Transcranial direct current stimulation modulates confirmation bias in instructed reinforcement learning

Nathan Tardiff & Sharon L. Thompson-Schill

Department of Psychology, University of Pennsylvania

BACKGROUND

- Reinforcement learning (RL) algorithms accord well with the functioning of neurobiological substrates of reward learning [1]. RL algorithms generally assume learning proceeds in an unbiased fashion based on reward prediction error. However, it has been recently shown that human subjects sometimes exhibit a form of confirmation bias in RL tasks. When given advice or instructions, subjects are biased toward following the instructions even when reward contingencies call them into question [2,3,4].
- Neuroimaging, genetic analyses, and computational modeling suggest a role for PFC in biasing instructed RL [4,5,6].

HYPOTHESES

If PFC is responsible for biasing instructed RL, the level of bias should vary with experimental manipulation of PFC function via transcranial direct current stimulation (tDCS).

- **Anodal stimulation should strengthen the bias by upregulating PFC.**
- **Cathodal stimulation should reduce the bias the downregulating PFC.**

METHODS

Subjects

N=52 (17 anodal, 18 cathodal, 17 sham) in between-subjects design
Performance cutoffs excluded 11 out of an initial sample of 63

Instructed probabilistic selection task

Training (feedback) 4 blocks, 20 of each training pair per block	カ ✓	ポ
Test (no feedback) 1 block, all possible pairings (AB, AC, AD, etc.) presented 6 times each	ゴ	セ

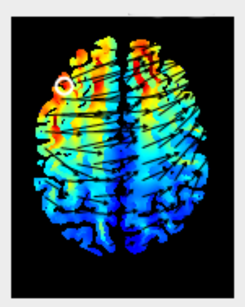
Training Pairs

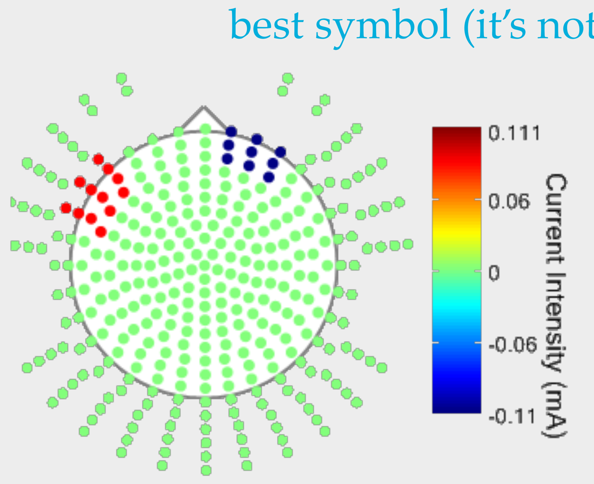
Uninstructed	カ A (80/20)	ポ B (20/80)
Instructed	ゴ C (60/40)	セ D (40/60)
Uninstructed	ネ E (60/40)	バ F (40/60)

Instruction: D is the best symbol (it's not!)

Stimulation

- F7-RSO
- 1mA
- Parameters chosen via current modeling to maximize current to DLPFC sites implicated in instructed RL
- Stimulation during training only (20 minutes)
- 30s ramp-up/ramp-down
- Sham: 30s stimulation





Analysis

- Mixed-effects (logistic) regression with random effects for all within-subjects factors
- Condition contrasts: anodal vs. sham, cathodal vs. sham

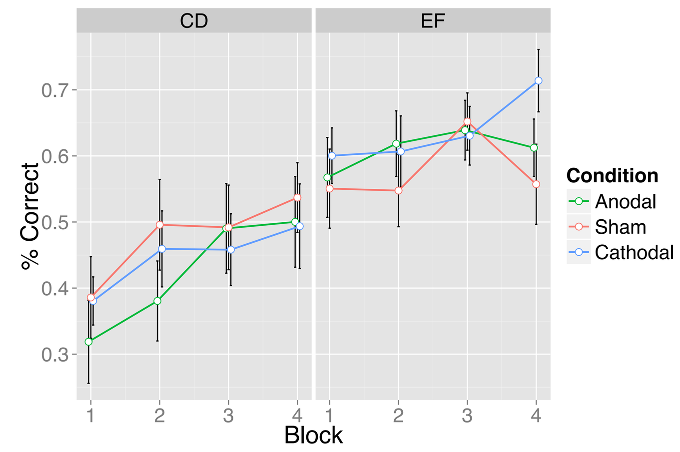
RESULTS

Training Phase

Instructed learning (CD vs. EF)

First block -> Last Block

- Confirmation bias effect replicated: Below-chance performance on instructed CD pair (z=−2.60, p=.009).
- Performance better on equivalently rewarded uninstructed EF pair (z=5.09, p<.001).
- No effect of stimulation (all ps>.12).

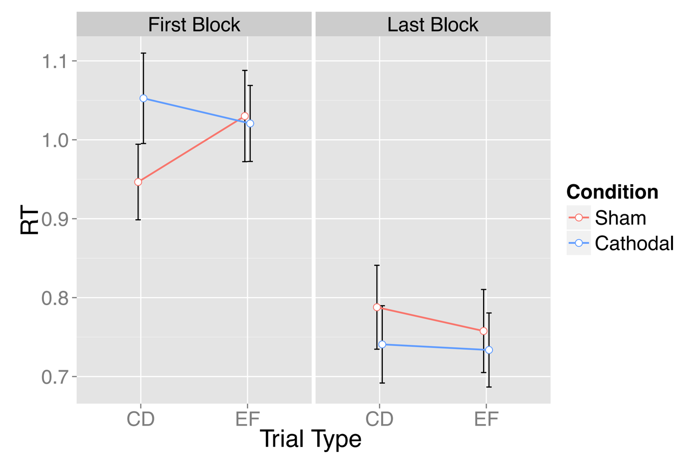


Learning Curve Analysis

- Anodal vs. Sham: Anodal more cubic learning on CD (z=−1.78, p=.07) , reflecting initial lower performance then rapid improvement.
- Anodal vs. Sham: Significant cubic trend when contrasting CD with EF (z=2.24, p=.03).

Reaction Time Analysis

- Slower responding on EF vs. CD (B=0.02, *t*(51.8)=1.84, p=.07) and greater speed-up from first block to last (B=−0.02, *t*(51.6)=−1.89, p=.06), reflecting effect of instruction on choice.
- Cathodal vs. Sham: No instruction effect for Cathodal. Condition x Trial Type x Block interaction (B=0.07, *t*(51.9)=2.45, p=.02). Cathodal reduced RT difference in first block on EF vs. CD (B=−0.11, *t*(51.8)=−2.32, p=.02). No interaction in last block (p>.96).
- No effect of stimulation on uninstructed trials (all ps>.14)

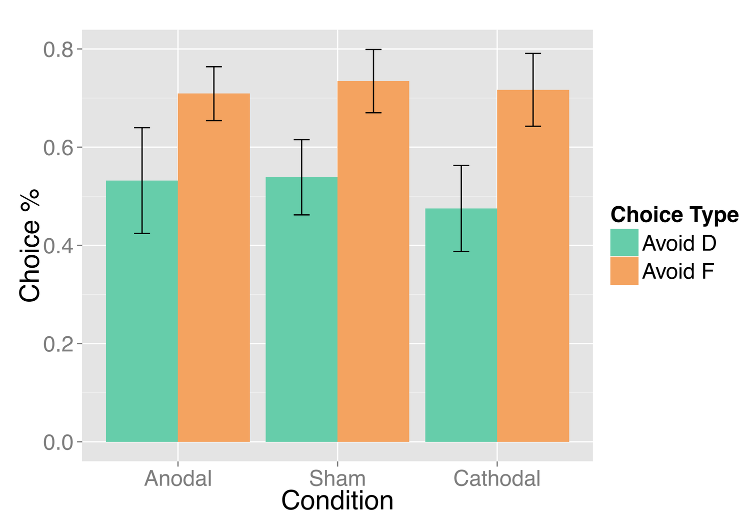


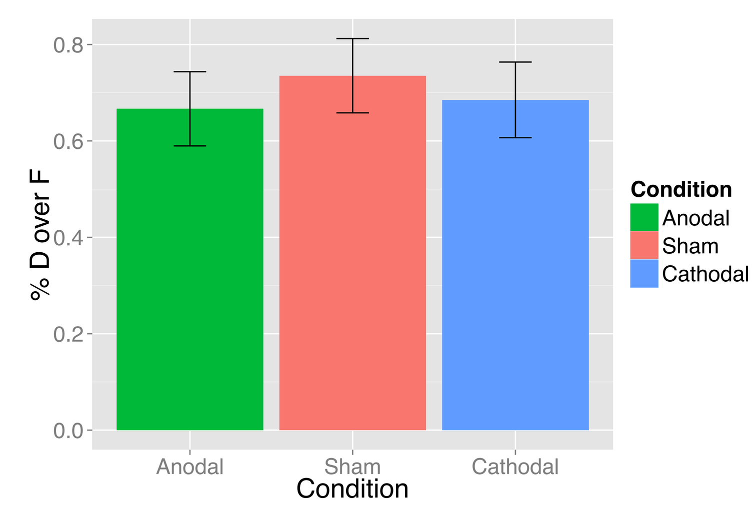
Uninstructed learning (AB, EF)

First block -> Last Block

- All groups learn to discriminate uninstructed pairs (z=8.66, p<.001) and learn across training (z=3.34, p<.001).
- No effect of stimulation (all ps>.13).

Test Phase





- Confirmation bias effect replicated. Avoid D < Avoid F (both rewarded 40/60) when paired with symbols not paired with during training (z=−3.59, p<.001). Favor D over F in head-to-head comparison (z=3.96, p<.001).
- No effect of stimulation (all ps>.44).

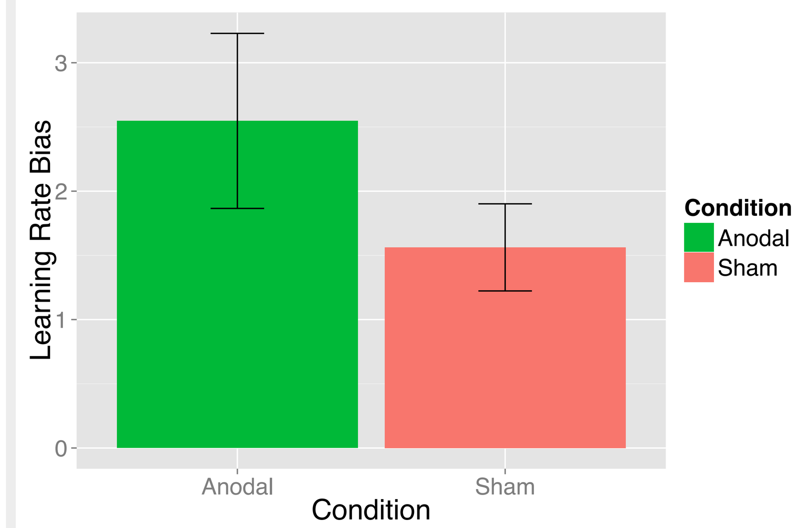
RESULTS

Modeling

$$Q_{t+1}(s) = Q_t(s) + \alpha*[r - Q_t(s)]$$

Q-learning models were augmented with various combinations of bias parameters (distortion of learning rate for D) or prior parameter (fitting initial Q value for D).

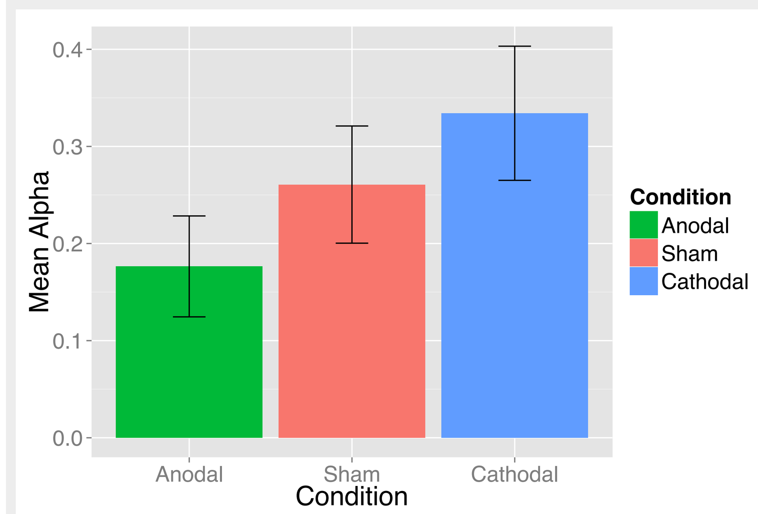
- Majority of subjects (36/52) best fit by an instructed model.
- Fit varied by group.

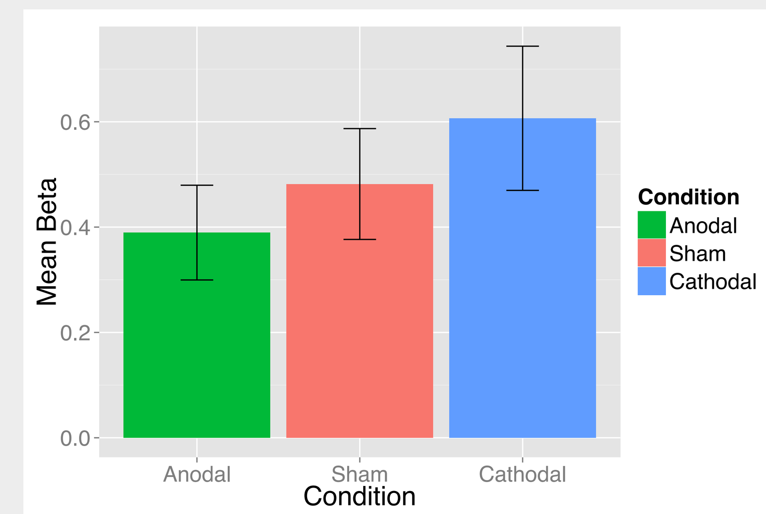


Group	Model	AIC Weight	BIC Weight
Anodal	Bias + Prior	1.0	0.0
	Prior	0.0	1.0
	Standard	0.0	0.0
Cathodal	Bias	.52	1.0
	Change Bias	.39	0.0
	Standard	0.0	0.0
Sham	Bias + Prior	.68	0.0
	Prior	.32	1.0
	Standard	0.0	0.0

Note: Only best fitting models for each group + Standard model shown.

- Elevated bias for anodal vs. sham (Bias + Prior model: W=205, p=.04).
- No effect on prior (p>.86).





- Averaged across models (top 3 + standard) clear ordering of learning rate and decision noise (temperature) by group.

CONCLUSIONS

- Results support hypotheses. Effects on learning were largely specific to instruction and support a role for PFC in biasing learning. This is consistent with a more general framework that posits that frontally-mediated control processes can have costs, as well as benefits, for learning [7].
- Accuracy and modeling point to increased bias under anodal stimulation.
- RT and modeling point to decreased role of prior for cathodal but intact bias, suggesting possibly separable mechanisms.
- Stimulation effects did not persist at test, indicating limited effect on reward learning.
- Future work should probe learning rate and decision noise modulations by tDCS in environments that would highlight these effects (e.g. explore/exploit tasks).

ACKNOWLEDGEMENTS

Funding for this research was provided by NIH grant #R01DC009209.

CONTACT

Thompson-Schill Lab >>
ntardiff@sas.upenn.edu

REFERENCES

1. Niv (2009)
2. Biele, Rieskamp, & Gonzalez (2009)
3. Delgado, Frank, & Phelps (2005)
4. Doll, Hutchinson, & Frank (2011)
5. Fouragnan et al. (2013)
6. Li, Delgado, & Phelps (2011)
7. Thompson-Schill, Ramscar, & Chrysikou (2009)