**Anodal tDCS over the medial prefrontal cortex enhances behavioral adaptation after punishments during reversal learning through increased updating of unchosen choice options**

**Abbreviated title:** Anodal tDCS of mPFC enhances behavioral adaptation

Martin Panitz1,2\*,Lorenz Deserno1,3\*, Elisabeth Kaminski1,4, Arno Villringer1,5,6, Bernhard Sehm1,7, Florian Schlagenhauf1,2

*1Max-Planck-Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany;*

*2Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany;*

*3Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, 97080 Würzburg, Germany;*

*4 Institute for General Kinesiology and Exercise Science, Faculty of Sport Science, University of Leipzig, Germany;*

*5Clinic for Cognitive Neurology, University Hospital Leipzig, 04103 Leipzig, Germany*

*6Mind Brain Body Institute at Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10099 Berlin, Germany*

*7Department of Neurology, Martin-Luther-University of Halle-Wittenberg, 06120 Halle (Saale), Germany*

\*equal contribution

**Corresponding author:** Martin Panitz,panitz@cbs.mpg.de

**Number of pages:** 33

**Number of tables:** 4

**Number of figures:** 4

**Number of words:** Abstract: 244 | Significance statement: 116 | Introduction: 534 | Discussion: 1489

**Funding, disclosure and acknowledgements**

This study was supported by the Max Planck Society and by grants from the German Research Foundation awarded to F.S. (SCHL 1969/2-2/5-1). The authors declare no competing financial interests.We thank all the participants who took part in this study. Furthermore, we thank K. Hudl, J. Watson, L. Luettgau and R. Menger for their assistance in recruitment and data acquisition. In addition, we thank S. L. Valk for help on designing the figures and helpful comments on an earlier version of the manuscript.

**Abstract**

The medial prefrontal cortex (mPFC) is assumed to be central for flexible behavioral adaptation. However, a potential causal relationship between mPFC activity and such behavior is incompletely understood. In the present study, we investigated whether transcranial direct current stimulation (tDCS) of the mPFC alters flexible behavioral adaptation during reward-based decision-making. We targeted a specific region of the mPFC previously associated with impaired flexible behavioral adaptation in alcohol-dependent patients. To this end, healthy human participants (n = 61; 30 female, 31 male) received anodal (n = 30) or cathodal (n = 31) tDCS versus sham tDCS while performing a reversal learning task. To assess the mechanisms of reinforcement learning (RL) underlying our behavioral observations, we applied models that varied with respect to the updating of the unchosen choice option.

We observed that anodal tDCS induced increased choice switching after punishments as compared to sham stimulation, while cathodal as compared to sham stimulation showed no effect on participants’ behavior. RL revealed increased updating of the respective unchosen choice option under anodal as compared to sham stimulation. This computational parameter of enhanced learning about unchosen choice options accounted well for the increase in choice switching after punishments under anodal as compared to sham tDCS. We demonstrate a mechanism of how the mPFC impacts on flexible reward-based behavioral adaptation through enhancing learning about unchosen choice options. The current findings provide a potential model for tDCS-interventions in clinical conditions related to flexible behavioral adaptation after punishments such as addiction.

**Significance Statement**

Flexible adaptive behavior is essential for survival and a core capacity of human beings. Altered control of behavioral adaptation has been linked to significant health problems such as addiction. We demonstrate that anodal tDCS of the mPFC results in increased choice switching after punishments in a reward-based reversal learning paradigm. Computational modeling revealed enhanced learning about the respective unchosen choice option under anodal stimulation relative to sham stimulation. By demonstrating a causal manipulation of reward-based learning, our research substantiates evidence for a crucial link of mPFC to adapt behavior after punishments. Moreover, our work might motivate future interventions in clinical populations in order to alleviate deficits in flexible behavioral adaptation after punishments for example in addiction.

**Introduction**

Flexible behavioral adaptation is a crucial capacity to survive in dynamic environments. To investigate behavioral adaptation, instrumental reversal learning tasks have been deployed in several studies (e.g., O’Doherty et al., 2001; Cools et al., 2002). One example to probe behavior in such tasks on a computational level is reinforcement learning (RL) from observed outcomes, such as rewards or punishments, to repeat or adjust choices (Sutton and Barto, 1998). However, humans can adapt their behavior flexibly by also taking into account unobserved outcomes. Hence, on a computational level, flexible behavioral adaptation during reversal learning can be enhanced by counterfactual inference about unchosen choice options (inference about the correlation of reward probabilities). This process constitutes an extension of incremental RL (Li and Daw, 2011; Reiter, 2016).

Previous work has linked behavioral adaptation during reversal learning to a fronto-striatal circuitry comprising the orbitofrontal cortex and the ventral striatum, as well as medial prefrontal regions (Cools et al., 2002; Izquierdo et al., 2017). The medial prefrontal cortex (mPFC) is suggested to play an essential role regarding performance monitoring during such tasks (Izquierdo et al., 2017). In particular, the ventromedial prefrontal cortex (vmPFC) has been linked to abstract inferences about higher-order structures during reversal learning (Hampton et al., 2006). It was proposed that the vmPFC encodes the value of the currently chosen option, whereas the anterior prefrontal cortex encodes the value of the unchosen choice option (Rushworth et al., 2011). In a previous study of our research group (Reiter et al., 2016), alcohol-dependent patients exhibited a well-known impairment in flexibly adjusting behavior as compared to healthy subjects. Computational modeling of behavior revealed reduced updating of the unchosen choice options after punishments as a process underlying the disrupted behavioral adaptation. This impairment was also related to reduced coding of neural error signals incorporating values of the unchosen choice option in the mPFC. However, from such clinical studies, insight is inherently limited to what extent change in mPFC activity is directly involved in modifying behavior.

In the current study, we aimed to investigate a causal link between mPFC activity and flexible behavioral adaptation in healthy adults. To this end, we applied transcranial direct current stimulation (tDCS) to the mPFC, targeting a coordinate reported to reflect inter-individual differences in abstract inference about unchosen choice options in our previous study (Reiter et al., 2016). TDCS is a non-invasive brain stimulation technique that modulates cortical excitability, allowing experimentally controlled conclusions about brain activity in a particular area and associated behavior (Nitsche et al., 2000; Kuo and Nitsche, 2015). While receiving tDCS, participants performed the same reversal learning task as used in our previous study in alcohol-dependent patients (Reiter et al., 2016). In addition to behavioral analyses of the observed behavior we sought to investigate the underlying mechanisms of flexible decision-making by the means of computational modeling of RL. Based on the findings of Reiter et al. (Reiter et al., 2016) together with the assumption that anodal stimulation increases and cathodal stimulation decreases excitability of the underlying cortical structures (Nitsche et al., 2000; Bestmann et al., 2015), we hypothesized that anodal tDCS of the mPFC would lead to increased and cathodal tDCS of the mPFC to decreased updating of the unchosen choice options in RL.

**Materials and Methods**

**Participants.**65healthy human participants were recruited. Before participation, volunteers underwent a medical examination to exclude any evidence for neurological diseases or contraindications such as intake of regular medication, including CNS-active drugs. Four participants were excluded from analyses due to inadequate task performance (see computational methods). Consequently, 61 participants were included into analyses, out of which 30 participants were part of an anodal tDCS group, receiving anodal and sham stimulation (15 female, age: *M* = 26.3, *SD* = 4.1, range = 20 – 35 years), and 31 participants were part of a cathodal tDCS group, receiving cathodal and sham stimulation (15 female, age: *M* = 27.0, *SD* = 3.2, range = 22 – 38 years). Both groups received sham and the respective *verum* stimulation in a fully counterbalanced within-subject design (**Figure 1C**). All participants gave written informed consent before the study and were financially compensated for participation. The study was approved by the local ethics committee of the University of Leipzig.

**Experimental design.**Participants were assigned to one of two groups, which did not differ regarding age, graduation or verbal intelligence (**Table 1**). Both groups were tested using a double-blind, sham-controlled, within-subject design. Importantly, groups differed with respect to the polarity of tDCS received during *verum* stimulation, i.e. anodal or cathodal stimulation, resulting in an anodal tDCS group (a-tDCS group) and a cathodal tDCS group (c-tDCS group). Intervals of as possible one week between testing sessions (testing interval: *M* = 7.4, *SD* = 2.0, range = 6 – 21 days) avoided potential carry-over effects of stimulation. Experiments were performed effectively during the same time in the same subjects between 8:30AM and 1:00PM to avoid potential interactions of daytimes. Gender was counterbalanced within groups as was intervention order (sham vs. *verum* stimulation).

**Decision-making task.** While receiving tDCS, participants performed a reward-based decision-making task (**Figure 1A/B**) as used previously (Reiter, 2016; Reiter et al., 2016, 2017). Two versions of the task with different stimuli were available and counterbalanced within groups. In 160 trials, participants had to choose between one of two cards, each represented by a different geometric symbol. Stimuli were randomly assigned to the left or right side. Within the stimulus presentation time (1.5 s), participants had to press a left or right button, after which the selected card was highlighted along with monetary win (10 Eurocent coin) or monetary loss (crossed 10 Eurocent coin) for 0.5 s. A fixation cross was presented during the exponential distributed, jittered inter-trial interval (min. 1.0 s, max. 12.5 s). If no response was given on time, the message “too slow” appeared. One of the two cards was associated with a high reward probability (80%) and a low punishment probability (20%), whereas the inverted reward probabilities pertained to the other card. Thus, outcome probabilities were perfectly anti-correlated, resulting in a simple higher-order structure of the task. For the first 55 trials (‘pre-reversal’ phase) the reward contingencies remained stable, whereas during the following 70 trials (‘reversal’ phase) the reward contingencies switched four times, alternating after 15 or 20 trials. During the last 35 trials (‘post-reversal’ phase) the reward contingencies were stable again. Due to the probabilistic nature of the task, feedback could be either informative or misleading (e.g. getting an informative reward vs. getting a misleading punishment when choosing the option with the 80% reward probability). Because feedback was drawn probabilistically, differences in proportion of the number of informative and misleading events between the two sessions of a participant were matched between individuals in both the a-tDCS and the c-tDCS group. On every testing day, participants got standardized instructions about the task with a subsequent training of 20 practice trials without any reversals using a different set of stimuli. Before practicing, participants were informed that one of the two cards had a superior chance of winning money and that they should try to win as much money as possible, as the earned money would be paid out at the end of testing. After the practice trials, instructions ended with a notification that the “better card” could change during the experiment and participants were told to be attentive to such changes. No other information on reversals or the anti-correlated task structure was provided.

**Transcranial direct current stimulation (tDCS).**For non-invasive brain stimulation, a direct current of 1 mA was delivered for twenty minutes using a battery-driven DC stimulator (neuroConn, Germany). Targetof our stimulation protocol was the mPFC based on its potential key role in enabling flexible value-based decision-making (Fellows and Farah, 2003, 2005, 2007; O’Doherty et al., 2003; Rushworth et al., 2009, 2011, 2012; O’Doherty, 2011; Glascher et al., 2012). Using neuronavigation (Brainsight Version 2; Rogue Research, Canada) and a standard brain, one stimulation electrode (4 x 4 cm, current density: .0625 mA/cm²) was placed over the mPFC region targeting MNI-coordinate: X = -8, Y = 62, Z = 12, reflecting peak statistics of inter-individual differences in coding of neural signatures incorporating values of the unchosen option, which were linked to inter-individual differences in behavior, reported in a previous study of our research group (Reiter et al., 2016). The reference electrode (10 x 10 cm, current density: .01 mA/cm2) was centrally placed on Cz according to the EEG 10-20 system. The small size of the stimulation electrode intentionally maximizes the current density in the cortical target region in order to promote the desired modulation of cortical excitability, whereas the large size of the reference electrode minimizes current density in the underlying cortex in order to enhance focality of tDCS (Nitsche and Doemkes, 2007). The electrodes were attached to the participants’ heads using elastic straps in a pair of saline-soaked sponges. To minimize impedance, the scalp was carefully cleaned with 70% isopropyl alcohol pads before attaching the electrodes to the head. The impedance of stimulation electrodes was always kept below 10 kΩ for each participant. To minimize side effects, such as itching or tingling sensations, current was ramped up and down for 30 s before and after stimulation. During *verum* stimulation, current was delivered for 20 min, while during sham stimulation current was maintained for only 30 s. Five minutes after the onset of stimulation, the experimental task was started. At the end of each testing session all participants were asked about their estimate whether sham or *verum* stimulation was administered. Successful blinding was tested by questionnaires applying binomial tests on false/right ratings at a *p*-level of .05.

**Behavioral data analysis.**Behavioral performance was quantified as choices of the stimulus with the 80% reward probability (correct choices) using MATLAB (The MathWorks, version 9.1.0.441655, R2016b) and was analyzed using a mixed-design ANOVA in JASP (JASP Team, JASP Version 0.12.1, 2020) with within-subjects factors stimulation (sham stimulation, *verum* stimulation), phase (pre-reversal, reversal, and post-reversal phase) and between-subjects factor group (a-tDCS group, c-tDCS group). Further, we investigated the effect of previous feedback on subsequent choices. To this end, we calculated stay-behavior depending on feedback (reward vs. punishment) in the previous trial (‘win-stay’ vs. ‘lose-stay’) and on the three phases of the task. Subsequently, we performed a mixed-design ANOVA using JASP on stay-behavior with within-subjects factors stimulation, feedback and phase, and between-subjects factor group. In case of violated assumptions of sphericity, we report Greenhouse-Geisser corrected *F*-values.

**Computational modeling of behavior.** To investigate the underlying mechanisms of flexible reward-based decision-making, we formalized learning and decision-making processes via computational models of RL. In RL, values are learned from past experiences and choices are made based on maximized expected value (Sutton and Barto, 1998). For all models, expectations are updated via a reward prediction error (RPE) δ:

(1)

with denoting the expectation for receiving a reward or punishment on trial and representing the actually received outcome. The RPE is used to update expectation of the next trial and is weighted by the learning rate :

(2)

So far, the agent updates expectations for the chosen stimulus only, which we refer to as single-update (‘SU’). An agent could also update expectations about the unchosen stimulus, thereby learning about the anti-correlated structure of the task. In such a model, which we refer to as double-update (‘DU’), an increase of the expectation for the chosen card connotes a decrease of the expectation for the unchosen card . The prediction error of the unchosen stimulus is computed as follows:

(3)

Thus, the expectation for the unchosen stimulus is updated following:

(4)

To account for possible inter-individual differences regarding the extent of updating the unchosen stimulus, we implemented a model where learning about the unchosen choice option is weighted by an additional parameter κ, which we refer to as individually double-update (‘iDU’) (Reiter, 2016; Reiter et al., 2016, 2017):

(5)

RL models could include one learning rate α, independent of the kind of feedback received (‘1α’). We additionally implemented models with separate learning rates for reward and punishment trials, αrew and αpun (‘2α’) (Reiter et al., 2016). For all models, in order to link learning to actual choices, we transformed learned expectations for each trial, i.e. , to choice probabilities by deploying a logistic function, the softmax:

(6)

The softmax equation includes the parameter β, which reflects inverse decision noise with highβchoices are being tightly determined by choice expectations and low β leading to more stochastic choices. Taken together, the model space was formed out of the factors ‘updating’ (SU, DU, iDU) and ‘learning rate’ (1α, 2α), resulting in a total of 6 models (SU-1α-β, DU-1α-β, iDU-1α-β, SU-2α-β, DU-2α-β, iDU-2α-β). See **Table 2** for a summary of model parameter priors.

Alternatively, we also analyzed a larger model space comprising additional hierarchical Bayesian learning models with a dynamic learning rate (specifically, the hierarchical Gaussian filter, HGF) (Mathys, 2011) as well as different decision models capturing choice repetition independent of learning as comparably applied in a study by Deserno et al. (Deserno et al., 2020). Model selection for that alternative model space replicated the findings of Deserno et al., but observed parameter differences regarding our tDCS intervention did not account well for the observed choice behavior. We thus refrained from reporting this in the current manuscript but the data can be made available upon request.

*Model fitting***.** Maximum-a-posteriori estimates of model parameters were derivedutilizing the HGF toolbox version 3 (included in the open source software collection TAPAS, <https://translationalneuromodeling.github.io/tapas/>). For optimization, a quasi-Newton algorithm was applied.

*Model selection***.**For relative model comparison, we applied random-effects Bayesian model selection (RFX-BMS; Stephan et al., 2009) using the VBA-toolbox (Daunizeau et al., 2014) to obtain the posterior probabilities (PP), exceedance probabilities (XP) and protected exceedance probabilities (PXP). PP represent the estimated model frequencies of the models considered. The XP represent the posterior probabilities that any given model is more frequent than all others. The PXP adjust XP for the possibility that differences can occur by chance (Rigoux et al., 2014). We examined whether model frequencies were different between the a-tDCS and the c-tDCS group by applying between-groups RFX-BMS, and whether model frequencies were equal across stimulation conditions across both groups by performing within-subject RFX-BMS. In case of model stability both between-groups and within-subjects, RFX-BMS on the pooled log-evidences was applied.

In addition, participants not fit better than chance by any model were identified using a binomial test on the percentage of choices explained by the model derived from negative log-likelihood relative to the number of trials (see Huys et al., 2012; Schlagenhauf et al., 2014) and excluded from the analysis (a-tDCS group: one participant, c-tDCS group : three participants). Excluding those subjects did not change the results.

To further demonstrate the validity of the model to capture the empirical data, we ran 10000 simulations of the task per tested session based on the inferred parameters of the winning model and subsequently - using the mean of the respective 10000 simulations - performed the same analysis on the simulated choice data as on the observed data.

*Parameter comparison.*Individually estimated parameters were tested forwithin-subject differences utilizing paired-samples *t*-tests or, in case of non-normally distributed parameters as examined by Shapiro-Wilk tests, non-parametric Wilcoxon signed-rank tests. Due to our a-priori hypothesis about the parameter κ we did not apply corrections for multiple comparisons.

**Neuropsychological testing.**To assess the influence of tDCS on general cognitive capacities, participants completed a working memory task, the Digit Span Backwards Test (Wechsler, 1955), and a task on cognitive speed, the Digit Symbol Substitution Test (DSST) (Wechsler, 1955). The two tasks were performed after completing the reversal learning task on both testing sessions. Using JASP, test scores were analyzed by a mixed-model ANOVA with within-subjects factor stimulation (sham stimulation, *verum* stimulation) and between-subjects factor group (a-tDCS group, c-tDCS group), respectively.

**Results**

*Stimulation protocol*. All participants tolerated the stimulation well and blinding was effective as participants could not detect whether they received sham or *verum* stimulation (a-tDCS group: sham stimulation proportion (false/right) = .50/.50, *p* = 1.0, anodal stimulation proportion (false/right) = .47/.53, *p* = .86; c-tDCS group: sham stimulation proportion (false/right) = .52/.48, *p* = 1.0, cathodal stimulation proportion (false/right) = .58/.42, *p* = .47).

*Behavior in the decision-making task*.No significant stimulation effects were observed on choosing the stimulus with higher reward probability, i.e. ‘correct choices’ (main effect of stimulation, *F*(1,59) = .03, *p* = .86; stimulation x group interaction, *F*(1,59) = .39, *p* = .53; stimulation x phase x group interaction, *F*(2,118) = .20, *p* = .82). As expected, a significant effect of phase indicated performance differences between the three phases of the task (main effect of phase, *F*(1.62,95.73) = 48.59, *p* < .001, partial η2 = .45; see **Figure 1D)**.

Regarding stay-behavior, participants stayed more with the previous choice after rewards compared with punishments (main effect of feedback, *F*(1,59) = 388.04, *p* < .001, partial η2 = .87). Further, we observed a significant stimulation (sham vs. *verum*) x feedback (reward vs. punishment) x group (a-tDCS vs. c-tDCS) interaction on stay-behavior (*F*(1,59) = 5.23, *p* = .026, partial η2 = .08). The stimulation x group interaction (*F*(1,59) = 3.29, *p* = .075) as well as all other remaining main effects or interactions were not significant (all other *p*-values > .16).

To follow-up on this three-way interaction, we performed a repeated-measures ANOVA on stay-behavior within each group. We observed a significant stimulation x feedback interaction in the a-tDCS group (*F*(1,29) = 5.40, *p* = .027, partial η2 = .16; for all *p*-values see **Table 3**), while we found no significant effects linked to stimulation in the c-tDCS group (all *p*-values > .30).

In the a-tDCS group, post-hoc paired-samples *t*-tests revealed that participants stayed less with the previous choice after punishments under anodal stimulation compared to sham stimulation (lose-stay: *t*(29) = 2.74, *p* = .010, Cohen’s *d* = .50; sham stimulation: mean = .59, *SD* = .14, anodal stimulation: mean = .53, *SD* = .19; see **Figure 2**) and unaffected stay-behavior after rewards (win-stay: *t*(29) = -.45, *p* = .66).

*Computational modeling*.Between-groups RFX-BMS showed a high probability that both groups had the same model frequencies (PP = .99). Further, in both groups, between-conditions RFX-BMS revealed strong evidence for model stability across stimulation conditions, i.e. within-subject model stability (a-tDCS group, XP = 1.00, PXP = .98; c-tDCS group, XP = 1.00, PXP = 1.00). RFX-BMS on the pooled log-evidences revealed that the model with individual double-updating and one learning rate was the relatively best fitting model (iDU-1α-β: PP=.39, XP=.72, PXP=.72; see **Figure 3**).

In line with our hypothesis, parameter comparison of the best fitting model (iDU-1α-β) revealed significantly increased updating of the unchosen choice option, represented by αuc (i.e. the product of κ times α), under anodal stimulation in the a-tDCS group (αuc: *W* = 350.00, *p* = .015, matched rank biserial correlation *r* = .51; κ: *t* = 2.16, *p* = .040, Cohen’s *d* = .39; all other *p*-values > .60; see **Table 4** and **Figure 4A**). Notably, the change on αuc induced by anodal stimulation was significantly correlated with the observed behavioral change on lose-stay induced by anodal stimulation in the a-tDCS group (ρ = -.48, *p* = .0083, **Figure 4B**). In the c-tDCS group, congruent to the absence of a behavioral effect, we observed no significant parameter differences (β: *W* = 156.00, *p* = .073; all other *p*-values > .20, **Table 4**).

When simulating choice data based on the inferred parameters, the best fitting iDU-1α-β reproduced the effect on lose-stay behavior in the a-tDCS group (stimulation x feedback interaction, *F*(1,29) = 3.43, *p* = .074, partial η2 = .11; see **Figure 4C**).

*Neuropsychology.*Analyzing the Digit Span Backwards Test revealed a main effect of group (*F*(1,59) = 5.24, *p* = .026, partial η2 = .08) comparing a-tDCS and c-tDCS group. However, this was qualified by a significant stimulation x group interaction (*F*(1,59) = 4.14, *p* = .047, partial η2 = .07). Repeated measures ANOVA within each group showed a significant main effect of stimulation in the a-tDCS group (*F*(1,29) = 5.72, *p* = .024, partial η2 = .17), whereas we found no significant effect of stimulation in the c-tDCS group (*F*(1,30) = .26, *p* = .61). In the a-tDCS group, post-hoc testing revealed an impairment in working memory as measured with the Digit Span Backwards Test under anodal compared to sham stimulation (*W* = 296.50, *p* = .031, matched rank biserial correlation *r* = .46, Wilcoxon signed-rank test; sham stimulation: mean = 8.43, *SD* = 2.64, anodal stimulation: mean = 7.70, *SD* = 2.20). We found no significant correlation between the change in working memory induced by anodal stimulation and either the behavioral change on lose-stay induced by anodal stimulation (*r* = .18, *p* = .35) or the change in the modeling parameter αuc (ρ = -.09, *p* = .64) in the a-tDCS group. We observed no significant effects regarding cognitive speed as measured with the Digit Symbol Substitution Test (all *p*-values > .70).

**Discussion**

In the present study, we examined the effect of tDCS over the mPFC on reward-based reversal learning in healthy adults. We observed that anodal tDCS modified participants’ behavioral adaption such that participants showed an increased tendency to switch after punishments. Our RL model accounted for this effect by enhanced learning about the respective unchosen choice option under anodal tDCS compared to sham stimulation. In contrast, we observed no behavioral effect for cathodal tDCS compared to sham stimulation.

We observed no significant differences regarding correct choices between stimulation conditions, but we found a feedback-specific effect on stay-behavior in the a-tDCS group. Thus, despite the effect on feedback-based behavior, tDCS did not have an effect to such an extent that overall performance was changed. This may be due to the investigated sample of largely high-educated participants, who performed the task very well. Anodal tDCS rather induced a subtle but highly specific alteration of how participants accomplished the task by inducing reduced lose-stay behavior. Comparison of computational parameters between the sham and anodal condition revealed an underlying process: increased updating of the unchosen choice option. We interpret this as an increase of inference about the alternative choice option under anodal stimulation. After a punishment, a higher parameter κ determines a sharper increased value for the unchosen choice option, thus, the forgone rewarding option becomes more attractive. Put differently, with a higher κ an agent becomes more sensitive to negative feedback through consideration of reward at the alternative choice, which results in an enhanced propensity to switch choices after punishments. At the same time, a higher κ should also lead to a propensity to stay more after rewards. However, in the present study, we found no significant difference regarding ‘win-stay’ in the a-tDCS group, although participants descriptively stayed more after rewards under anodal stimulation. However, win-stay is overall very high, rendering it unlikely to detect within-subject stimulation changes.

A study by Beharelle et al. (Beharelle et al., 2015) showed that tDCS over the right frontopolar cortex altered the offset between exploration and exploitation during a reward-based decision-making task. Interestingly, the placement of the electrodes was rather similar to our study. They found that anodal tDCS lead to higher sensitivity to recent negative PEs on exploitative choices, resulting in participants more likely to explore and choose one of two alternative options. Intuitively, this finding fits quite well to our behavioral finding that participants showed a propensity to switch more after punishments under anodal stimulation. This similarity is remarkable despite the differences in tasks and modeling. In another study evaluating the association between decision-making and frontal cortex activity, Boorman et al. (Boorman et al., 2009) found that during a two-armed bandit task activity of bilateral frontopolar cortex (FPC) was associated with the so-called “relative unchosen probability”. This refers to the representation of the relation between unchosen and chosen choice probabilities, hereby gathering inference about the relative advantage in favor of a switch to the alternative choice option. They reported an effect of the “relative unchosen probability” in the FPC to be accompanied by a higher probability to switch to the better choice option. While the task and modeling were more specifically designed as compared to our simple reversal learning task, the findings resonate well with our modeling approach regarding learning about the unchosen choice options, which was elevated under anodal stimulation. Indeed, it is possible that our tDCS intervention also modulated cortical activity in the FPC, thereby possibly enhancing constantly made inferences about a potential favorable switch in behavior. Given that current density is proposed to have its peaks not under the center but beneath the edges of electrodes (Kronberg and Bikson, 2012), it is possible that cortical structures found by Boorman et al., particularly the left FPC, might have been significantly affected by our tDCS intervention. Our modeling revealed a value-based learning effect. This is in contrast to a study by Hämmerer et al. (Hämmerer et al., 2016), where anodal tDCS over the vmPFC increased randomness of choices during value-based decision-making, behaviorally reflected by a reduced percentage of correct choices. Different characteristics of the task design likely contribute to this divergence (task with slowly drifting ‘random walks’ reward probabilities). Moreover, the tDCS stimulation protocol (e.g. direct current of 2 mA, reference electrode below the inion) of the study by Hämmerer et al. also differed.

Our finding of altered lose-stay behavior is in line with previous research on behavioral adaptation in regard to serotonergic transmission (Cools et al., 2011). Den Ouden et al. (den Ouden et al., 2013) related lose-stay behavior to a genetic polymorphism encoding the serotonin transporter, whereas a polymorphism associated with the dopamine transporter influenced perseveration after reversals. It appears plausible that the current intervention affected the serotonergic circuit, given the role of mPFC in serotonergic processing, e.g. afferences from the mPFC to the dorsal raphe nucleus (Celada et al., 2001). Previous studies suggest that anodal tDCS enhances functions of serotonergic transmission (Das et al., 2016) and evidence has been found that extracellular serotonin levels affect anodal (and cathodal) tDCS effects on motor cortex excitability (Nitsche et al., 2009). Thus, in the present study, anodal tDCS might have led to altered functioning of the serotonergic system.

Anodal stimulation led to an impairment in working memory in the a-tDCS group examined via the Digit Span Backwards Test. Notably, this effect was not related to the behavioral effect seen during reversal learning. Working memory capacity has been associated with anticorrelated activity between mPFC and dlPFC (Keller et al., 2015), thus, anodal tDCS of mPFC might have altered this balance in our participants.

In the current study, we did not observe effects of cathodal stimulation on decision-making behavior. Considering the results of numerous tDCS studies, the initial presumption that anodal stimulation causes excitatory and cathodal stimulation causes inhibitory effects has to be handled with caution (Bestmann et al., 2015; Parkin et al., 2015). Particularly regarding executive cognitive domains, inhomogeneous effect sizes and directionalities between anodal and cathodal tDCS have been reported (Jacobson et al., 2012), supporting the hypothesis that the effects of tDCS on executive functions are multifaceted. Regarding decision-making, analogous to our study, Bogdanov et al. (Bogdanov et al., 2015) and Soutschek et al. (Soutschek et al., 2018) both found significant behavioral effects for anodal but not for cathodal tDCS.

With often limited possibilities in regard to the treatment of neuropsychiatric conditions, tDCS serves as a promising easy-to-handle, inexpensive and well-tolerated tool for therapeutic settings. Indeed, tDCS previously showed beneficial effects in neuropsychiatric populations (Boggio et al., 2008; Klauss et al., 2014; Batista et al., 2015; Reinhart et al., 2015a, 2015b). Our stimulation target in the mPFC was based on the representations of neural activation differences found in alcohol-dependent patients compared to healthy participants (Reiter et al., 2016). In the study by Reiter et al., on the computational level, alcohol-dependent patients showed an impaired integration of unchosen choice options during reversal learning as reflected in the same parameter as affected by anodal tDCS in the present study. Our results suggest a possible potential for anodal tDCS over the mPFC to ameliorate deficits in flexible behavioral adaptation after punishments in alcohol-dependent patients. However, it should be noted that the observed behavior of the alcohol-dependent patients was much more severely impaired when contrasted to the behavior in our present tDCS study. This may limit or facilitate therapeutic potential. In line with this, the means of the derived learning parameters as well as mean age were considerable different in both alcohol-dependent patients and healthy control subjects as compared to participants in the present study. Further research investigating healthy subjects in different age ranges is needed before eventually testing clinical samples in order to unveil the possible therapeutic potential of a tDCS intervention regarding flexible behavioral adaptation.

Our study indicates a causal relationship between mPFC activity and specific decision-making processes. Further research combining stimulations protocols with concurrent neuroimaging methods is warranted to better understand the neural signatures underlying the behavioral changes observed in our intervention. Further, although the current target region in the mPFC was carefully selected based on previous research, the location is challenging to stimulate via tDCS. Before the applied current reaches the target region, it passes through cortical regions lying anterior and slightly dorsal to the target region, implicating that these regions also got significantly affected by the tDCS intervention. Moreover, we did not stimulate an active control region in order to better understand the regional functional specificity of tDCS effects over the mPFC (Polanía et al., 2018).

In sum, the results of our study point towards a mechanism of how the mPFC governs choice switching after punishments during reversal learning taking place through enhanced learning about the unchosen choice option. We demonstrate that flexible adaptive behavior and an herewith associated specific aspect of learning are directly malleable by anodal tDCS. Our observations could serve as a model for further research eventually aimed at alleviation of suffering in patients exhibiting alterations in flexible behavioral adaptation.

**References**

Batista EK, Klauss J, Fregni F, Nitsche MA, Nakamura-Palacios EM (2015) A randomized placebo-controlled trial of targeted prefrontal cortex modulation with bilateral tDCS in patients with crack-cocaine dependence. Int J Neuropsychopharmacol 18:1–11.

Bestmann S, de Berker AO, Bonaiuto J (2015) Understanding the behavioural consequences of noninvasive brain stimulation. Trends Cogn Sci 19:13–20.

Boehme R, Deserno L, Gleich T, Katthagen T, Pankow A, Behr J, Buchert R, Roiser JP, Heinz A, Schlagenhauf F (2015) Aberrant Salience Is Related to Reduced Reinforcement Learning Signals and Elevated Dopamine Synthesis Capacity in Healthy Adults. J Neurosci 35:10103–10111.

Bogdanov M, Ruff CC, Schwabe L (2015) Transcranial Stimulation Over the Dorsolateral Prefrontal Cortex Increases the Impact of Past Expenses on Decision-Making. Cereb Cortex 27:1094–1102.

Boggio PS, Sultani N, Fecteau S, Merabet L, Mecca T, Pascual-Leone A, Basaglia A, Fregni F (2008) Prefrontal cortex modulation using transcranial DC stimulation reduces alcohol craving: A double-blind, sham-controlled study. Drug Alcohol Depend 92:55–60.

Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS (2009) How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. Neuron 62:733–743.

Brunoni AR, Vanderhasselt MA (2014) Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: A systematic review and meta-analysis. Brain Cogn 86:1–9.

Celada P, Puig M V, Casanovas JM, Guillazo G, Artigas F (2001) Control of dorsal raphe serotonergic neurons by the medial prefrontal cortex: Involvement of serotonin-1A, GABA(A), and glutamate receptors. J Neurosci 21:9917–9929.

Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J Neurosci 22:4563–4567.

Cools R, Nakamura K, Daw ND (2011) Serotonin and dopamine: Unifying affective, activational, and decision functions. Neuropsychopharmacology 36:98–113.

Das S, Holland P, Frens MA, Donchin O (2016) Impact of transcranial direct current stimulation (tDCS) on neuronal functions. Front Neurosci 10:550.

Daunizeau J, Adam V, Rigoux L (2014) VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. PLoS Comput Biol 10(1): e1003441.

den Ouden HEM, Daw ND, Fernández G, Elshout JA, Rijpkema M, Hoogman M, Franke B, Cools R (2013) Dissociable Effects of Dopamine and Serotonin on Reversal Learning. Neuron 80:1090–1100.

Deserno L, Boehme R, Mathys C, Katthagen T, Kaminski J, Stephan KE, Heinz A, Schlagenhauf F (2020) Volatility Estimates Increase Choice Switching and Relate to Prefrontal Activity in Schizophrenia. Biol Psychiatry Cogn Neurosci Neuroimaging 5:173–183.

Fellows LK, Farah MJ (2003) Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. Brain 126:1830–1837.

Fellows LK, Farah MJ (2005) Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. Cereb Cortex 15:58–63.

Fellows LK, Farah MJ (2007) The role of ventromedial prefrontal cortex in decision making: Judgment under uncertainty or judgment per se? Cereb Cortex 17:2669–2674.

Glascher J, Adolphs R, Damasio H, Bechara A, Rudrauf D, Calamia M, Paul LK, Tranel D (2012) Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. Proc Natl Acad Sci 109:14681–14686.

Hämmerer D, Bonaiuto J, Klein-Flügge M, Bikson M, Bestmann S (2016) Selective alteration of human value decisions with medial frontal tDCS is predicted by changes in attractor dynamics. Sci Rep 6:25160.

Hampton AN, Bossaerts P, O’Doherty JP (2006) The Role of the Ventromedial Prefrontal Cortex in Abstract State-Based Inference during Decision Making in Humans. J Neurosci 26:8360–8367.

Huys QJM, Eshel N, O’Nions E, Sheridan L, Dayan P, Roiser JP (2012) Bonsai trees in your head: How the pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS Comput Biol 8(3): e1002410.

Izquierdo A, Brigman JL, Radke AK, Rudebeck PH, Holmes A (2017) The neural basis of reversal learning: An updated perspective. Neuroscience 345:12–26.

Jacobson L, Koslowsky M, Lavidor M (2012) TDCS polarity effects in motor and cognitive domains: A meta-analytical review. Exp Brain Res 216:1–10.

Keller JB, Hedden T, Thompson TW, Anteraper SA, Gabrieli JDE, Whitfield-Gabrieli S (2015) Resting-state anticorrelations between medial and lateral prefrontal cortex: Association with working memory, aging, and individual differences. Cortex 64:271–280.

Klauss J, Penido Pinheiro LC, Silva Merlo BL, Correia Santos G de A, Fregni F, Nitsche MA, Miyuki Nakamura-Palacios E (2014) A randomized controlled trial of targeted prefrontal cortex modulation with tDCS in patients with alcohol dependence. Int J Neuropsychopharmacol 17:1793–1803.

Kronberg G, Bikson M (2012) Electrode assembly design for transcranial Direct Current Stimulation: a FEM modeling study. Conf Proc IEEE Eng Med Biol Soc 2012:891–895.

Kuo MF, Nitsche MA (2015) Exploring prefrontal cortex functions in healthy humans by transcranial electrical stimulation. Neurosci Bull 31:198–206.

Li J, Daw ND (2011) Signals in Human Striatum Are Appropriate for Policy Update Rather than Value Prediction. J Neurosci 31:5504–5511.

Mathys C (2011) A Bayesian foundation for individual learning under uncertainty. Frontiers in Human Neuroscience 5:39.

Nitsche M a, Nitsche M a, Paulus W, Paulus W (2000) Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J Physiol 527 Pt 3:633–639.

Nitsche M, Doemkes S (2007) Shaping the effects of transcranial direct current stimulation of the human motor cortex. J Neurophysiol:3109–3117.

Nitsche MA, Kuo MF, Karrasch R, Wächter B, Liebetanz D, Paulus W (2009) Serotonin Affects Transcranial Direct Current-Induced Neuroplasticity in Humans. Biol Psychiatry 66:503–508.

O’Doherty J, Critchley H, Deichmann R, Dolan RJ, O&apos;Doherty J, Critchley H, Deichmann R, Dolan RJ (2003) Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. J Neurosci 23:7931–7939.

O’Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 4:95–102.

O’Doherty JP (2011) Contributions of the ventromedial prefrontal cortex to goal-directed action selection. Ann N Y Acad Sci 1239:118–129.

Parkin BL, Ekhtiari H, Walsh VF (2015) Non-invasive Human Brain Stimulation in Cognitive Neuroscience: A Primer. Neuron 87:932–945.

Polanía R, Nitsche MA, Ruff CC (2018) Studying and modifying brain function with non-invasive brain stimulation. Nat Neurosci 21:174–187.

Raja Beharelle A, Polania R, Hare TA, Ruff CC (2015) Transcranial Stimulation over Frontopolar Cortex Elucidates the Choice Attributes and Neural Mechanisms Used to Resolve Exploration-Exploitation Trade-Offs. J Neurosci 35:14544–14556.

Reinhart RMG, Zhu J, Park S, Woodman GF (2015a) Medial-Frontal Stimulation Enhances Learning in Schizophrenia by Restoring Prediction Error Signaling. J Neurosci 35:12232–12240.

Reinhart RMG, Zhu J, Park S, Woodman GF (2015b) Synchronizing theta oscillations with direct-current stimulation strengthens adaptive control in the human brain. Proc Natl Acad Sci U S A 112:9448–9453.

Reiter AMF (2016) The Feedback-related Negativity Codes Components of Abstract Inference during Reward-based Decision-making. J Cogn Neurosci 28:1127–1138.

Reiter AMF, Deserno L, Kallert T, Heinze H-J, Heinz A, Schlagenhauf F (2016) Behavioral and Neural Signatures of Reduced Updating of Alternative Options in Alcohol-Dependent Patients during Flexible Decision-Making. J Neurosci 36:10935–10948.

Reiter AMF, Heinze HJ, Schlagenhauf F, Deserno L (2017) Impaired Flexible Reward-Based Decision-Making in Binge Eating Disorder: Evidence from Computational Modeling and Functional Neuroimaging. Neuropsychopharmacology 42:628–637.

Rigoux L, Stephan KE, Friston KJ, Daunizeau J (2014) Bayesian model selection for group studies - Revisited. Neuroimage 84:971–985.

Rushworth MF, Mars RB, Summerfield C (2009) General mechanisms for making decisions? Curr Opin Neurobiol 19:75–83.

Rushworth MFS, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: One or many serial or parallel systems? Curr Opin Neurobiol 22:946–955.

Rushworth MFS, Noonan MAP, Boorman ED, Walton ME, Behrens TE (2011) Frontal Cortex and Reward-Guided Learning and Decision-Making. Neuron 70:1054–1069.

Schlagenhauf F, Huys QJM, Deserno L, Rapp MA, Beck A, Heinze HJ, Dolan R, Heinz A (2014) Supplement: Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. Neuroimage 89:171–180.

Soutschek A, Kang P, Ruff CC, Hare TA, Tobler PN (2018) Brain Stimulation Over the Frontopolar Cortex Enhances Motivation to Exert Effort for Reward. Biol Psychiatry 84:38–45.

Soutschek A, Ugazio G, Crockett MJ, Ruff CC, Kalenscher T, Tobler PN (2017) Binding oneself to the mast: stimulating frontopolar cortex enhances precommitment. Soc Cogn Affect Neurosci 12:635–642.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017.

Sutton RS, Barto AG (1998) Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press.

Wechsler D (1955) Manual for the Wechsler Adult Intelligence Scale. Oxford, Engl Psychol Corp.

**Tables**

|  |  |  |  |
| --- | --- | --- | --- |
|  | a-tDCS group (n = 30) | c-tDCS group (n = 31) | Test statistic |
| Age | 26.3 ± 4.1 | 27.0 ± 3.2 | *t*(59) = -.71, *p* = .48 |
| Gender | 15 female / 15 male | 15 female / 16 male | χ²(1) = .02, *p* = .90 |
| Smoking (smoking / non-smoking) | 8 / 22 | 3 / 28 | χ²(1) = 2.98, *p* = .084 |
| Graduation (0 = none, 1 = secondary modern school-leaving certificate, 2 = intermediate school-leaving certificate, 3 = university entrance qualification) | 3.0 ± 0 | 3.0 ± 0 |  |
| German vocabulary test (verbal intelligence) | 34.1 ± 2.5 | 33.6 ± 2.4 | *t*(59) = .88, *p* = .38 |
| Neo neuroticism | 14.0 ± 4.1 | 12.7 ± 3.8 | *t*(59)= 1.31, *p* = .20 |
| Neo extraversion | 20.4 ± 4.1 | 21.6 ± 3.2 | *t*(59) = -1.29, *p* = .20 |
| Neo openness to experience | 22.1 ± 3.8 | 20.5 ± 3.7 | *t*(59) = 1.71, *p* = .092 |
| Neo agreeableness | 20.5 ± 3.8 | 19.2 ± 3.8 | *t*(59) = 1.30, *p* = .20 |
| Neo conscientiousness | 23.4 ± 3.3 | 24.6 ± 3.6 | *t*(59)= -1.36, *p* = .18 |
| BIS-11 | 61.9 ± 8.0 | 59.3 ± 8.7 | *t*(59) = 1.19, *p* = .24 |
| UPPS premeditation | 32.6 ± 4.4 | 32.1 ± 3.5 | *t*(59) = .49, *p* = .62 |
| UPPS urgency | 34.6 ± 4.7 | 35.7 ± 4.5 | *t*(59) = -.97, *p* = .33 |
| UPPS sensation seeking | 26.9 ± 7.0 | 27.1 ± 8.1 | *t*(59) = -.10, *p* = .92 |
| UPPS perseverance | 30.6 ± 4.8 | 32.6 ± 3.8 | *t*(59) = -1.82, *p* = .074 |
| UPPS total | 124.7 ± 13.2 | 127.5 ± 11.1 | *t*(59) = -.92, *p* = .36 |
| BDI | 6.1 ± 6.1 | 4.9 ± 4.5 | *t*(59) = .87, *p* = .39 |
| STAI | 37.6 ± 8.5 | 35.8 ± 6.9 | *t*(59) = .91, *p* = .37 |

**Table 1. Sample characteristics.** Data are reported asgroup means with standard deviations, unless indicated otherwise.

|  |  |  |
| --- | --- | --- |
|  | Prior Mean | Prior Variance |
| *Learning models* |  |  |
| α(1α) | .5 | 1 |
| αrew (2α) | .5 | 1 |
| αpun (2α) | .5 | 1 |
| κ(iDU) | .1 | 1 |
| *Decision model* |  |  |
| β | 1 | 1 |

**Table 2. Parameter prior means and variances used in computational models.**

|  |  |
| --- | --- |
| stimulation | *F*(1,29) = 3.64, *p* = .066, partial η2 = .11 |
| feedback | *F*(1,29) = 185.02, *p* < .001, partial η2 = .86 |
| phase | *F*(2,58) = .29, *p* = .75, partial η2 = .01 |
| stimulation x feedback | *F*(1,29) = 5.40, *p* = .027, partial η2 = .16 |
| stimulation x phase | *F*(1.61,46.65) = 2.22, *p* = .13, partial η2 = .07 |
| feedback x phase | *F*(2,58) = .34, *p* = .72, partial η2 = .01 |
| stimulation x feedback x phase | *F*(2,58) = 2.90, *p* = .063, partial η2 = .09 |

**Table 3. Behavioral data analysis.** Repeated measures ANOVA on stay-behavior in the a-tDCS group with factors stimulation (sham vs. *verum*), feedback (reward vs. punishment) and phase (pre-reversal vs. reversal vs. post-reversal).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | a-tDCS group | | | c-tDCS group | | |
|  | Sham stimulation | Anodal stimulation | Test statistic | Sham stimulation | Cathodal stimulation | Test statistic |
| *Learning parameters* | | | | | | |
| α | .44 ± .16 | .47 ± .19 | *W* = 255.00, *p* = .66 | .50 ± .17 | .48 ± .17 | *W* = 186.00, *p* = .23 |
| κ | .33 ± .12 | .41 ± .15 | *t* = 2.16, *p* = .040,  Cohen’s *d* = .39 | .36 ± .13 | .37 ± .13 | *W* = 263.00, *p* = .78 |
| αuc (i.e. κ\*α) | .14 ± .06 | .20 ± .13 | *W* = 350.00, *p* = .015,  rank-biserial *r* = .51 | .18 ± .10 | .18 ± .11 | *W* = 232.00, *p* = .77 |
| *Decision parameter* | | | | | | |
| β | 5.98 ± 3.00 | 5.58 ± 2.33 | *W* = 210.00, *p* = .66 | 6.62 ± 3.82 | 5.60 ± 2.25 | *W* = 156.00, *p* = .073 |

**Table 4. Within-subject comparison of model parameters of the best fitting iDU-1α-β model.** Means and standard-deviations; paired-samples *t*-tests (*t*-statistics, *p*-values, in case of statistical significance effect size by Cohen’s *d*) or, in case of non-normality, Wilcoxon signed-rank tests (*W*-statistics, *p*-values, in case of statistical significance effect size by the matched rank biserial correlation).

**Figure legends**

**Figure 1. Task and study design; behavioral results on correct choices. A)** Reward-based decision-making task. Exemplary trial sequence. **B)** Anti-correlated structure of the task. One card had a reward probability of 80% along with a 20% probability for a punishment, with reciprocal reward probabilities for the other stimulus. Reward contingencies were stable during the first 55 trials (‘pre-reversal’ phase) and the last 35 trials (‘post-reversal’ phase), whereas during the intermediate 70 trials reward contingencies switched every 15 respective 20 trials (‘reversal’ phase). **C)** Study design. Two independent groups (a-tDCS group, c-tDCS group) were recruited and tested in a within-subject design, undergoing interventions in fully balanced order. **D)** Behavioral data results on correct choices, i.e. choosing the card with 80% reward probability. In both groups, a significant effect of phase was present, but no significant effect of stimulation was observed. Bar plot figures denote the mean with standard errors of the mean as error bars and individual data points (gray: sham stimulation, red: anodal stimulation, blue: cathodal stimulation).

**Figure 2. Behavioral data results for feedback-specific stay-behavior.** In the a-tDCS group **(A)**, participants stayed significantly less after punishments during anodal stimulation compared to sham stimulation, while no significant effects were observed in the c-tDCS group **(B)**. Displayed are percentage of staying with the same card after receiving a punishment or reward for sham and stimulation condition in both groups as well as the individual difference scores between stimulation (anodal resp. cathodal) and sham condition. Bar plot figures denote the mean with standard errors of the mean as error bars and individual data points, asterisk indicates a significant difference at p < .05.

**Figure 3. Bayesian Model Selection.** Random-effects Bayesian model selection (RFX-BMS) revealed that a model with individual double-updating and one learning rate was the relatively best fitting model (iDU-1α-β). We show the posterior probabilities and the protected exceedance probabilities of the 6 models. Abbreviations: SU = single-update; DU = double-update; iDU = individually-weighted double-update; 1α = one learning rate, 2α = separate learning rates for rewards and punishments, β = decision parameter representing inverse decision noise.

**Figure 4. Computational modeling. A)** Within-subject parameter comparison. The parameter αuc was significantly higher under anodal stimulation compared to sham stimulation in the a-tDCS group. Small circles indicate individual values; error bars denote standard errors of the mean; asterisk indicates a significant difference at p < .05. Δ αuc, i.e. the difference score between anodal and sham stimulation regarding αuc. Triangles display individual values; error bars denote standard errors of the mean. **B)** Correlation between computational learning parameter change and observed behavioral change. The change on the parameter αuc was correlated with the behavioral change on lose-stay induced by anodal stimulation in the a-tDCS group. **C)** Results for feedback-specific stay-behavior on simulated data based on the inferred parameters in the a-tDCS group. The effect of less staying with the previous choice after punishments during anodal stimulation as compared to sham stimulation in the empirical data of the a-tDCS group was reproduced when simulating new data based on the inferred parameters of the iDU-1α-β (10000 simulations per subject and stimulation condition). Bar plot figures denote the mean with standard errors of the mean as error bars and simulated data points.