Title: Modulating human learning using transcranial direct current stimulation

Student: Evgeny Gluzman

Student Number: 14013666

Supervisor: Sven Bestmann

Department: Institute of Neurology

Word count:

Section	Word count
Abstract	183
Introduction	750
Methods & Materials	3748
Results	
Discussion	1462
Limitations of methods	750
Conclusion	Included under discussion

ANAT0021: MSc Neuroscience Research Project

DECLARATION OF OWNERSHIP

This submission is a result of my own work.

All help and advice, other than that received by tutors, has been acknowledged, and primary and secondary sources of information have been properly attributed.

Should this statement prove to be untrue, I recognise the right of the Board of Examiners to recommend what action should be taken in line with the University's regulations.

I acknowledge that UCL use the Turn It In® plagiarism detection system and that my work will be submitted to Turn It In®.

Name in block capitals EVGENY GLUZMAN

Signed (typed acceptable) Gluzman

Dated 13.08.2019

Declaration of Ownership form is completed in advance and bound into the project after the title page.

Abstract

Humans are capable of advanced decision-making which involves generalising and establishing causal relationships from limited data. This capacity is believed to be supported by prefrontal regions such as ventromedial prefrontal cortex (vmPFC) and orbitofrontal cortex (OFC). Previous research has investigated the contribution of vmPFC to decision-making using transcranial direct current stimulation (tDCS), a noninvasive brain stimulation method. Hämmerer, et al. (2016) found that delivering tDCS over prefrontal cortex reduced the accuracy with which human participants made choices when performing a reinforcement learning (RL) task. In order to investigate the contribution of vmPFC and OFC to decision-making, the same stimulation protocol was applied to participants performing a novel RL task. The RL model was modified so that it was possible to assess the effectiveness with which participants were able to use relational structure to make accurate choices. In this task, tDCS has reduced the accuracy with which participants made choices, consistent with previous findings. However, contrary to predictions stimulation did not affect participants' ability to make decisions based on relational rules. Potential explanations for these findings are discussed in the context of learning theory.

Introduction

Humans have a remarkable capacity to learn and make effective decisions in novel environments. Young children are able to infer causal relationships from very limited data (Gopnik, et al. 1997). Elucidating the mechanism for the ability to use relational structure in this way has been one of the central challenges in understanding intelligent behaviour. Throughout this report, relational structure is defined as the knowledge of causal relationships in the physical environment. A proposed explanation for the ability to make inferences that go beyond instrumental learning is that of a cognitive map. Tolman (1948) famously observed that rats are capable of finding new routes after previous ones were blocked without prior experience with them. He proposed that this behaviour could be supported by a model of the environment, a neural representation of relationships in the outside world. In the following century, several types of cells have been discovered which may facilitate the existence of cognitive maps such as place cells (O'Keefe & Nadel, 1978) and grid cells (Hafting, et al. 2005). There is mounting evidence that similar mechanisms may be found in non-spatial domains. For example, place cells in rat hippocampus have been found to fire an additional action potential for locations where there is reward (Hok, et al., 2007). In humans, functional magnetic resonance imaging (fMRI) experiments have also found increased activity in the hippocampus depending on proximity to goal.

Another set of brain regions theoretically responsible for complex decision-making is located in the prefrontal cortex (PFC). The scientific literature connecting the PFC to decision-making is quite extensive (Miller & Cohen, 2011). Walton, et al. (2010) ablated the OFC of monkeys performing a task that required them to keep track of the changing value of rewards. After the surgical procedure, monkeys were significantly impaired on the task and appeared unable to adjust their choices following reversals of the highest value stimulus. The lesioned monkeys relied on recent reinforcement history to make decisions rather than understanding of the task structure. This suggests that OFC may be an important region for keeping track of action-outcome contingencies. The vmPFC has also been identified as a key region, with fMRI studies showing grid-like representations similar to grid cells found in the entorhinal cortex (Constantinescu, et al. 2016).

An influential approach to studying decision-making is reinforcement learning (RL). RL is a field of study and a method for modeling how agents can learn to make optimal decisions based on feedback from the environment (Sutton & Barto, 1998). There has been a range of experimental findings suggesting RL is an effective method for describing human decision-making, such as reward prediction errors (the difference between expected and actual outcome) showing strong correlations to dopamine levels in the midbrain (Dolan and Dayan, 2013). In the field of RL, the problem of utilising relational structure to make effective decisions is known as model-based RL. It has recently been proposed that the ventral PFC, comprising OFC and ventrolateral prefrontal cortex facilitates this process by providing a basis for computations involving relational structure (Behrens, et al. 2018).

The reinforcement learning framework has recently been applied to understanding neurological and psychiatric disorders such as Parkinson’s disease (Shiner, et al., 2012) and depression (Huys, et al., 2013). Thus, the prospect of modulating RL parameters seems a promising avenue for research. A potential tool that may be used for this purpose is transcranial direct current stimulation (tDCS), a noninvasive brain stimulation method. tDCS acts by passing a small current between electrodes positioned on the scalp. Recently, Hämmerer, et al. (2016) demonstrated that tDCS reduced choice accuracy in an RL task. The stimulation electrode in their experiment was positioned over vmPFC, which is a region believed to be important for model-based decision-making. Therefore, this study aimed to extend their experiment to investigate the effect of tDCS on model-based RL.

A novel behavioural task was designed allowing the investigation of use of relational structure. Participants were required to choose between two cards based on three associated cues presented on screen at the same time. Two of the cards were related with a rule, which was useful for making optimal decisions. The hypothesis for this experiment was that prefrontal tDCS will reduce the overall accuracy of participant choices, following the Hämmerer, et al. (2016) study. An additional hypothesis was that tDCS would have an effect on the participant's ability to use relational structure to play the game effectively. This would affect parameters that we introduce specifically to model participant learning of the relational rules – the cross-terms.

Materials & Methods

Participants

14 participants (8 = Female) were recruited using the SONA participant recruitment system at University College London. Participants were recruited using a publically accessible mailing list at UCL. 3 participants were removed from the main analysis due to stimulation ending prematurely because of high impedance and one participant did not understand the task after completing training.

Behavioural task

Task

All participants were presented with a reinforcement learning task in which they played a card game, choosing between Red and Blue cards based on one of three cues presented on screen (see figure 3A). The probability of receiving a reward for choosing Red or Blue was either 90% or 10%. After participants made their choice between Red and Blue using keyboard keys ‘C’ and ‘V’ respectively, the outcome (Red or Blue) was determined based on the reward probability. If the outcome matched participant choice, a coin symbol appeared in place of the chosen card. If the outcome and choice did not match, a cross was displayed in place of the chosen card. Participants were awarded +10 points for a successful outcome, and +0 for an unsuccessful outcome. Their total score was displayed in the corner of the screen.

The probability of being rewarded on each trial was dependent on the currently presented cue (e.g. cue A, B or C). The reward probabilities associated with a particular cue remained constant for 10-15 trials and then reversed. Participants had to remember what were the highest value choices on presentation of each cue (such as A - Red, C - Blue) in order to play the game effectively. In addition, two of the cue cards were always related with a rule. The probabilities of receiving reward for choosing Red or Blue were opposite for the pair of cue cards. Knowing this rule allowed participants to effectively choose between Red and Blue after a reversal happened, because they did not need to have received feedback on the currently presented card, only on its pair.

There were two experimental blocks. In one block, participants were told which pair of cues A, B or C were opposite of each other. This will be further referred to as the “use” block. In the other block, participants were presented with a different set of cues X, Y and Z and instructed to learn which pair of cues were opposite. This will be referred to as the “generalisation” block. This difference in instruction appeared crucial as it has previously been reported that performance on an RL task is sensitive to instruction (Baram, personal communication). Each block consisted of 250 trials. Participants completed both blocks during a training session, another two blocks while receiving stimulation for the full duration and a further two blocks while receiving 30 seconds of stimulation (control condition). Condition, block and pair of correlated cues were counterbalanced across participants.

The experiment was conducted over two days. During day 1 participants were instructed and completed the training condition of the experiment (250 trials of use, 250 trials of generalisation). After completing the training, they played the game for another two blocks (in random order) while receiving either the full stimulation (“stimulation”) or 30 seconds (“control”). Participants were compensated 10£ per hour of the experiment and were awarded an additional 5£ for obtaining a score of at least 1400 and guessing the rule correctly during the generalisation block of the stimulation and control conditions (up to a 10£ bonus in total).

Reward schedule

The reward schedule was pre-determined and generated individually for each participant, condition and block. Reward probabilities for Red and Blue switched every 10-15 trials (see figure 1). The trials at which the reversals happened were the same for the related pair of cues but were determined independently for the unrelated pair of cues. In total, there were 20 reversals of reward probabilities for each of the three cues per block. There was no correlation between the related cues and unrelated cues ($p < .05$).

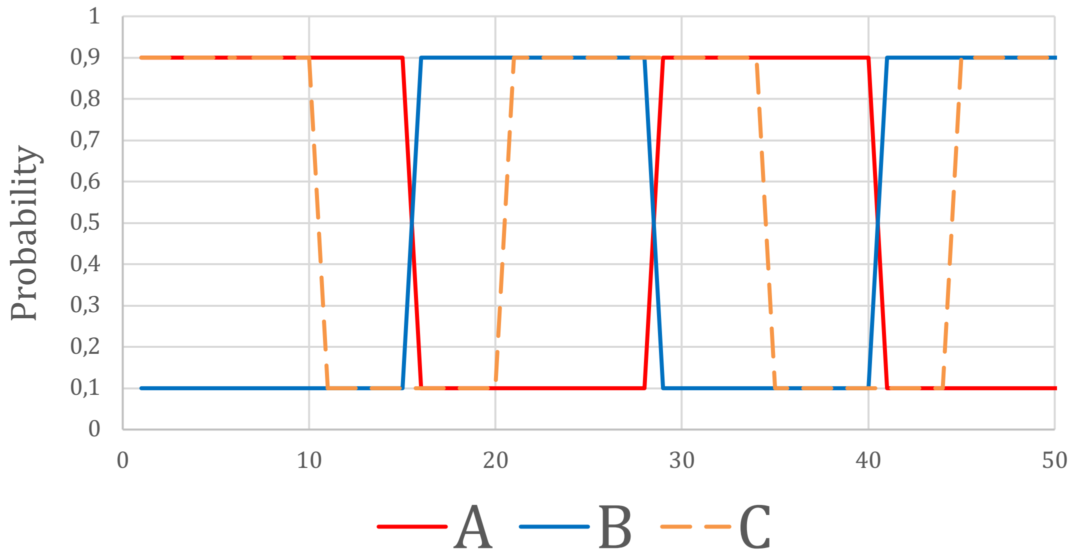


Figure 1. An illustration of the reward schedule used for one of the blocks in the experiment. First 50 trials are shown. Coloured lines track the probability of participant being reward for choosing cues A, B or C.

<u>Instruction</u>

Instruction was a key element in this experiment as previous research has shown that performance is highly sensitive to instruction (Baram, et al. 2019, unpublished). Participants were given a written instruction to read and listened to a verbal tutorial with the assistance of PowerPoint slides (see supplementary information).

Stimulation protocol

Participants received 1 mA tDCS either for 20 minutes (stimulation condition) or 30 seconds (control condition). Stimulation was applied using a battery-driven stimulator (neuroConn). 2x2 electrodes were positioned in such a way that the anode was over Fpz and the cathode just below the inion following Hämmerer, et al. (2016) [see figure 2]. A conductive electrode paste was used to ensure current flow. The stimulation started with a fade in period of 15s where the current gradually increase to its full intensity, and finished with a fade out period of 15s.The stimulator was switched off as soon as participants completed the full task. All participants completed the task in less than 20 minutes, so participants in the stimulation group received tDCS for the full duration of the task.

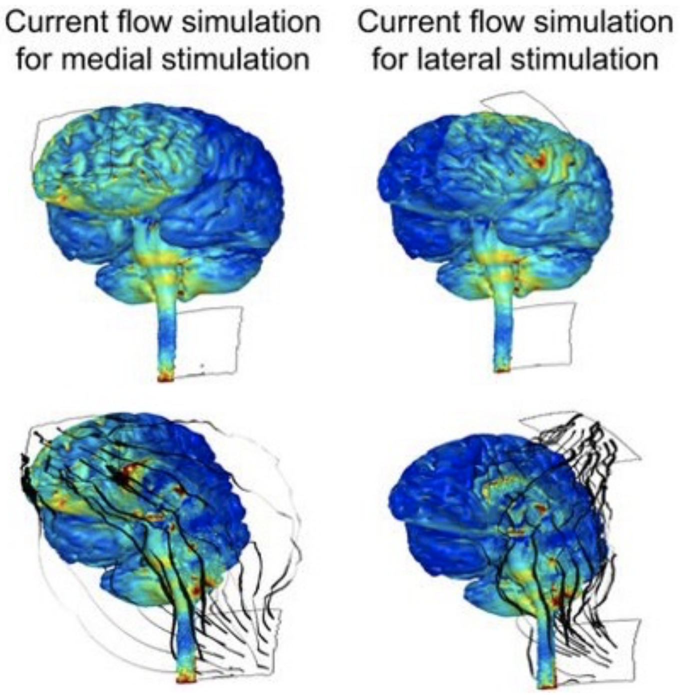

Current flow simulation for medial stimulation
Current flow simulation for lateral stimulation

Figure 2. Current flow estimates for electrodes position over Fpz and inion (left). Right shows current flow estimates for an addition control used by Hämmerer, et al. (2016) which was not included in this experiment. This is discussed in the limitations of methods section.

Reinforcement learning model

A standard Resorla-Wagner learning rule was fitted to the data (Rescorla & Wagner):

$$\hat{r}^{S}{}_{t} = \hat{r}^{S}{}_{t-1} + \alpha\epsilon^{S}{}_{t};\ \epsilon^{S}{}_{t} = y_t - \hat{r}^{S}{}_{t-1}$$

Where $\hat{r}^{S}{}_{t}$ is the outcome probability at trial t for state S ∈ {A, B, C}, (in this case, one of the three cues), $\epsilon^{S}{}_{t}$is the prediction error and α is the learning rate. The prediction error is equal to the difference of the outcome *y* on current trial (1 for Red, 0 for Blue) and outcome probability on previous trial $\hat{r}^{S}{}_{t-1}$. The model was modified to include cross-terms (Baram, 2016, unpublished transfer report). These terms allowed the model to update the value of choosing red on presentation of a cue based on the outcomes of trials when the other cues were presented:

$$\begin{aligned}\hat{r}^{A}{}_{t} &= (1-\alpha)\hat{r}^{A}{}_{t-1} + \alpha y_t \\ \hat{r}^{B}{}_{t} &= (1-\alpha|H_{AB}|)\hat{r}^{B}{}_{t-1} + \alpha H_{AB} y_t \\ \hat{r}^{C}{}_{t} &= (1-\alpha|H_{AC}|)\hat{r}^{C}{}_{t-1} + \alpha H_{AC} y_t\end{aligned}$$

Where $|H_{AB}|$,$|H_{AC}|$,$|H_{BC}|$ are the cross-terms, varying between 1 and -1. Participant choices were modelled using a sigmoidal probability distribution:

$$P(choice\ =\ Red)\ =\ (1 + e^{-\beta(2\hat{r}^{S}{}_{t}-1)})^{-1}$$

Where $\beta$was a free parameter. The parameters β, α and the three cross-terms were fitted to data using the function fmincon in MATLAB.

An experimental pilot was conducted to make sure participants were capable of learning the task and the cross-terms were working as intended. Participants (N = 14) played through the training section of the experiment completing the use and generalisation blocks. The modified RL model could be fit to the data, and cross-terms for the correlated cue pair were lower suggesting the cross-terms worked as intended. Most participants accurately reported what the correlated pair of cues was in blocks where they have not been told which cues were related.

Data analysis

<u>MANOVA</u>

A repeated measures multivariate analysis of variance (MANOVA) was conducted in SPSS in order to determine if condition and block had a significant effect on participant behaviour. Condition was entered into the model as a factor with levels training, stimulation and control. Additionally, block was entered as a two-level factor (use or generalisation). The MANOVA was used to determine if condition and block affected a range of dependent variables (see table 1). In order to control for the effect of stimulation day on performance, stimulation order (stimulation on day 1 followed by control on day 2, or the reverse) was entered into the model as a between-subjects variable.

Shapiro-Wilk tests were conducted and determined that almost all DVs were normally distributed (59/66 measurements were not significant at p > .05 level). Given that MANOVA is robust to violations of the normality assumption and that the measurements that were not normally distributed belonged to different DVs, MANOVA was chosen as the appropriate statistical test. Sphericity assumption was violated for 2 out of 11 DVs, so test statistics for these variables are reported with Greenhouse-Geisser correction.

Name of DV	How was the variable calculated	Description
Accuracy	The RL model was fit to data, resulting in parameters alpha, beta, and three cross-terms. The resulting parameters were then used to obtain the probability of choosing Red on each trial. Choices were simulated according to these probabilities. Accuracy (%) was then obtained by comparing actual participant choices to simulated choices made by the RL agent.	This variable provided an estimate of participants' ability to make effective decisions based on their changing estimates of the value of choosing Red or Blue on each trial.
% optimal choices	Participant choices were compared to objectively highest value choices on each trial. The objectively highest value choice was the card that led to reward with 90% probability. % of optimal choices is the % of choices made by participants which were the same as objectively optimal choices.	This provided a measure of how good were participants at the task overall, and how effectively participants kept track of the reversals of the highest value choice. Because of the stochastic nature of the task participants were rewarded for making a suboptimal choice, or were not rewarded for making a correct choice on 10% of trials. This DV takes into account whether participants made the correct choice, regardless of whether they were rewarded
Reaction time (RT)	The time it took for participants to respond with Red or Blue was recorded. Trial duration was limited to 4s.	Recording reaction time allows us to analyse whether differences in accuracy between conditions occurred due to RT differences.
Score	Participants were awarded 10 points for correct answers, and 0 for incorrect. Score = number of correct answers * 10, with maximum possible score of 2500.	Score provides a direct measure of how well participants performed at the task
Cross-term for correlated cue pair	Fitting the RL model to participant choices and outcomes resulted in three cross-terms for each pair of cues. The cross-term corresponding to the correlated pair of cues was used in the	Examining this cross-term allowed for the investigation of participants' ability to learn and reliance on relational structure.

	analysis.	
Cross-term sum	The moduli of the three cross-terms obtained by fitting the RL model were added together.	This DV provided a measure of how much relational structure participants learned across all three cue pairs, regardless of whether they thought the correlation was positive or negative.
Cross-term difference	The cross-terms for the unrelated pairs of cues were subtracted from the cross term for the correlated cue pair, then the average of these values was taken.	This DV allowed to compare participants estimates of correlation of the related and unrelated cues. This DV was necessary to factor in the cross-terms for unrelated pairs when comparing across participants. That is because the unrelated pairs varied across participants (either AB, BC or AC).
Cross-term optimality	The cross-terms obtained by fitting the RL model to participant choices were subtracted from cross-terms from fitting the model to objectively highest value choices. Then, the sum of moduli of these differences was taken.	This DV provided an estimate of how well participants have used the relational structure compared to an optimal RL agent that always made the best choice.
Accuracy after reversal	This was calculated by identifying trials which happened after a reversal, and after participants received feedback on one of the cues from the correlated pair. Participant choices on these trials were compared to objectively highest value choices.	This DV allowed to compare use of relational structure across conditions without relying on cross-terms
α	Learning rate was calculated by fitting it to the RL model that was previously specified	The learning rate captured how much participants updated their value estimates following feedback
β	Inverse temperature parameter was calculated by fitting it to the RL model	The temperature parameter reflected how random participant choices were.

Table 1. An outline of the dependent variables (DVs) that were entered into the MANOVA.

Regression

A multinomial logistic regression was conducted in order to determine the effect of previous trial outcomes on participant choices on current trial. The predicted categorical DV was choice on one of the cues from the correlated pair (1 for Red, 0 for Blue). 18 IVs were created corresponding to six presentations of each cue prior to the current trial. These IVs coded the outcome on these trials (1 for Red, -1 for Blue). The first six IVs corresponded to trials n-1,...,n-6 when the same cue as the one predicted was presented. The next six IVs corresponded to previous six outcomes on the correlated cue. These IVs were an indicator of how much relational structure was used, as they reflected the impact of feedback on one cue on choices when its correlated pair was presented. The last six IVs corresponded to previous outcomes on the unrelated cue. Data from all participants within a condition x block group (e.g. stim_gen) were concatenated and rearranged so that the predicted choices were always on a cue from the correlated pair, the first six IVs were outcomes on the same cue, the next six on its pair, and the last six on the unrelated cue. The regression analysis was conducted using the mnrfit function in MATLAB. The regression model could be described with the following equation:

$$Y^{A} = \beta_{0} X_{0} + \beta^{A}_{n-1} X^{A}_{n-1} + \beta^{B}_{n-1} X^{B}_{n-1} + \beta^{C}_{n-1} X^{c}_{n-1} + , \ldots.., \beta^{A}_{n-6} X^{A}_{n-6} + \beta^{B}_{n-6} X^{B}_{n-6} + \beta^{C}_{n-6} X^{c}_{n-6}$$

Where Y^A is the outcome when cue A was presented, X_0 is a constant, X^S_n is outcome on trial n on one of the cues S = {A, B, C}. In order to compare the effect of outcomes on choices between experimental conditions, an additional regression was conducted with an IV coding for condition (1 for stimulation, -1 for control). Interaction terms between condition and previous outcomes were also added to the analysis.

Results

Main results

Participants were presented with a reinforcement learning task in which they had to choose between Red and Blue cards based on associated cues (see methods). The cues were related with a rule they were required to keep track of. Each participant completed three conditions of the task - training, stimulation and control, with two blocks of 250 trials (use and generalisation) in each. Task performance was analysed by fitting a modified RL model to the data (see methods). Fitting the model to data resulted in parameters characterising participant behaviour: the learning rate A, the inverse temperature parameter B, and three cross-terms modelling relationships between each pair of the three cues (eg AB, AC, BC). Participant behaviour was then analysed using a repeated-measures ANOVA comparing task performance and resulting parameters across the three groups (training, stimulation and control) and two blocks (use and generalisation).

Overall performance on the task

Participants were able to learn and complete the behavioural task successfully. 8/14 participants correctly reported which pair of cues were correlated after completing the training session, with 11/14 guessing correctly after day 1 of testing, and 10/14 during day 2. Overall, condition had a significant effect on task performance as revealed by the multivariate ANOVA (F(22, 26) = 2.91, p = .005, partial η2 = .71). There was no overall difference between use and generalisation blocks (F(22, 26) = 0.55, p = .797, partial η2 = .86). Between-subjects tests have also shown that the testing order (whether participants received stimulation on day 1 or day 2) did not have a significant effect on any of the DVs (all p < .05).

Univariate tests were conducted to find out which DVs were significantly different across conditions. The average score achieved by participants was significantly different between conditions (F(2, 22) = 6.25, p = .007, partial η2 = .36). Simple contrasts revealed that this was due to participants scoring higher during stimulation compared to training (F(1, 11) = 11.45, p = .005, partial η2 = .49), but there was no significant difference between control and training (F(1, 11) = 3.62, p = .082, partial η2 = .53) or stimulation and control (F(1, 11) = 1.87, p = .199, partial η2 = .15) (see figure 3B). Average reaction time was also significantly different across the three groups (F(2, 22) = 28.50, p < .001, partial η2 = .72). Simple contrasts showed that participants took longer to respond during training compared to stimulation (F(1, 11) = 30.50, p < .001, partial η2 = .72) and control (F(1, 11) = 38.36, p = <.001, partial η2 = .76) conditions. There was no difference in average reaction time between stimulation and control (F(1, 11) = 0.89, p = .366, partial η2 = .08) (see figure 3C).

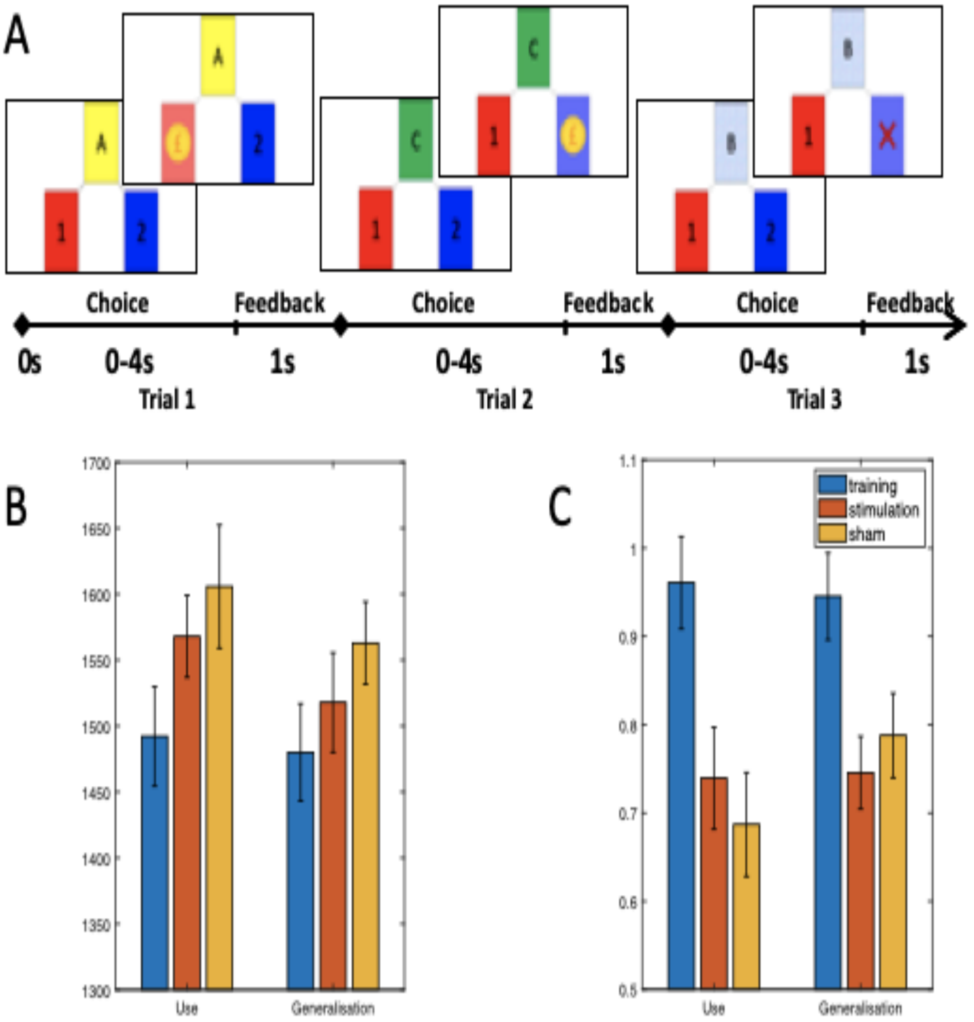


Figure 3. ***A.*** *Representation of the experimental task. Participants chose between Red and Blue cards based on associated cue cards that were presented at the same time. Upon making a choice, participants received feedback on their choice, a coin for correct or a red cross for incorrect. Participants had 4 seconds to make a choice and feedback was fixed at 1s.* ***B.*** *Scores obtained by participants in each condition and block.* ***C****. Average reaction time.* Error bars show +-1 SE.

Participants made more accurate choices in the control condition compared to stimulation and training (see figure 4A). Accuracy was calculated by comparing participant choices to the performance of an RL agent that made decisions based on value estimates and fitted-cross terms (see methods). A univariate test confirmed that accuracy was significantly different across the three conditions (F(2, 22) = 6.94, p = .005, partial η2 = .39). Simple contrasts revealed that this was due to worse performance during training compared to control (mean difference - 6.8%, F(1, 11) = 11.68, p = .006, partial η2 = .52) and stimulation compared to control (mean difference - 5.7%, F(1, 11) = 7.27, p = .021, partial η2 = .40), but not due to differences between training and stimulation (F(1, 11) = 0.22, p = .648, partial η2 = .02).

In addition, the proportion of objectively highest value choices made by participants was estimated. The objectively highest value choice was simply the option (Red or Blue) which was associated with a 90% probability of reward on each trial. The proportion of optimal choices was significantly different between conditions (F(2, 22) = 5.26, p = .014, partial η2 = .32). Participants performed worse during training compared to control (F(1, 11) = 12.55, p = .005, partial η2 = .53), and compared to stimulation (F(1, 11) = 11.68, p = .006, partial η2 = .52), but there was no significant difference between stimulation and control (F(1, 11) = 1.84, p = .200, partial η2 = .133) (see figure 4B).

Finally, the fitted parameters alpha and beta were not significantly different between conditions. (F(2, 22) = 1.20, p = .319, partial η2 = .10) and (F(2, 22) = 0.29, p = .754, partial η2 = .03) respectively.

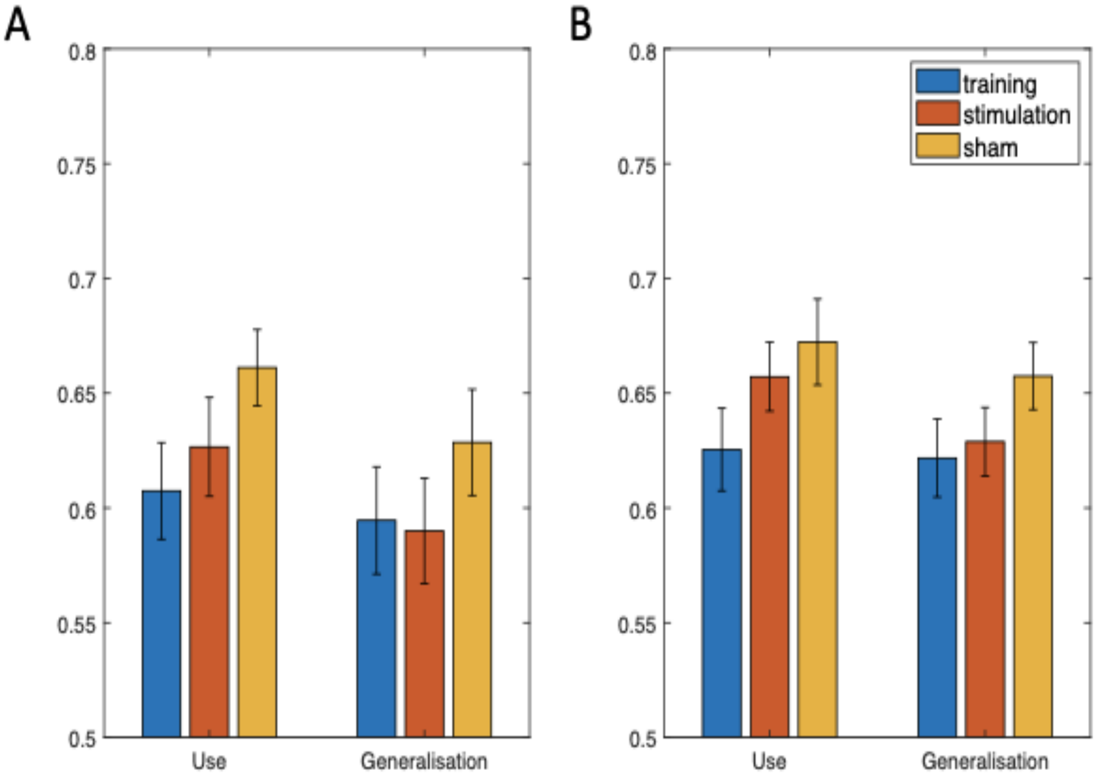


Figure 4. ***A.*** *Accuracy with which participants made choices compared to a RL agent that was making choices based on value estimates.* ***B.*** *Proportion of choices made by participants that were objectively the highest value choice. Error bars show +-1 SE.*

Use of relational structure

In order to estimate how much the task structure guided participant behaviour cross-terms were fit to the data. Overall, the mean of cross-terms for the correlated pair of cues was -0.496 compared to 0.025 for pairs of uncorrelated cues. This suggests that participants have learned the task structure successfully, using feedback from trial outcomes on one of the cues to make choices on its correlated pair. The mean of cross-terms corresponding to the correlated cue pair were compared across conditions (see figure 5A). Overall, there was no significant difference between training, stimulation and control in the value of the correlated cross-term (F(2, 22) = 1.92, p = .171, partial η2 = .15). In order to take into account the values of the cross-terms for the uncorrelated pairs of cues, the mean of differences between cross-terms was calculated (see methods). This was also not significantly different across conditions (F(2, 22) = 2.30, p = .124, partial η2 = .17) [see figure 5B].

In addition, the absolute values of fitted cross-terms were added together to estimate how much relational structure participants learned across all cue pairs. Participants had the same total cross-term value across conditions (F(2, 22) = 0.42, p = .663, partial η2 = .04) [see figure 5C]. To evaluate the ability of participants to effectively learn the correlations between pairs of cues, the cross-terms obtained by fitting the model to human behaviour were compared to the performance of an optimal agent that always made objectively highest value choices. There was no significant effect of condition on this variable (F(2, 22) = 0.39, p = .597, partial η2 = .03) [see figure 5D].

Finally, in order to analyse participants' use of relational structure without relying on the cross-term analysis, their choices on trials following a reversal of the relative values of choosing Red and Blue were analysed. This was done by calculating choice accuracy on trials following reversal where participants had not yet received feedback on the value of Red and Blue for the presented cue, but could rely on previous feedback for its correlated pair to make a decision. Accuracy after reversal also wasn't significantly different across conditions (F(2, 22) = 1.39, p = .271, partial η2 = .11).

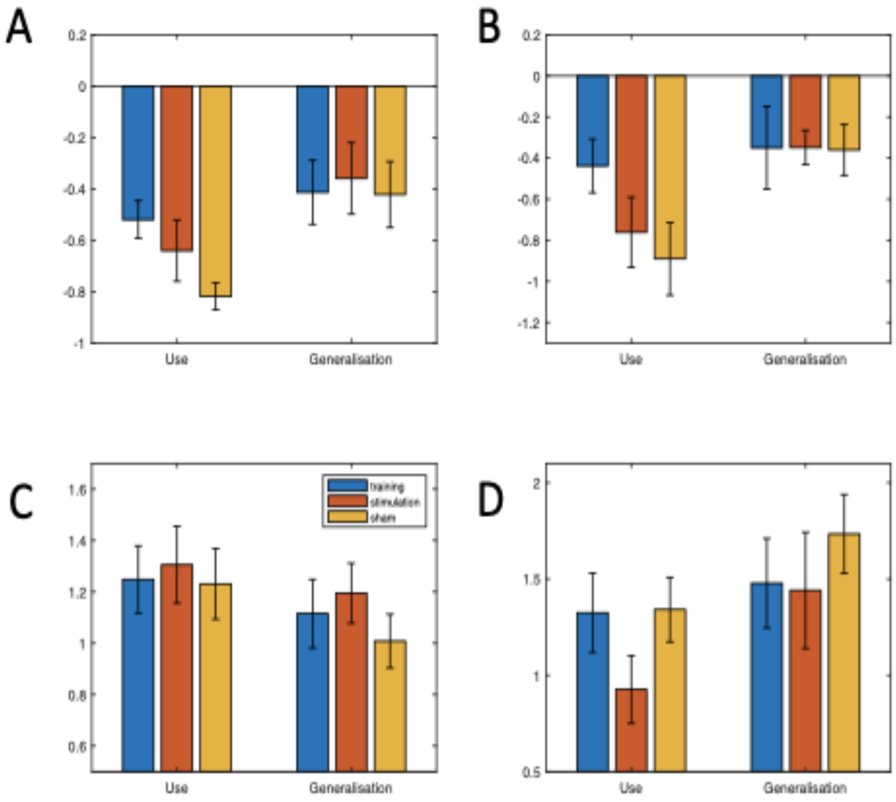


Figure 5. ***A.*** *The values of the cross-terms corresponding to the correlated pair of cues, averaged across participants.* ***B.*** *Mean of differences between correlated and uncorrelated cross-terms.* ***C****. Sum of moduli of cross terms.* ***D.*** *Differences between cross-terms fitted to participant data, and an optimal agent. (see methods for details of calculation.*

Effect of previous trial outcome on participant choices

In order to determine the impact of previous trial outcomes on participant decisions, a multinomial logistic regression was conducted. 18 independent variables were constructed representing outcome of previous six trials for each cues. These were used to predict participant choice (Red or Blue) when one of the cues was presented. The first 6 predictors corresponded to previous 6 outcomes on the same cue, the following 6 to outcomes on its correlated pair and the other 6 to outcomes on the unrelated cue. Six separate logistic regression analyses were conducted for each group. Participant choices on current trial were significantly predicted by outcomes on the previous trial n-1 on the same cue (see table 2) in all conditions. Additionally, participant choices were predicted by outcomes on trial n-2 on the correlated cue.

The stimulation and control conditions were compared with the help of an additional IV representing condition. An interaction terms was created by multiplying outcome on trial n-1 for the correlated cue with the variable representing condition. A logistic regression with these additional variables was conducted. Condition significantly predicted participant choices (β = -0.08, p = .03), while the interaction term for condition x trial n-1(B) did not (β = -0.07, p = .09). Additionally, a simplified logistic regression model was created which only included outcomes on previous trials on the correlated cues, condition and the interaction term. In this model, the interaction term of condition and trial n-1(B) significantly predicted participant choice (β = -0.08, p = .04).

IV	training_use	training_gen	stimulation_use	stimulation_gen	control_use	control_gen
Intercept	-0.14	-0.20	-0.24	-0.15	-0.21	0.05
n-1(A)	**0.48**	**0.67**	**0.60**	**0.44**	**0.70**	**0.62**
n-2(A)	**0.23**	0.08	-0.03	-0.03	0.02	-0.15
n-3(A)	-0.04	0.06	-0.08	-0.03	-0.12	0.12
n-4(A)	-0.05	-0.07	0.03	-0.02	-0.06	0.00
n-5(A)	-0.11	0.04	-0.03	-0.03	-0.26	-0.13
n-6(A)	0.02	0.02	-0.14	-0.13	-0.10	0.01
n-1(B)	**-0.53**	**-0.45**	**-0.83**	**-0.75**	**-0.90**	**-0.69**
n-2(B)	-0.06	-0.10	**-0.25**	-0.21	-0.17	-0.18
n-3(B)	0.05	-0.08	0.12	0.00	0.00	**0.22**
n-4(B)	0.05	**0.15**	**0.22**	0.01	0.07	0.08
n-5(B)	-0.02	-0.13	0.08	0.05	-0.12	0.14
n-6(B)	0.14	-0.01	-0.02	0.13	-0.05	0.00
n-1(C)	0.08	-0.09	-0.19	-0.20	-0.10	-0.03
n-2(C)	-0.05	-0.04	0.09	0.07	0.13	0.03
n-3(C)	-0.10	0.07	-0.05	-0.12	-0.19	0.05
n-4(C)	0.00	0.05	0.03	-0.07	0.01	-0.10
n-5(C)	-0.05	**-0.23**	0.06	-0.01	0.08	0.06
n-6(C)	-0.03	-0.05	**-0.24**	-0.07	0.06	-0.10

Table 2. Table showing regression coefficients for each independent variable, condition, and block. Coefficients that were significant at the p < .05 level are highlighted in bold.

Discussion

tDCS modulates choice accuracy in an RL task

Analysis of experimental results suggests that the behavioural task worked as intended, and participants were able to keep track of changing reward probabilities. Almost all participants correctly guessed which pair of cues was related with a rule, suggesting that participants understood the structure of the task. Prefrontal tDCS impaired participants' ability to make accurate choices. Participants were 5.7% less accurate in their decisions while receiving stimulation compared to the control condition. This is comparable to the results obtained by Hämmerer, et al. (2016) who found that prefrontal tDCS reduced choice accuracy by 8% using the same electrode montage. Their analysis also relied on the standard RL model based on the Rescorla-Wagner rule, with the only difference being that the present experiment included novel cross-term equations used to estimate learned relations between cue cards. Our study provides further support for the hypothesis that prefrontal tDCS affects choice accuracy in RL tasks.

The observed effect agrees with a broad range of evidence suggesting prefrontal regions play a key role in decision-making. The vmPFC has been considered responsible for comparing choice options and selecting the highest value choice (O'Doherty, 2011). The reduction in accuracy observed in this experiment may be caused by tDCS interfering with efficient vmPFC function. However, making precise estimates of the effect of tDCS on neural activity is challenging as its mechanism of action has not yet been fully elucidated. The current state of knowledge is that tDCS tonically depolarises or hyperolarises the resting membrane potential of neurons (Nitsche, et al., 2008). In an attempt to provide a robust theoretical description linking physiological and behavioural effects of tDCS, computational neurostimulation was recently introduced as an approach to modelling tDCS (Bonaiuto & Bestmann, 2015). Computational neurostimulation comprises models of brain stimulation which bridge population-level neural dynamics with observed behavioural effects. Hämmerer, et al. (2016) attempted to provide such a description by constructing a biophysical attractor model (BAM) simulating the effect of stimulation on vmPFC neurons (see figure 6). In this model, membrane depolarisation at the anodal site increases choice stochasticity by increasing the impact of noise on competing populations of pyramidal cells.

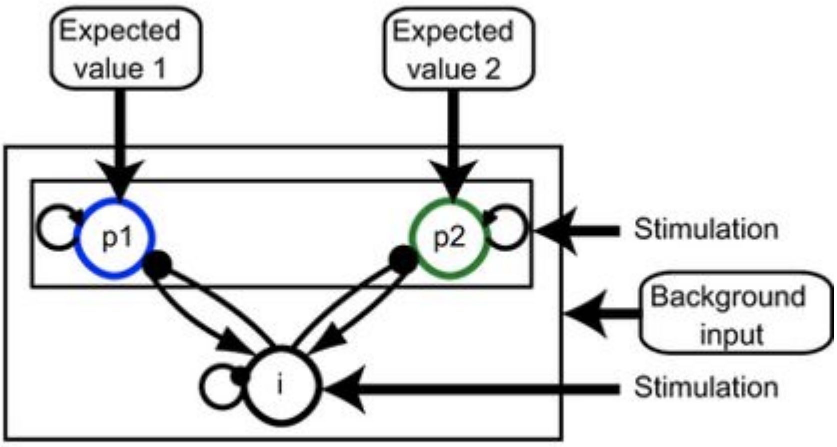


Figure 6. A model of the effect of stimulation on competitive neural dynamics in the vmPFC. Pyramidal cells p1, p2 and inhibitory interneurons i exhibit recurrent excitation of the neural populations they belong to. They also compete via lateral inhibition. The dynamic interaction of these pools of neurons form an attractor network with multiple stable attractor states. An increase in the activity of p1 relative to p2 leads to a change of state, allowing for the comparison of the expected values of different actions. Stimulation depolarises the membranes of p1, p2 and i and injects noise into the system, increasing choice stochasticity. (figure from Hämmerer, et al. 2016).

The observed effect of stimulation on accuracy may be explained by the perturbation of attractor dynamics described in this model. However, the results of the present experiment do not completely agree with Hämmerer, et al. (2016). In their study, the reduction in accuracy was attributed to increased choice stochasticity represented by β, which was significantly different between conditions. This experiment did not find such a difference. There are multiple potential explanations for this observation. Modifying the RL model to include cross-terms could have made the parameter fitting process noisier, resulting in less accurate estimates of β. It is also possible that the reduction in accuracy Hämmerer, et al. (2016) observed was not caused by increased choice stochasticity as represented by β. Decision computations occur across multiple scales (Hunt, et al. 2014) and it is possible that the observed reduction in accuracy is due to an impairment in the estimation of values of different actions, rather than in their comparison. β is only relevant in the latter case, as it controls the effect of value estimates on choice, but not the estimates themselves.

No effect of tDCS on use of relational structure

Contrary to the experimental hypothesis, there was no observed effect of prefrontal tDCS on participants' ability to use relational structure to make choices. This may have been due to the analysis not being sensitive enough to reveal a genuine effect of stimulation, or lack of effect of stimulation. While the standard error for choice accuracy was 0.02 it was 0.10 for cross-terms. This suggests that a small effect of stimulation may not have been detected by the analysis due to noise. The results of one of the logistic regression analyses also support this possibility. However, the regression coefficient for the interaction term comparing experimental conditions was small. The significance of the regression coefficient for the interaction term may also reflect an effect of a general reduction in accuracy, rather than the ability to use the correlation rule to make choices. If participants were less accurate while receiving stimulation, the outcomes of previous trials would have had less of an effect on their choices even if they did rely on relational structure to make the decisions. Additionally, when more regressors were added to the model the interaction term was non-significant. Finally, the fact that tDCS did not affect the accuracy of participant choices immediately following a reversal suggests that the lack of observed effect of stimulation is not due to cross-term analysis failing to detect it.

This lack of effect may be explained by considering current theoretical accounts of decision-making. Decision computations are highly distributed, and the computation of choice value is not localised to any specific brain region (Hunt & Hayden, 2017). In fMRI experiments, up to 30% of the brain shows activity correlated to various aspects of decision-making (Vickery & Chun, 2011). Therefore, a slight reduction in the efficiency of decision computations in the OFC and vmPFC might not be sufficient to induce a measurable change in participants' ability to use relational structure. tDCS affects both excitatory and inhibitory neurons, which may not cause the overall output of the network to change. The contribution of other regions may be sufficient to maintain effective decision-making.

An emerging theme in neuroscience research is that cognitive impairments that appear localised turn out to be a consequence of a graded reduction in general capacity to process information. The present study was partly motivated by Walton, et al. (2010) observation that OFC lesions impair decision-making following reversals. However, their study also found that increasing the overall difficulty of the task by reducing reward probabilities has led to a similar impairment compared to the lesion condition. This suggests that a reduction in the ability to keep track of action-outcome contingencies may be the consequence of a general reduction in resources available for computation. It is possible that while tDCS affected the neuronal populations which contribute to the use of relational structure, this was simply not enough to elicit a discrete effect on behaviour.

This hypothesis could potentially be investigated in a further experiment by manipulating the difficulty and attentional requirements of the task. Decision-making is closely related to attention (Niv, et al. 2015). Participants in the present experiment were highly motivated to pay attention to correlations between cues due to the requirement to identify the related cue pair. It is possible that once participants were told the rule (or learned it in the generalisation block) it was actively maintained with the help of working memory or semantic control processes, with the subtle effect of tDCS in the vmPFC having no measurable impact on use of relational structure. An overall decrease in reward probabilities could have changed how taxing the task is on the multiple systems subserving goal-directed decision-making. As cross-terms in this analysis showed that participants used relational structure even in the few cases where they did not guess which pair of cues were correlated, another potential modification to the experiment would be not instructing participants that there are relationships between cues in this experiment. The requirement to keep track of relational structure without conscious effort may allow the investigation of tDCS effects on use of relational structure under different experimental conditions.

Limitations of methods

Individual differences in current flow

One major limitation of this study was that stimulation parameters were the same for all participants. Individual differences in brain morphology affect the flow of current through the brain and could result in a different effect of tDCS. Current approaches to tDCS allow the modelling of current flow through brain tissue. This is the approach that Hämmerer, et al. (2016) took in their study (see figure 2). An estimate of current flow may be used to adjust stimulation parameters, such as the magnitude of the current. However, this approach considerably limits the pool of potential participants as a brain scan is required for current flow modelling. Due to time constraints and the high cost of an fMRI scan, such modelling was not conducted.

Sample size

Sample size is an important factor which is one of the determinants of the statistical power of an experiment (Cohen, 1969). The number of participants who participated in this experiment (N = 14) is comparable to the sample size in the Hämmerer et al. (2016) experiment (N = 16). However, a power analysis conducted in G*power software suggests 17 participants would have been required to detect a medium effect size (0.5) with statistical power of 0.8. If the observed effect was small (0.2), detecting the effect with this level of power would require a sample size of 36. This leaves the possibility that tDCS had a small effect on participants' ability to use relational structure, but the study was not sufficiently powered to detect it. However, the observed effect on accuracy was not small ($F(1, 11) = 7.27$, $p = .021$, partial η2 = .40) suggesting that some effects of tDCS on decision-making could have been detected with the present sample size. It was not plausible to test 36 participants twice given the time constraints, but this problem could be addressed in a further study using the same experimental paradigm.

Control condition did not receive full duration of tDCS

Another important limitation of this study was that control condition involved only 30 seconds of stimulation, while the stimulation condition received the full duration (up to 20 minutes). tDCS could be felt by participants as a slight tingling sensation on their skin. Therefore, the observed effect on accuracy could have been a placebo effect, or due to participants being distracted by the current. In order to control for this possibility, an additional condition could have been included where participants receive the full duration of the stimulation at a different electrode position. However, choosing such a control site proved difficult, as a broad range of brain regions have been shown to play an important role in model-based decision-making. For example, we could not have used the lateral prefrontal cortex control site used by Hämmerer, et al. (2016) because of its proximity to the parietal cortex, which has been found to play an important role in model-based decision-making (Baram, personal communication, 2019).

The decision to not include an additional control was not entirely unjustified, as Hämmerer et al. (2016) found no significant difference between the control condition receiving stimulation at the lateral site and no stimulation. Additionally, in the present experiment most participants did not report noticing any differences between the stimulation condition and control. Participants often adapt to the sensation of being stimulated (Nitsche, et al., 2008) and can't reliably tell if they are being stimulated. However, in order to completely remove the possibility that the observed effects were due to the sensation of being stimulated, an additional control site could have been identified. In that case it would also be necessary to conduct current flow simulations to ensure that current does not pass through regions which contribute to decision-making.

Parameter fitting not sensitive enough

Finally, it is possible that the reinforcement learning model fitted to the data did not represent participant behaviour accurately enough. Fitting cross-terms to objectively optimal choices did not always result in non-zero cross-terms for the unrelated cues. In addition, the fmincon function only allowed for cross-terms to be fitted to the data in each block. Hence, there was only one cross-term to be compared between conditions, excluding the possibility of analysing changes in how participants used relational structure as they progressed through the trials in each block. Using a different function to fit these parameters could have allowed the tracking of how cross-terms changed on each trial, letting the experimenter observe any effects of stimulation which are specific to phases of the behavioural task.

Conclusion

In conclusion, tDCS significantly reduced the accuracy of participant choices in a reinforcement learning task requiring them to keep track of relational structure. This is consistent with previous findings by Hämmerer, et al. (2016). The stimulation did not have an effect on participants' ability to learn or use a correlation rule relating pairs of cue cards to make effective decisions, as indicated by parameters fitted to a modified reinforcement learning model that was used in this study. This could have been be due to the overall difficulty level of the task, or the task engaging a range of conscious, deliberate processes and requiring participants to focus. These parameters can be manipulated in a future experiment in order to probe the effect of tDCS on human learning.

References

Bonaiuto, J. & Bestmann, S. Understanding the nonlinear physiological and behavioural effects of tDCS through computational neurostimulation. *Prog. Brain Res.* In press (2015).

Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z.What is a cognitive map? organizing knowledge for flexible behavior. Neuron. 2018 10 24; 100(2):490-509

Constantinescu, A.O., O'Reilly, J.X., and Behrens, T.E.J. (2016). Organizing conceptual knowledge in humans with a gridlike code. Science 352, 1464– 1468.

Gopnik, A. A. N. Meltzoff, Words, Thoughts, and Theories
(MIT Press, Cambridge, MA, 1997).

Hämmerer, D., Bonaiuto, J., Klein-Flugge, M., Bestmann, S. Selective alteration of human value decisions with medial frontal tDCS is predicted by changes in attractor dynamics *Scientific Reports* **volume6**, Article number: 25160 (2016)

Hafting, T., Fyhn, M., Molden, S., Moser, M.B., and Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806.

Hok, V. Pierre-Pascal Lenck-Santini, Sébastien Roux, Etienne Save, Robert U. Muller, Bruno Poucet, Goal-Related Activity in Hippocampal Place Cells, Journal of Neuroscience 17 January 2007, 27 (3) 472-482; DOI: 10.1523/JNEUROSCI.2864-06.2007

Hunt, L. T., Dolan, R. J., & Behrens, T. E. J. (2014). Hierarchical competitions subserving multi-attribute choice. Nature Neuroscience, 17(11), 1613–1622. doi:10.1038/nn.3836

Hunt, L. T., & Benjamin Y. Hayden, A distributed, hierarchical and recurrent framework for Shiner T., Ben Seymour, Klaus Wunderlich, Ciaran Hill, Kailash P. Bhatia, Peter Dayan, Raymond J. Dolan, Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease, Brain, Volume 135, Issue 6, June 2012, Pages 1871–1883, Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. Annual Review of Neuroscience, 24, 167-202.

Niv, Y, Reka Daniel, Andra Geana, Samuel J. Gershman, Yuan Chang Leong, Angela Radulescu, Robert C. Wilson, Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms
Journal of Neuroscience 27 May 2015, 35 (21) 8145-8157; DOI: 10.1523/JNEUROSCI.2978-14.2015

Nitsche & Paulus, W. Transcranial direct current stimulation – update 2011. Restor. Neurol. Neurosci. 29, 463–492 (2011).

O'Doherty, J. P. Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Ann. N. Y. Acad. Sci.* 1239, 118–129 (2011).

O'Keefe, J., and Nadel, L. (1978). The Hippocampus as a Cognitive Map (Oxford University Press).

Sutton, R., and Barto, A. (1998). Introduction to Reinforcement Learning (MIT Press).

Viard, A., Doeller, C. F., Hartley, T., Bird, C. M., & Burgess, N. (2011). Anterior hippocampus and goal-directed spatial decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(12), 4613–4621. doi:10.1523/JNEUROSCI.4640-10.2011

Vickery, T. J., Chun, M. M. & Lee, D. Ubiquity and specificity of reinforcement signals throughout the human brain. Neuron 72, 166–177 (2011).

Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rudebeck, P.H., and Rushworth, M.F.S. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron 65, 927–939.

Supplementary information

Code

All code for this experiment will be made available at https://github.com/e-gluzman.

Participant instructions

Earn up to 25£ for playing a card game while your brain is being stimulated! (learning and decision-making experiment)

TRAINING

In this experiment you will be making decisions between options in order to get as many points as possible (maximize reward). The experiment is broken up into trials – each trial is a couple of seconds long. There will be two parts to the training session – each approximately 10 minutes long. You will be reimbursed more for your participation if you perform better at the task.

In this experiment, on any given trial, you will be presented with one of three 'Cue Cards' at the top of the screen – A, B or C. You will then have to choose between a Red card and a Blue card. One of them has a 90% chance of winning you a reward, and the other has 10%, depending on the Cue Card. The aim of the game is to maximize reward. In order to do so, you need to choose the card with the highest change of winning - Red or Blue. Getting a reward will award you +10 points, if you do not get a reward you will not get any points.

In order to score as many points as possible, you have to remember associations between Cue Cards and Red or Blue cards. You can learn these associations from previous trials. For example, if the Cue Card is A and you are rewarded for choosing Red, remember the association A – red. It may be best to choose Red when you see A next time and, for example, Blue when you see B. These associations change! Sometimes Red will stop being the best choice for A and you will need to choose Blue when you see A.

Important hint: the Cue Cards will be related with a rule. In this experiment, there will always be two Cues which are the opposite of each other. For example, A can be always the opposite of C. In this case if you know that A is currently associated with a 90% chance of reward on Red, you can infer that C is currently associated with a 90% change of reward on Blue. Next time you see card C you do not have to guess as you know Blue will give you the highest chance of reward. Use this rule to maximise your score. For example, if you notice a change in one Cue Card, you can infer that a related Cue Card will also have changed. There is an additional reward if you perform really well!

Part 1.

In this part of the experiment, I will tell you at the start which two Cue Cards are the opposite of each other (anticorrelated). The other Cue Card will be unrelated. Use this rule to keep track of the changes in which Cue cards are related to which colour (Red/Blue).

Part 2. After you finish part 1, I will have to quickly start the next part.

This part of experiment is the same, except I will not tell you which Cue Cards are the opposite of each other! You will be presented with new cards X, Y, Z and have to try and learn which two Cue Cards are opposite (anticorrelated). You have to figure out which of the cue cards XYZ are related and tell me at the end of the experiment.

Important:

You will get a lot of errors in this experiment. This is important for you to figure out when the associations have changed, and you need to change your strategy. Don’t get discouraged and keep playing!

Do not spend too much time on each trial. You have a limited amount of time to complete the experiment, so you should not spend more than ~2 seconds on making a choice. It is ok to occasionally stop and think, but most of the time you should be making choices quickly.

There are no other relationships in this experiment except for two of the cue cards being the opposite of each other. Focus on this rule to maximise reward.

One of the cue cards will be unrelated to the other two. However, since it also has a 90% chance of rewarding you for choosing Red or Blue it may seem related at some points of the experiment. Do not get distracted by this – the anticorrelated cue cards will always be the opposite of each other, so you can tell them apart.

Please remember which Cue cards are the opposite of each other, so you can tell me specifically at the end – e.g. “Y and Z”.

STIMULATION

After you complete training, you will be asked to play the same game while receiving brain stimulation. This involves passing a weak (1 mA) current between electrodes at the front and at the back of your head. This kind of stimulation is safe when delivered in a lab environment. You may feel a tingling sensation on your skin as the current gradually increases, but you will soon adapt to the sensation and won't feel it anymore.

In order to set up the stimulation, I will have to take a couple of measurements of your head with a tape in order to determine electrode position. I will also clean the electrode positions with a cotton pad and use a special electrode paste to attach the electrodes. I will then wrap a tape around your head to secure the electrodes in their positions. This will be done before the training session.

I will then set up the game you played during training and turn on the stimulation. The rules will be the same, but the two cue cards which are the opposite of each other will not necessarily be the same as during training. There will again be two parts to the experiment. In one part, I will tell you which Cue cards are related, and you will have to use this rule to obtain the highest reward. In another part you will have to figure out which Cues are the opposite of each other by playing the game and tell me at the end. You will be reimbursed 10 pounds for the full experiment. If you get a high score and guess which Cues are anticorrelated, you will be reimbursed an additional 5 pounds!

If at any point you feel irritated by the tingling, or uncomfortable for some other reason, please let me know and I will stop the experiment.

DAY 2 stimulation

After you complete training and stimulation today, you will have to come back for another 30 minute session of stimulation within a few days. This will be exactly the same as today, except the pair of Cues which are opposite of each other may be different. You will be paid 5 pounds for this session. If you get a high score and guess which Cues are anticorrelated, you will be reimbursed an additional 5 pounds!

After you complete both days of the experiment you will be reimbursed 15 pounds + up to 10 pounds more depending on how you performed. Please come back for day 2 of the experiment – we cannot use your data if you don't complete day 2!

Tutorial slides (see next).

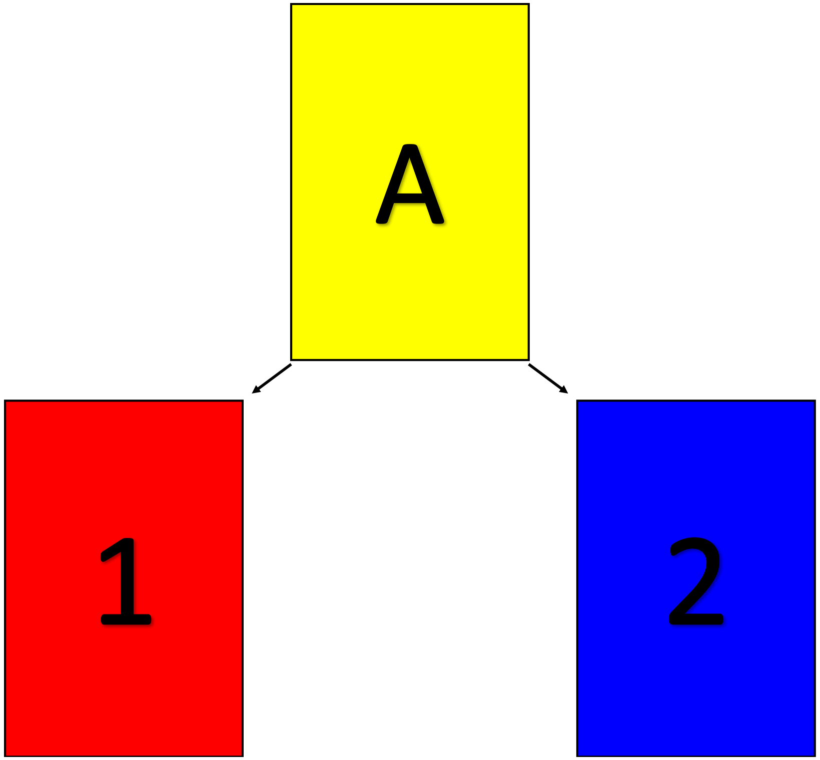
In this experiment you will be presented with three cards. The top card – in this case A,B or C is a cue card.

You have to pick between Red and Blue cards depending on the current cue.

Each cue is associated with a different probability of reward – when Red gives you a reward with 0.9 probability, Blue is the opposite of that – 0.1

You have to remember which choice is better when each of the cues is presented – eg “pick Red on A, pick Blue on B”

You were presented cue A. You need to decide whether Red or Blue will most likely lead to a reward based on prior information

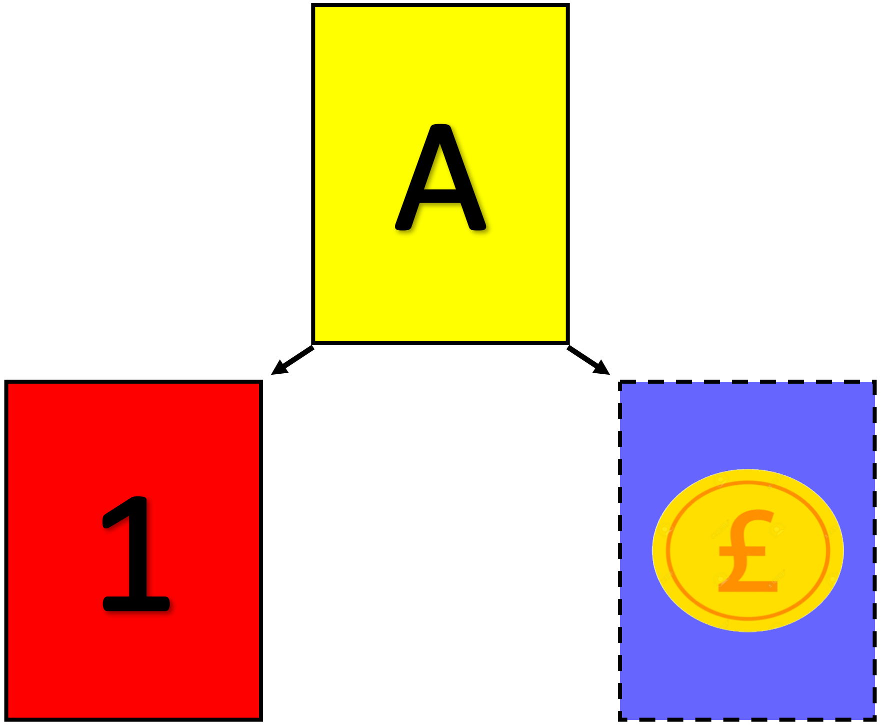


A

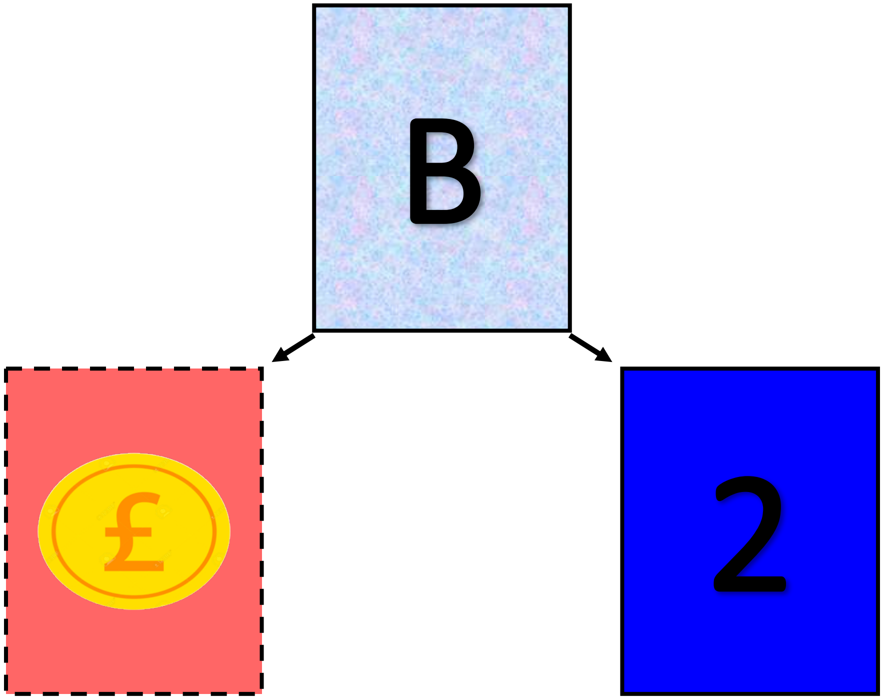
1

2

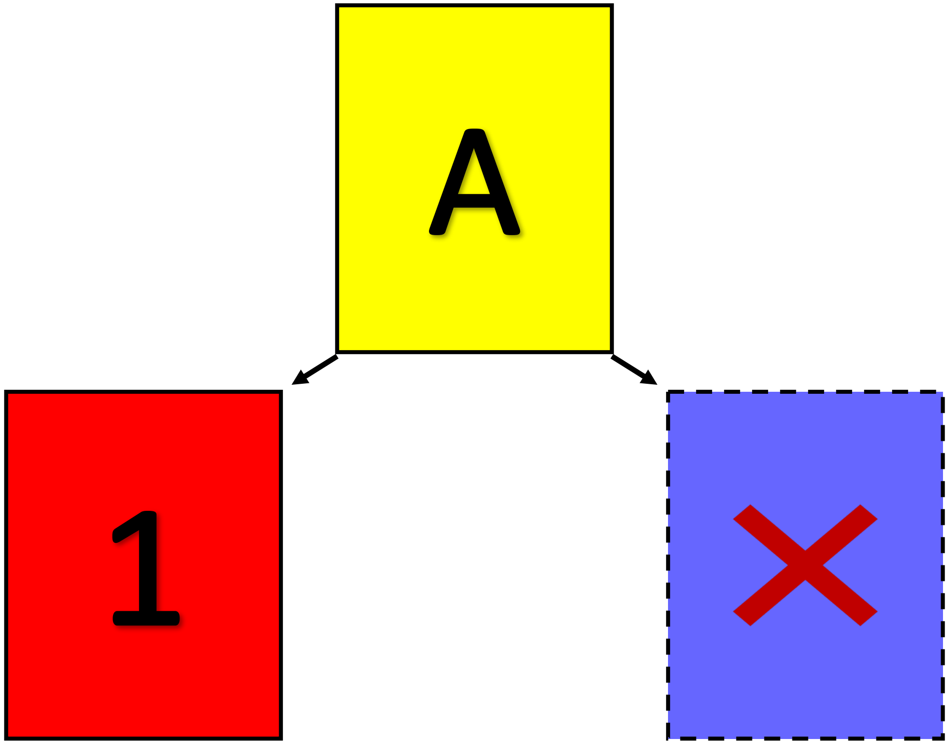
You can choose blue by pressing “v” – you have been rewarded! This means blue is a good choice on A



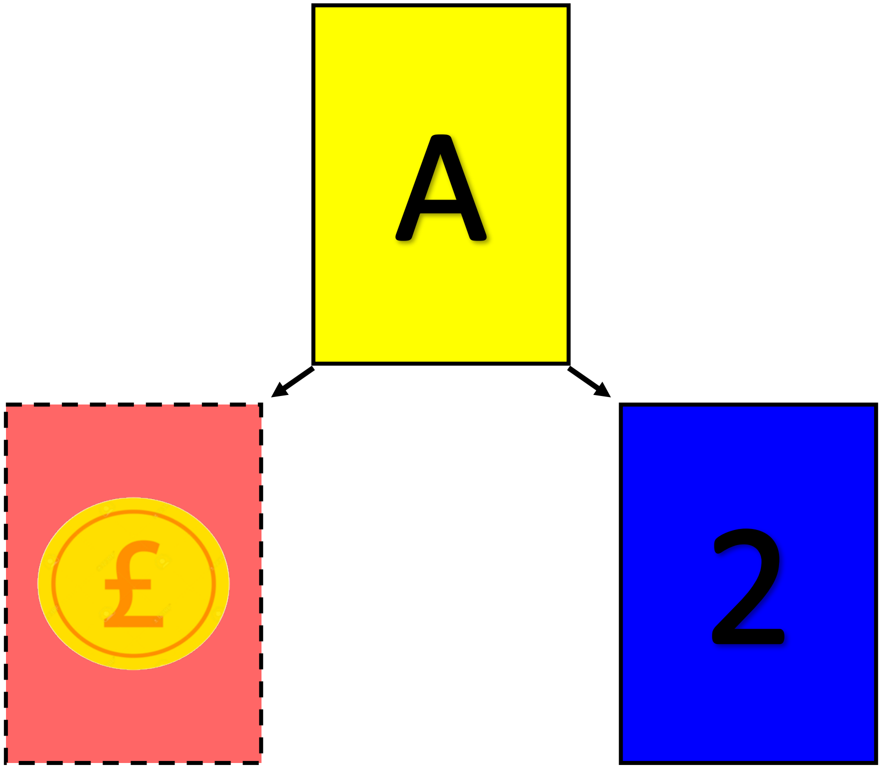
Some cues can be related to each other. For example, B can always be the opposite of A. Since you were rewarded on A-blue you can choose B-red



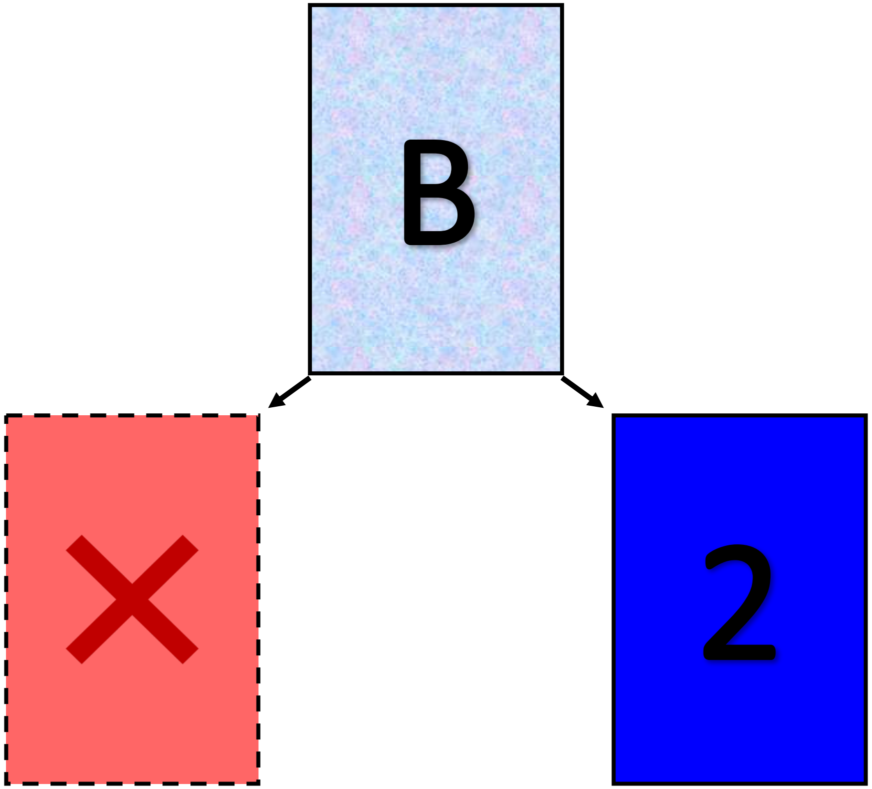
However, cues change the probabilities they are associated with sometimes! A is not blue anymore, so if you choose blue you will most likely not get a reward



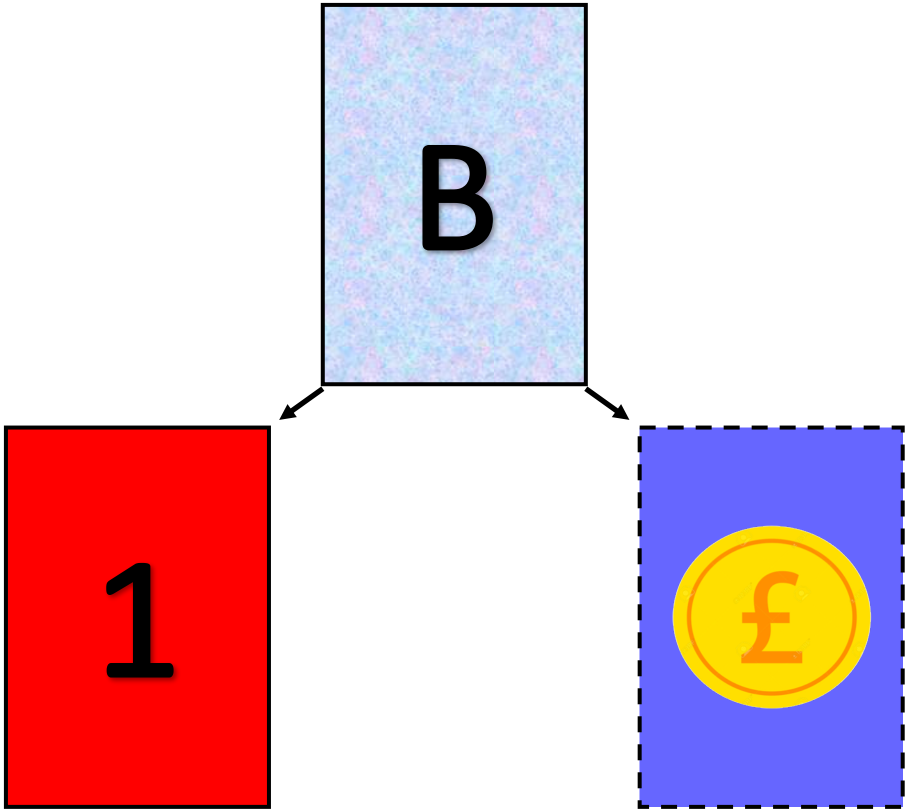
In this case you are better off choosing Red when you see A, until it changes again



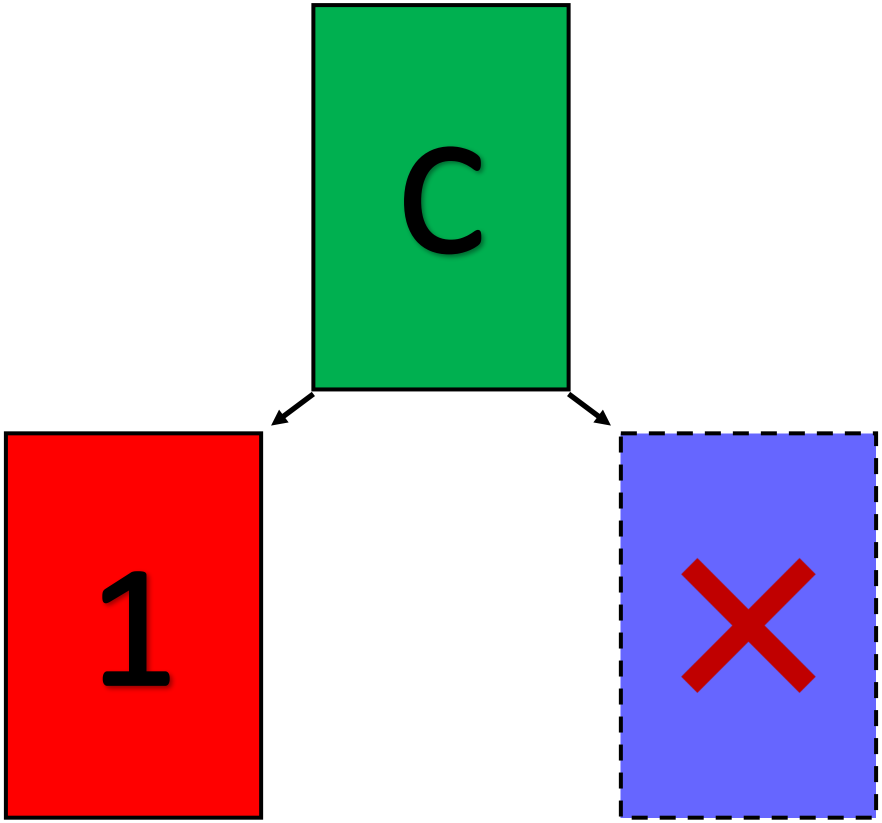
If A and B are related with a rule, and A changes from Blue to Red, B will also change from Red to Blue.



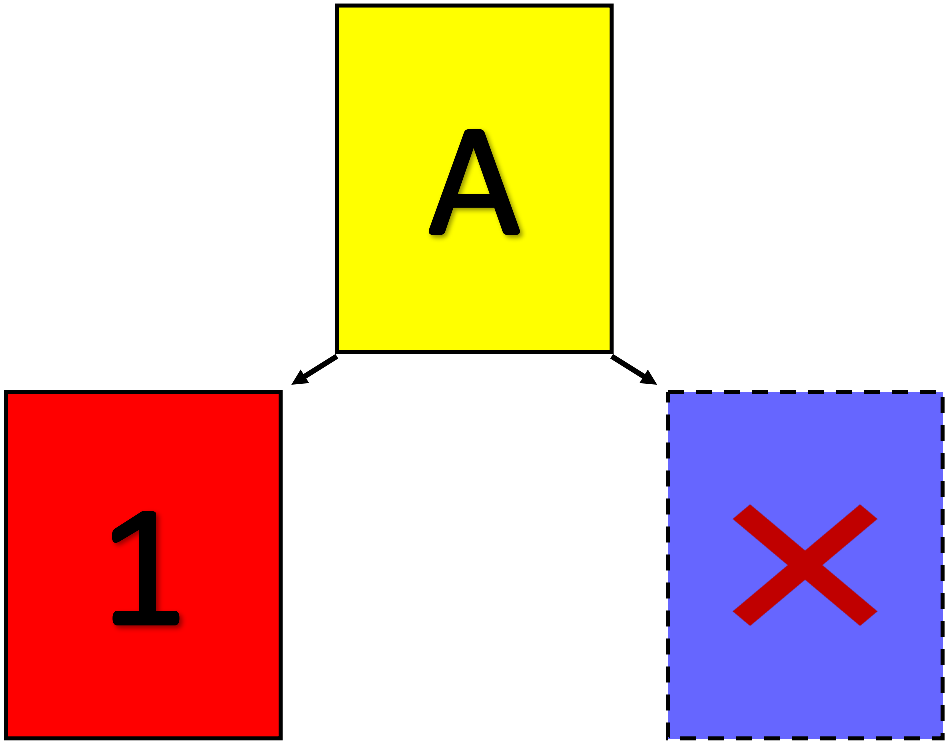
If A is now on Red, B must be on Blue!

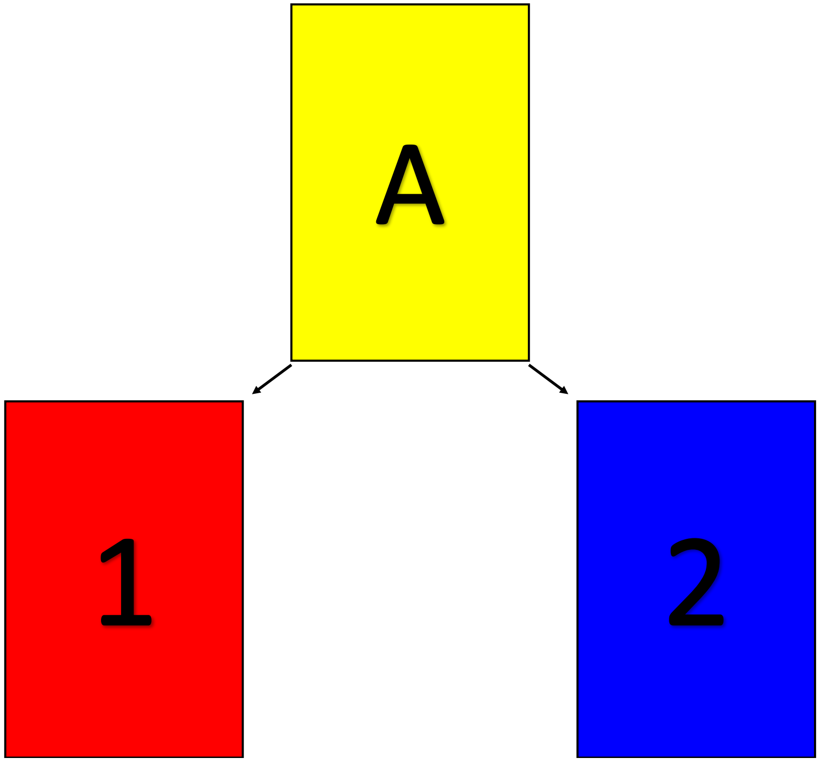


There is also a third cue. Cues don't have to be related to each other. In fact, two cues will always be related and one unrelated.



Because of the nature of the task, sometimes you will not get a reward even if you choose correctly. A reward is given with probability 0.9 if you chose the high value option and 0.1 if you chose the low value option.





A

1

2

Keys: “C” “V”