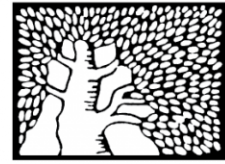


מכון ויצמן למדע

WEIZMANN INSTITUTE OF SCIENCE



Inhibitory and excitatory mechanisms in the human cingulate-cortex support reinforcement learning

Document Version:

Accepted author manuscript (peer-reviewed)

Citation for published version:

Bezalel, V, Paz, R & Tal, A 2019, 'Inhibitory and excitatory mechanisms in the human cingulate-cortex support reinforcement learning: A functional Proton Magnetic Resonance Spectroscopy study', *NeuroImage (Orlando, Fla.)*, vol. 184, pp. 25-35. <https://doi.org/10.1016/j.neuroimage.2018.09.016>

Total number of authors:

3

Digital Object Identifier (DOI):

[10.1016/j.neuroimage.2018.09.016](https://doi.org/10.1016/j.neuroimage.2018.09.016)

Published In:

NeuroImage (Orlando, Fla.)

License:

CC BY-NC

General rights

@ 2020 This manuscript version is made available under the above license via The Weizmann Institute of Science Open Access Collection is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognize and abide by the legal requirements associated with these rights.

How does open access to this work benefit you?

Let us know @ library@weizmann.ac.il

Take down policy

The Weizmann Institute of Science has made every reasonable effort to ensure that Weizmann Institute of Science content complies with copyright restrictions. If you believe that the public display of this file breaches copyright please contact library@weizmann.ac.il providing details, and we will remove access to the work immediately and investigate your claim.

Inhibitory and excitatory mechanisms in the human cingulate-cortex support reinforcement learning

Vered Bezalel^{1,2}, Rony Paz^{1*}, Assaf Tal^{2*}

¹ Department of Neurobiology, Weizmann institute of Science, Israel

² Department of Chemical Physics, Weizmann institute of Science, Israel

* These authors contributed equally to the work

correspondence should be addressed to R.P. or A.T.

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 magnetic-resonance-spectroscopy (¹H-MRS) to examine these mechanisms during reinforcement learning with three different conditions: high cognitive load (uncertainty); probabilistic discrimination learning; and a control null-condition. Subjects 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, suggesting 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, excitatory and inhibitory levels were correlated during high cognitive load. The result suggests that availability of dACC inhibitory resources enables successful learning. Our approach establishes a novel way to examine the contribution of the balance between excitation and inhibition to learning and motivation in behaving humans.

Introduction

Studies in animals have highlighted the importance of excitation and inhibition for reinforcement learning[1-3]. The balance between them (E/I balance) is critical and maintained under most conditions, yet the exact ratio is highly dynamic[4, 5], and variations support information processing and learning[4, 6]. Although it is much harder to assess excitation and inhibition in humans, the main contributors: Glutamate and GABA, can be accurately and reliably measured through Proton Magnetic Resonance Spectroscopy (¹H-MRS) [7-9]. Indeed, MRS-observed neurotransmitter levels reflect task-related activity with studies demonstrating sensitivity to baseline [10-15] and rapidly-modulating levels [16-18]. Most studies measured concentrations during rest and correlated it with later/prior behavior[13-15, 19-21], yet some even measured during behavior[17, 18, 22]. A few studies have correlated baseline (rest) levels with subsequent/prior learning metrics[11, 12, 23], and one has even examined changes during motor learning[16]. However, none have quantified neurotransmitter modulations during active reinforcement-learning.

Error-based learning involves the dorsal-anterior-cingulate-cortex (dACC), which mediates motivation, cognition and action[24-28]. The dACC plays a crucial role in reward-guided decision making, as it promotes action-outcome associations and monitors goal-directed behaviors [27, 29]. In particular, dACC activation is modulated by requirements for cognitive control [30-32], and is involved in monitoring choice outcome in uncertain environments [33-35], as well as biases decisions that require high mental effort [36-38]. However, the contribution of excitation-inhibition mechanisms to these functions in the human ACC remains poorly understood.

We used ¹H-MRS 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. We hypothesized that levels of neurotransmitters during the high-cognitive-load, and potentially during active-Null conditions, would predict and enable successful discrimination learning.

Results

A two-alternative forced choice task was conducted using a block design paradigm while subjects ($n=31$) laid in an MRI scanner (Fig.1A). During each trial, subjects heard two pure-frequency tones, each to a different ear, in pseudorandom order. Following a visual go-signal comprised of two arrows, the subjects chose a tone by pressing right or left button. The outcome was one of three: a monetary gain (+2), a loss (-2), or neutral feedback (0). There were three experimental blocks: (1) the Null condition where the subjects received neutral (0) outcome, independent of their choice; (2) the uncertainty condition, 50/50 probability for loss/gain, independent of the participant's choice; and (3) the probabilistic discrimination condition, in which one of the tones was paired with a monetary gain in 80% of the trials and monetary loss in 20%, and vice versa for the other tone. All the subjects underwent a rest scan before the experimental blocks. Additionally some subjects ($n= 18$) also underwent a rest scan after the experimental blocks.

Learning occurs under probabilistic discrimination, but not under uncertainty or Null

Over the course of the discrimination block, subjects gradually increased their preference to the tone that was associated with higher gain ('better-option'), such that by the end of this block, the mean selection probability exceeded 70% (Fig.1B,C). In comparison, no preference to either one of the options was observed during the other two conditions, in which the subjects presented a stable mean probability of 0.5 ± 0.04 (uncertainty condition) and 0.48 ± 0.03 (null condition) to select each option (Fig.1B,C). This was the case for both groups of subjects, when the uncertainty condition was first (Fig.1B-left, Fig.1C-top, $n=20$), and when the Null condition was first (Fig.1B-right, Fig.1C-bottom, $n=11$).

The preference to the better-option in the discrimination condition was also reflected by the averaged selection probability during two epochs, namely, the first and last 10 trials of the block (Fig. 1D). A difference in selection probability between conditions was observed ($p<0.00001$, $F(2, 60) = 18.3$, two way repeated measures ANOVA; mean selection probability of the first and second epochs: 0.5, 0.48 and 0.65, in uncertainty, null and discrimination conditions, respectively), arising from a significant preference to the better-option in the discrimination condition when compared to the uncertainty and null conditions ($p<0.001$ and $p<0.00001$ respectively, post hoc Tukey-Kramer). Additionally, we also observed a difference between the first and second epochs ($p<0.001$, $F(1, 30) = 19.2$, two way repeated measures ANOVA) that originated from a preference to the better-option in the second epoch ($p<0.001$, post hoc Tukey-Kramer). An interaction between these effects was observed ($p<0.0001$, $F(2, 60) = 12.2$, two way repeated measures ANOVA), deriving from the increased preference to the higher gain

tone in the second epoch of the discrimination condition compared to its first epoch ($p < 0.00001$, post hoc Tukey-Kramer). Examination of the performance in the second epoch revealed a preference to the high-gain-tone in the discrimination condition compared to the null and uncertainty conditions (both $p < 0.00001$ post hoc Tukey-Kramer). However, there was no difference in the average selection probability between the conditions in the first epochs of the blocks ($p > 0.05$, post hoc Tukey-Kramer).

We conclude that subjects learned the task in the probabilistic discrimination condition, behaved at chance level in the Null condition (no outcomes), and also behaved at chance level under uncertainty (when rewards and punishments occur, but there is no reliable information to learn, leading to a high-cognitive-load).

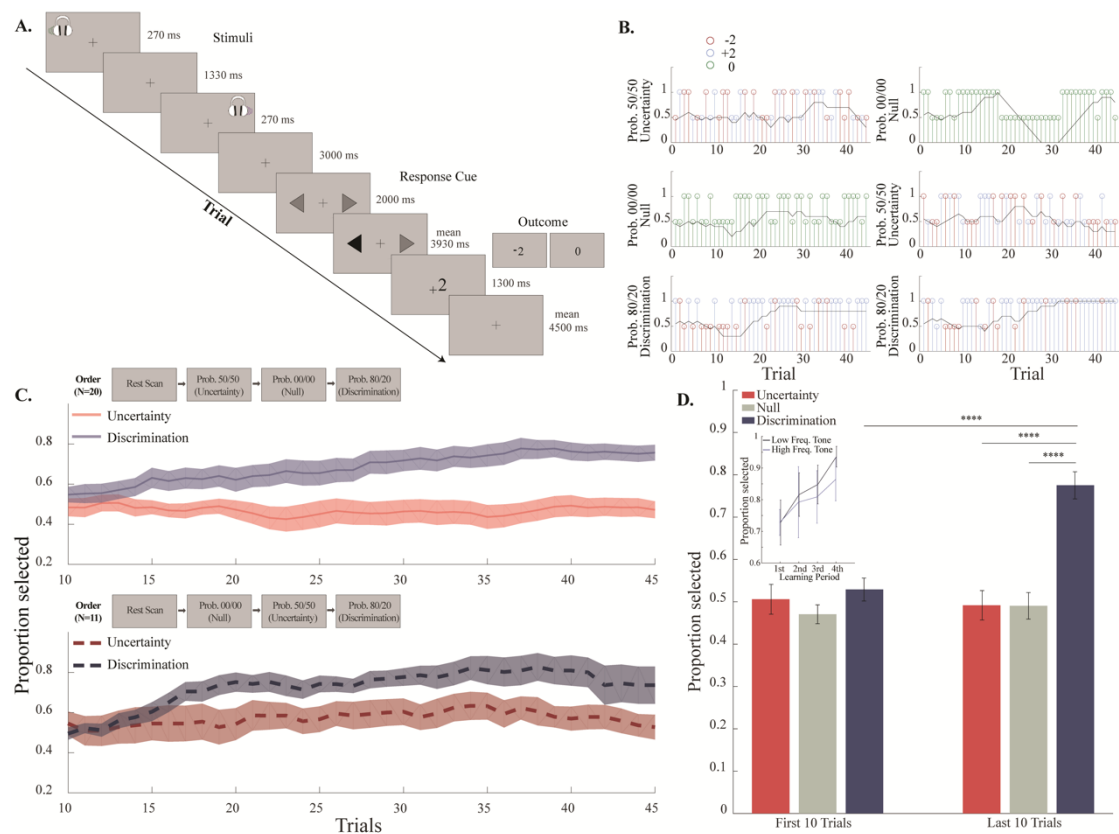


Figure 1. Experimental design and behavioral results **A.** In each trial, subjects were exposed to two pure tones that were played out in succession to different ears. The response cue, presented by two gray opposing arrows, instructed to choose between the tones (left or right button press). Following 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 subjects (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 μ (“uncertainty condition”), 00/00 probability with consistent 0 μ reward (“Null condition”), and 80/20 probability to lose or gain 2 μ (“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 \pm shaded SEM). Top: average across subjects that were

exposed to the first ordering (n=20): MRS rest scan, followed by the uncertainty, Null, and discrimination conditions. Bottom: average across subjects 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).

Measuring neurotransmitter concentrations in the dACC using ¹H-MRS

The MRS technique requires long scanning sessions in order to achieve good signal to noise ratio. Therefore, we defined the region-of-interest (the dACC) a-priori based on anatomical maps (Fig.2A). Please see methods for full description of how we measure and quantify GABA and Glx concentrations (Fig.2B). In addition, we conducted two tests to assess whether results might be confounded by the order of experimental conditions and/or the stability of metabolite levels over the long scanning time.

First, we assessed putative temporal metabolic drift by examining metabolites levels during the rest scan, which is the only scan type that was not influenced by task manipulation but could be affected by time. Metabolites levels were compared in subjects that performed the rest scan block twice: at the beginning and at the end of the experiment (Fig.2C, n=18). There was no difference between the levels of GABA/Water ($t(17) = 0.46$, $p > 0.05$, paired sample t-test) or Glx/Water ($t(17) = 1.06$, $p > 0.05$, paired sample t-test). We conclude that the time did not influence metabolite levels and is not a major contributor to the results.

To further validate that the order of conditions was not a factor, we had two groups of subjects, one performed the uncertainty before the Null condition (n=20), and the other performed the Null condition before the uncertainty (n=11), and compared levels of GABA and Glx during the Null condition in the two experimental orders. Because we already established that metabolite levels during rest are stable from start to end, we standardized GABA and Glx levels during the Null condition to the initial rest level (Eq. (1)). We found no difference (Fig.2D) between the signals of GABA (Wilcoxon $p > 0.05$, $Z = 0.14$) or Glx (Wilcoxon $p > 0.05$, $Z = 0.56$).

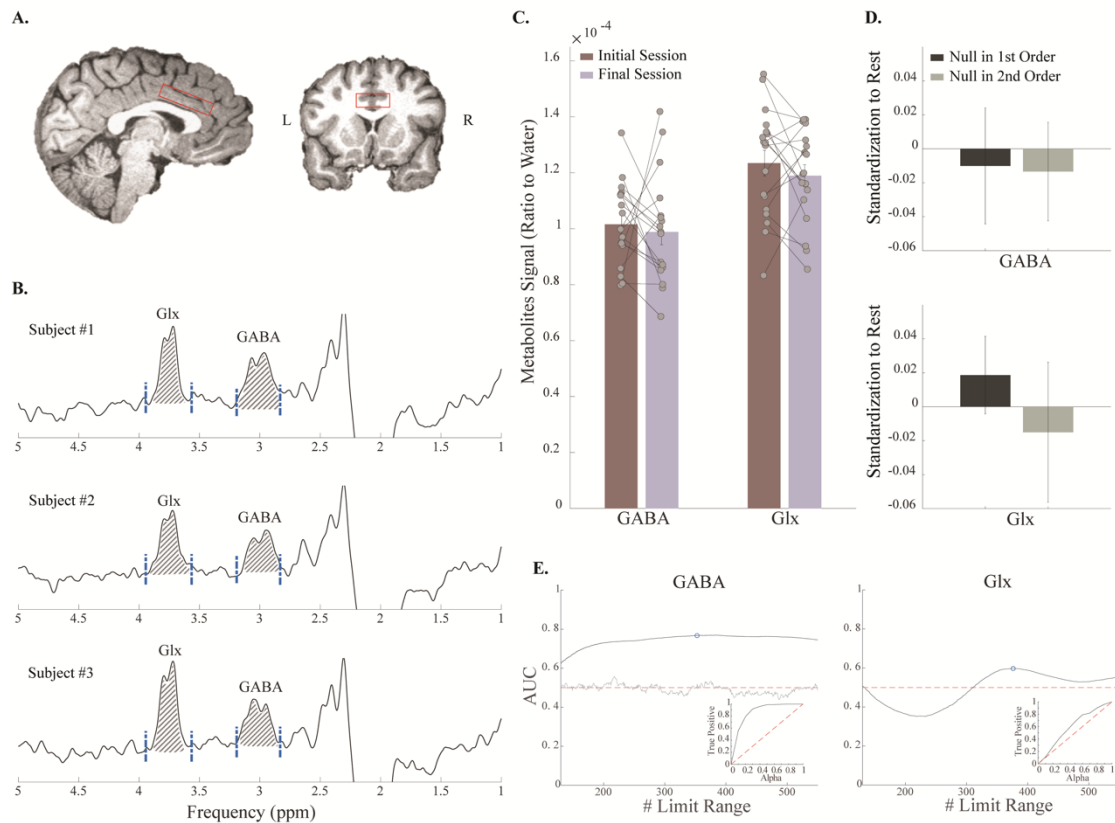


Figure 2. MRS measurements of GABA and Glx concentrations are stable. **A.** A $40 \times 25 \times 10 \text{ mm}^3$ voxel was placed on the midline of the dACC, chosen anatomically, presented in sagittal (left) and coronal (right) T1-weighted anatomical MPRAGE scans. **B.** Sample spectra from three different subjects (144 averages, 9:57 minutes), displaying the peaks of GABA and Glx at 3.01 and 3.75 ppm respectively. Blue dashed lines represent the integration limits that were used to quantify each metabolite. The integrated area under the curve, used for quantification, is illustrated by diagonal lines. **C.** Eighteen subjects went through an additional final session of rest scan after completion of the task. Comparison between metabolites levels in the initial and in the final rest scans, revealed no difference over time in GABA/Water and Glx/Water levels ($p > 0.05$, paired sample t-test). Levels of metabolites in single subjects are represented by connected dots (mean \pm SEM). **D.** Order of experimental conditions did not affect metabolite levels in the Null condition. Compared between the first ($n=20$) and the second ($n=11$) orderings for both GABA and Glx standardized to rest (Wilcoxon $p > 0.05$). (mean \pm SEM). **E.** Chosen integration limits for each metabolite were validated to be robust by an optimization process. We varied the limits and used a ROC curve analysis (insets) to choose the best limits. Shown are examples of the area-under-curve (AUC) for each pair of tested integration limits, with the best integration limits for each metabolite circled in blue. The grey line is based on simulated random noises as control. The results are highly similar to the limits chosen traditionally (usually based on visual considerations alone).

Increased GABA during uncertainty condition only

Differences between the conditions of the task were found when examining GABA/Water and Glx/GABA levels (Fig. 3A; $p < 0.05$ with $F(2, 60) = 3.6$ and $F(2, 60) = 4.1$ respectively, one way repeated measures ANOVA). These differences originated from elevated GABA/Water and reduced Glx/GABA levels during the uncertainty condition compared to the discrimination condition ($p < 0.05$, post hoc Tukey-Kramer). Glx/GABA displayed a marginal non-significant trend for the difference between the Null condition to the uncertainty condition ($p = 0.052$, post hoc Tukey-Kramer), but not for GABA/Water levels ($p > 0.1$, post hoc Tukey-Kramer). No

differences were found for GABA/water and Glx/GABA between the Null and discrimination conditions ($p > 0.1$, post hoc Tukey-Kramer). Finally, Glx/Water did not exhibit any differences between the conditions ($p > 0.1$, $F(2, 60) = 0.5$, one way repeated measures ANOVA).

For robustness, we repeated these comparisons after standardization to rest levels, for Glx, GABA and Glx/GABA (Eq. (11)). This yielded similar results to the ones obtained using the raw data, mainly presented a significant difference between the uncertainty to the discrimination condition in GABA and in Glx/GABA levels. When standardized to rest, GABA and Glx/GABA levels were found to vary significantly between the uncertainty and discrimination conditions (Fig. 3B; $p < 0.05$ with $F(2, 60) = 3.5$ and $F(2, 60) = 3.8$ respectively, repeated measures one way ANOVA; $p < 0.05$, post hoc Tukey-Kramer). A non-significant trend was exhibited by Glx/GABA when comparing the uncertainty and Null conditions ($p = 0.086$, post hoc Tukey-Kramer), but no differences were found when examining GABA ($p > 0.05$, post hoc Tukey-Kramer). Additionally, no difference was found between the Null condition to the discrimination condition in GABA level or in Glx/GABA ($p > 0.05$, post hoc Tukey-Kramer). No difference was found between the conditions in the Glx levels ($p > 0.05$, $F(2, 60) = 0.6$, one way repeated measures ANOVA).

Furthermore, we standardize directly between conditions (see methods), and used one sample t-test with Bonferroni adjusted alpha level of 0.015 (0.05/3). The comparison between the uncertainty to the discrimination conditions (Fig. 3C) yielded a significant standardized GABA and Glx/GABA scores ($t(30) = 3.0$, $p = 0.005$, and $t(30) = -2.64$, $p = 0.013$, respectively; one sample t-test). As expected, Glx did not change significantly in this comparison ($t(30) = -0.56$, $p > 0.015$, one sample t-test). The comparisons between the null to the uncertainty condition and between the null to the discrimination condition were non-significant in all measurements ($p > 0.015$, one sample t-test, data not shown).

It was tempting to hope that Glx/GABA results have an additional component to the GABA findings alone, thereby providing evidence to a specific E/I balance mechanism. To assess whether the significant differences in Glx/GABA between the uncertainty and the discrimination conditions could be explained by GABA, Glx or their combined effect, we implemented repeated measures ANOVA on GABA/water with and without Glx/water as a varying covariate. The resulting effect size increased from $\eta^2 = 0.24$ without the covariate to $\eta^2 = 0.26$ with the covariate. Subsequently, the square of each effect size was modeled as an index of correlation between the learning conditions to the GABA measurement, and a test for the difference between correlations was performed on the Fisher-transformed correlations. The test revealed no-difference between correlations, meaning Glx measurement was not found to add any significant information ($Z = 0.07$, $p > 0.05$).

We conclude that the major changes between conditions occur due to changes in GABA concentrations mainly.

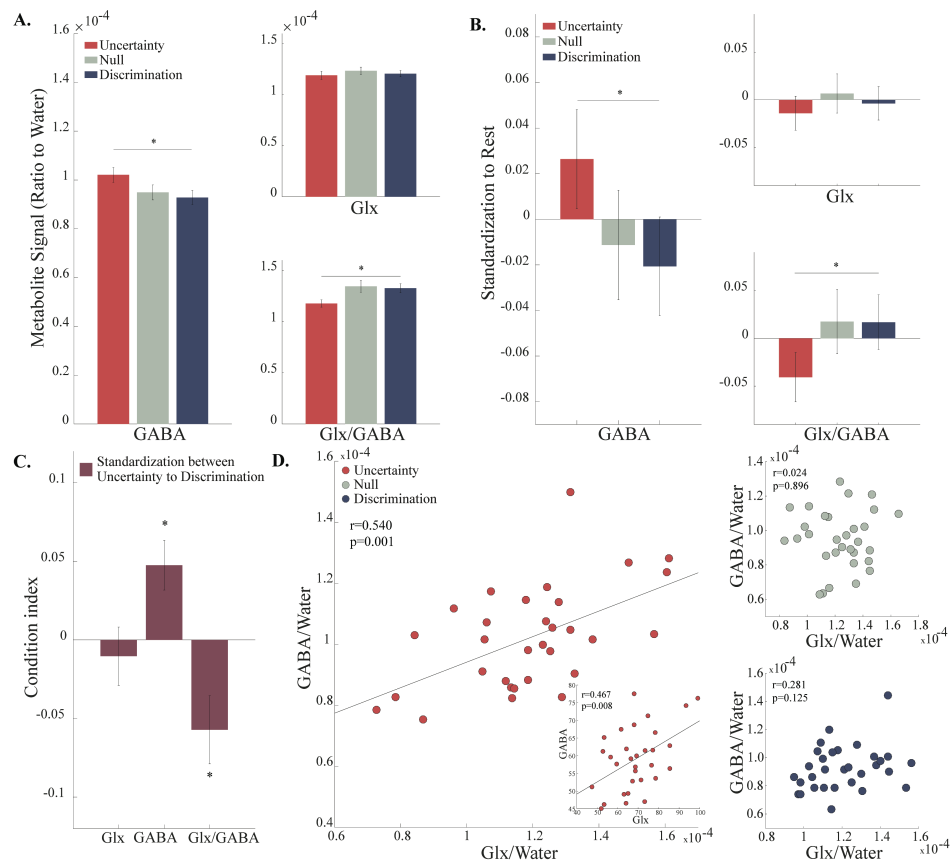


Figure 3. Higher GABA levels observed during the uncertainty condition. A,B, Metabolites levels quantified as ratio to water (A) or standardized to rest (B) reveal a consistent result: higher GABA levels and lower Glx/GABA levels in the uncertainty condition, yet similar lower levels during the Null and discrimination conditions ($p < 0.05$, repeated measures one way ANOVA; $*p < 0.05$, post hoc Tukey-Kramer). See text for detailed statistics. C. Standardization between pairs of task conditions reveal significant differences in GABA and Glx/GABA levels when comparing between the uncertainty and the discrimination conditions ($*p < 0.015$, one sample t-tests with Bonferroni adjusted alpha; error bars represent the SEM). D. A positive correlation between Glx/Water and GABA/Water levels during the uncertainty condition only ($r = 0.540$, $p = 0.001$). Inset shows the same without normalization by water signal ($r = 0.467$, $p = 0.008$).

GABA and Glx levels are correlated during uncertainty condition

Although Glx was not found to significantly contribute to the differences observed between the conditions in Glx/GABA, and these were driven mainly by GABA changes – we aimed to further analyze the relationships between Glx and GABA. To do so, we calculated Pearson's correlation coefficient between Glx/Water and GABA/Water in each of the conditions (Fig. 3D). A positive correlation was found during the uncertainty condition ($r = 0.540$, $p = 0.001$), but not during the Null ($r = 0.024$, $p > 0.05$) or discrimination ($r = 0.281$, $p > 0.05$) conditions. This persisted when examining the correlation between GABA and Glx without the contribution of

water (Fig. 3D inset $r=0.467$, $p=0.008$), and again there were no significant correlations in the other two conditions ($p>0.05$ in both, $r=-0.12$ for Null condition; $r=0.26$ for discrimination condition). A regression of GABA on Glx levels showed that although they increased in correlation, the E/I balance does change ($\beta = 0.4182$, significantly different than 1, $p<0.01$).

We propose that changes in Glx, although too small to show significant variations across conditions, still maintain a balance with the GABA signal when it varies extensively and significantly i.e. when GABA levels are especially high as in the uncertainty condition.

GABA levels in the Null condition predict learning performance

We tested whether GABA and Glx levels during each experimental condition were directly related to subjects' behavioral performance during the discrimination condition, as a proxy for their learning ability (Fig.4A). To do so, we calculated Pearson's correlation coefficients between the behavioral measures during the discrimination condition to GABA/water and Glx/water levels during the different conditions.

We used two behavioral measures. First, we approximated the learning-rate as the trial when a subject reached performance criterion of 0.7, which is similar to the overall mean selection probability (0.69). There was a negative correlation (Fig.4B) between the learning-rate and GABA/water levels during the Null condition ($r=-0.550$, $p=0.001$), but not during the discrimination ($r=-0.020$, $p=0.888$) or uncertainty ($r=0.096$, $p=0.618$) conditions. Glx/water levels in the different experimental conditions did not reveal significant correlations with learning rate ($p>0.05$ in all conditions, $r=-0.07$ for Null condition; $r<0.001$ for uncertainty condition; $r=-0.1$ for discrimination condition).

Second, we assessed behavioral performance by the proportion of selection of the better-option. We found a positive correlation (Fig.4C) between the proportion and GABA/water levels during the Null condition (Fig.4C; $r=0.418$, $p=0.019$); but not during the other two conditions ($r= -0.004$, $p=0.980$ for uncertainty; $r=0.067$, $p=0.717$ for discrimination). As before, the selection probability was uncorrelated with Glx/water levels in any of the conditions ($p>0.05$ in all conditions; $r=0.03$, 0.09 and 0.15 for the null, uncertainty and discrimination conditions, respectively).

We conclude that baseline levels of available GABA, i.e. in a Null condition, predict better performance in a subsequent discrimination learning task.

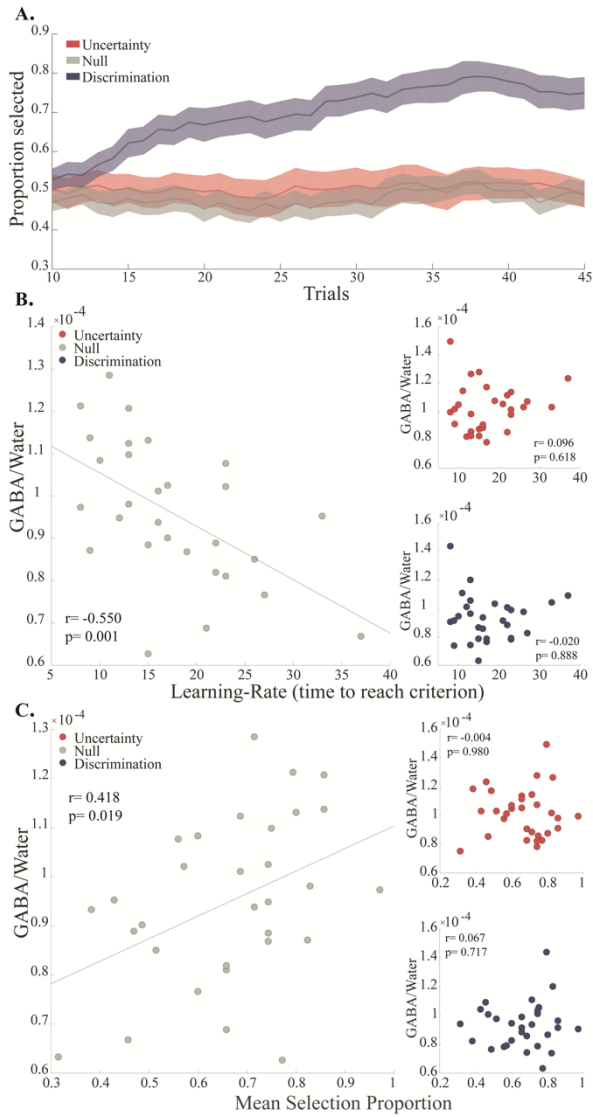


Figure 4. GABA levels in the Null condition predict behavioral performance during discrimination learning condition. **A.** Behavioral performance across all conditions. Gradual increase in selection probability of the better-option (high-gain tone) is observed during the discrimination condition. Averaged across all subjects ($n=31$) in 10-trial running window (mean \pm SEM). **B.** Negative correlation between the learning-rate, quantified as the trial # when a 0.7 criterion is reached during the discrimination condition, and GABA/water levels during the Null condition ($r=-0.550$, $p=0.001$). **C.** Positive correlation between mean selection of the better-option during the discrimination condition and GABA/water levels during the Null condition ($r=0.418$, $p=0.019$).

Discussion

The paradigm presented here simultaneously examined excitatory and inhibitory modifications in the dACC during learning. Behaviorally, the participants performed the task as expected, and learned to prefer the option that predicted higher gain in the discrimination condition, while lacked any preference in the two other conditions. These behavioral differences were expressed as increased GABA and decreased Glx/GABA levels during the uncertainty condition – when rewards and punishments are abundant, but there is no information to learn from. Although Glx was not significantly altered across conditions and did not seem to contribute beyond GABA to the difference in Glx/GABA across the conditions, further analysis revealed that Glx and GABA were positively correlated in the uncertainty condition only. This further supports the notion that when GABA is significantly increased, Glx levels do try to maintain a balance. Finally, examination of the relationship between GABA levels to the behavioral measures revealed that higher GABA levels in the Null condition predicted better learning, both when measured as time to reach criterion (learning rate), and when measured as final performance.

As previously shown, E/I balance is essential for cortical network stability [4, 5, 39, 40]. The precision of the balance depends on the degree of correlation between excitation and inhibition, ranging from a global balance in the absence of correlated inputs to a detailed balance for strong correlations [41]. The tight correlation, observed during the uncertainty condition, suggests a more homeostatic activity, in which specific functional patterns are amplified [42] and promote the occurrence of a distinctive functional activity. Moreover, we observed elevated GABA levels that can be interpreted as increase in inhibition which occurred in situations that are known to modulate the activity of the dACC, and consist of uncertain rewards [33-35, 43], high mental effort [36-38, 44], and requirements for cognitive control [30-32]. Additionally, a line of evidence linked the inhibitory system with achievement of efficient performance during high cognitive load [45, 46]. Therefore, presumably, during a need to make a decision in highly demanding cognitive load, specific functional patterns are amplified in the dACC and an increased inhibition occurs. This inhibition is required in order to achieve efficient and better performance. Similarly, individual differences in inhibition levels in the dACC might be linked to each individual experienced cognitive load, and reflect mental effort, uncertainty or a need for cognitive control.

In a neutral situation, such as the Null condition, the influences of external cognitive aspects of learning like incentives (rewards/punishments) are likely eliminated, exposing the underlying intrinsic motivational or attention-oriented processes, or simply the availability of the relevant neurotransmitter. Therefore, the correlation between inhibition in the dACC during neutral situation and later performance when learning is possible, might reflect the connection between the cognitive load and the motivational processes that take place during learning [47-50], which

in their turn, lead to a better performance [51]. The fact that we did not identify any changes during learning itself (the discrimination condition) remain to be explored further. A possible interpretation is that learning occurred too fast for the long scan process required for MRS to detect changes beyond the signal-to-noise ratio; yet another possibility is that once uncertainty is reduced (because the subjects unveil a pattern to learn), less and less inhibitory mechanisms are required to maintain high cognitive load, attention, or uncertainty. A final possibility we cannot rule-out, is that the number of errors as well as punishments diminishes quickly, and the dACC signal diminishes with it. If either of the above is indeed the case, a design with much slower learning rate might allow detection of GABA levels in the dACC that are inversely-correlated with the learning performance.

Overall, the effects demonstrated here emphasize the importance of the dACC in processes of decision making and expand the understanding regarding its involvement in high cognitive demanding situations, and how it can later enable better learning. Importantly, it paves the way to further use of magnetic-resonance-spectroscopy (MRS) in identifying specific neural mechanisms and the use of neurotransmitters in brain networks during active behavior, and next steps should examine further regions, either by a-priori selection as done here, or by development of better sequences that allow simultaneous measurements from several regions. Combined with the paradigm we used here, it unveils the contribution of excitatory and inhibitory modifications during active learning in humans.

Acknowledgments: We thank Dr. Edna Furman-Haran, Nachum Stern and Fanny Attar for assisting with the MRI procedures. This work was supported by ISF #26613 and ERC-2016-CoG #724910 grants to R. Paz, and the Minerva Foundation for A. Tal. A. Tal also acknowledges the support of the Monroy-Marks Career Development Fund and the historic generosity of the Harold Perlman Family.

Methods

Participants

All studies were performed in accordance to the procedures approved by the Internal Review Board of Haemek Medical Center (Afula, Israel). Thirty-seven right-handed healthy subjects (age range 22–39 years; median age 26 years; 21 females) participated in the experiment after written informed consent was obtained. Three subjects were excluded from the experiment due to distorted signals, resulting from magnetic field drift, improper shimming or motion. Three additional subjects were excluded due to recurrent GABA and Glx outlier levels (more than 2.5 standard deviations).

Behavioral Paradigm

A two-alternative forced choice task was conducted using a block design paradigm (Fig. 1A), while subjects laid in an MRI scanner. During each trial, subjects heard two short (270 ms) pure-frequency tones which were played out in succession and in a random order, each to a different ear. Subsequently, a visual response cue appeared, comprised of two arrows, corresponding to the two sides the tones were played to. The subjects chose their preferred tone by selection of the right or left remote buttons of a response box, corresponding to the side their preferred tone was played to. Following the selection, the chosen arrow was marked in black and subsequently an outcome screen appeared, presenting the response for the last selection, which was one of three possible outcomes: a monetary gain (+2), loss (-2) or neutral feedback (0). The paradigm consisted of three experimental conditions, each implemented within a separate block composed of 45 trials and lasting 12 minutes: (1.) the null condition with a consistent 0 reward (00/00 probability of loss/gain); (2.) the uncertainty condition (50/50 loss/gain probability, independent of the participant's choice); and (3.) the discrimination condition (80/20 loss/gain probability), in which one of the tones was paired with a monetary gain in 80% of the trials it was chosen and monetary loss in 20% of the trials it was chosen, and vice versa for the other tone. Stimuli were generated by MATLAB (R2015b, The MathWorks, Natick, USA) using the Psychophysics Toolbox extension[52]. Each of the conditions was assigned one of three pairs of tone frequencies, mixed between subjects in pairing order. The tones' frequencies were kept at a 4:7 ratio to facilitate their distinction. A fourth frequency pair was assigned to a short initial training session.

Pretest A pre-test was conducted prior to the experiment to examine innate bias to low or high frequency tones. We compared between two groups with $n=5$ subjects each, in one the low frequency tones and in the other the high frequency tone was associated with the higher monetary gain in the discrimination condition (0.8 probability to gain money). The trials were divided into four periods, and the averaged selection probability was calculated for the tone that

was associated with higher monetary gain. No bias was evident, and learning was similar to both options (Fig. 1D inset; Wilcoxon $p > 0.05$ in each one of the periods).

MR Scanning Protocol

MR data was collected using a 3 Tesla Tim Trio scanner (Siemens, Erlangen), with a 12-channel receiver head coil. Body coil with a peak B_1 of 19 μT was used for transmission. Anatomical images were acquired using three-dimensional T_1 -weighted MPRAGE sequence ($1 \times 1 \times 1 \text{ mm}^3$ voxels, TR/TE/TI = 2300/2.98/990 ms, 176 slices, FA=9°, in-plane FOV=256×256 mm², TA=4:44 min). The sagittal and coronal T_1 -weighted anatomical images enabled localization of a $40 \times 25 \times 10 \text{ mm}^3$ ^1H -MRS voxel on the midline of the dACC (Fig. 2A). In order to enhance magnetic field homogeneity, first and second order shim adjustments were performed using the default Siemens shimming tool, yielding water line widths of 6-7 Hz. The GABA and Glx ^1H -MRS spectra were acquired using the spectral editing sequence Mescher-Garwood Point RESolved Spectroscopy (MEGA-PRESS)[8]. Each MEGA-PRESS scanning block (TR/TE=2000/68 ms, 2048 complex FID points, 2 kHz bandwidth, 16-step phase cycle) consisted of a metabolite scan (144 averages, TA=9:57 min) that was followed by a water reference scan (16 averages, TA=1:00 min). The individual coils' spectra were phased and weighted by their signal to noise ratios using the reference coil sensitivity maps before combining their respective spectra.

Procedure

Participants underwent initial training composed of four trials in front of a computer to ensure the task was correctly understood. Upon completion, they entered the magnet head first supine, while being visually monitored for awareness. A response box was handed to all subjects, and a projector was used for visual feedback while lying in the scanner. Anatomical images were acquired and enabled localization of the ^1H -MRS voxel in the dACC. Shimming was carried out, followed by a ^1H -MRS “rest” scan, during which the subjects were asked to focus on a fixation cross placed at the center of the screen. Following the rest scan, the participants began the experimental session in which, at the onset of each experimental condition block, a MEGA-PRESS metabolite scan was initiated, followed by a water reference scan. The responses of the subjects, along with the rewards they received, were recorded and labeled (Fig. 1B). The subjects were divided into two cohorts, each exposed to a different ordering of the experimental conditions (Fig. 1C): The first cohort (20 subjects) started with two blocks of the uncertainty condition, continued with one block of the null condition, and concluded with two blocks of the discrimination condition. The second cohort (11 subjects) started with one block of the null condition, followed by two blocks of the uncertainty condition and concluded with two blocks of the discrimination condition. Eighteen subjects (seven from the first cohort

and all subjects from the second cohort) also performed an additional and final rest scan after concluding the task.

Behavioral Analysis

The behavioral data was analyzed concentrating on the first block of each condition, using in-house MATLAB scripts. In order to examine the dynamics of the subjects' choices over the course of each experimental condition's block, we performed a moving average of the selection probability with a window width of 10 trials. Additionally, to assess behavioral performance per condition, we calculated the averaged selection probability of the high gain tone. To probe the learning rate in the discrimination condition, we calculated the mean overall selection probability, and used the first trial in which each subject reached this selection probability.

¹H-MRS Processing and Quantification

MEGA-PRESS spectra were analyzed using in-house MATLAB scripts. Spectra were zero-filled 8-fold, apodized using an exponentially time decaying function with a linewidth of 3.2 Hz, Fourier transformed and phased. Difference spectra were generated by subtraction of the alternating ON and OFF spectra, which were subsequently aligned to the NAA methyl singlet at 2.01 ppm.

Metabolite quantification was achieved via peak integration with the integration limits 3.56 to 3.94 for Glx and 2.83 to 3.19 for GABA (Fig. 2B). The signals of GABA's methylene group+ macromolecular contributions at 3.01 ppm [9, 53], and the Glx complex at 3.75 ppm are represented by:

$$S \propto C \cdot V \cdot (1 - f_{CSF}) \cdot B_{1-} \cdot f_R$$

where C is the metabolite's concentration, V the voxel volume, B_{1-} the receiver coil sensitivity, f_{CSF} the fraction of CSF in the voxel, and f_R a factor accounting for T_1 and T_2 relaxation. Taking the ratio of each metabolite to the reference water signal (GABA/water, Glx/water) removed common factors such as B_{1-} and V, and reduced signal variability. The remaining factors, such as f_R and f_{CSF} , as well as the macromolecular contamination, remained unaccounted for, but were assumed constant throughout the paradigm when interpreting our results. Since our analysis focused on intra-subject changes to metabolite levels during the learning task, these inter-subject sources of variability did not bias our conclusions. To further factor-out these constant elements we have also examined the standardized differences:

$$f_{A,B}^{met} = \left(\frac{S_A^{met}}{S_A^{water}} - \frac{S_B^{met}}{S_B^{water}} \right) / \left(\frac{S_A^{met}}{S_A^{water}} - \frac{S_B^{met}}{S_B^{water}} \right), \quad (1)$$

where met =Glx, GABA and A,B=null, discrimination, uncertainty or rest. The expression in Eq. (1) is independent of f_{CSF} , f_R , B_{1-} and any other factor assumed unchanged between conditions. This expression, when standardized to rest, served as standardization to baseline activity, and

when was conducted between two conditions was reflecting a direct comparison between conditions of the task.

Validating robustness of integration limits

In order to validate the robustness of choosing the integral limits for each metabolite, we further carried an optimization process, varying the integration limits, from 3.01 ± 0.35 ppm for GABA and 3.755 ± 0.35 ppm for Glx, and then doing multiple comparisons to choose the best 'classifier'. In accordance to Eq. (1), the receiver operating characteristic (ROC) curve of each classifier was calculated three times, testing for a difference between two conditions of the task in each time (Fig. 2C inset). The areas under the curve (AUCs) of all the tested classifier were compared (Fig. 2C) and the classifiers with the highest AUC were selected in each comparison type. Subsequently, for each metabolite, the averages of the highest AUCs were set as the chosen integration limits: 3.56 to 3.94 for Glx and 2.83 to 3.19 for GABA. As a control, we performed this process on two sets of random noises. The AUCs for all the tested pairs were close to 0.5 (in gray line, Fig. 2C), implicating that no significant data could be found in case of noise, and that the integration limits that showed significance for the real data, indeed included important information. We note that the process resulted in integration limits that are highly similar to the ones chosen by an experienced observer and/or in similar studies that take such limits a-priori.

Statistical tests

Changes to within-subject behavioral and metabolic measures were compared using two- or one-way repeated measure analyses. Tucky-Kramer tests were conducted for post-hoc paired comparisons. One sample t-tests was conducted in order to assess normalized metabolic levels significance, while paired sample t-tests was conducted for temporal metabolic drift analyzing. To analyze behavioral pretest data and in order to assess conditions ordering effect, we performed unpaired group comparisons using nonparametric Wilcoxon rank sum tests. Pearson's correlation coefficient was conducted in order to test correlations between metabolites levels and behavioral performance. Unless otherwise stated, significance level was set to $P < 0.05$.

References

1. Johansen, J.P., Cain, C.K., Ostroff, L.E., and LeDoux, J.E. (2011). Molecular mechanisms of fear learning and memory. *Cell* *147*, 509-524.
2. Myhrer, T. (2003). Neurotransmitter systems involved in learning and memory in the rat: a meta-analysis based on studies of four behavioral tasks. *Brain Research Reviews* *41*, 268-287.
3. Kelley, A.E. (2004). Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neuroscience & biobehavioral reviews* *27*, 765-776.
4. Isaacson, J.S., and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron* *72*, 231-243.
5. Treviño, M. (2016). Inhibition controls asynchronous states of neuronal networks. *Frontiers in synaptic neuroscience* *8*.
6. Letzkus, J.J., Wolff, S.B., Meyer, E.M., Tovote, P., Courtin, J., Herry, C., and Lüthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* *480*, 331.
7. Bottomley, P.A. (1987). Spatial localization in NMR spectroscopy in vivo. *Annals of the New York Academy of Sciences* *508*, 333-348.
8. Mescher, M., Merkle, H., Kirsch, J., Garwood, M., and Gruetter, R. (1998). Simultaneous in vivo spectral editing and water suppression. *NMR in Biomedicine* *11*, 266-272.
9. Mullins, P.G., McGonigle, D.J., O'gorman, R.L., Puts, N.A., Vidyasagar, R., Evans, C.J., and Edden, R.A. (2014). Current practice in the use of MEGA-PRESS spectroscopy for the detection of GABA. *Neuroimage* *86*, 43-52.
10. Donahue, M.J., Near, J., Blicher, J.U., and Jezzard, P. (2010). Baseline GABA concentration and fMRI response. *Neuroimage* *53*, 392-398.
11. Jocham, G., Hunt, L.T., Near, J., and Behrens, T.E. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nature neuroscience* *15*, 960-961.
12. Levar, N., van Leeuwen, J.M., Puts, N.A., Denys, D., and van Wingen, G.A. (2017). GABA concentrations in the anterior cingulate cortex are associated with fear network function and fear recovery in humans. *Frontiers in human neuroscience* *11*.
13. Muthukumaraswamy, S.D., Edden, R.A., Jones, D.K., Swettenham, J.B., and Singh, K.D. (2009). Resting GABA concentration predicts peak gamma frequency and fMRI amplitude in response to visual stimulation in humans. *Proceedings of the National Academy of Sciences* *106*, 8356-8361.
14. Northoff, G., Walter, M., Schulte, R.F., Beck, J., Dydak, U., Henning, A., Boeker, H., Grimm, S., and Boesiger, P. (2007). GABA concentrations in the human anterior cingulate cortex predict negative BOLD responses in fMRI. *Nature neuroscience* *10*, 1515.
15. Yoon, J.H., Grandelis, A., and Maddock, R.J. (2016). Dorsolateral Prefrontal Cortex GABA Concentration in Humans Predicts Working Memory Load Processing Capacity. *Journal of Neuroscience* *36*, 11788-11794.

16. Floyer-Lea, A., Wylezinska, M., Kincses, T., and Matthews, P.M. (2006). Rapid modulation of GABA concentration in human sensorimotor cortex during motor learning. *Journal of neurophysiology* *95*, 1639-1644.
17. Hasler, G., van der Veen, J.W., Grillon, C., Drevets, W.C., and Shen, J. (2010). Effect of acute psychological stress on prefrontal GABA concentration determined by proton magnetic resonance spectroscopy. *American Journal of Psychiatry* *167*, 1226-1231.
18. Mangia, S., Tkáč, I., Gruetter, R., Van de Moortele, P.-F., Maraviglia, B., and Uğurbil, K. (2007). Sustained neuronal activation raises oxidative metabolism to a new steady-state level: evidence from ¹H NMR spectroscopy in the human visual cortex. *Journal of Cerebral Blood Flow & Metabolism* *27*, 1055-1063.
19. Schmitz, T.W., Correia, M.M., Ferreira, C.S., Prescott, A.P., and Anderson, M.C. (2017). Hippocampal GABA enables inhibitory control over unwanted thoughts. *Nat Commun* *8*, 1311.
20. Sumner, P., Edden, R.A., Bompas, A., Evans, C.J., and Singh, K.D. (2010). More GABA, less distraction: a neurochemical predictor of motor decision speed. *Nat Neurosci* *13*, 825-827.
21. Boy, F., Evans, C.J., Edden, R.A., Singh, K.D., Husain, M., and Sumner, P. (2010). Individual differences in subconscious motor control predicted by GABA concentration in SMA. *Curr Biol* *20*, 1779-1785.
22. Michels, L., Martin, E., Klaver, P., Edden, R., Zelaya, F., Lythgoe, D.J., Luchinger, R., Brandeis, D., and O'Gorman, R.L. (2012). Frontal GABA levels change during working memory. *PLoS One* *7*, e31933.
23. Sampaio-Baptista, C., Filippini, N., Stagg, C.J., Near, J., Scholz, J., and Johansen-Berg, H. (2015). Changes in functional connectivity and GABA levels with long-term motor learning. *Neuroimage* *106*, 15-20.
24. Dayan, P., and Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Curr. Opin. Neurobiol.* *18*, 185-196.
25. Haber, S.N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* *35*, 4.
26. Heilbronner, S.R., and Hayden, B.Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annual Review of Neuroscience*, Vol 39 *39*, 149-170.
27. Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., and Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. *Nature neuroscience* *9*, 940.
28. Lee, D., Seo, H., and Jung, M.W. (2012). Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci* *35*, 287-308.
29. Kolling, N., Wittmann, M.K., Behrens, T.E., Boorman, E.D., Mars, R.B., and Rushworth, M.F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature neuroscience* *19*, 1280-1285.
30. Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science* *306*, 443-447.
31. Shenhav, A., Cohen, J.D., and Botvinick, M.M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature neuroscience* *19*, 1286-1291.

32. Sheth, S.A., Mian, M.K., Patel, S.R., Asaad, W.F., Williams, Z.M., Dougherty, D.D., Bush, G., and Eskandar, E.N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioral adaptation. *Nature* *488*, 218.
33. Amiez, C., Joseph, J.-P., and Procyk, E. (2005). Reward encoding in the monkey anterior cingulate cortex. *Cerebral cortex* *16*, 1040-1055.
34. Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience* *10*, 1214.
35. Payzan-LeNestour, E., Dunne, S., Bossaerts, P., and O'Doherty, J.P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron* *79*, 191-201.
36. Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E., and Rushworth, M.F. (2009). Effort-based cost-benefit valuation and the human brain. *Journal of Neuroscience* *29*, 4531-4541.
37. Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., and Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *Journal of Neuroscience* *30*, 14080-14090.
38. Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., and Rushworth, M.F. (2006). Separate neural pathways process different decision costs. *Nature neuroscience* *9*, 1161.
39. Yizhar, O., Fenno, L.E., Prigge, M., Schneider, F., Davidson, T.J., O'Shea, D.J., Sohal, V.S., Goshen, I., Finkelstein, J., and Paz, J.T. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* *477*, 171-178.
40. Van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* *274*, 1724-1726.
41. Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* *334*, 1569-1573.
42. Murphy, B.K., and Miller, K.D. (2009). Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* *61*, 635-648.
43. Walton, M.E., Croxson, P.L., Behrens, T.E., Kennerley, S.W., and Rushworth, M.F. (2007). Adaptive decision making and value in the anterior cingulate cortex. *Neuroimage* *36*, T142-T154.
44. Hosking, J.G., Cocker, P.J., and Winstanley, C.A. (2014). Dissociable contributions of anterior cingulate cortex and basolateral amygdala on a rodent cost/benefit decision-making task of cognitive effort. *Neuropsychopharmacology* *39*, 1558.
45. Aron, A.R. (2007). The neural basis of inhibition in cognitive control. *The neuroscientist* *13*, 214-228.
46. Medalla, M., and Barbas, H. (2009). Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. *Neuron* *61*, 609-620.
47. Shenhav, A., Botvinick, M.M., and Cohen, J.D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* *79*, 217-240.

48. Etzel, J.A., Cole, M.W., Zacks, J.M., Kay, K.N., and Braver, T.S. (2015). Reward motivation enhances task coding in frontoparietal cortex. *Cerebral Cortex* *26*, 1647-1659.
49. Braver, T.S., Krug, M.K., Chiew, K.S., Kool, W., Westbrook, J.A., Clement, N.J., Adcock, R.A., Barch, D.M., Botvinick, M.M., and Carver, C.S. (2014). Mechanisms of motivation–cognition interaction: challenges and opportunities. *Cognitive, Affective, & Behavioral Neuroscience* *14*, 443-472.
50. Anselme, P. (2010). The uncertainty processing theory of motivation. *Behavioural brain research* *208*, 291-310.
51. Cerasoli, C.P., Nicklin, J.M., and Ford, M.T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological bulletin* *140*, 980.
52. Brainard, D.H., and Vision, S. (1997). The psychophysics toolbox. *Spatial vision* *10*, 433-436.
53. De Graaf, R.A. (2013). *In vivo NMR spectroscopy: principles and techniques*, (John Wiley & Sons).