Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The effect of high-definition transcranial direct current stimulation to the pre-supplementary motor area on habit: A randomized cross-over experiment using reinforcement learning in healthy subjects

Vasundhra Teotia
OCD Clinic, National Institute of Mental Health and Neuro Sciences (NIMHANS)
Karthik Sheshachala
OCD Clinic, National Institute of Mental Health and Neuro Sciences (NIMHANS)
Srinivas Balachander
OCD Clinic, National Institute of Mental Health and Neuro Sciences (NIMHANS)
Vani Holebasavanahalli Thimmashetty
OCD Clinic, National Institute of Mental Health and Neuro Sciences (NIMHANS)
Ganesan Venkatasubramanian
OCD Clinic, National Institute of Mental Health and Neuro Sciences (NIMHANS)
YC Janardhan Reddy
OCD Clinic, National Institute of Mental Health and Neuro Sciences (NIMHANS)
Janardhanan C. Narayanaswamy
jana.narayanaswamy@monash.edu

Monash University

Research Article

Keywords: Habit, Goal-directed behavior, HD-tDCS, pre-supplementary motor area, reinforcement learning, compulsivity

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11348334/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Compulsivity is a transdiagnostic marker in various psychiatric disorders and is postulated to be mediated by a shift from goal-directed to habitual behavior. Pre-supplementary motor area (pre-SMA) functioning has been linked to habit learning. We examined the montage and polarity-specific effects of high-definition transcranial direct current stimulation (HD-tDCS) on pre-SMA in modulating habitual learning in healthy subjects. We hypothesized that anodal stimulation would lead to more goal-directed responses in a reinforcement learning task and vice versa with cathodal stimulation.

Materials and methods

28 consenting healthy volunteers were administered anodal, cathodal, and sham HD-tDCS targeting pre-SMA while they performed a computer-delivered sequential reinforcement learning task using a single-blind, randomized cross-over design. Logistic regression was performed to assess the effects of stimulation on habit-driven/model-free (MF) vs. goal-directed/model-based (MB) responses in task performance. A hybrid Bayesian model-fitting technique was also used to generate MB and MF parameters.

Results

The participants showed a hybrid pattern of learning. Cathodal stimulation reduced stay behavior irrespective of previous reward or previous transition [B= -0.165, p = 0.018]. Further, cathodal stimulation was significantly associated with the MF parameter (0.805, p = 0.009), while no association between anodal or sham stimulation with any of the outcomes was found.

Conclusion

Our study provides preliminary evidence that habitual and goal-directed learning could be modulated by transcranial direct current stimulation applied to the pre-SMA. Further exploration is needed to understand the underpinnings of habits and their role in disorders and to translate tDCS as a treatment modality for disorders of compulsivity.

The study was registered with the Clinical Trials Registry- India (CTRI/2021/04/032753) on 12th April 2021

Introduction

Behavior can be guided by either the future outcome or a preceding stimulus, distinguishing them into goal-directed or habitual behavior, respectively (Balleine and Dezfouli 2019). Goal-directed behaviors are intentional, sensitive to the value of the prospective goals, more accurate, and require effort and attention(Balleine and Dezfouli 2019). Habits are repetitive responses to environmental stimuli or

preceding actions that are automatically triggered, inflexible, and reactive, with insensitivity to outcome devaluation and action-outcome contingencies(Wood and Rünger 2016). Habits are essential for daily routines because they expedite actions and decrease the working memory load. However, they lie on a spectrum ranging from the good ones to ones that are performed unintentionally, as in the case of habit slips that emerge accidentally while one is distracted, under stress or time pressure, fatigued, sleep-deprived, or when the working memory is taxed(Wood et al. 2022). At the extreme end of the spectrum are maladaptive appetitive and avoidance habits, that are compulsively performed, with a disproportionate amount of time spent on them(Robbins et al. 2019), and underlie various psychiatric disorders(Gillan et al. 2016).

Three fundamental processes are known to strengthen habit formation, these include repetition, rewards, and the interval type of reinforcement schedule (Robbins and Costa 2017). These processes have been used to design experiments, such as outcome devaluation tasks to study habit formation. In these tasks, subjects are variably trained, to gain a reward (to test appetitive habits) (Tricomi et al. 2009; de Wit et al. 2012a) or receive a punishment (to test avoidance habits) (Gillan et al. 2014). They are subsequently subjected to an outcome devaluation procedure that tests the intactness of the knowledge of action-outcome contingencies and whether the outcome's incentive value guided the behavior (Adams and Dickinson 1981; Adams 1982). Continued automatic responding despite the devaluation indicates habits. Advances in computational modeling techniques have helped to dissociate between goal-directed and habitual behavior by drawing parallels between goal-directed and model-based (MB) reinforcement learning(works on an internal representation or cognitive map of the task structure), and habitual with model-free (MF) reinforcement learning(repeating previously rewarded actions) especially using the two-step sequential learning tasks(Daw et al. 2011).

Animal studies have demonstrated that the dorsolateral striatum(corresponding to the primate putamen) is responsible for habitual behavior and the dorsomedial striatum(corresponding to the primate caudate nucleus) is responsible for goal-directed behaviors (Dolan and Dayan 2013). Studies have also implicated a dissociation of goal-directed and habitual behavior in the cortical regions in rodents, with the prelimbic cortex, M2 cortical regions(analogous to primate preSMA) involved in goal-directed behaviors (Corbit and Balleine 2003; Gremel and Costa 2013), and the infralimbic cortex controlling habitual behaviors (Killcross and Coutureau 2003; Coutureau and Killcross 2003; Dolan and Dayan 2013). Recent task-based functional magnetic resonance imaging (fMRI) studies in humans have implicated ventromedial prefrontal cortex(vmPFC)[medial prefrontal cortex(mPFC) and orbitofrontal cortex(OFC)] in encoding action-related value signals(goal-directed behaviors) (Huang et al. 2020) and the right posterior putamen/globus pallidus in encoding habitual responding (de Wit et al. 2012b; Huang et al. 2020).

There is no consensus about the cortical area involved in habitual behaviors in humans. The pre-SMA has been suggested to be involved in acquiring new action sequences(Nakamura et al. 1998), stimulus-response associations, inhibition of response(Nachev et al. 2008), and habit formation, i.e. initiation of chunking of action sequences(Kennerley et al. 2004). Pre-SMA also has a role in switching actions

especially when the action-outcome contingencies change and habitual responding is no longer beneficial(Isoda and Hikosaka 2007).

There is emerging research using non-invasive brain stimulation techniques modulating cortical excitability to find the causal role between brain regions and their functions of perception, cognition, and motor behaviors(Filmer et al. 2014). Transcranial direct current stimulation (tDCS) is one such non-invasive brain stimulatory method wherein a weak current(1-2mA) is passed across the cortex using two electrodes-anode and cathode-placed on the scalp(Chase et al. 2020). Stimulation via these electrodes leads to neuromodulatory effects by either increasing the cortical excitability(anode) or decreasing it(cathode)(Nitsche and Paulus 2000). Anodal tDCS leads to the facilitation of behavioral performance, whereas cathodal stimulation leads to impaired performance. The high-definition tDCS(HD-tDCS) is a recent modification of the traditional tDCS which delivers more focal stimulation to a selected brain region(DaSilva et al. 2015) and results in better adaptive neural plasticity.

Pre-SMA is a promising target for non-invasive neuromodulation as it is easily localised, present superficially, and has a potential role in habit formation. Modulation of pre-SMA with HD-tDCS may be a way to examine the role of this region in habitual behavior. Such exploration has implications for understanding if this trans-diagnostic feature(Gillan et al. 2016) could be altered in psychiatric disorders of compulsivity and addictions.

Thus, in this study, we aimed to examine the potential role of pre-SMA HD tDCS in modulating the habitual learning of information in healthy subjects. We hypothesized that anodal stimulation of pre-SMA with HD-tDCS would result in more MB or goal-directed behavior, whereas cathodal stimulation would lead to MF or habitual behavior.

Materials and methods

Participants

The participants of the study were healthy individuals (n = 28) without any history of lifetime psychiatric (including substance use disorders) or chronic medical conditions. The sample size was estimated for repeated measures ANOVA model for 3 measurements with an expected medium effect size of 0.25, an alpha of 0.05, and an estimated power of 0.80. Only right-handed healthy individuals in the age group of 18–40 years, who could understand and speak English and had a formal education of at least 12 years with no Axis I psychiatric disorder or contraindications to receiving tDCS were included in the study. The recruitment process occurred from April 2021 to October 2021. All the participants provided written informed consent. The study was initiated after obtaining ethical approval from the institute's ethics committee. The study was registered with the Clinical Trials Registry- India (CTRI/2021/04/032753).

Assessments

The volunteers were screened to rule out any Axis I disorders using the MINI International Neuropsychiatric Interview(MINI) 5.0(Sheehan et al. 1998; Pradeep et al. 2018). A checklist of tDCS contraindications(Brunoni et al. 2011) was administered to evaluate if the volunteer was fit to participate in the trial. None of the participants had Axis I psychiatric illness or contraindication to receive HD-tDCS. The volunteers were confirmed to be right-handed by applying the modified Edinburgh Handedness Inventory (EHI) to avoid any confounding effects of lateralisation(Oldfield 1971). Further, self-report measures of habit using the Creatures of Habit Scale (COHS)(Ersche et al. 2017) were used to identify the participant's propensity for habits and the Obsessive-Compulsive Inventory-Revised(OCI-R) was applied to measure compulsive tendencies(Foa et al. 2002).

Set-up and HD-tDCS procedure

The participants who met the criteria were instructed to wash their hair before the procedure to avoid any impedance due to excessive sebum on the scalp or hair products and were asked to abstain from using any stimulants like caffeine 2 hrs before the session. All except one participant underwent three sessions of 20 minutes, each separated by 48 hours (one participant underwent the third session after 60 hours). The experiment was conducted in a quiet, well-lit room with the participant seated on a comfortable chair facing a blank wall with the HD-tDCS device placed on a table just behind the participant out of view. A laptop with a 15-inch monitor with adequate brightness displaying the task was placed in front of the participant. The head measurements were taken in accordance with the International 10–10 EEG system and using a suitably sized EEG cap, the central electrode was placed at the FCz corresponding to the preSMA and the position of the return electrodes corresponded to the FC3, FC4, Fz and CPz. The Soterix M×N9 HD-tES system delivered the anodal stimulation of + 2.00mA through the central electrode and cathodal stimulation of -0.5mA through the return electrodes. For sham stimulation, an auto-sham feature exists on the Soterix device which simulates a real waveform, wherein at the start and end of the stimulation, a ramp up and down upto 2mA occurs for the first/last 20 seconds and the rest of the 20 minutes no current is delivered. This is done by pressing the button indicating "sham". The duration of the stimulation was set to 20 minutes. Impedance values were checked to be within the adequate range of 0.0K to 350.0K(Villamar et al. 2013). For the first 10 minutes of the session, the participant went through the tutorial of the task and any doubts about the task (described below) were cleared. After 10 minutes of receiving the stimulation, the participant was verbally cued to start the task. The order of the type of stimulation, i.e., anodal, cathodal, or sham was randomized for each participant and was not revealed to them. The randomization sequence was generated using a computerized algorithm. At the end of the third session, the participant was asked to guess which of the sessions they received a sham stimulation. Hence, all subjects received three sessions of HD-tDCS – one each of anodal, cathodal, and Sham type – delivered in a randomized, single-blind fashion while performing the habit task.

Task

The Spaceship task (reinforcement learning task) was adapted from Decker et al(Decker et al. 2016) which was a modification of the task by Daw et al 2011(Daw et al. 2011). The tutorial before the task described the cover story and explained probabilistic rewards and transitions and 20 trials were given before the main task started. The main task consisted of 200 trials in three blocks spaced by two breaks with the main goal being to earn the maximum amount of “space treasures”. The task consisted of a first-stage choice, a transition and a second stage choice. In the first stage choice, the participant had to choose between two spaceships (Blue or Green) which would travel from Earth to either of the two planets (Red or Purple). Each spaceship had a higher chance of going to either of the planets, i.e., the Blue spaceship made a common transition to the Red planet with a 70% probability and a rare transition to the Purple planet with a 30% probability. Conversely, the Green spaceship made a common transition to the Purple planet (70%) and a rare transition to the Red planet (30%). The second stage choice involved choosing between the two alien stimuli that were presented on either of the planets that the transition occurred to. After making the second stage choice, the participant would be presented with a space treasure which is the reward, or an empty circle based on a slowly drifting probability between 0.2 and 0.8 as the trials advanced, and this was expected to motivate the participants not to stick to one choice so that their total rewards can be maximised. Each choice stimulus moved to the centre of the screen and was highlighted by a dashed square box for 1 second. The reward was presented for 1 second and the intertrial interval also lasted for 1 second. The participant had 3 seconds to make each choice else an “X” appeared over the stimuli and that trial was not rewarded. The missed trials were omitted from the analysis.

An MB (goal-directed) learner is expected to choose a subsequent first stage by integrating the state transitions, i.e., rare or common, and the rewards obtained in the previous trial, whereas an MF (habitual) learner would choose independent of the transition type, i.e., a reward on the previous trial would lead to repetition of the first stage choice. For example, if selecting the Green spaceship lead to a rare transition, i.e., to the Red planet and a reward was earned on choosing an alien, an MB learner on the subsequent first stage choice would choose the Blue spaceship to increase their chance of returning to that rewarded state(red planet with the alien that would give a treasure). This choice is made by considering the transition state, there is a reward-by-transition interaction effect. On the contrary, an MF learner would continue to choose the Green spaceship irrespective of the transition type. Here, the main effect is by reward. Hence, this task enables us to distinguish the type of strategy used, i.e. MB or MF.

Statistical analysis

The sample characteristics were described using descriptive statistics like mean and standard deviation. Correlations between socio-demographic variables, task performance and behavioral measures, i.e. OCI-R and COHS total scores were performed. The task data was extracted using the MATLAB scripts.

Regression analysis

Logistic regression was performed to assess the effects of Pre-SMA HD-tDCS stimulation on habit-driven vs. goal-directed behavior in task performance. The probability of staying with the first stage

choice as a function of the previous reward (coded as a binary outcome – reward or no reward), type of stimulation (Anodal, Cathodal and Sham), and previous transition type (coded as common or rare) was estimated using a generalized linear mixed-effects model using the lme4 package on R 4.2.2 software(R Core Team 2022). The main effect of reward on the probability of staying on the first stage choice predicted an MF learning strategy whereas a reward by transition interaction predicted the MB learning strategy. The first or second stage choice trials where the participants did not respond were excluded from the analysis along with the first 9 trials for all subjects. The p-value threshold was kept at < 0.05.

Reinforcement Learning (RL) model fitting

A disadvantage of logistic regression analysis is that it only accounts for the influence of reinforcement on the immediately previous trial. In contrast, a Bayesian computational model-fitting technique takes into consideration the entire preceding history of rewards. Hence, similar to the hybrid model described in Gershman et al(2013)(Otto et al. 2013), and Decker et al (2016)(Decker et al. 2016) we fit our subjects' choices to a weighted combination of an MF SARSA(λ) temporal difference algorithm and an MB "tree-search" reinforcement learning algorithm. For each trial, this algorithm computes the first-stage action value. We used an open source code(Nussenbaum et al. 2020) to obtain model parameters on MATLAB, 2023b(The MathWorks Inc. 2023). The model generated six parameters(see supplementary material) of which relevant to our primary question were MB inverse temperature parameter (β_{MB}) and MF inverse temperature parameter(β_{MF}). The values of these parameters have a linear relationship to the type of learning strategy being employed. Therefore, greater the value of β_{MB} or β_{MF}, more the employment of a MB or MF learning strategy respectively. Subsequently, we performed a linear regression analysis to assess the effect of the type of stimulation and order of stimulation on the MF and MB inverse temperature parameters.

Results

We recruited twenty-eight participants. The sample consisted of an equal number of males (n = 14) and females (n = 14) who were on average 26.5 years old [SD = 1.79] with similar educational attainment [Years of education = 19.28 ± 0.97]. The mean score for OCI-R was 7.21 ± 5.97 and COHS was 80.25 ± 17.14.

Significant correlations were present between OCI-R total scores and COHS total scores. COHS significantly correlated with age and years of education(Table 1). Age, sex, years of education, and total scores on OCI-R and COHS did not significantly correlate with total treasures won (see Table 1).

Table 1
Non-parametric correlations between socio-demographic variables, task performance, and behavioral measures

Age	**Age**	**YOE[a]**	**TTW[b]**	**OCI-R[c]**	**COHS[d]**
	1.000				
YOE[a]	.486**	1.000			
TTW[b]	-0.118	-0.055	1.000		
OCI-R[c]	0.145	-0.117	-0.037	1.000	
COHS[d]	− .323**	− .376**	0.058	.546**	1.000

[**. Correlation is significant at the 0.01 level (2-tailed).]

[a]YOE - Years of Education; [b]TTW- total treasures won ; [c]OCI-R - Obsessive compulsive Inventory - Revised ; [d]COHS - Creature of Habit Scale

The pattern of recruitment of the learning strategy with the stimulation types has been represented qualitatively in Fig. 2. The participants showed a significant main effect of reward [regression coefficient = 0.917, $p < 0.001$] i.e. MF behavior. Significant negative interactions [$Pr(>|z|) = -0.678$, $p < 0.001$] were also found for being rewarded on previous rare transitions with the probability of staying with the first stage choice (0.241) (see Table 2). Hence, the behavior was hybrid, comprising of both MF and MB control across all three types of stimulation.

Further, cathodal stimulation harmed the probability of first stage stay irrespective of the reward and reward by transition [$Pr(>|z|)= -0.165$, $p = 0.018$] as compared to anodal (0.032) and sham stimulation types (Fig. 2). Stimulation type did not significantly alter the interaction between the probability of staying and being rewarded on the previous rare transition (see Table 2).

Table 2
Logistic regression analysis for effects of main reward and transition on the probability of staying

Predictor	Estimate	Std. Error	Z-value	P-value
(Intercept)	1.505	0.212	7.114	< 0.001
Order	0.223	0.028	7.943	< 0.001
Reward	0.917	0.127	7.226	< 0.001**
Reward x Rare Transition	-0.678	0.154	-4.414	< 0.001**
Anodal tDCS	0.032	0.07	0.453	0.651
Cathodal tDCS	-0.165	0.07	-2.371	0.018*
Reward x Anodal tDCS	-0.003	0.07	-0.04	0.968
Reward x Cathodal tDCS	-0.001	0.069	-0.015	0.988
Reward x Rare Transition x Anodal tDCS	0.053	0.121	0.438	0.662
Reward x Rare Transition x Cathodal tDCS	-0.029	0.118	-0.249	0.803
(Significance codes: P < 0.0001'***', P < 0.001 '**', P < 0.01 '*')				

To get more robust results, we used a computational reinforcement learning model to obtain the model-free (β_{MF}) and model-based (β_{MB}) parameters and assess the effects of stimulation on these. We found a significantly positive regression coefficient for the effect of cathodal stimulation on the model-free parameter(β_{MF}) (0.805, p = 0.009) (Table 3.2, Fig. 3). Moreover, on analysing the effect of the order of stimulation on these parameters, we found no significant effect of session number on β_{MB} and β_{MF} (Table 3.1 and 3.2).

Table 3.1
Linear Regression analysis for effects of stimulation and order on model-based inverse temperature parameter

(Intercept)	Estimate	Std.Error	df	t value	Pr(>|t|)
	2.52085	**0.50914**	**79.99947**	**4.951**	**< 0.0001*** **
Order	-0.01483	0.17541	53	-0.085	0.933
Stimulation type Anodal	-0.16538	0.34931	53	-0.473	0.638
Stimulation type Cathodal	0.29644	0.3466	53	0.855	0.396
(Significance codes: P < 0.0001'***', P < 0.001 '**', P < 0.01 '*')					

Table 3.2
Linear Regression analysis for effects of stimulation and order on model-free inverse temperature parameter

(Intercept)	Estimate	Std.Error	df	t value	Pr(>|t|)
	2.0047	**0.3992**	**73.847**	**5.022**	**< 0.0001*** **
Order	0.1009	0.1509	53	0.669	0.506
Stimulation type Anodal	0.583	0.3005	53	1.94	0.0577
Stimulation type Cathodal	0.8053	0.2981	53	2.701	0.00926**
(Significance codes: P < 0.0001'***', P < 0.001 '**', P < 0.01 '*')					

Knowledge of task structure

To assess the knowledge of the state-transition structure, at the end of the 200 trials, a question was displayed on the screen("Which spaceship travelled to the red planet most of the time?"). 5 out of 28 participants answered incorrectly during session 1. Therefore, 82% of the participants were aware of the transition structure in session 1.

If the knowledge of task structure were intact we would expect the participants would be slower in responding during the second stage state if it was a rarer transition and faster if it was a common transition. Hence, we also evaluated the response times on the second stage choice as a function of the transition type and stimulation. We found that participants responded slowly on a rarer transition (0.145, $p < 0.0001$) however, there was no significant effect of stimulation on second stage response times (Table 4).

Table 4
Linear regression coefficients for effect of transition and stimulation type on second stage response times

(Intercept)	Estimate	Std.Error	df	t value	Pr(>|t|)
	0.542	**0.0281**	**31.471**	**19.253**	**< 0.0001**
Rare transition	0.145	0.0149	16546.85	9.714	< 0.0001
Stimulation type Anodal	0.003	0.0115	16546.7	0.246	0.806
Stimulation type Cathodal	0.009	0.0115	16546.54	0.828	0.407

Fidelity of Blinding

Finally, the participants were also asked to guess whether they received sham or verum stimulation based on the sensation, at the end of the third session and it was seen that 7 out of 28 participants were able to correctly guess that they received sham stimulation, i.e. 25% of the participants.

Discussion

Our results suggest that the PreSMA HD-tDCS has a differential and polarity/stimulation-dependent effect on habit learning/habitual responses. While cathodal stimulation increased model free or habitual responses, anodal and sham stimulation did not produce an effect. The focus of this study was to investigate whether habits could be modulated by HD-tDCS applied corresponding to the pre-SMA region. The subcortical structure, putamen, has been consistently identified to underlie habitual behavior(Tricomi et al. 2009), however, this structure cannot be targeted using the non-invasive brain stimulation methods. Moreover, the cortical structures for habits in humans remain to be concurred upon. The supplementary motor cortex (SMC) comprising the pre-SMA, SMA proper and supplementary eye field area (SEF) has efferents to the putamen (Nachev et al. 2008). Choosing the target site of stimulation as pre-SMA was based on its role in action sequences, response inhibition, switching actions(Nakamura et al. 1998; Isoda and Hikosaka 2007) and studies showing its possible role in goal-directed behaviors(Nachev et al. 2007; Gremel and Costa 2013). However, there is also hyperactivity seen in the pre-SMA during habitual or compulsive behaviors, in patients with OCD(Banca et al. 2015). Whether this hyperactivity is present primarily due to increased habitual behaviors or secondary to goal-directed deficits remains a question. Since pre-SMA is involved in action sequences(important for goal-directed behaviors, skills, and habits), it is unclear whether it strengthens one system or weakens the other, or it parallelly modulates both model-based and model-free processes. Hence, it was important to evaluate the role of pre-SMA in habits and goal-directed behaviors.

We observed that the participants displayed a hybrid pattern of behavior across all types of stimulation. Further, it was also seen that applying cathodal HD-tDCS to the central pre-SMA reduced the probability of repeating a first stay choice or “stickiness”, and increased MF/habitual learning strategy. There was no change in the MB learning strategy following any stimulation condition. Anodal and sham stimulation had no effects.

The logistic regression analysis showed that our participants had a hybrid pattern of learning, i.e, both reward and reward by transition interaction influenced the probability of first stage stay. This is similar to the pattern of learning recruited by healthy adult participants shown in many studies(Feher da Silva and Hare 2020). It also showed us that the probability of repeating a first-stage choice was significantly reduced with cathodal stimulation. However, the regression analysis did not demonstrate an increase in MF behavior via cathodal stimulation found through our computational modeling technique. Non algorithmic regression analysis of reinforcement learning tasks relies only on the effects of reward and reward transition interactions to explain MB and MF agents. They also use only history from immediate previous trials. This may be a simplistic way of describing complex human decision-making processes and may not robustly capture the underlying component processes(Collins and Cockburn 2020) in learning. Algorithmic reinforcement learning models on the other hand inform us about the learning rates, stickiness, reward prediction errors and exploration-exploitation trade-offs(inverse temperature parameters) that are used to estimate values of first stage choices over the full learning history.

Although our regression analysis did not capture the increase in employment of an MF learning strategy with cathodal stimulation shown through our computational modeling technique, it did reveal a distinct behavioral shift. Participants exhibited a greater propensity to switch their choices independent of reward or reward and transition interactions with cathodal stimulation compared to anodal or sham stimulation. The observed tendency aligns with the findings related to the stickiness parameter(detailed in the supplementary material) as estimated through model fitting, where it was seen that cathodal stimulation significantly reduced “stickiness” to a choice. The stickiness parameter tells us how likely an individual would repeat their choice. Reduced stickiness also indicates a stochastic choice selection(Nussenbaum and Hartley 2019). It is seen that adults compared to children or adolescents tend to be more sticky and children may avoid repeating choices irrespective of how rewarding the choice may be(Nussenbaum and Hartley 2019). PreSMA has been shown in previous studies to be activated when conflicting cues are presented and resolves conflict by helping choose an appropriate action (Nachev et al. 2008; Horga et al. 2011). It is possible that cathodal stimulation by inhibition decreases this ability to resolve the conflict creating an unstable representation of outcome probability hence increasing stochastic choice behavior.

Given the fact that cathodal stimulation has the property of inhibiting the underlying structure, as anticipated there was a shift towards MF(habitual) behavior with just one session. We also know that the pre-SMA has connections to the putamen (Nachev et al. 2008) which shows increased activity during habitual behaviors(Tricomi et al. 2009). It is possible that inhibiting the preSMA increased putamen activity leading to habitual responses with cathodal stimulation. Contrary to the hypothesis, anodal stimulation did not show an effect opposite to what was seen with cathodal stimulation, i.e., anodal stimulation failed to tip the balance of responding towards a more goal-directed one (MB). We speculated on a few possibilities that could explain these results. Firstly, habitual behaviors are formed and become automatic over some time with repetition hence multiple sessions may be required to induce long-term potentiation or neuroplastic changes to revert from habitual to goal-directed behaviors in healthy individuals. Second, we recruited healthy volunteers who did not show any shift in the model of learning in both the sham and anodal sessions of stimulation. This could be explained by a ceiling effect of the task performance by the healthy volunteers limiting the effect of tDCS on further task performance.

Thirdly, there is still a limited understanding of the mechanism of how tDCS affects neural activity. For example, in a study, cathodal stimulation led to a 38–57% decrease in activation of SMA whereas anodal stimulation led to a minimal increase in activation of 5% attributing it to a ceiling effect(Baudewig et al. 2001). A linear relationship between the percentage of neural activity and a reduction or increase in behavioral output is not certain(Chase et al. 2020). This could also possibly explain our findings of an effect on the learning seen with cathodal and not with anodal stimulation. The neural activity could further depend on whether tDCS is being applied during a motor task or a cognitive one, with polarity specific inhibition or stimulation being seen more consistently in motor studies(Jacobson et al. 2012). Given that pre-SMA is located more rostrally(towards the prefrontal regions), and has a role in more complex cognitive functions compared to the SMA which is located closer to the motor cortex and has a

role in motor actions (Nachev et al. 2008), the homogenous nature of neural activity with stimulation of preSMA is uncertain.

Each region/neural network can play a role in multiple functions, hence to optimize the effect of tDCS for a particular function, some protocols target specific neural networks during their activation while performing that function, known as activity selectivity (Bikson et al. 2013). In our study, we applied an online stimulation(stimulation during task performance) to target preSMA's function in reinforcement learning selectively. An absence of the reduction of a MF process by anodal stimulation cannot prove a lack of neuronal activity in the pre-SMA in MF learning, since it was seen that cathodal stimulation enhanced MF learning. Furthermore, verum stimulation had no discernible increase or decrease on MB learning parameters, prompting questions about whether the pre-SMA serves as a cortical controller solely for MF processes rather than MB ones.

Previous neurostimulation studies targeting different areas such as dlPFC and ventrolateral prefrontal

cortex(vlPFC) have shown that inhibiting these areas via theta burst stimulation using transcranial magnetic stimulation(TMS) or cathodal tDCS disrupts the MB processes whereas anodal stimulation does not show any effects(Smittenaar et al. 2013, 2014; Weissengruber et al. 2019). This has been attributed to an already optimal performance in specific task conditions(where model-based processes are essential compared to model free ones)(Weissengruber et al. 2019). Additionally, the dlPFC and vlPFC regulate model based and MF learning depending on the working memory(Smittenaar et al. 2013) and through an arbitration mechanism(Weissengruber et al. 2019) respectively. These studies provide evidence of involvement of dlPFC and vlPFC in the model-based processes albeit in a different manner from PreSMA. The contribution of complex interactions of preSMA with these regions and the mechanism through which it plays a role in MF and MB processes still needs to be understood.

A strength of this study is that HD-tDCS was used in place of conventional tDCS which would provide a more focal stimulation to the targeted area through the use of smaller electrodes. Additionally, focal stimulation of a target region using a 4X1 ring electrode configuration of HD-tDCS obviates the need for a simultaneous return electrode placement in another region (based on conventional tDCS method). This study used a sham-controlled design, and the uniformity of experimental conditions was maintained for all participants. The participants were comparable in terms of age and education status. We also evaluated if compulsivity correlated with task performance. The baseline propensity for habits was also assessed through the COHS scale which did not correlate with overall treasures won in the task. It is important to deliver the right task instructions and assess if participants understand task structures to prevent them from mistakenly being classified as hybrid learners. This is because it is seen in a previous study that improving task instructions can lead to MB behaviors(Feher da Silva and Hare 2020). The instructions in our task were uniform and delivered through the computer. The participants' understanding of the task was assessed through knowledge of state-transitions. Overall, 82% participants had explicit awareness of the transition structure. Objectively as well, the response times on

second stage choice were determined to assess this knowledge. This is similar to the percentages of adults' knowledge of the task structure described in previous studies(Nussenbaum et al. 2020).

The participants were unaware of the stimulation type and their guess of whether it was sham or verum stimulation was assessed. The fact that less than 50% of the participants were able to accurately guess the stimulation type, suggests adequacy of blinding of the participants. To the best of our knowledge, this is the first study to evaluate the role of neuromodulation of pre-SMA in habitual behavior. There were a couple of limitations of the study. It was a single-blind study. However, as the task was computer delivered, the responses machine logged, and participants blinded to stimulation type helped to partly overcome this limitation. A decline in motivation by the third session, probably because the same task was used in all the sessions, might have played a role in task performance which could have been assessed, or secondary rewards(such as those used in appetitive tasks or monetary rewards based on the number of space treasures collected) (Tricomi et al. 2009) could have been introduced to improve the overall motivation. That MF learning might not be completely representative of habits(Gillan et al. 2015; Collins and Cockburn 2020) was another drawback to measuring the effect of stimulation on habits. The recent Reaction Time (RT) paradigms can help to more accurately deconstruct the two types of behavior(Hardwick et al. 2019; Luque et al. 2020), suppress goal-directed behaviors and bring out the latent habits(Hardwick et al. 2019; Luque et al. 2020). Using MRI-guided localization will increase the accuracy of the targeting of the correct area. Targeting the pre-SMA unilaterally (either left or right) might give us newer insights about the involvement of pre-SMA in habitual learning. These aspects can be considered in future studies.

Conclusion

Transcranial electrical stimulation can be used as a potential tool to explore habit related cognitive functions. The study explores the malleability of the transdiagnostic construct of habitual responses and cognitive inflexibility using non- invasive neuromodulation methods. Demonstration of the effect of verum stimulation in the healthy population indicates potential applicability in the disease populations. The disorders where this finding may have translational implications include addiction and disorders of compulsivity such as obsessive compulsive and related disorders and eating disorders. This may help us understand how the construct of habit can play a role in pathogenesis of disorders of compulsivity and in evolving neurobiologically informed treatment options.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

J.C.N. conceptualized the study, provided resources and supervised the work. V.T. conducted the study, collected data and wrote the main manuscript text. V.H.T provided supervision during the experiments. S.B. and K.S. performed the statistical analysis. S.B. and V.T prepared figures 1-3. All authors reviewed the manuscript.

Data Availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

References

1. Adams CD (1982) Variations in the Sensitivity of Instrumental Responding to Reinforcer Devaluation. Q J Exp Psychol Sect B 34:77–98. https://doi.org/10.1080/14640748208400878
2. Adams CD, Dickinson A (1981) Instrumental Responding following Reinforcer Devaluation. Q J Exp Psychol Sect B 33:109–121. https://doi.org/10.1080/14640748108400816
3. Balleine BW, Dezfouli A (2019) Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits. Front Psychol 10:2735. https://doi.org/10.3389/fpsyg.2019.02735
4. Banca P, Voon V, Vestergaard MD et al (2015) Imbalance in habitual versus goal directed neural systems during symptom provocation in obsessive-compulsive disorder. Brain J Neurol 138:798–811. https://doi.org/10.1093/brain/awu379
5. Baudewig J, Nitsche MA, Paulus W, Frahm J (2001) Regional modulation of BOLD MRI responses to human sensorimotor activation by transcranial direct current stimulation. Magn Reson Med 45:196–201. https://doi.org/10.1002/1522-2594(200102)45:2%3C196. ::aid-mrm1026%3E3.0.co;2-1
6. Bikson M, Name A, Rahman A (2013) Origins of specificity during tDCS: anatomical, activity-selective, and input-bias mechanisms. Front Hum Neurosci 7:688. https://doi.org/10.3389/fnhum.2013.00688
7. Brunoni AR, Amadera J, Berbel B et al (2011) A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation. Int J Neuropsychopharmacol 14:1133–1145. https://doi.org/10.1017/S1461145710001690
8. Chase HW, Boudewyn MA, Carter CS, Phillips ML (2020) Transcranial direct current stimulation: a roadmap for research, from mechanism of action to clinical implementation. Mol Psychiatry 25:397–407. https://doi.org/10.1038/s41380-019-0499-9
9. Collins AGE, Cockburn J (2020) Beyond dichotomies in reinforcement learning. Nat Rev Neurosci 21:576–586. https://doi.org/10.1038/s41583-020-0355-6

10. Corbit LH, Balleine BW (2003) The role of prelimbic cortex in instrumental conditioning. Behav Brain Res 146:145–157. https://doi.org/10.1016/j.bbr.2003.09.023
11. Coutureau E, Killcross S (2003) Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav Brain Res 146:167–174. https://doi.org/10.1016/j.bbr.2003.09.025
12. DaSilva AF, Truong DQ, DosSantos MF et al (2015) State-of-art neuroanatomical target analysis of high-definition and conventional tDCS montages used for migraine and pain control. Front Neuroanat 9:89. https://doi.org/10.3389/fnana.2015.00089
13. Daw ND, Gershman SJ, Seymour B et al (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027
14. de Wit S, Standing HR, Devito EE et al (2012a) Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. Psychopharmacology 219:621–631. https://doi.org/10.1007/s00213-011-2563-2
15. de Wit S, Watson P, Harsay HA et al (2012b) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. J Neurosci Off J Soc Neurosci 32:12066–12075. https://doi.org/10.1523/JNEUROSCI.1088-12.2012
16. Decker JH, Otto AR, Daw ND, Hartley CA (2016) From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. Psychol Sci 27:848–858. https://doi.org/10.1177/0956797616639301
17. Dolan RJ, Dayan P (2013) Goals and habits in the brain. Neuron 80:312–325. https://doi.org/10.1016/j.neuron.2013.09.007
18. Ersche KD, Lim T-V, Ward LHE et al (2017) Creature of Habit: A self-report measure of habitual routines and automatic tendencies in everyday life. Personal Individ Differ 116:73–85. https://doi.org/10.1016/j.paid.2017.04.024
19. Feher da Silva C, Hare TA (2020) Humans primarily use model-based inference in the two-stage task. Nat Hum Behav 4:1053–1066. https://doi.org/10.1038/s41562-020-0905-y
20. Filmer HL, Dux PE, Mattingley JB (2014) Applications of transcranial direct current stimulation for understanding brain function. Trends Neurosci 37:742–753. https://doi.org/10.1016/j.tins.2014.08.003
21. Foa EB, Huppert JD, Leiberg S et al (2002) The Obsessive-Compulsive Inventory: development and validation of a short version. Psychol Assess 14:485–496
22. Gillan CM, Morein-Zamir S, Urcelay GP et al (2014) Enhanced avoidance habits in obsessive-compulsive disorder. Biol Psychiatry 75:631–638. https://doi.org/10.1016/j.biopsych.2013.02.002
23. Gillan CM, Otto AR, Phelps EA, Daw ND (2015) Model-based learning protects against forming habits. Cogn Affect Behav Neurosci 15:523–536. https://doi.org/10.3758/s13415-015-0347-6
24. Gillan CM, Robbins TW, Sahakian BJ et al (2016) The role of habit in compulsivity. Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol 26:828–840. https://doi.org/10.1016/j.euroneuro.2015.12.033

25. Gremel CM, Costa RM (2013) Premotor cortex is critical for goal-directed actions. Front Comput Neurosci 7:110. https://doi.org/10.3389/fncom.2013.00110
26. Hardwick RM, Forrence AD, Krakauer JW, Haith AM (2019) Time-dependent competition between goal-directed and habitual response preparation. Nat Hum Behav 3:1252–1262. https://doi.org/10.1038/s41562-019-0725-0
27. Horga G, Maia TV, Wang P et al (2011) Adaptation to conflict via context-driven anticipatory signals in the dorsomedial prefrontal cortex. J Neurosci Off J Soc Neurosci 31:16208–16216. https://doi.org/10.1523/JNEUROSCI.2783-11.2011
28. Huang Y, Yaple ZA, Yu R (2020) Goal-oriented and habitual decisions: Neural signatures of model-based and model-free learning. NeuroImage 215:116834. https://doi.org/10.1016/j.neuroimage.2020.116834
29. Isoda M, Hikosaka O (2007) Switching from automatic to controlled action by monkey medial frontal cortex. Nat Neurosci 10:240–248. https://doi.org/10.1038/nn1830
30. 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. https://doi.org/10.1007/s00221-011-2891-9
31. Kennerley SW, Sakai K, Rushworth MFS (2004) Organization of action sequences and the role of the pre-SMA. J Neurophysiol 91:978–993. https://doi.org/10.1152/jn.00651.2003
32. Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex N Y N 1991 13:400–408. https://doi.org/10.1093/cercor/13.4.400
33. Luque D, Molinero S, Watson P et al (2020) Measuring habit formation through goal-directed response switching. J Exp Psychol Gen 149:1449–1459. https://doi.org/10.1037/xge0000722
34. Nachev P, Kennard C, Husain M (2008) Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci 9:856–869. https://doi.org/10.1038/nrn2478
35. Nachev P, Wydell H, O'neill K et al (2007) The role of the pre-supplementary motor area in the control of action. NeuroImage 36 Suppl 2T155–163. https://doi.org/10.1016/j.neuroimage.2007.03.034
36. Nakamura K, Sakai K, Hikosaka O (1998) Neuronal activity in medial frontal cortex during learning of sequential procedures. J Neurophysiol 80:2671–2687. https://doi.org/10.1152/jn.1998.80.5.2671
37. Nitsche MA, Paulus W (2000) Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J Physiol 527 Pt 3:633–639. https://doi.org/10.1111/j.1469-7793.2000.t01-1-00633.x
38. Nussenbaum K, Hartley CA (2019) Reinforcement learning across development: What insights can we draw from a decade of research? Dev Cogn Neurosci 40:100733. https://doi.org/10.1016/j.dcn.2019.100733
39. Nussenbaum K, Scheuplein M, Phaneuf CV et al (2020) Moving developmental research online: Comparing in-lab and web-based studies of model-based reinforcement learning. Collabra Psychol 6. https://doi.org/10.1525/collabra.17213. Article 17213

40. Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113. https://doi.org/10.1016/0028-3932(71)90067-4
41. Otto AR, Gershman SJ, Markman AB, Daw ND (2013) The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci 24:751–761. https://doi.org/10.1177/0956797612463080
42. Pradeep BS, Gururaj G, Varghese M et al (2018) National Mental Health Survey of India, 2016 - Rationale, design and methods. PLoS ONE 13:e0205096. https://doi.org/10.1371/journal.pone.0205096
43. R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
44. Robbins TW, Costa RM (2017) Habits. Curr Biol CB 27:R1200–R1206. https://doi.org/10.1016/j.cub.2017.09.060
45. Robbins TW, Vaghi MM, Banca P (2019) Obsessive-Compulsive Disorder: Puzzles and Prospects. Neuron 102:27–47. https://doi.org/10.1016/j.neuron.2019.01.046
46. Sheehan DV, Lecrubier Y, Sheehan KH et al (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 59 Suppl 20:22–33;quiz 34–57
47. Smittenaar P, FitzGerald THB, Romei V et al (2013) Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron 80:914–919. https://doi.org/10.1016/j.neuron.2013.08.009
48. Smittenaar P, Prichard G, FitzGerald THB et al (2014) Transcranial direct current stimulation of right dorsolateral prefrontal cortex does not affect model-based or model-free reinforcement learning in humans. PLoS ONE 9:e86850. https://doi.org/10.1371/journal.pone.0086850
49. The MathWorks Inc (2023) MATLAB version 23.2.0.2428915 (R2023b) Update 4. The MathWorks Inc., Natick, Massachusetts
50. Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225–2232. https://doi.org/10.1111/j.1460-9568.2009.06796.x
51. Villamar MF, Volz MS, Bikson M et al (2013) Technique and considerations in the use of 4x1 ring high-definition transcranial direct current stimulation (HD-tDCS). J Vis Exp JoVE e50309. https://doi.org/10.3791/50309
52. Weissengruber S, Lee SW, O'Doherty JP, Ruff CC (2019) Neurostimulation Reveals Context-Dependent Arbitration Between Model-Based and Model-Free Reinforcement Learning. Cereb Cortex N Y N 1991 29:4850–4862. https://doi.org/10.1093/cercor/bhz019
53. Wood W, Mazar A, Neal DT (2022) Habits and Goals in Human Behavior: Separate but Interacting Systems. Perspect Psychol Sci J Assoc Psychol Sci 17:590–605. https://doi.org/10.1177/1745691621994226

54. Wood W, Rünger D (2016) Psychology of Habit. Annu Rev Psychol 67:289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Figures

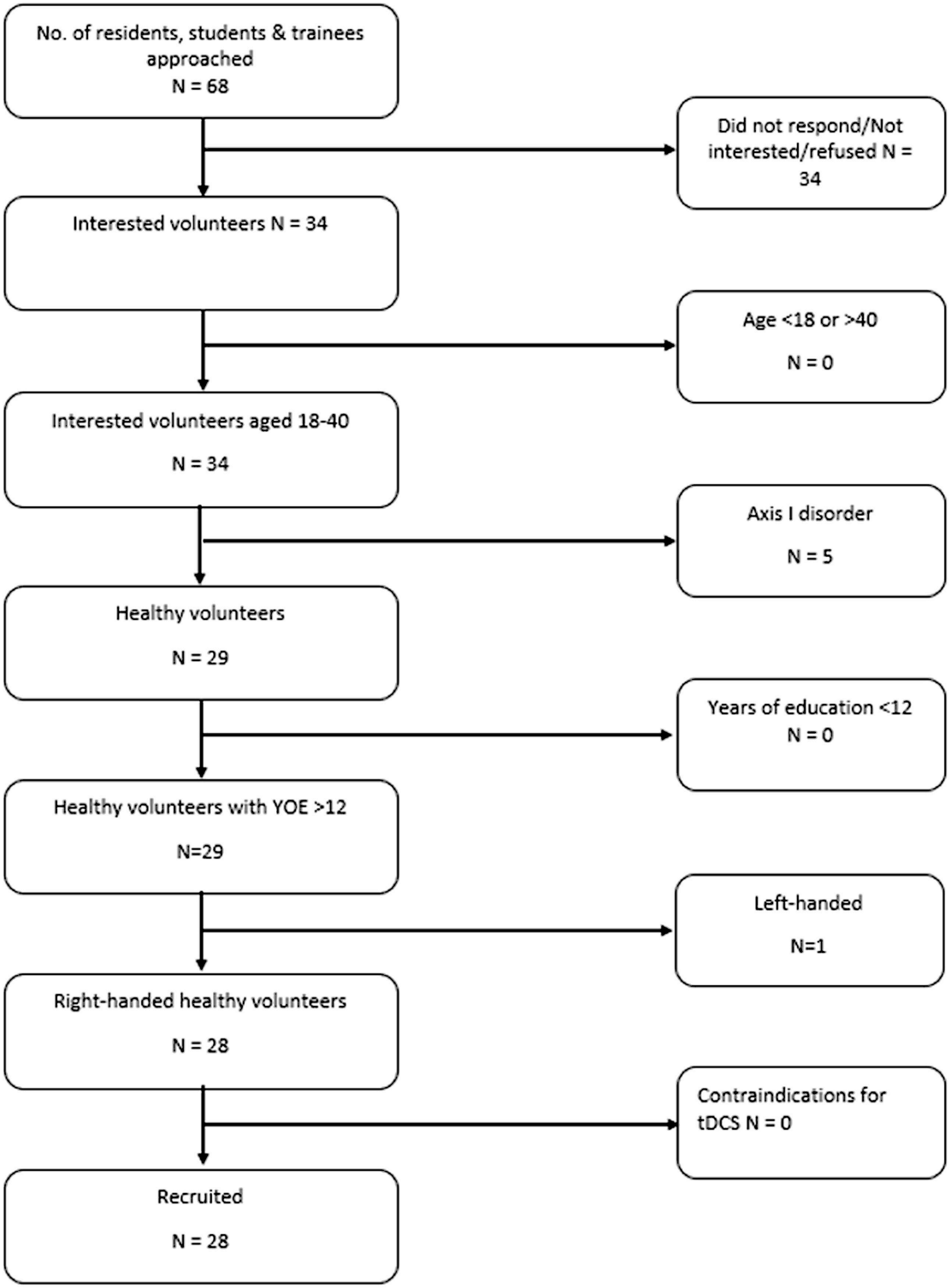


Figure 1

Flowchart of participant recruitment

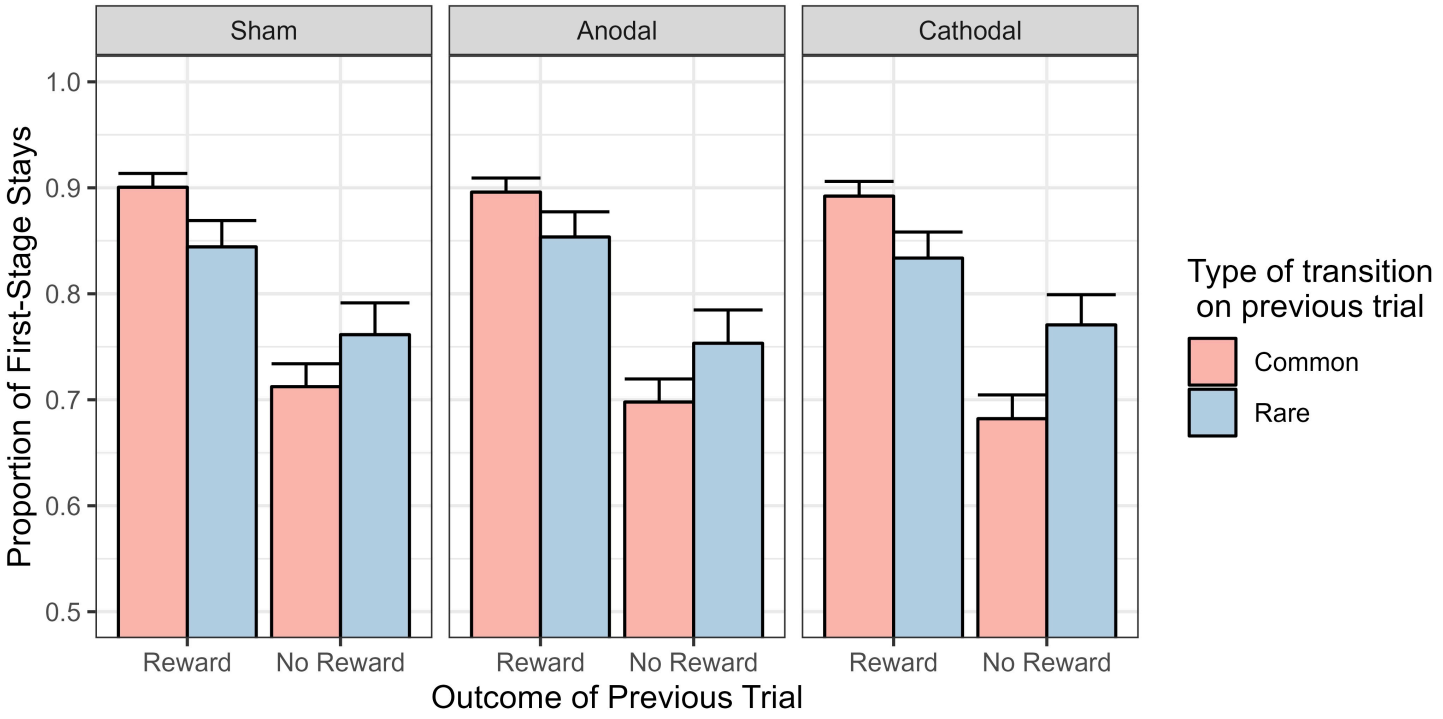


Figure 2

Probability of first stage stay based on the outcome of the previous trial and transition type, graphed for each stimulation type. The error bars represent ±1 SEM

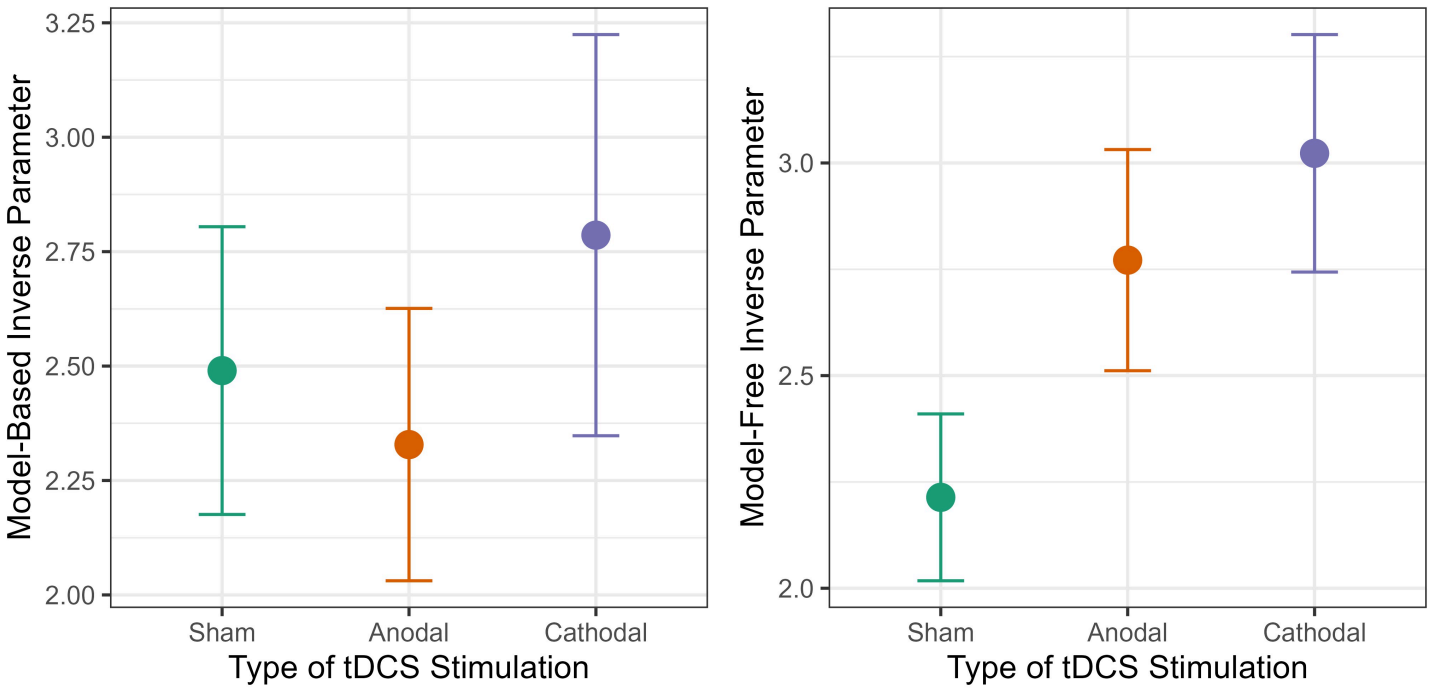


Figure 3

Effect of stimulation type on model-based and model-free inverse temperature parameters. Points represent mean values and error bars represent ±1 SEM

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx