028.
[51] Fellows I. Deducer: A Data Analysis GUI forR. J Stat Softw 2012;49. doi:10.18637/jss.v049.i08.
[52] Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ. Modeling Avoidance in Mood and Anxiety Disorders Using Reinforcement Learning. Biol Psychiatry 2017;82:532–9.
[53] Lee Y, Kroemer NB, Oehme L, Beuthien-Baumann B, Goschke T, Smolka MN. Lower dopamine tone in the striatum is associated with higher body mass index. Eur Neuropsychopharmacol 2018;28:719–31.
[54] Horstmann A, Fenske WK, Hankir MK. Argument for a non-linear relationship between severity of human obesity and dopaminergic tone. Obes Rev 2015;16:821–30.
[55] Cools R, Lewis SJG, Clark L, Barker RA, Robbins TW. L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson’s disease. Neuropsychopharmacology 2007;32:180–9.
[56] Vo A, Hiebert NM, Seergobin KN, Solcz S, Partridge A, MacDonald PA. Dopaminergic medication impairs feedback-based stimulus-response learning but not response selection in Parkinson’s disease. Front Hum Neurosci 2014;8:784.
[57] Vo A, Seergobin KN, Morrow SA, MacDonald PA. Levodopa impairs probabilistic reversal learning in healthy young adults. Psychopharmacology 2016;233:2753–63.
[58] Cools R. Dopaminergic modulation of cognitive function-implications for l-DOPA treatment in Parkinson’s disease. Neurosci Biobehav Rev 2006;30:1–23.
[59] Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, et al. Dopamine and performance in a reinforcement learning task: evidence from Parkinson’s disease. Brain 2012;135:1871–83.
[60] Beeler JA, Daw N, Frazier CRM, Zhuang X. Tonic dopamine modulates exploitation of reward learning. Front Behav Neurosci 2010;4:170.
[61] Eisenegger C, Naef M, Linssen A, Clark L, Gandamaneni PK, Müller U, et al. Role of dopamine D2 receptors in human reinforcement learning. Neuropsychopharmacology 2014;39:2366–75.
[62] Swart JC, Froböse MI, Cook JL, Geurts DE, Frank MJ, Cools R, et al. Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. Elife 2017;6. doi:10.7554/eLife.22169.
[63] Cools R, Altamirano L, D’Esposito M. Reversal learning in Parkinson’s disease depends on medication status and outcome valence. Neuropsychologia 2006;44:1663–73.
[64] Frank MJ. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci 2005;17:51–72.
[65] De Deurwaerdère P, Di Giovanni G, Millan MJ. Expanding the repertoire of L-DOPA’s actions: A comprehensive review of its functional neurochemistry. Prog Neurobiol 2017;151:57–100.
[66] Lyubashina O, Panteleev S. Effects of cervical vagus nerve stimulation on amygdala-

evoked responses of the medial prefrontal cortex neurons in rat. Neurosci Res 2009;65:122–5.
[67] Averbeck BB, Costa VD. Motivational neural circuits underlying reinforcement learning. Nat Neurosci 2017;20:505–12.
[68] Dayan P, Yu AJ. Phasic norepinephrine: a neural interrupt signal for unexpected events. Network 2006;17:335–50.
[69] Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP. The Neural Representation of Unexpected Uncertainty during Value-Based Decision Making. Neuron 2013;79:191–201.
[70] Jepma M, Murphy PR, Nassar MR, Rangel-Gomez M, Meeter M, Nieuwenhuis S. Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. PLoS Comput Biol 2016;12:e1005171.
[71] Klok MD, Jakobsdottir S, Drent ML. The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. Obes Rev 2007;8:21–34.
[72] Hevey D, Thomas K, Laureano-Schelten S, Looney K, Booth R. Clinical Depression and Punishment Sensitivity on the BART. Front Psychol 2017;8:670.