J Cogn Enhanc
DOI 10.1007/s41465-017-0030-7

ORIGINAL ARTICLE

Transcranial Direct Current Stimulation (tDCS) of the Anterior Prefrontal Cortex (aPFC) Modulates Reinforcement Learning and Decision-Making Under Uncertainty: a Double-Blind Crossover Study

Elias P. Casula[1,2] · **Giulia Testa**[3] · **Patrizia S. Bisiacchi**[2,4] · **Sara Montagnese**[4,5] · **Lorenza Caregaro**[4] · **Piero Amodio**[4,5] · **Sami Schiff**[4,5]

Received: 9 September 2016 / Accepted: 8 May 2017

Abstract Reinforcement learning refers to the ability to acquire information from the outcomes of prior choices (i.e. positive and negative) in order to make predictions on the effect of future decision and adapt the behaviour basing on past experiences. The anterior prefrontal cortex (aPFC) is considered to play a key role in the representation of event value, reinforcement learning and decision-making. However, a causal evidence of the involvement of this area in these processes has not been provided yet. The aim of the study was to test the role of the orbitofrontal cortex in feedback processing, reinforcement learning and decision-making under uncertainly. Eighteen healthy individuals underwent three sessions of tDCS over the prefrontal pole (anodal, cathodal, sham) during a probabilistic learning (PL) task. In the PL task, participants were invited to learn the covert probabilistic stimulus-outcome association from positive and negative feedbacks in order to choose the best option. Afterwards, a probabilistic selection (PS) task was delivered to assess decisions based on the stimulus-reward associations acquired in the PL task. During cathodal tDCS, accuracy in the PL task was reduced and participants were less prone to maintain their choice after positive feedback or to change it after a negative one (i.e., win-stay and lose-shift behavior). In addition, anodal tDCS affected the subsequent PS task by reducing the ability to choose the best alternative during hard probabilistic decisions. In conclusion, the present study suggests a causal role of aPFC in feedback trial-by-trial behavioral adaptation and decision-making under uncertainty.

Keywords aPFC · Reinforcement learning · Decision-making · tDCS · Trial-by-trial adaptive behavior

Abbreviations

aPFC	Anterior prefrontal cortex
vmPFC/ mOFC	Ventromedial prefrontal cortex/medial orbitofrontal cortex
tDCS	Transcranial direct current stimulation

Introduction

In a natural environment, the outcomes of our choices are frequently characterized by some degrees of uncertainty. Reinforcement learning refers to the ability to learn from the outcomes of prior choices (i.e., positive/negative) in order to make predictions on the effect of future decisions and adapt the behavior basing on past experiences (Noonan et al. 2012). This process seems to involve two different neural systems: (1) the striatal dopaminergic system, which plays a key role in the formation of stimulus-response associations (Frank and Claus 2006) and (2) the anterior prefrontal cortex (aPFC), including the lateral orbitofrontal cortex (lOFC) and the ventromedial prefrontal cortex/medial orbitofrontal cortex

✉ Sami Schiff
sami.schiff@unipd.it

[1] Non-invasive Brain Stimulation Unit, Department of Behavioural and Clinical Neurology, Santa Lucia Foundation IRCCS, Rome, Italy

[2] Department of General Psychology, University of Padova, Padova, Italy

[3] Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy

[4] Centro Interdipartimentale di Ricerca sulla Modellistica delle Alterazioni Neuropsichiche in Medicina Clinica (CIRMANMEC), University of Padova, Padova, Italy

[5] Department of Medicine, University of Padova, Padova, Italy

Published online: 24 May 2017

(vmPFC/mOFC), which is involved in the representation of event value implicated both in the acquisition and in the subsequent valuation of stimulus-response contingencies (Kringelback 2005; Frank and Claus 2006). In particular, the vmPFC/mOFC seems to be involved in value comparison during decision-making (Rushworth et al. 2011; Walton et al. 2011), whereas the lOFC is more concerned with reward-credit assignment during reward association learning (for a review see Noonan et al. 2012).

The role of the aPFC in reinforcement learning and decision-making has been highlighted in studies showing that lesion to this area impaired the ability to learn associations between reward outcomes and choice options (Rolls et al. 1994; Fellows and Farah 2003, 2007). However, a specific role of aPFC in feedback processing, behavioral adaptation, and decision-making under uncertainty remains to be elucidated in humans. In this direction, Tsuchida and co-workers, by adopting a probabilistic "non-deterministic" reversal-learning task, showed that vmPFC/mOFC-damaged patients are impaired in flexible learning from probabilistic feedback. In particular, these patients are less able to learn from positive/negative monetary feedback and tend to change their choices more frequently after a positive feedback (Tsuchida et al. 2010). In another study, Wheeler and Fellows (2008) investigated how positive and negative feedbacks guide reinforcement learning and decision-making under uncertainty in patients with vmPFC/mOFC lesions. The authors adopted a two-phase probabilistic reinforcement learning task (i.e., learning phase and testing phase) developed by Frank et al. (2004) and Frank (2005)) to test the role of dopaminergic pathways in reinforcement learning in patients with Parkinson's disease (see also Volpato et al. 2016). Results showed that patients with vmPFC/mOFC lesions were impaired in both reinforcement learning (learning phase) and in the ability to use negative feedback to make decisions in a novel context (testing phase). In contrast, these patients were as good as both healthy controls and patients with non-vmPFC/mOFC lesions in making decisions derived from positive feedback (Wheeler and Fellows 2008).

Recently, non-invasive brain stimulation techniques able to modulate cortical excitability of specific brain area have been adopted to investigate the causal role of the lateral prefrontal cortex (LPFC) in feedback processing during reinforcement learning, using the same paradigm adopted by Wheeler and Fellows (2008). Ott et al. (2011) showed that continuous theta burst stimulation over the left dorsolateral prefrontal cortex (DLPFC) do not influence reinforcement learning per se, rather modulates choices based on positive feedback in the testing phase, enhancing the probability of choosing the "better option" against the others. Another study by Turi et al. (2015), using transcranial direct current stimulation (tDCS) showed that modulation of the left DLPFC increases maladaptive trial-by-trial behavioral shifting during learning, since participants "stayed less" on their choices, even after receiving positive or negative feedback. In addition, the stimulation of DLPFC is associated with increased randomness of choices during learning and decreased accuracy for choosing the better option in the testing phase.

Although previous studies in brain-damaged patients suggested a key role of the aPFC in reinforcement learning and feedback processing, non-invasive brain stimulation studies focused only on the role of DLPFC. In the present study, tDCS was applied over the aPFC (Karim et al. 2010; Chib et al. 2013; Manuel et al. 2014) during a probabilistic reinforcement learning task (Frank et al. 2004; Frank 2005; Wheeler and Fellows 2008) with the aim to investigate the causal role of this area in (1) probabilistic reinforcement learning, (2) trial-by-trial behavioral adaptation in response to positive/negative feedback, and (3) decision-making under uncertainty.

Methods

Participants

Eighteen healthy volunteers (6 males; age mean ± standard deviation, 25.2 ± 3.8 years) were recruited to participate to a double-blind crossover-designed tDCS study. Each participant was requested to undergo three experimental sessions (specified above) and was informed about the duration of each session. All participants were checked for tDCS exclusion criteria and had no history of neurological, psychological (see paragraph 2.4), or other relevant medical disorders (Nitsche et al. 2003). Also, none of them took psychotropic drugs. All the participants took part in the experiment voluntarily, after providing written informed consent. All participants were Italian and naive to Japanese language. The University Hospital of Padua Ethical Committee approved the experimental procedure.

Procedure and Task

The experiment took place in a quiet dimly lit room. Participants sat in a comfortable armchair in front of an 18-in CRT monitor controlled by a Pentium IV computer. The task was divided into two parts: a "learning phase" and a "test phase." During the learning phase, participants performed a probabilistic learning (PL) task, while, during the test phase, they performed a probabilistic selection (PS) task. The three experimental sessions were scheduled in different days separated by at least 48 h. During each experimental session, participants received anodal, cathodal, or sham tDCS, with the stimulation order randomized and counter-balanced across participants (Fig. 1a).

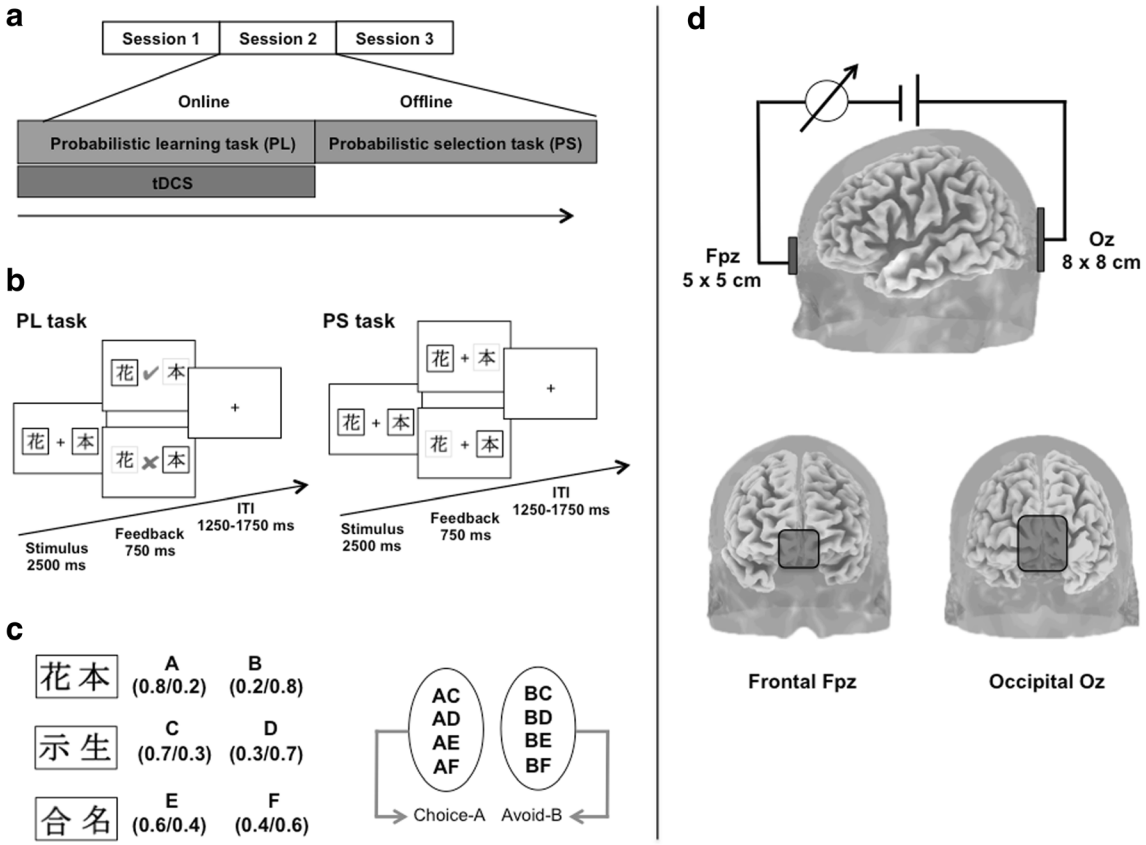

Fig. 1 **a** Experimental design. **b** Trial time line for the probabilistic learning (PL) and the probabilistic selection (PS) task. **c** Stimuli adopted during the PL task with their associated feedback probability (*left*) and the new-pairs of stimuli (choose-A and avoid-B) adopted in the PS task (*right*). **d** tDCS electrode montage: active electrode was placed over the Fpz site, and reference electrode was placed over Oz

Probabilistic Learning Task

In the learning phase, participants were requested to perform a PL task. During the PL task, three fixed pairs of Japanese Hiragana ideograms (AB, CD, EF) were randomly and equally presented across four blocks, 80 times (Fig. 1b, left). Each block comprised 60 trials. All the stimuli were presented on a white background. Each of the six ideograms was associated with a different probability (i.e., $0 \leq p(x) \leq 1$) of receiving positive (i.e., $A = 0.8$; $B = 0.2$; $C = 0.7$, $D = 0.3$, $E = 0.6$, and $F = 0.4$) and negative feedbacks (i.e., $A = 0.2$; $B = 0.8$; $C = 0.3$, $D = 0.7$, $E = 0.4$, and $F = 0.6$). Thus, each of the three fixed pairs was composed of two ideograms with complementary probability of being associated with positive (i.e., $AB = 0.8/0.2$; $CD = 0.7/0.3$, and $EF = 0.6/0.4$) and negative feedbacks (i.e., $AB = 0.2/0.8$; $CD = 0.3/0.7$, and $EF = 0.4/0.6$) (Fig. 1c, left). In each trial, ideograms were displayed 4° of visual angle on the left and on the right of a fixation point, in the middle of a white square with the side subtending 2° of visual angle outlimited by a thin blue border. In the three experimental sessions, three different sets of ideograms were adopted in a randomized order between participants. In each trial, participants were instructed to choose between the two ideograms presented on the screen. They were informed that for each pair of ideograms, the choice of one of the two would be more frequently associated with a positive feedback, while the other would be more frequently associated with a negative feedback. Thus, participants were instructed to press with the left/right hand the keys "Z"/"M" on the computer keyboard, depending on whether they thought the left or the right stimulus was more frequently associated with the "correct" (i.e., positive) feedback. After each choice, a probabilistic visual feedback appeared on the center of the screen showing a positive feedback (i.e., a green tick) or a negative feedback (i.e., a red cross) if the response was correct or incorrect, respectively. During the experimental session, they had to work out the best alternative in each pair of ideograms by using the history of positive/negative feedback. Before starting the experiment, a practice block was performed to ensure participants had understood the task. The maximum reaction time for each trial was 2.5 s. If no response was detected, a question mark appeared on the center of the screen and the trial was considered null. We used three different ideogram pairs in each session to avoid participants to use the associations learned in the previous sessions. Thus, each participant saw a total of 18 ideograms, 6 for each session. Trial sequence was completely

randomized in every session. The probabilistic association between each ideogram and positive/negative feedback was predetermined in the trial sequence within each block. In this way, every ideogram was associated with both positive and negative feedbacks in the trial list with the frequency defined by the experimental design. Thus, in each block, the reward structure completely matched the probability defined by the experimenter in each session.

Probabilistic Selection Task

After the PL task, a PS task was delivered to evaluate the probabilistic positive/negative-ideogram associations derived from the feedbacks received during the learning phase (Fig. 1b, right) by using eight new pairs of ideograms. The new pairs AC, AD, AE, AF, BC, BD, BE, and BF were obtained by coupling the highest positive cumulative feedback ideogram A and the lowest positive cumulative feedback B with all the others. Two blocks each of 88 trials presented randomly composed the PS task. This procedure allows evaluation of whether participants learned more from positive ("choose-A") or negative ("avoid-B" condition) feedback to their responses (Fig. 1c, right). Old pairs (AB, CD, EF) were also presented to evaluate explicit learning ("old-pairs" condition). In this test phase, participants did not receive any feedback to their responses and they were invited to make their choices using the probabilistic rules implicitly acquired during the PL task.

Transcranial Direct Current Stimulation

Electricity for tDCS was generated by a portable, battery-supplied direct current (DC) stimulator (HDCStim, Newronika s.r.l., Milan, Italy) which was previously programmed by one of the experimenter (S.S.) using a touchscreen LCD connected to the stimulator before every session. Stimulation started after the PL task instructions were provided and finished approximately at the end of the probabilistic learning task (Fig. 1a). DC stimulation was delivered with two saline-soaked surface sponge electrodes, with the active electrode (25 cm^2) placed over Fpz and the reference electrode (64 cm^2) over Oz, according to the 10-20 EEG-sys (Jasper 1958; Fig. 1d). This montage was chosen to maximize the anterior-posterior current flow through the aPFC more specifically within the vmPFC/mOFC and to minimize the lateral shunting along the skin (Fumagalli et al. 2010). A montage with a large reference electrode joined with a small one over target site was chosen to increase the effectiveness of stimulation under the target electrode (Ly et al. 2016).

During active stimulation, participants received a constant current of 2 mA for 25 min. The current was always ramped up over the first 30 s of stimulation and down over the last 30 s. For sham stimulation, the electrode positioning was the same but the current was slowly ramped down after 30 s of stimulation with random polarity in order to preserve the initial tingling sensation. Before and after the stimulation, a visual analog scale (VAS) ranging from 0 to 10 was filled in by the participants to evaluate the perception of pain/discomfort relating to the stimulation and changes in nine domains of mood, namely happy/sad; calm/restless; fast mind/slow mind; apathetic/dynamic; confused/lucid; strong/weak; satisfied/unfulfilled; worried/unconcerned; and tense/relaxed (Fregni et al. 2008). During the entire session, the experimenter remained in the room but away from the participant's visual field.

Data Analysis

Response accuracy and reaction times (RTs) to the PL and the PS task were separately considered and analyzed as dependent variables in a within-subjects design. Performance in PL task was analyzed with a 3 × 3 × 4 repeated measures ANOVA with type of stimulation (anodal, cathodal, and sham), pair condition (AB, CD, EF), and block (1st, 2nd, 3rd, 4th) as within-subject factors.

In order to evaluate the effects of the type of stimulation on feedback processing and behavioral adaptation, we analyzed the occurrence of (1) "win-stay" responses, defined as the percentage of trials in which participants chose the same stimulus after having received a positive feedback the last time that the chosen stimulus was presented, and (2) "lose-shift" responses, defined as the percentage of trials in which participants changed their choice of stimulus after having received a negative feedback the last time that the chosen stimulus was presented (Evenden and Robbins 1983; Frank and Kong 2008). Such analysis was performed with a 3 × 2 × 4 repeated measures ANOVA with type of stimulation (anodal, cathodal, sham), trial-by-trial behavioral adaptation (win-stay, lose-shift), and block (1st, 2nd, 3rd, 4th) as within-subject factors.

Performance in PS task was first analyzed with a 3 × 3 repeated measures ANOVA, with type of stimulation (anodal, cathodal, sham) and condition (choose-A, avoid-B, old-pairs) as within-subject factors. In addition, to evaluate the ability to make correct choices based on the rules learned during the PL task, a distinction within the new pairs of the PS task was made: "easy" probabilistic choices, occurring when the mostly rewarded ideogram (i.e., A) was paired with ideograms associated frequently with negative feedback (i.e., D, F) and when the mostly punished ideogram (i.e., B) was paired with ideograms frequently associated with positive feedback (i.e., C, E); and "hard" probabilistic choices, occurring when the mostly rewarded ideogram (i.e., A) was paired with ideograms frequently associated with positive feedback (i.e., C, E) and when the mostly punished ideogram (i.e., B) was paired with ideograms frequently associated with negative feedback (i.e., D, F). Responses to these pairs were analyzed using a 3 × 2 repeated measures ANOVA with type

of stimulation (anodal, cathodal, sham) and choice difficulty (easy, hard) as within-subject factors.

Prior to undergoing ANOVA procedures, normal distribution was assessed by Shapiro-Wilks' test for all the variables. Level of significance was set at $\alpha = .05$. Sphericity of data was tested with Mauchly's test before performing statistical analysis. When sphericity was violated (i.e., $p < 0.05$), the Greenhouse-Geisser correction was used. Post hoc comparisons were performed using the Bonferroni correction. Friedman's test was used for each VAS to evaluate the presence of mood and sensation changes due to tDCS.

Results

The analyses on the VAS did not reveal any significant difference in mood or sensations before and after the three types of stimulation (all *p*s > 0.05).

The analysis on PL task accuracy revealed a significant main effect of block [$F_{(3,51)} = 25.67$; $p < 0.0001$; $\eta^2_p = 0.60$], showing a significantly lower accuracy in the first block, compared to all the others (Fig. 2a; all *p*s < 0.0001), and the effect of pair condition [$F_{(2,34)} = 17.63$; $p < 0.0001$; $\eta^2_p = 0.51$], with lower accuracy for the EF pair compared to both the AB and CD pairs (post hoc *p*s < 0.0001). The significant effect of type of stimulation [$F_{(2, 34)} = 3.8$; $p < 0.03$; $\eta^2_p = 0.19$] reveals lower accuracy during cathodal compared to sham stimulation (Fig. 2b; $p = 0.03$). No significant interactions were detected on this analysis (all *p*s > 0.37).

The same analysis applied on RTs revealed a significant main effect of block [$F_{(3, 51)} = 27.92$; $p < 0.00001$; $\eta^2_p = 0.62$] and pair condition [$F_{(2, 34)} = 4.74$; $p = 0.015$; $\eta^2_p = 0.22$]. Again, post hoc analysis revealed the expected longer RTs in the first block compared to the other blocks, and for the EF pair compared to the CD pair (post hoc *p*s < 0.0002). Mean RTs in the PL task did not reveal any differences between cathodal, anodal, and sham stimulation and no interaction including the type of stimulation [all *p*s > 0.19].

The analysis on feedback processing during the PL task revealed a significant main effect of behavior [$F_{(1, 17)} = 94.87$; $p < 0.00001$; $\eta^2_p = 0.85$] with higher occurrence of win-stay behavior compared to lose-shift and revealed a significant main effect of type of stimulation [$F_{(2, 34)} = 3.55$; $p = 0.04$; $\eta^2_p = 0.17$], revealing lower accuracy for cathodal stimulation compared to sham. Interaction between type of stimulation and behavior was also found to be significant [$F_{(2,34)} = 3.72$; $p = 0.03$; $\eta^2_p = 0.18$] although post hoc analysis did not show a significant difference between the occurrence of win-stay and lose-shift behaviors in the three types of stimulation (see Table 1). A significant interaction between behavior and block was also present [$F_{(3, 51)} = 20.75$; $p < 0.0001$; $\eta^2_p = 0.55$] showing an increase in win-stay and a decrease in lose-shift during learning process (i.e., from the 1st block to the 4th; post hoc *p*s < 0.001).

The analysis of response accuracy in the PS task showed a main effect of condition [$F_{(2, 34)} = 10.13$; $p = 0.0005$; $\eta^2_p = 0.37$], revealing an expected higher accuracy in the old-pairs condition compared to the other two (post hoc *p*s < 0.002). In contrast, the main effect of the type of stimulation did not reach statistical significance [$F_{(2, 34)} = 2.32$; $p = 0.11$; $\eta^2_p = 0.11$] and no interaction between type of stimulation and condition was found ([$F_{(4, 68)} = 0.33$; $p = 0.8$; $\eta^2_p = 0.015$]; see Table 1).

The analysis of RTs obtained from the PS task showed a significant main effect of condition [$F_{(2, 34)} = 18.54$; $p < 0.00004$; $\eta^2_p = 0.52$], revealing slower overall RTs during

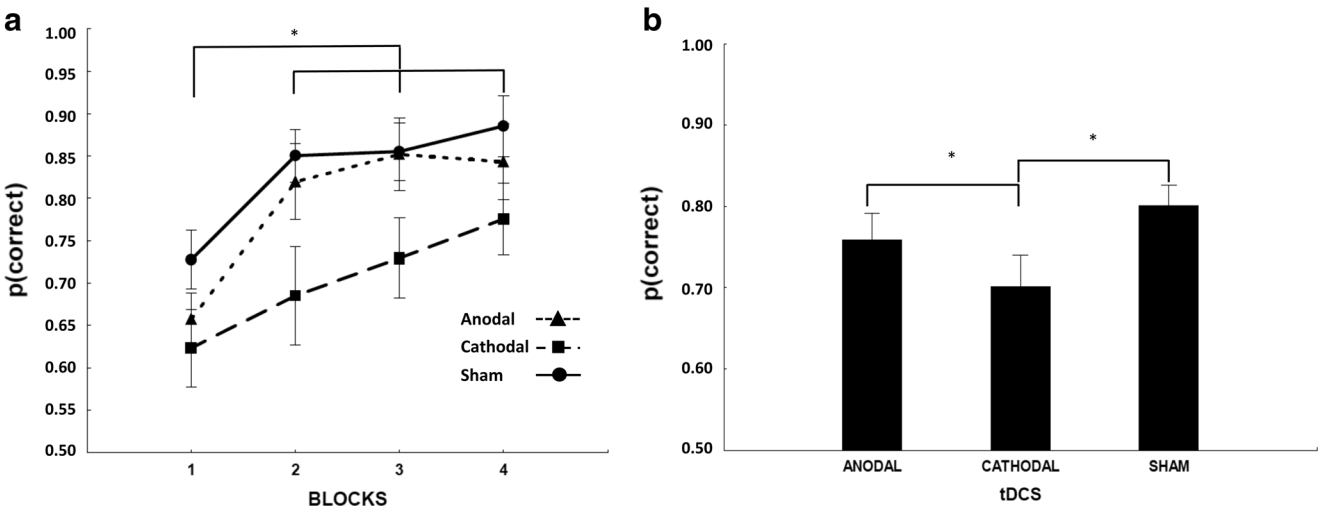

Fig. 2 **a** Percentage of accuracy during the PL task for the different experimental blocks in the three tDCS sessions (anodal, cathodal, and sham). **b** Overall percentage of accuracy in the three tDCS sessions. During cathodal stimulation, accuracy was significantly reduced compared to the other conditions. *Vertical bars* describe the standard error. Bonferroni post hoc test: *$p < .05$. *Vertical bars* indicate mean ± standard error

Table 1 Trial-by-trial adaptive behaviors in the PL task for anodal, cathodal, and sham tDCS. Mean ± SD

	tDCS		
	Anodal	Cathodal	Sham
Win-stay (%)	51.89 ± 13.62	49.9 ± 12.86	56.3 ± 9.54
Lose-shift (%)	13.05 ± 7.97	11.99 ± 8.34	10.06 ± 6.94

the avoid-B condition compared to both choose-A and old-pairs (post hoc ps < 0.0001). No effects of type of stimulation or interaction between type of stimulation and condition were present (post hoc ps > 0.4) (Table 2).

The analysis on hard and easy probabilistic decisions revealed a main effect of choice difficulty [$F_{(1,\ 17)} = 118.81$; $p < 0.00001$; $\eta^2_p = 0.87$], showing an expected significantly higher accuracy for easy pairs. No main effect of stimulation was present; however, a significant interaction of type of stimulation and choice difficulty was found [$F_{(2,\ 34)} = 4.95$; $p < 0.02$; $\eta^2_p = 0.23$]. As shown in Fig. 3, post hoc analyses revealed that during the PS task, responses to hard pairs of stimuli, but not to easy pairs, were significantly less accurate after anodal compared to cathodal stimulation (post hoc $p = 0.026$). Analysis on RTs did not reveal any significant effect (all ps > 0.1).

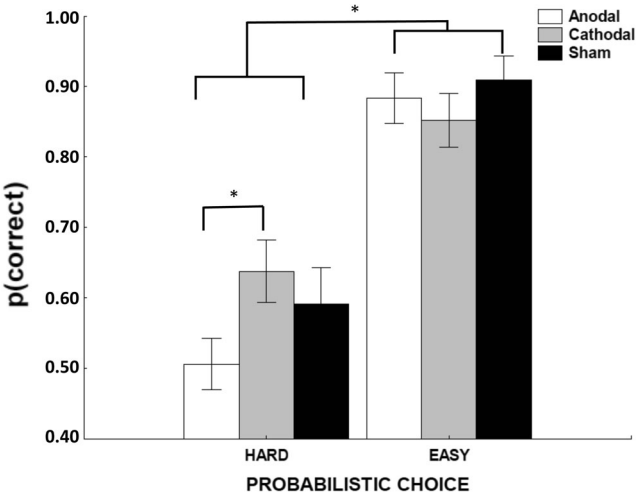

Fig. 3 Percentage of accuracy during the PS task divided for hard (i.e., AC, AE, BD, and BF) and easy probabilistic choices (i.e., AD, AF, BC, and BE). Dispersion expressed as standard error. *Vertical bars* indicate mean ± standard error

Discussion

The aims of the present study were to evaluate the causal role of the aPFC in (1) probabilistic reinforcement learning, (2) trial-by-trial behavioral adaptation in response to positive/negative feedbacks, and (3) decision-making under uncertainty. Here, we adopted a two-phase probabilistic learning task (Frank et al. 2004) to investigate reinforcement learning and the ability to use previously learned stimuli-reward/punishment associations to make decisions in a completely new context with various degrees of probabilistic difficulty.

In the learning phase (PL task), cathodal stimulation of aPFC was found to reduce response accuracy compared to both anodal and sham stimulation. The analysis of the trial-by-trial behavioral adaptation highlighted a modulation of the percentage of both win-stay and lose-shift behavior, depending on the stimulation condition. Importantly, we observed that cathodal stimulation did not influence the learning performance per se, since no interaction between the type of stimulation and block was detected. The reduction in overall accuracy and frequency of win-stay and lose-shift responses is likely produced by an impairment of the selection process as a result of cathodal stimulation, demonstrating a causal role of aPFC in modulating adaptive behavior. Notably, modulation of the trial-by-trial behavioral adaptation was not affected from the type of feedback received during the PL task. On the other hand, in terms of retrieval of new stimulus-response associations, learning seems to be preserved.

Table 2 Accuracy in the three conditions of PS task for anodal, cathodal, and sham tDCS. Mean ± SD

	tDCS		
	Anodal	Cathodal	Sham
Choose-A (%)	76 ± 24	59 ± 21	79 ± 25
Avoid-B (%)	63 ± 22	59 ± 21	70 ± 26
Old-pairs (%)	78 ± 25	70 ± 26	83 ± 26

The testing phase (PS task) allowed a further evaluation of the aPFC role in response selection during decision-making under uncertainty. Participants had to choose the best alternative without receiving any feedback. In this phase, old pairs of ideograms (learned during the PL task), as well as new pairs, were presented to evaluate the tendency to select and/or avoid the better (i.e., the ideogram A) and the worst (i.e., the ideogram B) alternative, respectively. Participants had to retrieve the probability of receiving positive/negative feedback associated with each of the ideograms presented in the PL task and to use these representations to make a decision on the best alternative to choose in a new context. It is important to take into account that during the PS task, participants did not receive any stimulation; thus, a significant effect of tDCS should be considered as an aftereffect.

The analysis of both response accuracy and RTs in the PS task did not reveal any significant effect of stimulation condition nor of the type of choice (i.e., choose-A or avoid-B). However, we observed a significant difference in accuracy between old-pairs and both choice-A and avoid-B conditions. Interestingly, when old pairs were presented, no effects of the type of stimulation were observed, suggesting that choices within ideogram pairs

presented during the PL task were not modulated by tDCS. Notably, response accuracy to old pairs was very high in the testing phase. In each session of the PL task, participants were exposed to the same ideogram pairs (i.e., AB, CD, EF—called old-pairs in the PS task) and each decision depended on both the items (Rothermund et al. 2015). Thus, while in the old-pairs condition the selected and the unselected ideograms are associated with the same action, in the choose-A/avoid-B conditions, participants had to infer the correct response basing on probabilistic associations between each ideogram and the frequency of the following positive/negative feedback. Thus, a difference in the accuracy between the old-pairs and the choose-A/avoid-B conditions is an expected result (Gozli et al. 2014). Following these considerations, it can be assumed that decision-making during the presentation of overlearned old pairs was mainly managed by the striatal sub-cortical dopaminergic system (Frank and Claus 2006). In this condition, response selection should be regulated by the automatic activation of specific stimulus-response associations, with no (or low) influence of the top-down modulation exerted by the aPFC. In contrast, when probabilistic decisions require transferring the probabilistic value associated with an ideogram in a new decision context (i.e., choose-A/avoid-B conditions), aPFC plays a key role during response selection. Along the same lines, Frank and Claus (2006) suggested a critical role of the aPFC in the maintenance of value representations in working memory, during decision-making. In addition, these authors suggest an anatomo-functional distinction between mOFC, mainly implicated in representing positive rewards, and lOFC, mainly implicated in representing punishments. Although the suggested role of the aPFC in reward-based decision-making is coherent with our results, the specificity of the mOFC in representing positive values seems not to be supported in the present study.

The absence of an effect of stimulation on the type of learning from positive (i.e., choose-A) or negative feedback (i.e., avoid-B) seems to be in contrast with the results of Wheeler and Fellows (2008) showing a decrease in avoid-B accuracy in patients with vmPFC/mOFC lesion. The authors suggested that the vmPFC/mOFC is involved in the ability to use information from negative feedbacks, rather than positive, to make decisions in a novel context. However, in literature, there is a lack of consensus on whether reward or punishment is encoded in the same or in different brain regions (Elliott et al. 2003; Knutson and Bossaerts 2007; Liu et al. 2007; Seymour et al. 2007). Along the same lines, our finding of a reduction of both win-stay and lose-shift during cathodal stimulation of the aPFC does not agree with the idea of a particular role of the medial part of OFC (i.e., vmPFC/mOFC) in representing reward, but not punishment, during decision-making (O'Doherty et al. 2001; Small et al. 2001; Kringelbach et al. 2003; Kringelbach and Rolls 2004; Bechara et al. 1994, 2000; Fellows and Farah 2003; Rolls and Grabenhorst 2008; Tsuchida et al. 2010).

According to a recent model (Noonan et al. 2010), a functional distinction between lateral and medial OFC has been proposed, suggesting that the vmPFC/mOFC is involved in value comparison during decision-making, whereas the lOFC is involved in reward-credit assignment during reward association learning (for a review see Noonan et al. 2012). Our results can be framed within this view, since stimulation of the aPFC (both anodal and cathodal) did not affect participant's ability to acquire ideogram value from the history of positive/negative feedbacks (PL task), a process linked with the activity of lOFC based on credit assignment hypothesis (Noonan et al. 2010). In addition, we found that tDCS over the aPFC, particularly targeting the vmPFC/mOFC, did not affect reinforcement learning neither the ability to make correct choices when old pairs were presented in the testing phase. However, the observed modulation of decision-making within the new pairs of ideograms, especially those based on positive feedbacks, suggests a specific role of vmPFC/mOFC in feedback processing and response choice, but not in the acquisition and the availability of previously learned stimulus-reward associations. Recently, Camille et al. (2011) proposed that the medial portion of the OFC is involved in stimulus-reward associations and decision-making, while action-reward evaluation is linked to the activity of the anterior cingulate cortex. In line with their results, the probabilistic task used in the present study allows the investigation of stimulus-reward, but not action-reward values, corroborating the idea that the vmPFC/mOFC is implicated in stimuli evaluation during response selection and decision-making, but not in credit assignment (Noonan et al. 2012).

In the present study, we further investigated the role of the aPFC in decision-making under uncertainty, by distinguishing hard and easy choices. Here, we assume that when the two ideograms presented within the new pairs have a similar probability of receiving positive/negative feedback, decisions are harder and require top-down support by the aPFC, and vice versa. In our study, after anodal stimulation, participants were less accurate in making hard decisions compared to cathodal stimulation, suggesting that this kind of stimulation of the aPFC interferes with decision-making when the detection of the best options is difficult. Since only hard choices, regardless of their type (i.e., choose-A, avoid-B), were affected by the stimulation, it can be assumed that the activity of the vmPFC/mOFC is related to the degree of difficulty of the decision process. This corroborates the finding of Rolls et al. (2010) who observed an fMRI cluster in the mOFC, whose activation was linearly correlated with the difficulty of decision in olfactory and pleasantness discrimination (Rolls et al. 2010).

Overall, our data support the hypothesis of a functional role of the human aPFC (1) in trial-by-trial behavioral adaptation based on feedback processing, affecting mainly win-stay behavior, and (2) in using the learned stimulus-reward associations in order to guide response selection under uncertainty in

a new decision context (O'Doherty et al. 2001; Small et al. 2001; Kringelbach et al. 2003; Kringelbach and Rolls 2003; Wallis 2007; Elliott et al. 2010). In contrast, the finding that tDCS over the aPFC does not affect the acquisition of new stimulus-reward associations based on both positive and negative feedbacks suggests that our stimulation protocol was ineffective in modulating reinforcement learning ability.

Based on this data, a first methodological consideration may be done. Although tDCS is considered a low spatial resolution technique, our data support a specific modulation of trial-by-trial response selection and uncertainty decision-making. In agreement with Noonan et al. (2012) hypothesis, the present results support the distinction between medial and lateral OFC, respectively, in response selection and in credit assignment during the acquisition of new stimulus-reward associations. Thus, we suggest that the electrode montage adopted here was able to effectively target the medial part of the OFC (i.e., vmPFC/mOFC), but not the lOFC. Further studies should be performed to evaluate this idea and to demonstrate that a more lateral stimulation of the aPFC is able to modulate credit assignment process, but not decision criteria, as previously suggested (Rushworth et al. 2011; Noonan et al. 2012). A second methodological consideration regards the online and offline effects of tDCS. In detail, cathodal stimulation was found to affect online trial-by-trial behavioral adaptation during the PL task, whereas anodal stimulation mainly affects offline harder decisions within a new decision context. It is generally assumed that cathodal tDCS reduces cortical excitability while anodal stimulation enhances cortical excitability (Nitsche and Paulus 2001). However, this dichotomy, which works well mainly within the motor cortex, was recently challenged (Reato et al. 2013), and it is less consistent when different areas are stimulated in studies investigating cognitive functioning (Jacobson et al. 2012). In the present study, the effect of stimulation seems to be subtractive in terms of behavioral advantage/disadvantage, both during cathodal (i.e., online) and after anodal (i.e., offline).

In conclusion, the present study further demonstrates in healthy volunteers a causal relationship between the aPFC and processes related to decision-making and behavioral adaptation. Unfortunately, the extension of the aPFC and its subdivision in medial OFC and lateral OFC requires the development of more precise experimental designs for dissociating the functional role of the different sub-regions in reinforcement learning and decision-making. Finally, the present work supports the use of tDCS in the investigation of mechanisms of decision-making and reinforcement learning within the aPFC in healthy participants, as well as in clinical populations with addiction (e.g., pathological gambling, smoke, alcohol, cocaine, food misuse—Boggio et al. 2008; Fregni et al. 2008; Fecteau et al. 2010; Levasseur-Moreau and Fecteau 2012) and in other psychopathological conditions (e.g., obsessive-compulsive disorder, depression, anxiety, and post-traumatic stress disorder—Volpato et al. 2013; Loo et al. 2012; Ferrucci et al. 2013; Marin et al. 2014), known to affect decision-making and reinforcement learning.

Acknowledgements The authors are grateful to Esmeralda Balistreri, Gloria Imperato, and Daniele Aprile for participants' recruitment and data collection.

Compliance with Ethical Standards All the participants took part in the experiment voluntarily, after providing written informed consent. The University Hospital of Padua Ethical Committee approved the experimental procedure.

References

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7–15.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10*, 295–307.

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

Camille, N., Tsuchida, A., & Fellows, L. K. (2011). Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. *The Journal of Neuroscience, 31*, 15048–15052.

Chib, V. S., Yun, K., Takahashi, H., & Shimojo, S. (2013). Noninvasive remote activation of the ventral midbrain by transcranial direct current stimulation of prefrontal cortex. *Trans Psychiatry, 3*, e268.

Elliott, R., Newman, J. L., Longe, O. A., & Deakin, J. W. (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. *The Journal of Neuroscience, 23*, 303–307.

Elliott, R., Agnew, Z., & Deakin, J. F. (2010). Hedonic and informational functions of the human orbitofrontal cortex. *Cerebral Cortex, 20*, 198–204.

Evenden, J. L., & Robbins, T. W. (1983). Dissociable effects of d-amphetamine, chlordiazepoxide and alpha-flupenthixol on choice and rate measures of reinforcement in the rat. *Psychopharmacology, 79*, 180–186.

Fecteau, S., Fregni, F., Boggio, P. S., Camprodon, J. A., & Pascual-Leone, A. (2010). Neuromodulation of decision-making in the addictive brain. *Substance Use & Misuse, 45*, 1766–1786.

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain, 126*, 1830–1837.

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se?. *Cerebral Cortex, 17*(11), 2669–2674.

Ferrucci, R., Bortolomasi, M., Fumagalli, M., Giacopuzzi, M., Mameli, F., Priori, A., Scelzo, E., & Vergari, M. (2013). Transcranial direct current stimulation (tDCS): a treatment for depressive illness. *Int Journ Psychiatry Clinic Practic, 15*, 22–23.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and non-medicated parkinsonism. *Journal of Cognitive Neuroscience, 17*, 51–72.

Frank, M. J., & Claus, E. D. (2006). Anatomy of decision: striato-orbitofrontal interactions in reinforcement learning, decision making and reversal. *Psychological Review, 113*, 300–326.

Frank, M. J., & Kong, L. (2008). Learning to avoid in older age. *Psychology and Aging, 23*, 392–398.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science, 306*, 1940–1943.

Fregni, F., Orsati, F., Pedros, W., Fectea, S., Tome, F. A., Nitsch, M. A., Mecca, T., Macedo, E., Pascual-Leon, A., & Boggio, P. S. (2008). Transcranial direct current stimulation of the prefrontal cortex modulates the desire for specific foods. *Appetite, 51*, 34–41.

Fumagalli, M., Vergari, M., Pasqualetti, P., Marceglia, S., Mameli, F., Ferrucci, R., Mrakic-Sposta, S., Zago, S., Sartori, G., Pravettoni, G., Barbieri, S., Cappa, S., & Priori, A. (2010). Brain switches utilitarian behavior: does gender make the difference? *PloS One, 5*(1), e8865.

Gozli, D. G., Moskowitz, J. B., & Pratt, J. (2014). Visual attention to features by associative learning. *Cognition, 133*, 488–501.

Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: a meta-analytical review. *Experimental Brain Research, 216*, 1–10.

Jasper, H. H. (1958). The ten–twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.

Karim, A. A., Schneider, M., Lotze, M., Veit, R., Sauseng, P., Braun, C., & Braibaumer, N. (2010). The truth about lying: inhibition of the anterior refrontal cortex improves deceptive behavior. *Cerebral Cortex, 20*, 205–213.

Knutson, B., & Bossaerts, P. (2007). Neural antecedents of financial decisions. *The Journal of Neuroscience, 27*, 8174–8177.

Kringelbach, M. L., & Rolls, E. T. (2003). Neural correlates of rapid context-dependent reversal learning in a simple model of human social interaction. *NeuroImage, 20*, 1371–1383.

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology, 72*, 341–372.

Kringelbach, M. L., O'Doherty, J., Rolls, E. T., & Andrews, C. (2003). Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cerebral Cortex, 13*, 1064–1071.

Kringelback, M. L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews. Neuroscience, 6*, 691–702.

Levasseur-Moreau, J., & Fecteau, S. (2012). Translational application of neuromodulation of decision-making. *Brain Stimulation, 5*, 77–83.

Liu, X., Powell, D. K., Wang, H., Gold, B. T., Corbly, C. R., & Joseph, J. E. (2007). Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. *The Journal of Neuroscience, 27*, 4587–4597.

Loo, C. K., Alonzo, A., Martin, D., Mitchell, P. B., Galvez, V., & Sachdev, P. (2012). Transcranial direct current stimulation for depression: 3-week, randomised, sham-controlled trial. *British J Psychiatry, 20*, 52–59.

Ly, V., Bergmann, T. O., Gladwi, T. E., Volman, I., Usberti, N., Cools, R., & Roelofs, K. (2016). Reduced affective biasing of instrumental action with tDCS over the prefrontal cortex. *Brain Stimulation, 9*, 380–387.

Manuel, A. L., David, A. W., Bikson, M., & Schinder, A. (2014). Frontal tDCS modulates orbitofrontal reality filtering. *Neuroscience, 265*, 21–27.

Marin, M. F., Camprodon, J. A., Dougherty, D. D., & Mohammed, R. M. (2014). Device-based brain stimulation to augment fear extinction: implications for PTSD treatment and beyond. *Depression and Anxiety, 31*, 269–278.

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology, 57*, 1899–1901.

Nitsche, M. A., Liebetanz, D., Anta, A., Lang, N., Tergau, F., & Paulus, W. (2003). Modulation of cortical excitability by weak direct current stimulation—technical, safety and functional aspects. *Supplements to Clinical Neurophysiology, 56*, 255–276.

Noonan, M. P., Walton, M. E., Behrens, T. E., Sallet, J., Buckley, M. J., & Rushworth, M. F. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 20547–20552.

Noonan, M. P., Kolling, N., Walton, M. E., & Rushworth, M. F. (2012). Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. *The European Journal of Neuroscience, 35*, 997–1010.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*, 95–102.

Ott, D. V., Ullsperger, M., Jocham, G., Neumann, J., & Klein, T. A. (2011). Continuous theta-burst stimulation (cTBS) over the lateral prefrontal cortex alters reinforcement learning bias. *NeuroImage, 57*, 617–623.

Reato, D., Gasca, D., Datta, A., Bikson, M., Marshall, L., & Parra, L. C. (2013). Transcranial electrical stimulation accelerates human sleep homeostasis. *PLoS Computational Biology, 9*(2), e1002898. doi:10.1371/journal.pcbi.1002898.

Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Progress in Neurobiology, 86*, 216–244.

Rolls, E. T., Hornak, J., Wade, D., & McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery, and Psychiatry, 57*, 1518–1524.

Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Choice, difficulty, and confidence in the brain. *NeuroImage, 53*, 694–706.

Rothermund, K., Wentura, D., & De Houwer, J. (2015). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 3*, 482–495.

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron, 70*, 1054–1069.

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *The Journal of Neuroscience, 27*, 4826–4831.

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C., & Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: from pleasure to aversion. *Brain, 124*, 1720–1733.

Tsuchida, A., Doll, B. D., & Fellows, K. F. (2010). Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *The Journal of Neuroscience, 30*, 16868–16875.

Turi, Z., Mittner, M., Opitz, A., Popkes, M., Paulus, W., & Antal, A. (2015). Transcranial direct current stimulation over the left prefrontal cortex increases randomness of choice in instrumental learning. *Cortex, 63*, 145–154.

Volpato, C., Piccione, F., Cavinato, M., Duzzi, D., Schiff, S., Foscolo, L., & Venneri, A. (2013). Modulation of affective symptoms and resting state activity by brain stimulation in a treatment-resistant case of obsessive-compulsive disorder. *Neurocase, 19*, 360–370.

Volpato, C., Schiff, S., Facchini, S., Silvoni, S., Cavinato, M., Piccione, F., Antonini, A., & Birbaumer, N. (2016). Dopaminergic medication modulates learning from feedback and error-related negativity in Parkinson's disease: a pilot study. *Frontiers in Behavioral Neuroscience*. doi:10.3389/fnbeh.2016.00205.

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience (Palo Alto, CA), 30*, 31–56.

Walton, M. E., Behrens, T. E., Noonan, M. P., & Rushworth, M. F. (2011). Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world. *Annals of the New York Academy of Sciences, 1239*, 14–24.

Wheeler, E., & Fellows, L. K. (2008). The human ventromedial frontal lobe is critical for learning from negative feedback. *Brain, 131*, 1323–1331.