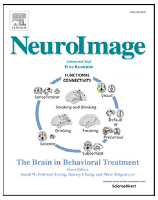
NeuroImage 184 (2019) 25–35

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



Inhibitory and excitatory mechanisms in the human cingulate-cortex support reinforcement learning: A functional Proton Magnetic Resonance Spectroscopy study

Vered Bezalel [a,b], Rony Paz [a,*,1], Assaf Tal [b,*,1]

[a] Department of Neurobiology, Weizmann Institute of Science, Israel
[b] Department of Chemical and Biological Physics, Weizmann Institute of Science, Israel

ARTICLE INFO

Keywords:
Anterior cingulate cortex
fMRS
GABA
E/I-balance
Reinforcement learning

ABSTRACT

The dorsal anterior cingulate cortex (dACC) is crucial for motivation, reward- and error-guided decision-making, yet its excitatory and inhibitory mechanisms remain poorly explored in humans. In particular, the balance between excitation and inhibition (E/I), demonstrated to play a role in animal studies, is difficult to measure in behaving humans. Here, we used functional magnetic-resonance-spectroscopy (^{1}H-fMRS) to measure the brain's major inhibitory (GABA) and excitatory (Glutamate) neurotransmitters during reinforcement learning with three different conditions: high cognitive load (uncertainty); probabilistic discrimination learning; and a control null-condition. Participants learned to prefer the gain option in the discrimination phase and had no preference in the other conditions. We found increased GABA levels during the uncertainty condition, potentially reflecting recruitment of inhibitory systems during high cognitive load when trying to learn. Further, higher GABA levels during the null (baseline) condition correlated with improved discrimination learning. Finally, glutamate and GABA levels were correlated during high cognitive load. These results suggest that availability of dACC inhibitory resources enables successful learning. Our approach helps elucidate the potential contribution of the balance between excitation and inhibition to learning and motivation in behaving humans.

1. Introduction

Studies in animals have highlighted the importance of excitation and inhibition for reinforcement learning (Johansen et al., 2011; Kelley, 2004; Myhrer, 2003). The balance between them (E/I balance) is critical and maintained under most conditions, yet the exact ratio is highly dynamic (Isaacson and Scanziani, 2011; Treviño, 2016), and variations in this ratio support information processing and learning (Isaacson and Scanziani, 2011; Letzkus et al., 2011). Although it is much harder to assess excitation and inhibition in humans, the main contributors: Glutamate and GABA, can be measured through Proton Magnetic Resonance Spectroscopy (^{1}H-MRS) (Bottomley, 1987; Mescher et al., 1998; Paul G Mullins et al., 2014). The exact interpretation of MRS-observed neurotransmitter levels remains an open question (Mangia et al., 2007a,b; Stagg et al., 2011). However, both Glutamate and GABA have been shown to reflect task-related activity with studies demonstrating sensitivity to baseline (Donahue et al., 2010; Jocham et al., 2012; Levar et al., 2017; Muthukumaraswamy et al., 2009; Northoff et al., 2007; Yoon et al., 2016) and rapidly-modulating levels (Floyer-Lea et al., 2006; Hasler, van der Veen, Grillon, Drevets and Shen, 2010; Mangia et al., 2007a,b). Most studies measured concentrations during rest and correlated it with later/previous behavior (Boy et al., 2010; Muthukumaraswamy et al., 2009; Northoff et al., 2007; Schmitz et al., 2017; Sumner et al., 2010; Yoon et al., 2016), yet some even measured during behavior (Hasler et al., 2010; Mangia et al., 2007a,b; Michels et al., 2012). A few studies have correlated baseline (rest) levels with subsequent/prior learning metrics (Jocham et al., 2012; Levar et al., 2017; Sampaio--Baptista et al., 2015), and one has even examined changes during motor learning (Floyer-Lea et al., 2006).

Error-based learning involves the dorsal-anterior-cingulate-cortex (dACC), which mediates motivation, cognition and action (Dayan and Niv, 2008; Haber and Knutson, 2010; Heilbronner and Hayden, 2016; Kennerley et al., 2006; Lee et al., 2012). The dACC plays a crucial role in reward-guided decision making, as it promotes action-outcome

---

* Corresponding authors. 234 Herzel St., Rehovot 7610001, Israel.
E-mail addresses: rony.paz@weizmann.ac.il (R. Paz), assaf.tal@weizmann.ac.il (A. Tal).

[1] These authors contributed equally to the work.

https://doi.org/10.1016/j.neuroimage.2018.09.016
Received 12 May 2018; Received in revised form 10 August 2018; Accepted 7 September 2018

associations and monitors goal-directed behaviors (Kennerley et al., 2006; Kolling et al., 2016). In particular, dACC activation is modulated by requirements for cognitive control (Ridderinkhof et al., 2004; Shenhav et al., 2016; Sheth et al., 2012), and is involved in monitoring choice outcome in uncertain environments (Amiez et al., 2005; Behrens et al., 2007; Payzan-LeNestour et al., 2013), as well as biases decisions that require high mental effort (Croxson et al., 2009; Prévost et al., 2010; Rudebeck et al., 2006). Its overall involvement in error detection and processing, as well as conflict monitoring, further link it to learning processes (Botvinick, 2007). There have been several studies which have examined the neurometabolic changes in the dACC during interference and Stroop tasks (Kuhn et al., 2016; Taylor et al., 2015); and the functional metabolic correlates of learning in hippocampal and striatal systems (Bell et al., 2018; Stanley et al., 2017). However, the contribution of excitation-inhibition mechanisms to learning processes in the human ACC remains poorly understood.

In the current study, we used functional ^{1}H-MRS (^{1}H-fMRS) during reinforcement learning in humans, and measured modulations in dACC levels of GABA and Glx while participants engaged in a learning task that compared three factors: full uncertainty (high-cognitive-load); probabilistic discrimination learning, and an active Null condition. Based on prior studies which demonstrated recruitment of the dACC during

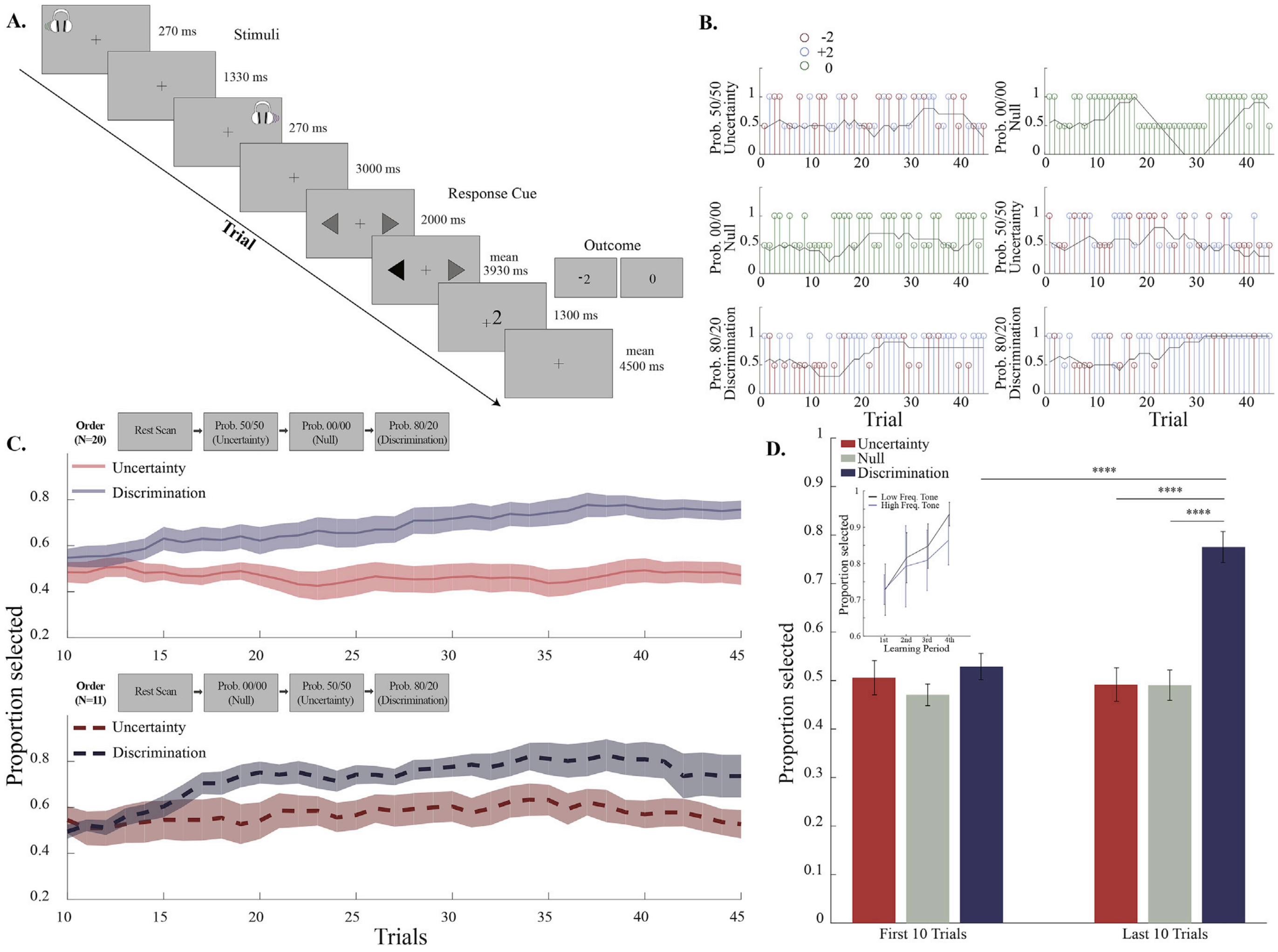


Fig. 1. Experimental design and behavioral results.

A. In each trial, participants were exposed to two pure tones that were played out in succession to different ears. The response cue, presented by two gray opposing arrows, instructed to choose between the tones (left or right button press). Following selection, the arrow corresponding to the chosen laterality was blackened, and the outcome screen indicated monetary gain, loss or neutral outcome (+2,-2 or 0).

B. Behavior of single two participants (rows), in the order of exposure to each of the 3 conditions (rows, top to bottom). The scanning session consisted of three scanning blocks, each attributed to one behavioral condition: 50/50 probability to lose or gain 2₪ ("uncertainty condition"), 00/00 probability with consistent 0₪ reward ("Null condition"), and 80/20 probability to lose or gain 2₪ ("discrimination condition"). Blue, red and green represent gain, loss and zero reward respectively. Taller stems represent selection of one tone, while shorter stems represent selection of the other tone (1 and 0.5 in the y axis respectively). The black line is the probability to choose the tone designated as the gain-tone in the 80/20 condition, averaged over a 10-trial moving window.

C. Tone selection probability during the uncertainty and discrimination conditions, presented separately for the two ordering of conditions (mean ± shaded SEM). Top: average across participants that were exposed to the first ordering (n = 20): MRS rest scan, followed by the uncertainty, Null, and discrimination conditions. Bottom: average across participants that were exposed to the second ordering of conditions (n = 11): MRS rest scan followed by the Null, uncertainty, and discrimination conditions. The graphs demonstrate a gradual increase in the high-gain-tone ('better-option') selection probability during the discrimination condition, irrespective of the ordering, and no selection preference during the uncertainty condition (see text for statistics).

D. High gain tone selection probability averaged across the first and last 10 trials in each condition (mean ± SEM). Participants chose the high gain tone more often in the last part of the discrimination condition compared to its first part, as well as to the last parts of the two other conditions (****$p < 0.0001$, post-hoc Tukey-Kramer). Inset presents pretest results (n = 10) showing that learning was similar regardless of whether a high frequency (light blue) or low frequency tone served as the better-option (Wilcoxon $p > 0.05$ in each one of the four points comparisons).