

Stimulation of the vagus nerve reduces learning

in a go/no-go reinforcement learning task

Anne Kühnel^{1*}, Vanessa Teckentrup², Monja P. Neuser², Quentin J. M. Huys^{3,4},
Caroline Burrasch², Martin Walter^{2,5-7}, & Nils B. Kroemer^{2*}

- ¹ Department of Translational Research in Psychiatry, 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 Division of Psychiatry, University College London, London
- ⁴ Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London
- ⁵ Otto-von-Guericke University Magdeburg, Department of Psychiatry and Psychotherapy, Germany
- ⁶ Clinical Affective Neuroimaging Laboratory, Magdeburg, Germany
- ⁷ 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

Word count: 4000 words

Abstract

When facing decisions to approach rewards or to avoid punishments, we often figuratively go with our gut, and the impact of metabolic states such as hunger on motivation are well documented. However, whether and how vagal feedback signals from the gut influence instrumental actions is unknown. Here, 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, participants after an overnight fast. First, mixed-effects logistic regression analysis of choice accuracy showed that tVNS acutely impaired decision-making, $p = .045$. Computational reinforcement learning models identified the cause of this as a reduction in the learning rate through tVNS ($\Delta\alpha = -0.092$, $p_{boot} = .002$), particularly after 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. Hence, tVNS appeared to influence learning rather than action execution. These 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, patiently resisting temptation to receive bigger returns later is often beneficial in the longer term. However, in certain bodily states, forfeiting immediate rewards can have negative long-term consequences. For instance, missing out on food in a hungry state [5] can increase the risk of starvation [1], which may profoundly hamper future prospects. Short-sightedness may hence address urgent dietary requirements at the expense of long-term opulence [6]. Despite its obvious evolutionary importance, little is known about how bodily homeostatic needs shape decision making in humans. 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. Along that pathway, vagal afferents have been shown to modulate dopaminergic [10,11], 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 signaling regulates food intake [16], and stimulation of

the vagus nerve has been associated with reduced food intake and decreased weight gain [20]. More broadly, memory function [21], cognitive flexibility [22], and mood [23] are also influenced by vagal signaling. Collectively, these results suggest that vagal signals conveying the metabolic state influence behavior, including reward seeking, via alterations in dopamine.

Until recently, research in humans has been limited by the invasive nature of cervical vagus nerve stimulation (VNS). Nevertheless, VNS has been related to a broad range of behavioral effects such as enhancing cognitive functioning [24] and memory retention [25,26]. VNS also has a role in the treatment of depressive and anxiety symptoms [25]. Lately, non-invasive transcutaneous VNS (tVNS) has become feasible, opening new avenues for research and treatment. It involves transcutaneous stimulation of the auricular branch of the vagus nerve behind the ear. This has been shown to affect projections to the NTS in preclinical studies [27], and 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. Therefore, tVNS provides a promising approach for the investigation of the link between vagal metabolic signals and reward-related decisions.

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. Since vagal signals also modulate the dopaminergic system, motivation and reward learning should be shaped by vagal stimulation as well. While response vigor has been linked to dopamine tone [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 also influenced by dopamine tone [39,40]. Perhaps counterintuitively, high dopamine tone reduces the constraint of actions imposed by previous rewards [40] as phasic signals are proportionally smaller (value theory, [41]) or choices rely less on learned values (thrift theory, [2]). Moreover, learning is differentially supported by dopamine depending on the specific task [42], with distinct but interacting circuits underlying learning from rewards or punishment. Consequently, vagal feedback may lead to changes in reward learning, response vigor, and action selection mediated by dopaminergic signaling. 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 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 and our recent work corroborating these predictions [40], we hypothesized that during tVNS participants' performance would be impaired, as increased dopamine tone reduces 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 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 heightened 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. For the current analysis, five participants had to be excluded (n=4: did not complete both task sessions, n=1: did not make any go response). Thus, we included 39 participants (23 female, $M_{age} = 25.5 \pm 4.0$ years; $M_{BMI} = 23.0 \pm 3.0$ kg/m²). 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) before 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%; $p_{\text{binomial}} = .049$) suggesting that blinding was effective. Sessions started between 7.00am and 11:00am and lasted about 2.5h in each. 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 targeting 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. We determined individual stimulation strength 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 0.1mA and increased by the experimenter in 0.1-0.2mA steps. Participants rated the sensation after every increment until ratings settled around 5 corresponding to “mild prickling”. Then, the stimulation continued throughout the task block according to the 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 (~20min) and an effort allocation task (~40min) before the learning task.

After completing state-related questions, participants received rewards according to task performance and compensation (either as 32€ or partial course credit). Both visits followed the same standardized protocol.

Paradigm

We hypothesized that tVNS affects 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 task. Due to the well-known characteristics of the dopaminergic circuit [44,45], we sought to disentangle effects of tVNS on action- or valence-dependent learning. To this end, we used a previously established go/no-go reinforcement learning task [42,43,46]. 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 complete a target detection task and either respond by pressing a button (go) or withhold their response (no-go). Third, they saw the outcome of the state-action combination, which was either a win (5 cents), punishment (-5 cents), or an omission (no win/punishment, 0 cents). Using trial and error, participants had to learn which action following each fractal was best in terms of maximizing wins or minimizing losses.

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 correct 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 task. In total, the task included 240 trials (60 per condition) and took 15min to be completed.

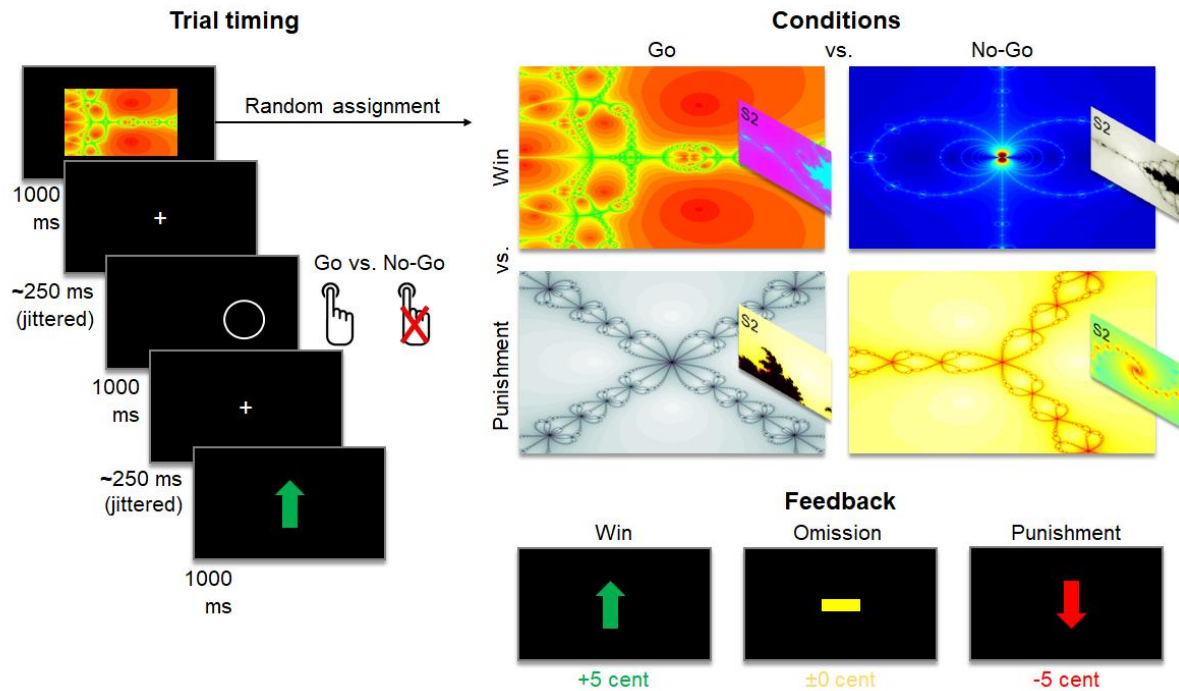


Figure 1: Schematic summary of the go/no-go reinforcement learning task. To maximize the total payoff, participants had to learn which action (go vs. no-go) during the target detection stage following a given fractal was associated with the best possible outcome (i.e., receiving reward or avoiding impending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

Data analysis

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

To estimate the effects of tVNS on choice accuracy (go vs. no-go), we defined a full mixed-effects analysis as implemented in HLM [47]. 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 \times 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 \times win, stimulation \times go, stimulation \times go \times win). Furthermore, we included a log-transformed trial regressor capturing improvements in accuracy across trials. 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 \times 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 (EM) in the computational model, we obtained empirical Bayes estimates, which take group-level distributions into account, as individual estimates of tVNS 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. [42,43] as standard model (for details and equations, see SI). Models were fit using hierarchical EM as described by Huys et al. [48]. Model fit was assessed using group-level integrated BIC (iBIC, [48]) 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 session. To ensure stability of individual estimates, we used 10 EM initializations to calculate the mean and the coefficient of variation. Furthermore, we assessed recovery of observed behavior based on simulations with estimated parameters.

Statistical analysis and software

We assessed all tVNS effects using a significance threshold of $p < .05$ (two-tailed) and corrected for multiple comparisons across the five parameters in the computational model analysis using Bonferroni correction. We also planned correction across 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 tVNS and sham condition were tested using bootstrapping (1000 resamples). We performed data analyses with Matlab v2016a (computational model) or HLM v7 (mixed-effects models) and data visualization with R v5.0.1 and Deducer [49].

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, 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 \times 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 (*approach reward*; *avoid punishment*) and instrumental 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 ($t = 0.78$, $p = .44$). As task-dependent improvement across sessions is common in reinforcement learning tasks, we controlled for stimulation order (sham/tVNS first) in the analyses. Order of stimulation modulated stimulation slopes ($t = -3.60$, $p < .001$) with stronger impairments in overall performance if tVNS was applied first. Notably, acute tVNS-induced reductions in performance did not lead to deficits in the second session with higher day-to-day improvements in the group that received tVNS first ($t = 2.05$, $p = .048$).

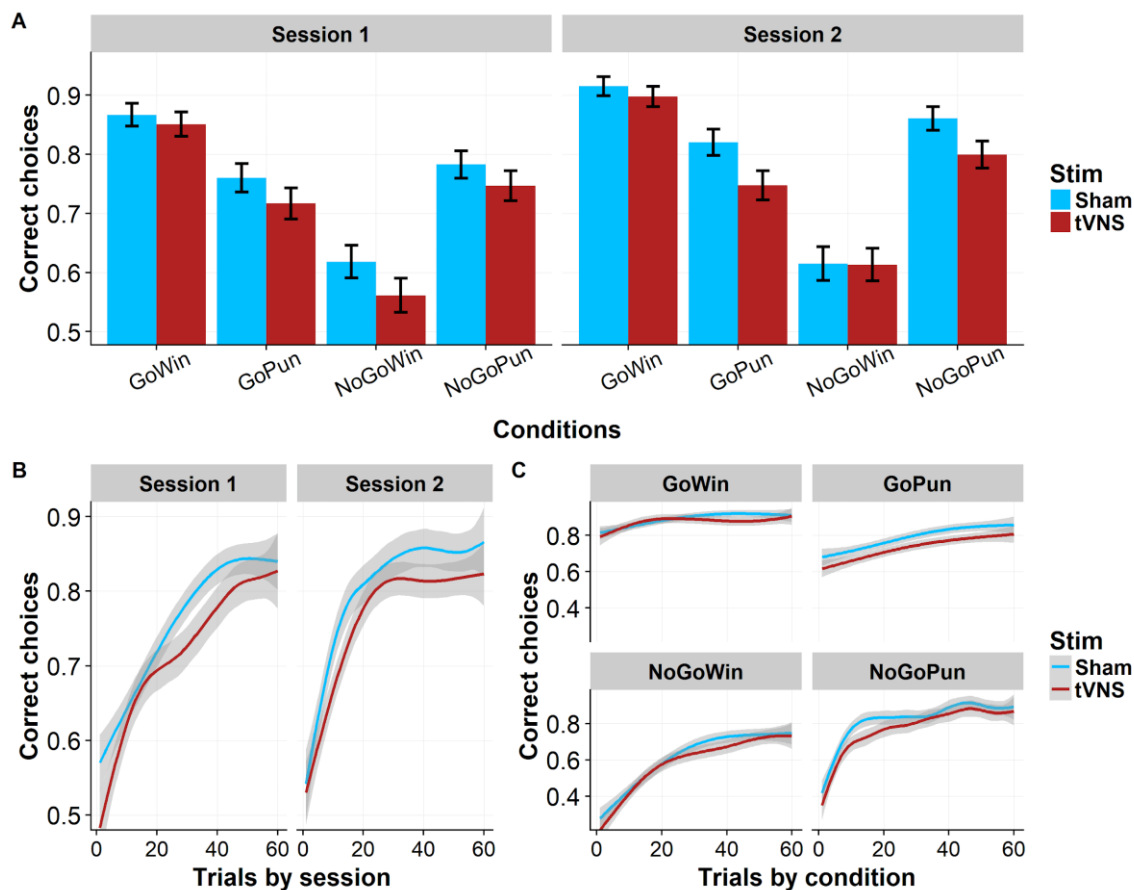


Figure 2: Choice accuracy is reduced in the tVNS condition compared to sham stimulation. A: Mean choice accuracy for tVNS and sham stimulation in each session and condition. Error bars depict 95% confidence intervals. B: Choice accuracy for tVNS and sham stimulation over trials separated by session indicate 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.

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 fitted a computational reward-learning model [42] using an EM algorithm to empirically regularize parameter estimates. We estimated five parameters controlling choices 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. Simulated data based on individually estimated parameters corresponded well with observed data (Figure 3C-E) and parameter estimates were sufficiently stable across the 10 reruns with median coefficients of variation between .002 and .030 for the five parameters.

Impaired performance during tVNS was mainly reflected in a reduced learning rate alpha (Figure 3A; $\Delta\alpha = -0.092$, $p = .009$, $p_{boot} = .002$, corrected for stimulation order: $t = -2.741$ $p = .009$; $\alpha_{Bonferroni} = .01$). Additionally, participant's choices in the tVNS condition were 'noisier' and less dependent on learned action values ($\Delta\xi = 0.035$, $p = .086$, $p_{boot} = .050$), although only nominally significant before correction for multiple testing.

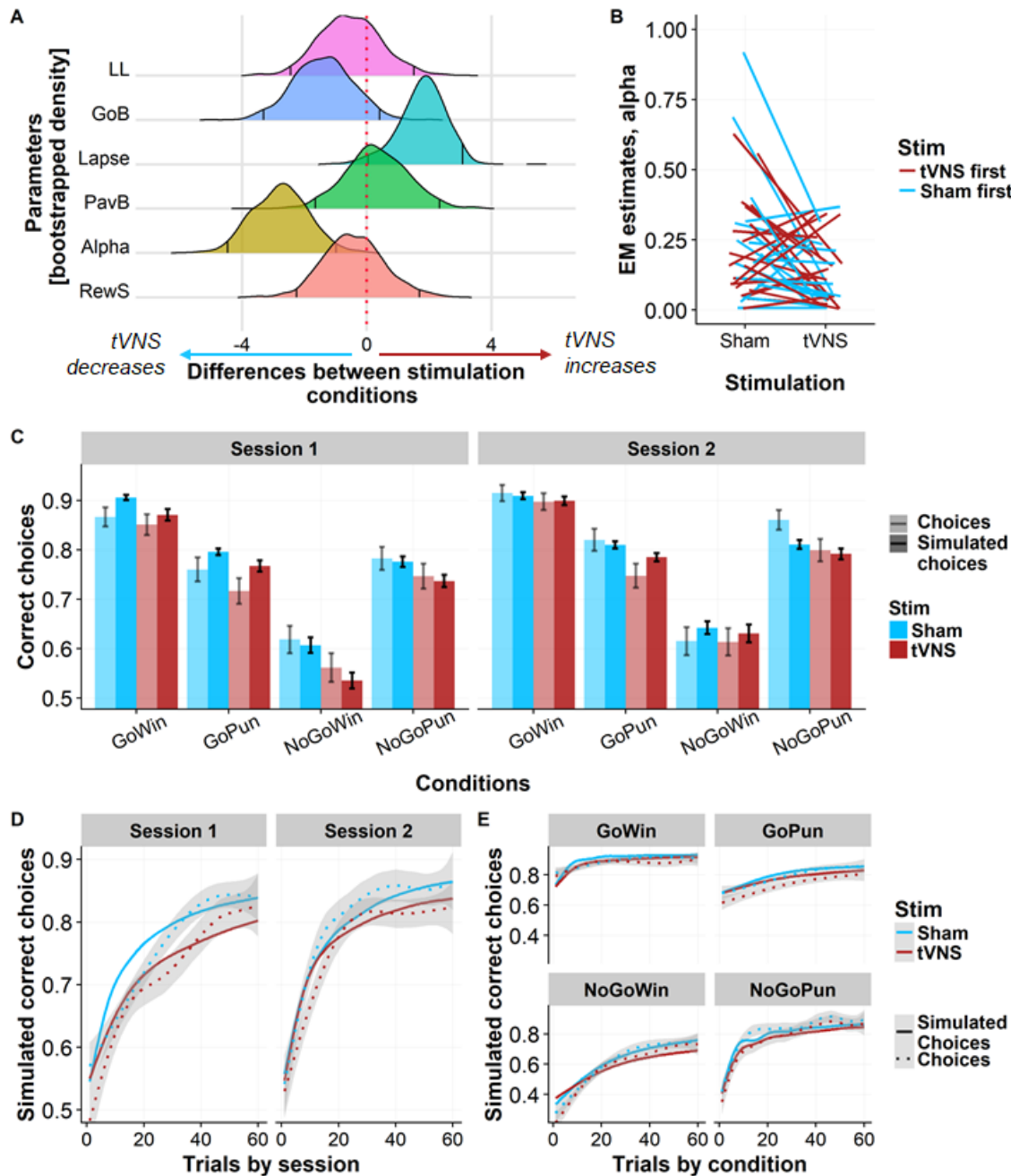


Figure 3: Reduced choice accuracy in the tVNS condition is driven by a reduced learning rate and increased choice stochasticity in the 5-parameter computational model. **A:** Bootstrapped density plots of the differences in individual parameter estimates between tVNS and sham stimulation. Lines indicate 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:** Mean choice accuracy simulated from individual parameter estimates recover stimulation effects in each session and condition. Error bars depict 95% confidence intervals. Transparent bars depict choices, dark bars simulated choices. **D:** Choices simulated from individual parameter estimates recover participants' choice patterns and stimulation effects for sessions. **E:** 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)

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. While the 6-parameter model provided a more parsimonious account at the group level ($\Delta\text{iBIC} = 263$), it only improved individual model fits for 27 out of 78 sessions. In line with the 5-parameter model, tVNS reduced learning rates. Interestingly, this effect was primarily driven by a decrease in the punishment condition (see SI).

We also explored more complex models by additionally separating reward sensitivity and/or Pavlovian bias for reward and punishment as previously described [46]. However, these models did not provide a more parsimonious account at the individual level compared to the simpler models and individual estimates became increasingly unstable across iterations precluding their use to reliably estimate within-subject stimulation effects (see SI).

tVNS effects are associated with body weight

Behavioral effects of dopamine increases are known to depend on baseline dopamine tone. One well-established correlate is BMI [50,51] and tVNS effects on general accuracy ($t = 1.987$, $p = .055$), as well as on 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.

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 response time were

observed ($t = 0.826$, $p = .414$, Figure S.2). This further corroborates that impaired performance was mediated by slowed learning and not by altered action selection.

Discussion

The vagus nerve rapidly transmits interoceptive signals to the brain. It thereby confers information such as metabolic state and 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 learning rates 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 needs.

Vagal feedback signals evoked by tVNS acutely impaired choice accuracy and reduced learning rates in valenced go-/no-go learning. This is in agreement with the value theory of dopamine and with 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 and to reduced action control via phasic outcome signals. Accordingly, reduced learning after L-DOPA administration has been reported in patients with Parkinson's disease [52,53] as well as healthy participants [54]. Similarly, reduced learning rates are in line with the dopamine overdose hypothesis [55], as tVNS reduced learning rates

more strongly in participants with a healthy weight, who were previously shown to have a higher dopamine tone in the striatum compared to overweight and obese individuals [50,51]. This may indicate that tVNS-induced increases in dopamine tone led to reduced learning rates predominantly in participants with optimal baseline dopamine function.

In addition to slower learning, impaired performance in the task may also be caused by increased choice stochasticity as predicted by the thrift hypothesis of dopamine [2]. Here, increased dopamine tone would indicate heightened average reward and energy availability [36], consequently leading to more exploration reflected in an uncoupling of learned value and choice [56]. In agreement with the thrift theory, we found that tVNS was associated with increased decision noise [57,58]. 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 Pavlovian or go biases nor response times in any condition. In previous studies, pharmacologically-induced increases in tonic dopamine modulated Pavlovian [43] or motivational biases [59] and differentially affected learning from rewards versus

punishments [60,61]. Although, we observed that punishment learning, but not reward learning was significantly reduced, 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 possible explanation is that tVNS affects multiple transmitter systems and their interplay might lead to mixed behavioral alterations. However, many dopaminergic drugs such as L-DOPA also act on other transmission systems [62] 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 VNS acutely reduces amygdala-evoked responses in the prefrontal cortex of rats [63]. 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 [64]. To conclude, tVNS appears to reduce the speed of 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 modulations of dopamine, our study has limitations. First, although we focused on dopamine, tVNS is not specifically targeting only one

neurotransmitter. tVNS has also been associated with heightened noradrenergic signaling via the locus coeruleus leading to improved memory performance by increasing arousal and attention [33]. Nonetheless, phasic noradrenaline signals have also been shown to track unsigned prediction errors (“surprise”) [65,66]. Surprise signals are critical for learning and, accordingly, treatment with an noradrenaline reuptake inhibitor was associated with comparable baseline-dependent changes in learning rates [67]. Since action contingencies are fixed throughout the task, it is not possible to 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 modulated tVNS 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 could indicate that vagal afferents modulate dopamine tone and, in accordance with the value theory of dopamine, 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 punishments may depend on the metabolic state signaled by the vagus nerve. Thereby, rapid learning which actions in each 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 was less pronounced in overweight participants with a lower baseline dopamine tone, which is

in line with the reported reduced sensitivity to peripheral metabolic feedback [68]. 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 [46,69]. More broadly, reduced dependence on learned contingencies may also offer the possibility to prevent over-reliance on learned action-outcome combinations and encourage exploration. In turn, this could lead to greater behavioral flexibility that may be advantageous in many naturalistic environments.

Acknowledgement

We thank Anni Richter for sharing the German task instructions with us. The study was supported by the University of Tübingen, Faculty of Medicine fortune grant #2453-0-0. VT, CB, & 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. QH is supported by the UCL NIHR BRC.

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 Computational Biology* 2015;11:e1004301. doi:10.1371/journal.pcbi.1004301.
- [4] Kroemer NB, Small DM. Fuel not fun: Reinterpreting attenuated brain responses to reward in obesity. *Physiology & Behavior* 2016;162:37–45. doi:10.1016/j.physbeh.2016.04.020.
- [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. doi:10.3389/fnint.2012.00049.
- [7] Veldhuizen MG, Babbs RK, Patel B, Fobbs W, Kroemer NB, Garcia E, et al. Integration of Sweet Taste and Metabolism Determines Carbohydrate Reward. *Current Biology* 2017;27:2476-2485.e6. doi:10.1016/j.cub.2017.07.018.
- [8] Lartigue G de. Role of the vagus nerve in the development and treatment of diet-induced obesity. *The Journal of Physiology* 2016;5791–815. doi:10.1113/JP271538@10.1111/(ISSN)1469-7793.StressEndocrinologyobesityanddiabetes.
- [9] Grill HJ, Hayes MR. Hindbrain Neurons as an Essential Hub in the Neuroanatomically Distributed Control of Energy Balance. *Cell Metabolism* 2012;16:296–309. doi:10.1016/j.cmet.2012.06.015.
- [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. doi:10.1126/science.1239275.
- [11] Han W, Tellez LA, Perkins MH, Perez IO, Qu T, Ferreira J, et al. A Neural Circuit for Gut-Induced Reward. *Cell* 2018;175:665-678.e23. 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 Research* 2006;1119:124–32. doi:10.1016/j.brainres.2006.08.048.
- [13] Hulseley DR, Hays SA, Khodaparast N, Ruiz A, Das P, Rennaker RL, et al. Reorganization of Motor Cortex by Vagus Nerve Stimulation Requires Cholinergic Innervation. *Brain Stimulation* 2016;9:174–81. doi:10.1016/j.brs.2015.12.007.
- [14] Ferreira JG, Tellez LA, Ren X, Yeckel CW, Araujo IE de. Regulation of fat intake in the absence of flavour signalling. *The Journal of Physiology* 2012;590:953–72. doi:10.1113/jphysiol.2011.218289.
- [15] de Araujo IE, Ferreira JG, Tellez LA, Ren X, Yeckel CW. The gut–brain dopamine axis: A regulatory system for caloric intake. *Physiology & Behavior* 2012;106:394–9. doi:10.1016/j.physbeh.2012.02.026.
- [16] de Lartigue G, Ronveaux CC, Raybould HE. Deletion of leptin signaling in vagal afferent neurons results in hyperphagia and obesity. *Molecular Metabolism* 2014;3:595–607. doi:10.1016/j.molmet.2014.06.003.
- [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. *Behavioral Neuroscience* 2008;122:1257–63. doi:10.1037/a0013111.
- [18] Palmiter RD. Is dopamine a physiologically relevant mediator of feeding behavior? *Trends in Neurosciences* 2007;30:375–81. doi:10.1016/j.tins.2007.06.004.

- [19] Palmiter RD. Dopamine Signaling in the Dorsal Striatum Is Essential for Motivated Behaviors. *Annals of the New York Academy of Sciences* 2008;1129:35–46. doi:10.1196/annals.1417.003.
- [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. doi:10.1016/j.appet.2010.06.008.
- [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. *Nature Communications* 2018;9:2181. 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. *Neurobiology of Learning and Memory* 2017;142:190–9. doi:10.1016/j.nlm.2017.05.006.
- [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. doi:10.1523/JNEUROSCI.0252-14.2014.
- [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. *PNAS* 2006;103:15641–5. doi:10.1073/pnas.0601977103.
- [25] Howland RH. Vagus Nerve Stimulation. *Curr Behav Neurosci Rep* 2014;1:64–73. doi:10.1007/s40473-014-0010-5.
- [26] Ghacibeh GA, Shenker JI, Shenal B, Uthman BM, Heilman KM. The Influence of Vagus Nerve Stimulation on Memory. *Cognitive and Behavioral Neurology* 2006;19:119. doi:10.1097/01.wnn.0000213908.34278.7d.
- [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 Neuroscience* 2013;14:85. doi:10.1186/1471-2202-14-85.
- [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. doi:10.1016/j.brs.2014.11.018.
- [29] Kraus T, Kiess O, Hösl K, Terekhin P, Kornhuber J, Forster C. CNS BOLD fMRI Effects of Sham-Controlled Transcutaneous Electrical Nerve Stimulation in the Left Outer Auditory Canal – A Pilot Study. *Brain Stimulation* 2013;6:798–804. doi:10.1016/j.brs.2013.01.011.
- [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. *Biological Psychiatry* 2016;79:266–73. doi:10.1016/j.biopsych.2015.03.025.
- [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. *Journal of Affective Disorders* 2016;205:319–26. doi:10.1016/j.jad.2016.08.003.
- [32] Jacobs HIL, Riphagen JM, Razat CM, Wiese S, Sack AT. Transcutaneous vagus nerve stimulation boosts associative memory in older individuals. *Neurobiology of Aging* 2015;36:1860–7. doi:10.1016/j.neurobiolaging.2015.02.023.
- [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. *Neurobiology of Learning and Memory* 2016;132:49–56. doi:10.1016/j.nlm.2016.05.007.
- [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. *European Neuropsychopharmacology* 2015;25:773–8. doi:10.1016/j.euroneuro.2015.03.015.
- [35] Sellaro R, van Leusden JWR, Tona K-D, Verkuil B, Nieuwenhuis S, Colzato LS. Transcutaneous Vagus Nerve Stimulation Enhances Post-error Slowing. *Journal of Cognitive Neuroscience* 2015;27:2126–32. doi:10.1162/jocn_a_00851.
- [36] Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* 2007;191:507–20. doi:10.1007/s00213-006-0502-4.

- [37] Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. *Science* 1997;275:1593–9. doi:10.1126/science.275.5306.1593.
- [38] Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience* 2013;16:966–73. doi:10.1038/nn.3413.
- [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. doi:10.1038/nature05051.
- [40] Kroemer NB, Lee Y, Pooeh S, Eppinger B, Goschke T, Smolka MN. L-DOPA reduces model-free control of behavior by attenuating the transfer of value to action. *NeuroImage* 2019;186:113–25. doi:10.1016/j.neuroimage.2018.10.075.
- [41] Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, et al. Mesolimbic dopamine signals the value of work. *Nature Neuroscience* 2016;19:117–26. doi:10.1038/nn.4173.
- [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. doi:10.1016/j.neuroimage.2012.04.024.
- [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. doi:10.1007/s00213-013-3313-4.
- [44] Frank MJ, Seeberger LC, O'Reilly RC. By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science* 2004;306:1940–3. doi:10.1126/science.1102941.
- [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. doi:10.1016/j.neuroimage.2014.12.070.
- [46] Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ. Modeling Avoidance in Mood and Anxiety Disorders Using Reinforcement Learning. *Biological Psychiatry* 2017;82:532–9. doi:10.1016/j.biopsych.2017.01.017.
- [47] Raudenbush S, Bryk A, Cheong Y, Congdon R, Du Toit M. HLM 7 [Computer software]. Lincolnwood, IL: Scientific Software International 2011.
- [48] 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 Computational Biology* 2011;7:e1002028. doi:10.1371/journal.pcbi.1002028.
- [49] Fellows I. Deducer: a data analysis GUI for R. *Journal of Statistical Software* 2012;49:1–15.
- [50] 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. *European Neuropsychopharmacology* 2018;28:719–31. doi:10.1016/j.euroneuro.2018.03.009.
- [51] Horstmann A, Fenske WK, Hankir MK. Argument for a non-linear relationship between severity of human obesity and dopaminergic tone. *Obesity Reviews* 2015;16:821–30. doi:10.1111/obr.12303.
- [52] 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. doi:10.1038/sj.npp.1301153.
- [53] 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. doi:10.3389/fnhum.2014.00784.
- [54] Vo A, Seergobin KN, Morrow SA, MacDonald PA. Levodopa impairs probabilistic reversal learning in healthy young adults. *Psychopharmacology* 2016;233:2753–63. doi:10.1007/s00213-016-4322-x.
- [55] Cools R. Dopaminergic modulation of cognitive function-implications for l-DOPA treatment in Parkinson's disease. *Neuroscience & Biobehavioral Reviews* 2006;30:1–23. doi:10.1016/j.neubiorev.2005.03.024.

- [56] 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. doi:10.1093/brain/aws083.
- [57] Beeler JA, Daw ND, Frazier CRM, Zhuang X. Tonic Dopamine Modulates Exploitation of Reward Learning. *Front Behav Neurosci* 2010;4. doi:10.3389/fnbeh.2010.00170.
- [58] 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. doi:10.1038/npp.2014.84.
- [59] 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 Sciences* 2017;6:e22169. doi:10.7554/eLife.22169.
- [60] 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. doi:10.1016/j.neuropsychologia.2006.03.030.
- [61] Frank MJ. Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience* 2005;17:51–72. doi:10.1162/0898929052880093.
- [62] De Deurwaerdère P, Di Giovanni G, Millan MJ. Expanding the repertoire of L-DOPA's actions: A comprehensive review of its functional neurochemistry. *Progress in Neurobiology* 2017;151:57–100. doi:10.1016/j.pneurobio.2016.07.002.
- [63] Lyubashina O, Panteleev S. Effects of cervical vagus nerve stimulation on amygdala-evoked responses of the medial prefrontal cortex neurons in rat. *Neuroscience Research* 2009;65:122–5. doi:10.1016/j.neures.2009.06.002.
- [64] Averbeck BB, Costa VD. Motivational neural circuits underlying reinforcement learning. *Nat Neurosci* 2017;20:505–12. doi:10.1038/nn.4506.
- [65] Dayan P, Yu AJ. Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network: Computation in Neural Systems* 2006;17:335–50. doi:10.1080/09548980601004024.
- [66] 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. doi:10.1016/j.neuron.2013.04.037.
- [67] Jepma M, Murphy PR, Nassar MR, Rangel-Gomez M, Meeter M, Nieuwenhuis S. Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. *PLOS Computational Biology* 2016;12:e1005171. doi:10.1371/journal.pcbi.1005171.
- [68] 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. *Obesity Reviews* 2007;8:21–34. doi:10.1111/j.1467-789X.2006.00270.x.
- [69] Hevey D, Thomas K, Laureano-Schelten S, Looney K, Booth R. Clinical Depression and Punishment Sensitivity on the BART. *Front Psychol* 2017;8. doi:10.3389/fpsyg.2017.00670.