The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change

E Fouragnan[#1,2], BKH Chau[#2,4], D Folloni[#2], N Kolling[2], L Verhagen[2], M Klein-Flügge[2], L Tankelevitch[2], GK Papageorgiou[2,5], JF Aubry[6], J Sallet[#2], MFS Rushworth[#2,3]

[1]School of Psychology, University of Plymouth, UK

[2]Wellcome Integrative Neuroimaging (WIN), Department of Experimental Psychology, University of Oxford, Oxford, UK

[3]Wellcome Integrative Neuroimaging (WIN), Centre for Functional MRI of the Brain (MRI), Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford

[4]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong

[5]McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

[6]Institut Langevin, Espci Paris, CNRS UMR 7587, INSERM U979, PSL Research University, France

[#] These authors contributed equally to this work.

Abstract

The neural mechanisms mediating sensory-guided decision making have received considerable attention but animals often pursue behaviors for which there is currently no sensory evidence. Such behaviors are guided by internal representations of choice values that have to be maintained even when these choices are unavailable. We investigated how four macaque monkeys maintained representations of the value of counterfactual choices– choices that could not be taken at the current moment but which could be taken in the future. Using functional magnetic resonance imaging, we found two different patterns of activity co-varying with values of counterfactual choices in a circuit spanning hippocampus, anterior lateral prefrontal cortex, and anterior cingulate cortex (ACC). ACC activity also reflected whether the internal value representations would be translated into actual behavioral change. To establish the causal importance of ACC for this

translation process, we used a novel technique, Transcranial Focused Ultrasound Stimulation, to reversibly disrupt ACC activity.

Introduction

Every day, chacma baboons, an old world primate, navigate to and from the safety of their sleeping post and distant foraging or watering sites[1]. The decision to move to alternative locations is not simply guided by accumulation of sensory evidence for that choice but by internal representation or memory of the alternative choice's value. The same is true when they move back towards the sleeping post in the evening. While sensory and associative decision making have been well-studied[2], less is known about how representations of counterfactual choices – choices not currently taken but which may be taken in the future – are held in memory and guide behavior.

In humans, lateral frontal polar cortex (lFPC) holds counterfactual information[3–5]. This may underlie its role in exploratory behavior[6]. However, many questions remain. First, some of the same studies report a similar pattern of activity in anterior cingulate cortex (ACC)[3,5,6]. Other studies have emphasized a related role for ACC in encoding the value of switching behavior and rejection of the default choice[7,8]. Here we introduce a simple paradigm that makes separation of the roles of the areas possible and distinguishes them from a third region: hippocampus. Within the hippocampal formation, the subiculum projects monosynaptically to ACC[9]. Information held in memory in such medial temporal structures may guide decision making[2]. Although little is known about whether or how activity in hippocampus encodes counterfactual choices, it is clear that hippocampal lesions disrupt switching between choices in other tasks[10].

We also address a second issue: whether macaques possess a brain region with a functional role corresponding to that of human lFPC. Human frontal polar cortex can be subdivided into lateral and medial sub-regions lFPC and mFPC[11,12]. While resting state connectivity pattern exhibited by human mFPC and macaque FPC are similar, human lFPC's more closely resembles macaque lateral prefrontal cortex (lPFC). It is therefore unclear if macaques hold counterfactual information as humans do and, if they can, whether it is mediated by macaque FPC or lPFC. We know that when macaques are given feedback about what would have happened had another choice been made, they use it to guide their next choice[13,14]. However, how information about the multiple counterfactual choices that typically exist in natural environments is retained while another choice is actually made is unknown.

Finally, our experiment allowed comparison of two fundamentally different ways in which counterfactual choice information might influence behavior. On the one hand, information about currently unavailable choices must be held if future behavior is to be accurate when that choice once again becomes available. This might be mediated by some combination of ACC/lPFC/lFPC. On the other hand, holding information about currently unavailable choices may impact on the current decision being made. We show that the second influence of counterfactual choice is mediated by a distinct neural circuit centered on ventromedial prefrontal / medial orbitofrontal cortex (vmPFC/mOFC).

Four macaques chose between pairs of abstract visual stimuli while in the MRI scanner (fig.1a, b). On each trial, the two stimuli available for choice (available options) were drawn from a set of three, each associated with distinct reward probabilities (fig.1a). The rewards were delivered probabilistically in a manner that fluctuated across the session, with two of the options reversing towards the middle of a session (fig.1c). Each stimulus' reward probability was uncorrelated from that of the others (<22% mean shared variance). On each trial one of the two available options was chosen by the monkey, the other was unchosen, and a third option was invisible and unavailable for choice. Both the unchosen option and the unavailable option can be considered counterfactual choices – although these choices were not made on the current trial, they might be made on a future occasion.

Behavioral analyses demonstrated that animals maintained representations of counterfactual choice values to guide future behavior on subsequent trials. We therefore used fMRI to test whether neural activity reflected counterfactual choice values according to one of several possible schemes. FMRI allowed us to search for activity related to counterfactual choice value throughout the brain. First, neural activity might represent the value of the unavailable option (Hypothesis 1, fig.1e). Alternatively, it might reflect the value of any counterfactual option – options that are currently unavailable for choosing, and options that are available on the current trial but which are unchosen. In such a scheme, it may not be important whether a counterfactual choice is unavailable or unchosen, however, if such a representation is to guide future behavior, then it should reflect the ranked values of the alternative options (Hypothesis 2, fig.1f). We also compared this with a third scheme in which an unavailable option's value had no influence on neural activity (Hypothesis 3, fig.1g). Notably such a coding scheme corresponds to the claim that ACC activity simply reflects decision difficulty[8,15]. According to this view, it is the difference in value between the choices available that determines decision difficulty (when the difference is large it is easy to identify the better choice but this is not the case when the difference is small). However, according to this view, an option not actually available does not affect difficulty of the current decision and therefore does not influence ACC.

In our animal model it was possible to investigate not just correlation between neural activity and behavior but the activity's causal importance for behavior[16]. We used transcranial focused ultrasound stimulation (TUS). Like transcranial magnetic stimulation (TMS), TUS can alter neural activity[17] but unlike TMS, it can even do so in relatively deep structures such as ACC[18]. The TUS 250 kHz ultrasound stimulation was concentrated in a cigar-shaped focal spot several centimeters below the focusing cone. A series of five experiments, each conducted in three macaques, has demonstrated that this protocol transiently, reversibly, reproducibly, and focally alters neural activity[17,18]. A similar TUS protocol altered saccade planning in macaques when applied to the frontal eye fields but not to a location 10-12mm distant[19]. Importantly the minimally invasive nature of the stimulation made it possible to examine not just a region of interest such as ACC but also a control region in the same animals and to do so without MRI incompatible implants. In the current study, consistent with our ranked counterfactual hypothesis (Hypothesis 2), ACC TUS impaired translation of counterfactual choice values into actual behavioral change.

Results

Animals learned option values and maintained them in memory without forgetting

To behave adaptively in this task, animals should estimate each option's reward probability and maintain these estimates in memory. If there are three options (A, B, C) then animals should retain what they have learned about option C even if subsequent trials involved presentation of only options A and B. The representations of C's value should then guide future decisions when C becomes available again. We therefore modeled animals' choices using a reinforcement learner[20,21] and tested whether the unavailable option's estimated reward probability (which in our experiment determines expected value) either decayed over time and/or became distorted to account for risk preference[22,23]. After simulating behavior with several reinforcement learning models (Methods and Supplementary fig.1), Bayesian model comparison revealed that monkeys did not forget unavailable option values nor distorted probability. Thus, animals learned the options' values and maintained them in memory without forgetting even when options were not available on a given trial.

To confirm the relationship between the better model's predictions and behavior, we compared choice probabilities predicted by the Maintain model and the actual recorded frequencies of animals' responses and found that the model matched behavior well (fig.1d; Pearson R^2=0.92). Having established the goodness of fit of the Maintain model to behavior, all further analyses were conducted using the expected values estimated with this model. To predict behavior as in humans and artificial decision making networks[24], estimates for the two available options were categorized as "high value" (HV) and "low value" (LV) and accuracy was categorically defined as HV selection. With these estimates, we found that the difference in value between the two available options (sometimes called "difficulty" as depicted in fig.1g) as well as the total value of available options were reliable predictors of animals' choice accuracy (value difference: Cohen's d=1.42; t_{24}=7.12, P=2.3x10^{-7}; total value: Cohen's d=0.82; t_{24}=4.10, P=4.04x10^{-4}) and reaction times (value difference: Cohen's d=-0.74; t_{24}=-3.68, P=0.001; total value: Cohen's d=-1.11; t_{24}=-5.54, P=1.07x10^{-5}; fig.1d).

Value associations of counterfactual options guide future choices

To guide future behavior, it is essential to retain counterfactual choice values in case these choices become available again in the future. There are at least two different ways that animals can maintain counterfactual information for future use. The first way is to consider which choices are available and which are not on each trial (Hypothesis 1; fig.1e)[25] and thus to categorize the options as "chosen", "unchosen" and "unavailable". A second way to describe the options (Hypothesis 2; fig.1f) is to think of both the unchosen and the unavailable options as alternative courses of action constituting the counterfactual choices – potential choices that were not, or could not, be taken on the current trial but which might be taken in the future. Animals might rank the expected value associated with the counterfactual options. Therefore, we characterized them as the "better" and "worse" counterfactual options irrespective of their availability. Finally we can test the hypothesis that animals only represent the difficulty of the current decision (Hypothesis 3; fig.1g)[15,26].

In line with the first hypothesis, we performed a logistic regression assessing whether the unavailable option's expected value influenced its future selection when it next reappeared on the screen. Decisions to select the previously unavailable option were strongly related to its expected value (One sample t-test on regression coefficients: Cohen's d=1.59; t_{24}=7.95, P=3.5x10^{-8}; fig.2a). A complementary analysis confirmed these results and showed that accuracy of the future choice was influenced by the currently unavailable option, particularly when its most recent expected value was the best of the three options (Cohen's d=1.06; t_{24}=5.32, P=1.87x10^{-5}; fig.2b) beyond the effect of the current Chosen and Unchosen options (Chosen: Cohen's d=0.98; P=5.04x10^{-5}; Unchosen: Cohen's d=-0.87; P=2.92x10^{-5}).

In line with the second hypothesis, we performed a series of analyses similar to those described above but replacing value estimates for the unavailable option by estimates for better and worse alternative choices. These analyses revealed animals' decisions to switch to the better counterfactual choice were influenced by its expected value (Cohen's d=1.23; t_{24}=6.16, P=2.32x10^{-6}) but this was not true for the worse counterfactual choice (Cohen's d=-0.09). In summary, the worse counterfactual had less of an influence on the decision to switch (fig.2c-d). Overall, the results demonstrate two ways of categorizing the choices made in the task: either by classifying them as "available" and "unavailable", or by considering the current chosen option in contrast to better and worse counterfactual choices. These frameworks guided analysis of fMRI data (fig.1e-g).

Hippocampal activity predicts successful future choices when the unavailable option becomes available again

Having established that animals not only represent choice value information that cannot be used on the current trial, but exploit this information on pending trials, the first fMRI-related analysis explored the extent to which neural activity reflected the expected value of the currently unavailable option (Hypothesis 1; fig.1e left panel). We tested for voxels across the whole brain where activity correlated with the trial-by-trial estimates of the unavailable option's expected value, particularly when the future selection was successful. We also included the expected value of the chosen and unchosen options as separate terms in the GLM (GLM1 in Methods). This analysis revealed one region in which the neural value coding of the unavailable option was different for successful future selection compared with unsuccessful future selection, surviving multiple correction (Z>3.1, whole-brain cluster-based correction P<0.001): right hippocampus (peak Cohen's d=0.72; Z=3.61, Caret-F99 Atlas (F99): x=16.5, y=-7.5, z=-12). At a lower threshold, we also found its bilateral counterpart: left hippocampus (peak Cohen's d=0.61; Z=3.05, F99 x=-14, y=-9, z=-12.5; fig.3a). There was, however, no significant relationship between hippocampal activity and the values of the choices that the monkeys were choosing between on the current trial (supplementary fig.2).

To illustrate the significant activity in bilateral hippocampal regions, we extracted the time course of the neural activation in two regions of interest (ROIs) (Methods, fig.3b left). Note that this analysis was performed for illustrative purposes only as the ROIs were formally linked to the comparison between correct and incorrect future selection used to establish the ROI location[27]. The activity pattern represented in this analysis is noteworthy as it shows

that the blood oxygenation level dependent (BOLD) signal in hippocampus is scaled by the expected value associated with the unavailable options only when the currently unavailable option is going to be chosen correctly on a future trial.

The hippocampus' role in maintaining information about currently unavailable choices may also encompass the prospect of rejecting the currently unavailable option if it is likely to be worse than the others[28]. To demonstrate this, we repeated the analysis in the trials preceding those in which the animal decided *not* to select a currently unavailable option. Critically, this analysis also revealed a greater BOLD signal for the value of the unavailable option on the current trial when this option was correctly rejected in the future compared to when it was incorrectly rejected (Leave-one-out peak selection: right Hippocampus: Cohen's d=0.59; t_{24}=2.96, P=0.006; left Hippocampus: Cohen's d=0.44; t_{24}=2.19, P=0.03; fig.3b right). In summary, hippocampal activity is scaled by the currently unavailable option's value more strongly (e.g. there is a stronger memory trace) when the next decision involving that option is going to be made correctly regardless of whether it is going to be chosen correctly (because it is highest in value) or rejected correctly (because it is lowest in value) in the future.

Finally, having established that hippocampal activity is related to memory of unavailable options we hypothesized that variation in such activity (at trial *t*) across sessions might predict variation in influence of the unavailable option's value on future accurate switching behavior (at *t+1*) (fig.2b). We found a significant correlation in the case of future decisions in which the unavailable option became accessible (Pearson R=0.43, P=0.03) but no correlation for the current decision while the unavailable option remained inaccessible (Pearson R=0.01; fig.3c). This result again suggests that the hippocampus is involved in future planning but not current, on-going decision making.

ACC ranks counterfactual options according to their expected value

The previous analysis was predicated on the idea that the brain maintains information in memory pertaining to currently unavailable choices while encoding what is relevant for the current decision elsewhere in the brain. Therefore, we next sought brain regions encoding the key decision variable – how much better is the currently chosen available option compared to the currently rejected available option. We searched for activity parametrically encoding the difference in value between the currently chosen and unchosen options (GLM2: chosen vs. unchosen expected values). Such a neural pattern, when locked to decision time, is sometimes referred to as a choice or value comparison signal. We found strong bilateral activations in a distributed network including ACC (peak Cohen's d=-0.75; Z=-3.75, F99 x=1, y=20.5, z=10.5), lPFC (right peak: Cohen's d=-0.92; Z=-4.61, F99 x=14.5, y=17.5, z=9.5; left peak: Cohen's d=-0.86; Z=-4.29, F99 x=-15, y=16, z=9.5) and ventromedial prefrontal cortex and adjacent medial orbitofrontal cortex (vmPFC/mOFC; peak Cohen's d=-0.80; Z=-4.01, F99 x=-5, y=14, z=2) encoding the (negative) difference in expected value between the chosen and unchosen options (fig.4a; |Z|>3.1, whole brain cluster-based correction P<0.001). In other words, activity in these areas increased as decisions became harder (e.g., because the subjective value of the chosen option became lower or the subjective value of the unchosen option became higher or both).

To first illustrate the relationship between option values and lPFC and ACC activity, we extracted BOLD time courses (using a leave-one-out cross-validation approach to avoid circularity of analyses) from ROIs over each region and performed further analyses (Methods). For each region, we found activity related to the difference between chosen and unchosen values was mainly driven by the negative relationship of the BOLD signal with the expected value of the chosen option (all $|Z|>3.1$ for the Chosen regressor); there was no significant activity for the Unchosen option. Importantly, the analysis contained an extra regressor representing the unavailable option's value, which also had no significant effect in ACC and lPFC. Importantly, the negative relationship between the ACC BOLD signal and the value of the chosen option may reflect the opportunity cost of switching away from the current choice.

Following this idea, in a second step, we tested whether the ACC might represent the possible alternatives that the animal might switch to in the future (Hypothesis 2). In this scheme, the two options not selected on the current trial, the unchosen option and the unavailable option, could both be considered counterfactual options that might be taken in the future and which could be ranked according to their expected value (GLM3: better vs. worse alternatives model, as per behavioral analyses). Using Bayesian statistics for each region within the same network (see Methods), we found that the activity pattern representing better and worse alternatives provided a significantly better account of neural activity in both ACC and lPFC compared to either the subjective choice comparison model (GLM2) or a third model (GLM4) that does not represent alternative options but rather the difficulty of selecting the current response (Hypothesis 3 in fig.1g) with $\varphi s > 0.95$ (fig.4b; see supplementary fig.3 and methods for full Bayesian Model Comparison[29]). Thus while ACC does not code for the value of the unchosen and unavailable options individually, it maintains a value of the best current alternative, and this effect is only visible in the data when the reference frame is altered from focusing on unchosen/unavailable to best alternative. One interpretation of the activity pattern is that it forecasts choosing the better of the counterfactual options during future decisions.

We directly tested this hypothesis using multiple regressions to investigate whether the activity in lPFC or ACC would predict upcoming switching behavior. For each ROI, we employed four regressors time-locked to the stimulus period of trial t, including i) the expected value of the better alternative if the future trial is a switch to that option; ii) the expected value of the better alternative if the future trial is a stay (i.e. repetition of the same choice as on current trial); iii) the expected value of the worse alternative if the future trial is a switch to that option; iv) the expected value of the worse alternative if the future trial is a stay. ACC activity predicted upcoming decisions to switch to the better and avoid the worse counterfactual (fig.4c; leave-one-out procedures for peak selection: post-hoc one sample t-tests: Best: Cohen's d=0.48; t_{24}=2.41; P=0.02, Worst: Cohen's d=-0.59; t_{24}=-2.94, P=0.007) but this was not true in lPFC (all Cohen's d<0.23, Ps>0.02). Such a pattern is consistent with a role for ACC in evaluating future strategies before execution[3,30–32]. By contrast, macaque anterior lPFC holds estimates of counterfactual choice values that are less immediately linked to behavior. Similarly, human frontal polar cortex activity reflects the values of alternative choice strategies in a manner that is also less immediately linked to behavior[26].

It has been suggested that ACC activity simply reflects decision difficulty[8,15] (fig.1g). When one option's value is much higher than the other option's, the decision is easy. But when the values of the two options are similar, the decision is difficult because it is hard to reject an alternative that is close in value. Our neural model comparison rejected this hypothesis (Supplementary fig.3c). Another possible index of decision difficulty is the reaction time (RT). We controlled for this in all our analyses by parametrically modulating the duration of the boxcar regressor locked at time of the decision by RT (regressor DEC in GLMs1, 2, 3, and 4).

ACC disruption impairs translation of counterfactual choice values into actual behavioral change

To test whether counterfactual choice value representations in ACC were causally important for effective behavioral switching, TUS was applied to the same ACC region. We previously demonstrated, using resting state fMRI (rs-fMRI) data that 40s sonification at 250 kHz reaches ACC and does so in a relatively focal manner having less effect on adjacent, even overlying, brain areas[18]. Here we provide an additional demonstration that ACC TUS increases activity correlation within the stimulated region but reduces correlation between the stimulated region and other regions (fig. 5a). Rs-fMRI scans were collected for two healthy animals (rs-fMRI from the two animals were acquired under no stimulation; rs-fMRI from one animal was acquired post ACC-TUS). As in previous investigations, the effects are specific to the stimulated area (fig. 5b). In two of the four macaques, the same stimulation was applied to ACC using MRI-guided frameless stereotaxy[19,33] immediately prior to nine testing sessions that were interleaved, across days, with nine control sessions in which no TUS was applied (fig.5a; Supplementary fig.4; Methods). We used a similar experimental design as in all previous fMRI sessions. There were clear differences in choice patterns between the ACC TUS and control conditions (fig.5c). For example, option 1 was often the best choice to take for most of the first part of the task (inset in fig.5c shows that this was the case for approximately the first 120 trials of the task). The frequency with which option 1 was chosen during this period was, however, reduced after TUS (Cohen's d=0.66; $t_{34}=1.92$; p=0.06). However, closer analysis revealed that option 1 was not always chosen less frequently after TUS. For example, the rate of choosing option 1 was unaffected on trials that followed those on which option 1 had previously been chosen (Cohen's d=0.36). The rate of choosing option 1 was, however, significantly reduced on trials that followed those on which it had previously been a counterfactual option – on trials on which it had previously been unavailable (Cohen's d=0.67; $t_{34}=1.97$; p=0.05, see fig.5d).

One possibility is that decisions are made differently after ACC TUS when they are difficult. Such a pattern of impairment would be expected by accounts of ACC function emphasizing monitoring the difficulty or conflict involved in action selection[8,15]. According to such accounts, decisions are difficult if the values of the options are similar. We therefore examined accuracy as a function of the difference in value between the best and worse available options (HV and LV), defined as the objective values (reward probability over the last ten trials). While once again we found evidence for a difference in ACC TUS versus control performance (Cohen's d=0.53; $t_{17}=2.31$, P=0.033) there was no evidence that TUS-induced impairment increased as difficulty increased (fig.5e; left hand side; see

supplementary fig.5d for analysis of accuracy using RL estimates); instead, if anything, the opposite was the case. In this respect the pattern of impairment is distinct to that seen after vmPFC/mOFC lesions when decision making is more impaired when decisions are difficult[34].

The fMRI analyses suggested ACC activity encodes the better counterfactual alternative but not the worse counterfactual alternative (fig.2f; 4b). Therefore, we examined whether ACC TUS diminished the influence of counterfactual options in general or diminished the influence of the better counterfactual option on behavior. We regressed the frequency with which monkeys switched, on one trial, onto the values of choices that, on a previous trial, had been counterfactual alternatives (fig.5d). As in previous analyses, without TUS, the value of the better counterfactual option significantly influenced the frequency with which monkeys subsequently switched to it (Cohen's d=1.57; t_{17}=6.7; P=3.62x10^{-6}) but this was not the case for the worse counterfactual option (Cohen's d=0.24, t_{17}=1.03, P=0.3). This was, however, not true for the TUS condition. When comparing control with TUS data, linear mixed-effect analysis revealed a significant difference between the effect of TUS and the influence of the best counterfactual values on switching (Cohen's d=0.70, t_{34}=2.05, P=0.04). The significant difference between the influence of the better and worse counterfactual option value on future switching behavior that was present in the baseline condition (post hoc test: Cohen's d=0.79; t_{17}=3.39; P=0.003) was abolished (Cohen's d=0.24; t_{17}=1.05; P=0.3) after ACC TUS (fig.5f).

We further hypothesized that this behavioral change would impact the monkeys' search strategies[7] and reduce the influence of entropy (the unpredictability of the environment; see Methods for computational definition of entropy) on their exploratory behavior[35]. In a running window analysis, we used the slope of entropy to predict the slope of cumulative stay choices (i.e. successive choices of the same option)[36]. As lower entropy favors exploiting knowledge to maximize gains and higher entropy favors exploring new options and discovering new outcomes, we expect to see a negative relationship between entropy and the frequency of stay choices. In the control condition, we found such a relationship (Cohen's d=-1.20; t_{28}=-6.59; P=3.77x10^{-7}) but this was not the case after ACC TUS (Cohen's d=0.04; t_{28}=0.22; P=0.82) (fig.5g). Note that, while local entropy and cumulative stay are negatively related to value difference (TUS-ACC: Cohen's d=-0.67; t_{28}=-3.65, P=0.001; SHAM-ACC: Cohen's d=-0.90; t_{28}=-4.95, P=3.17x10^{-5} supplementary fig.5a&b), we did not find any difference in the nature of the relationship between SHAM and TUS conditions (local entropy and value difference: Cohen's d=-0.03; t_{34}=-0.11, P=0.91; cumulative stay and value difference: Cohen's d=-0.28; t_{34}=0.83, P=0.41).

In a final TUS experiment, to control for the anatomical specificity of the observed effects, we examined the effect of TUS to lateral orbitofrontal cortex (lOFC) in four macaques, a brain region also associated with distinct aspects of reward-guided learning and decision making[37,38] (Methods). LOFC-TUS, however, had no impact on the way in which counterfactual choice value was translated into subsequent actual behavioral switching (supplementary fig.6). There was no difference for the effect of the best counterfactual on switching behaviours between the TUS-lOFC and SHAM-lOFC (Cohen's d=0.19; t_{19}=0.58, P=0.56; similarly if we only apply the test to the same two animals that had been examined

in the TUS-ACC experiment: Cohen's d=0.21; t_9=0.46, P=0.66). Further direct comparisons between TUS-lOFC and TUS-ACC showing significant differences between the two types of TUS are reported in supplementary figure 6. Additionally, there was no difference between the strength of the relationship between entropy and cumulative stay in TUS-lOFC and SHAM-lOFC condition (Cohen's d=0.32; t_{19}=0.99, P=0.33).

The unavailable option value affects the current value comparison via vmPFC/mOFC

One other area, vmPFC/mOFC, also carried a choice value comparison signal (fig.4a and fig.6b). This pattern of decision-related fMRI activity in vmPFC/mOFC has been reported previously in macaques[38]. Given vmPFC/mOFC's importance for many aspects of decision making[34,38], it is noteworthy that unlike ACC, vmPFC/mOFC activity reflecting better and worse counterfactual values did not predict behavioral switches on future trials (as per results presented in fig4c). Instead, vmPFC/mOFC is concerned with the decision being taken now rather than in the future. In the following analyses, however, we tested whether the value of the unavailable option was associated with any other impact on vmPFC/mOFC.

We first assessed whether the unavailable option's value was associated with any variation in monkeys' choices between available options. We computed accuracy (HV selection) and used a logistic regression to predict this categorical variable as a function of the unavailable option's value (including HV and LV in the model). Our results show that the higher the value of the unavailable option, the better animals were at discriminating between the two available options (Cohen's d=0.76; t_{24}=3.79; P=0.0005; similar results were obtain using a mixed-effect logistic regression model including sessions and animals as random effects using the *lmer4* package in the R environment: $\chi^2_{(1)}$=25.78; P<0.001). To illustrate this effect, we represented frequency of choosing an option (for example the Right option) as a function of the value difference between the two available options (Right-Left option values) for two different levels of the unavailable option values (high vs. low; median split). Importantly, although the unavailable option can never be chosen, its value is associated with a change in the efficiency of choice behavior (fig.6a; Cohen's d=-0.53; t_{24}=-2.66, P=0.01; see supplementary fig.7 for individual animal details), relative choice curves were steeper when the unavailable option had high versus low values.

To examine vmPFC/mOFC activity, we used a literature-based ROI selection (in area 11m/11; fig.5b, left). We focused on activity reflecting the value difference guiding decisions between available options (chosen value–unchosen value) and binned it according to the value of the unavailable option (low: 0-33%; middle: 33-66%; high: 66-100% percentiles of unavailable option value). The vmPFC/mOFC response to the chosen value–unchosen value difference was modulated by the currently unavailable option's value (linear mixed-effect analysis: Cohen's d=-1.15; t_{10}=-4.01, P=0.002; fig.6b, right panel), in exactly the same way as behavior. Normally vmPFC/mOFC activity reflects the value of the chosen option with a negative sign (fig.4b and fig.6d); as the chosen option's value falls and choosing it becomes more difficult, there is more activity in vmPFC/mOFC. This negative signal was diminished when the unavailable option value was very low and decisions between available options were less accurate. In summary, low (high) value unavailable options were associated with weaker (stronger) vmPFC/mOFC value comparison signals and

weaker (stronger) current decision accuracy. Importantly, the same analysis in the ACC and lPFC (both hemispheres) shows that the other areas behave differently and did not represent such modulation of value comparison by the unavailable option (all Ps>0.25).

To further test the strength of the link between the contextual factor's impact on the current decision and its neural impact in vmPFC/mOFC we exploited variability in the behavioral effect across sessions. We hypothesized that variation across sessions in the size of the contextual influence on vmPFC/mOFC would be related to variation in behavioral accuracy. To test this hypothesis, we first performed a partial regression analysis to reveal the uncontaminated effect of the contextual effect associated with the unavailable option's value on accuracy after controlling for the effects of the available options' values (Cohen's d=0.56; t_{24}=2.84, P=0.008; see fig.6c). Separately, we extracted the contextual effect associated with the unavailable option's value-related signal change across sessions (time course analysis performed with the GLM2, see fig.6d for illustration of the chosen and unavailable options). Sessions with a greater contextual impact on the value-related signal in the vmPFC/mOFC also exhibited a higher contextual impact on accuracy in the current trial (Pearson R=0.58, P=0.002, see fig.6e).

Discussion

Decision making is not only guided by accumulation of sensory evidence in favor of one choice over another but also by the values associated with choices that are currently unavailable but stored in memory[2]. It is both essential and a burden to store currently unavailable choice values when other choices are actually being taken at the current point in time. On the one hand, it is essential to retain unavailable choice values to *guide future behavior*; choices that are currently unavailable may be taken in the future if they become available again, if the value of the choice currently taken diminishes, if the current choice is no longer available, or if the value of the unavailable choice exceeds that of other alternatives offered in the future. On the other hand, holding information about unavailable choice values is a burden because it *distracts from the current choice* to be taken. Our results demonstrate that the value of a currently unavailable option is represented in the hippocampus (fig.3) where it is isolated from the values of the choices immediately available; currently available choice options have little effect on hippocampal activity (Supplementary fig.2). In accordance with several previous studies from our laboratory[7,24,34,39] and others[40,41] an area in mOFC/vmPFC is important for comparing the values of potential choices during the decision process. If, however, information about the currently unavailable option (or potentially some other factor that is correlated with the unavailable option's value but which is equally irrelevant to current performance) impacts on mOFC/vmPFC (fig.6) then this distracts animals from the current choice to be taken. By contrast translating the currently unavailable choice's value into a counterfactual plan that can be executed in the future depends on ACC (figure 4c). In line with this account ACC TUS disrupts the influence that counterfactual choice values have on behavioral switching (fig.5f) but it does not impact on the disruptive effect associated with an unavailable option's value on the current choice that is being made (supplementary fig.7). More broadly our results are in accordance with a view that decision making is not accomplished by any single area in isolation but by multiple areas such as mOFC/vmPFC and ACC on the basis of

different criteria[42,43]. ACC is especially concerned with signaling the value of behavioral change and alternative courses of action[7,44,45].

Like ACC, lPFC held counterfactual choice values. In this respect, lPFC activity resembles that seen in or near human FPl[3–6,11,12]. The cytoarchitecture of the macaque lPFC region studied here is not homologous with human FPl cytoarchitecture[46]. There are therefore two ways in which the current findings might be related to previous findings in humans. First, the encoding of counterfactual choice values in humans may have been incorrectly attributed to FPl and ought to be attributed to a specialized part of area 46 located in anterior prefrontal cortex that is distinct to more posterior regions 9/46v and 9/46d[47]. Alternatively, FPl may be a comparatively new and specialized region in humans. While we know that human FPl and FPm share cytoarchitectonic features it is possible that some of the circuit level interactions and functions of macaque 46 are associated with FPl in humans[11]. When species diverge over the course of evolution, what was originally a single area may become duplicated in one species but not another and connections previously associated with another area may become associated with the new area[48].

Notably, while lPFC held counterfactual choice values in a relatively straightforward manner that was unaffected by the likelihood that they would influence a change in behavior, this was not the case in ACC (fig.4). By contrast, both fMRI and TUS results suggest ACC is concerned with the translation of counterfactual information into a change of behavior.

ACC and lPFC have both been linked to the use of counterfactual information in macaques in previous neurophysiological recording studies[13,14]. One advantage of the approach taken in the present study is that we were able to record activity from both regions simultaneously and from the hippocampus and vmPFC/mOFC. The previous studies focused on the use of counterfactual feedback – after making a choice. By contrast, here we focus on how this information is held at the time of decision making while another choice is actually taken. In addition, we consider how counterfactual information is held even when a choice is temporarily unavailable.

While hippocampus, dlPFC, and ACC hold information about currently unavailable choices to guide future behavioral change, other mechanisms associated with vmPFC/mOFC have been linked to comparison of the values of specific choice options on the current trial (Figure 7). Information about currently unavailable choices is not relevant for such a mechanism but if it impinges on it then it *distracts from the current choice* to be taken. Although the presence of high value distracting information can impair decision making via a process of divisive normalization of choice values[39,49] so can distracting low value choice information[39]. The two effects may depend on the distinct manner in which choices are encoded in intraparietal cortex and vmPFC/mOFC respectively and it is possible that they may even act to cancel one another in many situations. However, manipulations to augment or diminish the influence of one mechanism or another may reveal one type of distracting influence more clearly. For example, while low value distractors may disrupt decision making via vmPFC/mOFC, in the absence of vmPFC/mOFC, the opposite effect prevails and decisions are particularly vulnerable to disruption by high value alternatives[34,50].

Online Methods

Subjects

Four male rhesus monkeys (Macaca mulatta) were involved in the experiment. They weighed 10.4–11.9 kg and were 7 years of age. They were group housed and kept on a 12 hr light dark cycle, with access to water 12–16 hr on testing days and with free water access on non-testing days. All procedures were conducted under licenses from the United Kingdom (UK) Home Office in accordance with the UK Animals (Scientific Procedures) Act 1986 and with the European Union guidelines (EU Directive 2010/63/EU).

Four animals were trained to perform the behavioural task in the MRI scanner. FMRI data from all four animals are reported. In a second part of the study we investigated the effect of TUS. Because of the positions of the head posts in two animals it was only possible to place the TUS cones to target ACC in two animals. It was, however, possible to apply TUS to the lateral location appropriate for targeting lOFC in all four animals.

Behavioral Training

Prior to the data acquisition, all animals were trained to work in an MRI compatible chair in a sphinx position that was placed inside a custom mock scanner simulating the MRI scanning environment. They were trained to use custom-made infra-red touch sensors to respond to abstract symbols presented on a screen and learned the probabilistic nature of the task until reaching a learning criterion. The animals underwent aseptic surgery to implant an MRI compatible head post (Rogue Research, Mtl, CA). After a recovery period of at least 4 weeks, the animals were trained to perform the task inside the actual MRI scanner under head fixation. The imaging data acquisition started once they performed at more than 70% accuracy (choosing the option with the highest expected value) for at least another three consecutive sessions in the scanner.

Experimental task

Animals had to choose repeatedly between different stimuli that were novel in each testing session (fig.1a). We used a probabilistic reward-based learning task inspired from tasks originally designed to study reinforcement learning. Choice options were allocated pseudo randomly to the right and left side of the screen and monkeys responded with a right or left infra-red sensor placed in front of each of their hands. The rewards were delivered probabilistically and the probabilities associated with the three options fluctuated during the entire session, with the probability of two of the options changing towards the middle of a session (fig.1c). Thus, the probability range for option A was [90% to 10%], the probability range for option B was [70% to 30%] and the probability range for option C was [10% to 90%]. Importantly, each day the task contained three choice stimuli, but only two of them were choosable on each trial (fig.1b). This manipulation alters the learning and decision task in two major ways. First, the subjects have to maintain in memory the value of the option that is not directly available. Second, it creates a horizon of choices that is not deterministic, as the animal cannot predict what option will be presented next. After making their decision, if an option selected led to a reward (as per the reward contingencies associated with each option), the unselected option disappeared and the chosen option remained on the screen and

a juice reward was delivered. If an option selected led to no-reward, no juice was delivered. The outcome phase lasted 1.5 seconds. Each reward was composed of two 0.6 ml drops of blackcurrant juice delivered by a spout placed near the animal's mouth during scanning. Each animal performed up to 200 trials per session. Each animal performed five to seven sessions in the MRI scanner. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[51]. The experiment was controlled by Presentation software (Neurobehavioral Systems Inc., Albany, CA).

Because very slow response trials may have been subject to interference in the choice selection process they were excluded from the fMRI analysis of choice selection (which was time-locked to the onset of stimulus presentation) or in the other behavioral analyses linked to these: trials with reaction times (RTs) more than 3 standard deviations from the log-transformed RT median were not included in the fMRI analysis (0.3% of trials were excluded in this way).

Reinforcement-learning algorithms

We used four reinforcement-learning algorithms (Maintain model, Decay model, Maintain model with distortion and the Decay model with distortion) to estimate trial-by-trial expected values associated with each option using animals' responses[52]. For all models, if stimulus A was selected on trial i, its value was updated via a prediction error, δ, as follows: $v_A(i+1) = v_A(i) + a.\delta(i)$ where a is the learning rate and the prediction error was given by $\delta(i) = r(i) - v_A(i)$. The values of the unselected stimulus (e.g. B) were not updated. The two first models differ in their assumptions of the stimulus that was not shown on that trial (e.g. C). In the Maintain model, the values of C were maintained at their current values such that $v_c(i+1) = v_c(i)$. In the Decay model, the values of C were updated as followed: $v_c(i+1) = v_c(i) + \gamma.(v_c(1) - v_c(i))$. The third and fourth model assumed that subjective value can be distorted by risk preference. Please note, however, that while probability distortion might make a reward probability appear higher or lower than it might otherwise be, it cannot lead to re-ordering of option values, as it is a strictly monotonic function. For these models[23,53,54], we fitted an additional free parameter η using the following equation:

$$w_A\left(v_A\right) = \frac{v_A^{\eta}}{\left[v_A^{\eta} + (1 - v_A)^{\eta)}\right]}, \text{ with } 0 < V < 1$$

To generate choices for both models, we first used a softmax procedure in which, on every trial, the probability of choosing stimulus A was given by: $PA(i) = \sigma(\beta(v_A(i) - v_B(i)))$ or $PA(i) = \sigma(\beta(w_A(i) - w_B(i)))$ for the distortion models where $\sigma(z) = 1/(1 + e^{-z})$ is the logistic function, and β the degree of stochasticity in making the decision. The model choice probabilities were then fitted against the discrete behavioral choices to estimate the free parameters (a, β, γ, η).

Model fitting

To estimate the free parameters (a, β, γ, η), we used a maximum likelihood estimation and a constrained non-linear optimization procedure (as implemented in *fmincon* in MATLAB)

separately for each session. The associated likelihood function was given by: $logL = \frac{\sum B_A \log P_A + \sum B_B \log P_B}{N_A + N_B}$ where N_A and N_B denote, the number of trials in which stimulus A and B were chosen, and B_A (B_B) equals 1 if A (B) was chosen on that trial, and 0 otherwise. We fitted this function similarly for the other two stimulus combinations (AC and BC) and found the optimal parameters by minimizing the sum of the three negative log-likelihoods.

Statistical analyses

For most analyses, we ran multiple linear or logistic regressions using Matlab (glmfit, robustfit). For logistic functions, we used a logit link with categorical predictors. All regressors were normalized (as in all fMRI regression analyses) in order to ensure between-model, between-session and between-modulator commensurability of the regression coefficients. For each session, we obtained one β regression weight for each regressor. These were then tested for statistical significance across all participants using either ANOVAs or t-tests. When assumptions about statistical tests were violated (data normality was tested by visually inspecting the residuals from the regressions), we transformed the data using a square root transform. All data were shown as mean with standard error of mean (mean ± SEM). Probabilities of $P < 0.05$ were considered as significant.

Reinforcement learning simulation

To characterize the effect of delay and probability distortion over the maintain model assumptions, we generated for each trial t the probability of choosing the best option according to the models, given the animals' history of choices and outcomes at trial t−1 and the individual best-fitting free parameters. We submitted all model-simulated choice probabilities to the same statistical analyses described below. In a first analysis (left panel in Fig S1c), we were interested in investigating whether the different models made distinct predictions as a function of the elapsed time since the unavailable option was last seen. To do so, we used both simulated and real choice data to compare switches to the unavailable option when the latter had been unavailable for 1, 2 or 3 consecutive trials. (Please note that the variance is significantly different in the three bins as the number of times that an option is the same for three consecutive trials is very limited (bin1: mean = 150; bin2: mean = 36, bin3: mean = 5). Secondly (right panel in Fig S1c), given the same model simulations, we investigated choice patterns before and after reversal. For this analysis, we looked at the choice frequency for each option before and after the 120th trial. Third (Fig S1d), the last feature of the data characterizing the task is the influence of valence (win/loss) on switch/stay pattern. We thus compare the frequency of switch behavior after a win/loss.

Imaging Data Acquisition

Awake-animals were head-fixed in a sphinx position in an MRI-compatible chair. We collected fMRI using a 3T MRI scanner and a four-channel phased array receive coil in conjunction with a radial transmission coil (Windmiller Kolster Scientific Fresno, CA). FMRI data were acquired using a gradient-echo T2* echo planar imaging (EPI) sequence with $1.5 \times 1.5 \times 1.5$ mm^3 resolution, repetition time (TR) = 2.28 s, Echo Time (TE) = 30 ms, flip angle = 90, and reference images for artifact corrections were also collected. Proton-density-weighted images using a gradient-refocused echo (GRE) sequence (TR = 10 ms, TE

= 2.52 ms, flip angle = 25) were acquired as reference for body motion artifact correction. T1-weighted MP-RAGE images ($0.5 \times 0.5 \times 0.5$ mm^3 resolution, TR = 2,5 ms, TE = 4.01 ms) were acquired in separate anesthetized scanning sessions.

fMRI data preprocessing

FMRI data were corrected for body motion artefacts by an offline-SENSE reconstruction method[55] (Offline_SENSE GUI, Windmiller Kolster Scientific, Fresno, CA). The images were aligned to an EPI reference image slice-by-slice to account for body motion and then aligned to each animal's structural volume to account for static field distortion[56] (Align_EPI GUI and Align_Anatomy GUI, Windmiller Kolster Scientific, Fresno, CA). The aligned data were processed with high-pass temporal filtering (3-dB cutoff of 100s) and Gaussian spatial smoothing (full-width half maximum of 3mm). The data that were already registered to each subject's structural space were then registered to the CARET macaque F99 template[57] using affine transformation.

fMRI data analysis

We employed a univariate approach within the general linear model (GLM) framework to perform whole-brain statistical analyses of functional data as implemented in the FMRIB Software Library[58,59]: $Y = X\beta + \varepsilon = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_N X_N + \varepsilon$ where Y is a T×1 (T time samples) column vector containing the times series data for a given voxel, and X is a $T \times N$ (N regressors) design matrix with columns representing each of the psychological regressors convolved with a hemodynamic response function specific for monkey brains[60,61]. β is a $N \times 1$ column vector of regression coefficients and ε a $T \times 1$ column vector of residual error terms. Using this framework we initially performed a first-level fixed effects analysis to process each individual experimental run which were then combined in a second-level mixed-effects analysis (FLAME 1 + 2) treating session as a random effects. For all analysis, we performed a cluster inference using a cluster-defining threshold of |Z| > 3.1 with a FWE-corrected threshold of P = 0.001. Time series statistical analysis was carried out using FMRIB's improved linear model with local autocorrelation correction. Applying this framework, we performed the GLMs highlighted below.

GLM1 – correct vs. incorrect future selection of the currently unavailable option

Our first fMRI analysis was designed to reveal the brain regions representing the value of the currently unavailable option to guide accurate future decision making. Specifically, locked to the decision time, we included a first boxcar regressor parametrically modulated by reaction times (RTs) to account for difficulty effects, as well as 2 boxcar regressors with a duration of 100 ms that were then convolved with the hemodynamic response function: 1) an modulated regressor indexing the occurrence of a decision (Dec; all event amplitudes set to one and the duration set to the RT for that trial), 2-3) two parametric regressors whose event amplitudes were modulated by the expected value of the unavailable option for i) future correct selection (unav$_{corr}$) and ii) future incorrect selection (unavincorr). Additionally, we included two fully parametric regressors whose event amplitudes were modulated by the expected value of the chosen (Ch) and unchosen (Unch) options that were available on the current trial. Locked to feedback time we included a binary regressor representing positive and negative feedback (+1/-1) and a categorical regressor representing right and left responses

(+1/-1), such as: $Y = \beta_1\, Dec + \beta_2\, unav_{cor} + \beta_3\, unav_{incor} + \beta_4\, Ch + \beta_5\, Unch + \beta_6\, Fbk + \beta_7\, Side + e$. Finally, to further reduce variance and noise in the BOLD signal, we add two unconvolved regressors locked at time of feedback and with a duration of a TR (2.28sec) for left and right responses (to capture variance in the BOLD signal caused by any field distortion coincident with responding), six nuisance regressors one for each of the motion parameters (three rotations and three translations), and extra single-trial nuisance covariates for abrupt changes in the BOLD signal.

GLM2 – Subjective choice comparison (Chosen option value vs. Unchosen option value)

Our second fMRI analysis was designed to reveal the brain regions representing the decision variable guiding choices between the options actually available on the current trial (Chosen option value-Unchosen option value). Locked to decision time, we included a first boxcar regressors parametrically modulated by RTs (to account for difficulty effects), as well as 3 boxcar regressors with a duration of 100 ms that were then convolved with the hemodynamic response function: 1) an modulated regressor indexing the occurrence of a decision (Dec; all event amplitudes set to one and the duration set to the RT for that trial), 2-4) three fully parametric regressors whose event amplitudes were modulated by the expected value of the chosen option (Ch), unchosen option (Unch) and unavailable option (Unav) and the same covariates of non-interest as described in GLM1: $Y = \beta_1\, Dec + \beta_2\, Ch + \beta_3\, Unch + \beta_4\, Unav + \beta_6\, Fbk + \beta_7\, Side + e$. In the third GLM (GLM3: counterfactual model), the Unchosen and Unavailable options were replaced by the Better and the Worse alternatives, in the fourth GLM (GLM4: difficulty model), the Chosen and Unchosen options were replaced by the High Value option and the Low Value option presented and finally, in the fifth GLM (GLM5: object identity model), the Chosen, Unchosen and Unavailable options were replaced by the values of Option 1, 2 and 3 (see figure 1, supplementary fig.3).

Neural model comparison

To assess goodness of fit at the neural level and avoid double dipping in favor of the hypothesis that we wanted to support (GLM3)[27], we first defined from GLM2, several ROIs within a network including all the brain areas that survived cluster level P < 0.001 (cluster-based correction) for the value comparison (chosen-unchosen) contrast. Within this network, we derived the log-evidence from GLM2, GLM3, GLM4 and GLM5. Log evidence was then fed into a Bayesian model selection random effects analysis (using the spm_BMS routine), which computed the exceedance probability of each GLM for each ROI. This analysis indicates which GLM best explained the neural data. We report the results for ACC, lPFC, and vmPFC/mOFC.

ROI analyses

We conducted analyses on ROIs defined as two-voxel radius spherical masks placed over the hippocampus (Right: x = 16.5, y = -7.5, z = -12; left: x = -14, y = -9, z = -12.5 CARET macaque F99 coordinates), ACC (x = 1, y = 20.5, z = 10.5), lPFC (x = 14.5, y = 17.5, z = 9.5), vmPFC/mOFC (x = -5, y = 14, z = 2). We used procedures now standardly employed in most human and animal neuroimaging studies[51,62,63] in which the mean and standard error (denoted in all figures by lines and shadings respectively) of all the within-subject b weights

were calculated across sessions for plotting the effect size time courses (each animal had a similar number of sessions).

Leave one out for ROI spatial peak selection and time-series group peak signal

We used two leave-one-out procedures to avoid circularity in our analyses. The first aimed at = identifying ROI peak voxels for the analyses of main effects for areas identified in all fMRI analyses. For each group level analyses, our procedure involved leaving one session out at a time. From the results of the remaining 24 sessions, we extracted local maxima of the relevant clusters and centered the ROIs for the left out session on the local maxima. We repeated this for all sessions. Therefore, the ROI selection was statistically independent from the data of the session that was subsequently analyzed in the ROI. We also used a leave-one-out procedure on the group peak signal to avoid potential temporal selection biases. For every session, we calculated the time course of the group mean beta weights of the relevant regressor based on the remaining 24 sessions. We then identified the (positive or negative) group peak of the regressor of interest within the analysis window of 1 to 6 seconds from decision onset. Then, we took the beta weight of the remaining subject at the time of the group peak. We repeated this for all subjects. Therefore, the resulting 25 "peak" beta weights were selected independently from the time course of the subject analyzed. We assessed significance using t-tests on the resulting beta weights.

Transcranial Focused Ultrasound Stimulation (TUS)

A single element ultrasound transducer (H115-MR, diameter 64 mm, Sonic Concept, Bothell, WA, USA) with a 51.74 mm focal depth was used with a coupling cone filled with degassed water and sealed with a latex membrane (Durex). The ultrasound wave frequency was set to the 250 kHz resonance frequency and 30 ms bursts of ultrasound were generated every 100 ms with a digital function generator (Handyscope HS5, TiePie engineering, Sneek, The Netherlands). Overall, the stimulation lasted for 40 s. A 75-Watt amplifier (75A250A, Amplifier Research, Souderton, PA) was used to deliver the required power to the transducer. A TiePie probe connected to an oscilloscope was used to monitor the voltage delivered. The recorded peak-to-peak voltage was constant throughout the stimulation session. Voltage values per session ranged from 128 to 136V and corresponded to a peak negative pressure of 1.152 to 1.292MP respectively measured in water with an in house heterodyne interferometer (see[64] for more details about the simulation protocol). Based on a mean 66% transmission through the skull[65], the estimated peak negative pressures applied ranged from 0.76 to 0.85 MPa at the target in the brain.

The transducer was positioned with the help of a Brainsight neuronavigating system (Rogue Research, Montreal, CA) so that the focal spot would be centered on the targeted brain region, namely the rACC (F99 coordinates x = 1, y = 20.5, z = 10.5) (identified according to coordinates of the maximum peak used in GLM2). The ultrasound transducer / coupling cone montage was directly positioned to previously shaved skin on which conductive gel (SignaGel Electrode; Parker Laboratories Inc.) had been applied. The coupling cone filled with water and gel was used to ensure ultrasonic coupling between the transducer and the animal's head.

A sham TUS condition (SHAM) was also implemented as a non-stimulation control. Sham sessions were interleaved with TUS sonication days and completely mirrored a typical stimulation session (setting, stimulation procedure, neuro-navigation, targeting of ACC, transducer preparation and timing of its application to the shaved skin on the head of the animal) except that sonication was not triggered.

To test for the specificity of TUS on the ACC, we collected 20 SHAM-lOFC and 20 TUS-lOFC (4 animals × 5 sessions) using the same experimental design as the TUS-ACC protocol. Two out of the four animals tested were also used in the TUS-ACC protocol. TUS and control days were interleaved in one of two pseudorandom orders that were counterbalanced across animals in each experiment. For example (T,T, R, S,S, R, T,T,T, R) where T, C, and R stand for TUS, sham, and rest days respectively – note a rest day always intervened at the point of transition between TUS and sham days. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[66]. Data collection and analysis were not performed blind to the conditions of the experiments.

Finally, given that the TUS procedure lasts for 40s and has a relatively sustained impact on neural activity, it will be possible in future experiments to examine the impact of ACC stimulation while recording activity from the ACC and interconnected areas either with fMRI or some other technique. However, if experiments of this type are to be attempted it will be possible to conduct them only after initially carrying experiments of the sort that we report here; it is necessary to establish the precise location of a neural signal before it can be targeted with the spatially focal TUS technique.

Entropy analyses

For the analyses presented in Fig. 5 (behavioral analysis of TUS data), we used a running window analysis with entropy defined as: $E(i) = \sum_{i=1}^{trials} p(x_{i,j}) \cdot log(p(x_{i,j}))$, in which $x_{i,j}$ is the probability that a given option j is associated with a positive feedback on trial i. We then used the slope of entropy (difference between the beginning and the end of a window of 20 trials) as a measure of environmental predictability. A positive change in entropy reflects that the environment is less and less predictable and should trigger exploration whereas a negative change in entropy should engage exploitative behavior. As a proxy for exploration/exploitation, we used the cumulative sum of stay behavior, which is simply a vector, keeping track of the number of times a choice has been chosen. Note that a consecutive stay for an option A that has been chosen on trial t could also include trials for which on the next trial (t +1) A would not be available but chosen on the subsequent trial (t+2).

vmPFC partial regression analysis

To test the strength of the link between the unavailable option's impact on the current decision and its neural impact in vmPFC/mOFC, we computed the accuracy residuals (Y*, from regressing accuracy against the values of the two available options omitting the unavailable one) and the unavailable residuals (X* from regressing the unavailable option value against the values of the two observable options) and then regressed Y* against X*[67] for each session separately (see average effect on Fig.6c).

Macaque rs-fMRI Data Acquisition, Preprocessing, and Analysis

Resting state fMRI (rs-fMRI) and anatomical MRI scans were collected for two healthy animals (rs-fMRI from the two animals were acquired under no stimulation; rsfMRI from one animal was acquired post ACC-TUS) under inhalational isoflurane anesthesia using a protocol which was previously proven successful[68,69] in preserving whole-brain functional connectivity as measured with BOLD signal. In the case of the TUS conditions, we used the same procedure as the one employed by[17,18]. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[66].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request

Code Availability Statement

The code to generate the results and the figures of this study are available from the corresponding author upon reasonable request.

References

1. Noser R, Byrne RW. Mental maps in chacma baboons: using inter-group encounters as a natural experiment. Anim Cogn. 2007; 10:331–340. [PubMed: 17372775]

2. Shadlen MN, Shohamy D. Decision Making and Sequential Sampling from Memory. Neuron. 2016; 90:927–939. [PubMed: 27253447]

3. Boorman ED, Behrens TE, Rushworth MF. Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex. PLOS Biol. 2011; 9:e1001093. [PubMed: 21738446]

4. Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS. How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. Neuron. 2009; 62:733–743. [PubMed: 19524531]

5. Scholl J, et al. The Good, the Bad, and the Irrelevant: Neural Mechanisms of Learning Real and Hypothetical Rewards and Effort. J Neurosci. 2015; 35:11233–11251. [PubMed: 26269633]

6. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441:876–879. [PubMed: 16778890]

7. Kolling N, Behrens TEJ, Mars RB, Rushworth MFS. Neural Mechanisms of Foraging. Science. 2012; 336:95–98. [PubMed: 22491854]

8. Kolling N, Behrens T, Wittmann M, Rushworth M. Multiple signals in anterior cingulate cortex. Curr Opin Neurobiol. 2016; 37:36–43. [PubMed: 26774693]

9. Aggleton JP, Wright NF, Rosene DL, Saunders RC. Complementary Patterns of Direct Amygdala and Hippocampal Projections to the Macaque Prefrontal Cortex. Cereb Cortex N Y N 1991. 2015; 25:4351–4373.

10. Jang AI, et al. The Role of Frontal Cortical and Medial-Temporal Lobe Brain Areas in Learning a Bayesian Prior Belief on Reversals. J Neurosci. 2015; 35:11751–11760. [PubMed: 26290251]

11. Neubert F-X, Mars RB, Thomas AG, Sallet J, Rushworth MFS. Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. Neuron. 2014; 81:700–713. [PubMed: 24485097]

12. Bludau S, et al. Cytoarchitecture, probability maps and functions of the human frontal pole. NeuroImage. 2014; 93(2 Pt):260–275. [PubMed: 23702412]

13. Abe H, Lee D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. Neuron. 2011; 70:731–741. [PubMed: 21609828]

14. Hayden BY, Pearson JM, Platt ML. Fictive Reward Signals in Anterior Cingulate Cortex. Science. 2009; 324:948–950. [PubMed: 19443783]

15. Kolling N, et al. Value, search, persistence and model updating in anterior cingulate cortex. Nat Neurosci. 2016; 19:1280. [PubMed: 27669988]

16. Miyamoto K, et al. Causal neural network of metamemory for retrospection in primates. Science. 2017; 355:188–193. [PubMed: 28082592]

17. Verhagen L, et al. Offline impact of transcranial focused ultrasound on cortical activation in primates. eLife. 2019; 8:e40541. [PubMed: 30747105]

18. Folloni D, et al. Manipulation of Subcortical and Deep Cortical Activity in the Primate Brain Using Transcranial Focused Ultrasound Stimulation. Neuron. 2019; 0

19. Deffieux T, et al. Low-intensity focused ultrasound modulates monkey visuomotor behavior. Curr Biol CB. 2013; 23:2430–2433. [PubMed: 24239121]

20. Fouragnan E, Queirazza F, Retzler C, Mullinger KJ, Philiastides MG. Spatiotemporal neural characterization of prediction error valence and surprise during reward learning in humans. Sci Rep. 2017; 7

21. Fouragnan E, Retzler C, Mullinger K, Philiastides MG. Two spatiotemporally distinct value systems shape reward-based learning in the human brain. Nat Commun. 2015; 6

22. Niv Y, et al. Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. J Neurosci. 2015; 35:8145–8157. [PubMed: 26019331]

23. Klein-Flügge MC, Bestmann S. Time-dependent changes in human corticospinal excitability reveal value-based competition for action during decision processing. J Neurosci Off J Soc Neurosci. 2012; 32:8373–8382.

24. Hunt LT, et al. Mechanisms underlying cortical activity during value-guided choice. Nat Neurosci. 2012; 15:470–476. [PubMed: 22231429]

25. Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci. 2008; 9:545–556. [PubMed: 18545266]

26. Kolling N, Wittmann M, Rushworth MFS. Multiple Neural Mechanisms of Decision Making and Their Competition under Changing Risk Pressure. Neuron. 2014; 81:1190–1202. [PubMed: 24607236]

27. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci. 2009; 12:535–540. [PubMed: 19396166]

28. Martin VC, Schacter DL, Corballis MC, Addis DR. A role for the hippocampus in encoding simulations of future events. Proc Natl Acad Sci. 2011; 108:13858–13863. [PubMed: 21810986]

29. Palminteri S, Khamassi M, Joffily M, Coricelli G. Contextual modulation of value signals in reward and punishment learning. Nat Commun. 2015; 6

30. Wittmann MK, et al. Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. Nat Commun. 2016; 7

31. Schuck NW, et al. Medial prefrontal cortex predicts internally driven strategy shifts. Neuron. 2015; 86:331–340. [PubMed: 25819613]

32. Procyk E, Tanaka YL, Joseph JP. Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. Nat Neurosci. 2000; 3:502–508. [PubMed: 10769392]

33. Paus T. Imaging the brain before, during, and after transcranial magnetic stimulation. Neuropsychologia. 1999; 37:219–224. [PubMed: 10080379]

34. Noonan MP, et al. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proc Natl Acad Sci U S A. 2010; 107:20547–20552. [PubMed: 21059901]

35. O'Reilly JX, et al. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc Natl Acad Sci U S A. 2013; 110:E3660–3669. [PubMed: 23986499]

36. Gallistel CR, Mark TA, King AP, Latham PE. The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. J Exp Psychol Anim Behav Process. 2001; 27:354–372. [PubMed: 11676086]

37. Rudebeck PH, Murray EA. The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. Neuron. 2014; 84:1143–1156. [PubMed: 25521376]

38. Papageorgiou GK, et al. Inverted activity patterns in ventromedial prefrontal cortex during value-guided decision-making in a less-is-more task. Nat Commun. 2017; 8

39. Chau BKH, Kolling N, Hunt LT, Walton ME, Rushworth MFS. A neural mechanism underlying failure of optimal choice with multiple alternatives. Nat Neurosci. 2014; 17:463–470. [PubMed: 24509428]

40. Rich EL, Wallis JD. Decoding subjective decisions from orbitofrontal cortex. Nat Neurosci. 2016; 19:973–980. [PubMed: 27273768]

41. Strait CE, Blanchard TC, Hayden BY. Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. Neuron. 2014; 82:1357–1366. [PubMed: 24881835]

42. Hunt LT, Hayden BY. A distributed, hierarchical and recurrent framework for reward-based choice. Nat Rev Neurosci. 2017; 18:172–182. [PubMed: 28209978]

43. Rushworth MFS, Kolling N, Sallet J, Mars RB. Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr Opin Neurobiol. 2012; 22:946–955. [PubMed: 22572389]

44. Hayden BY, Pearson JM, Platt ML. Neuronal basis of sequential foraging decisions in a patchy environment. Nat Neurosci. 2011; 14:933–939. [PubMed: 21642973]

45. Kolling N, Scholl J, Chekroud A, Trier HA, Rushworth MFS. Prospection, Perseverance, and Insight in Sequential Behavior. Neuron. 2018; 99:1069–1082.e7. [PubMed: 30189202]

46. Mackey S, Petrides M. Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. Eur J Neurosci. 2010; 32:1940–1950. [PubMed: 21050280]

47. Sallet J, et al. The Organization of Dorsal Frontal Cortex in Humans and Macaques. J Neurosci. 2013; 33:12255–12274. [PubMed: 23884933]

48. Krubitzer L. The magnificent compromise: cortical field evolution in mammals. Neuron. 2007; 56:201–208. [PubMed: 17964240]

49. Louie K, Khaw MW, Glimcher PW. Normalization is a general neural mechanism for context-dependent decision making. Proc Natl Acad Sci. 2013; 110:6139–6144. [PubMed: 23530203]

50. Noonan MP, Chau BKH, Rushworth MFS, Fellows LK. Contrasting Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision-Making in Humans. J Neurosci Off J Soc Neurosci. 2017; 37:7023–7035.

51. Chau BKH, et al. Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. Neuron. 2015; 87:1106–1118. [PubMed: 26335649]

52. Sutton, R. Reinforcement Learning: An Introduction. MIT Press; 1998.

53. Hunt LT, et al. Mechanisms underlying cortical activity during value-guided choice. Nat Neurosci. 2012; 15:470–476. [PubMed: 22231429]

54. Farashahi S, Azab H, Hayden B, Soltani A. On the Flexibility of Basic Risk Attitudes in Monkeys. J Neurosci Off J Soc Neurosci. 2018; 38:4383–4398.

55. Kolster H, et al. Visual Field Map Clusters in Macaque Extrastriate Visual Cortex. J Neurosci. 2009; 29:7031–7039. [PubMed: 19474330]

56. Kolster H, Janssens T, Orban GA, Vanduffel W. The retinotopic organization of macaque occipitotemporal cortex anterior to V4 and caudoventral to the middle temporal (MT) cluster. J Neurosci Off J Soc Neurosci. 2014; 34:10168–10191.

57. Van Essen DC, et al. Mapping visual cortex in monkeys and humans using surface-based atlases. Vision Res. 2001; 41:1359–1378. [PubMed: 11322980]

58. Smith SM, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004; 23(Suppl 1):S208–219. [PubMed: 15501092]

59. Fouragnan E, et al. Reputational priors magnify striatal responses to violations of trust. J Neurosci Off J Soc Neurosci. 2013; 33:3602–3611.

60. Nakahara K, Hayashi T, Konishi S, Miyashita Y. Functional MRI of Macaque Monkeys Performing a Cognitive Set-Shifting Task. Science. 2002; 295:1532–1536. [PubMed: 11859197]

61. Kagan I, Iyer A, Lindner A, Andersen RA. Space representation for eye movements is more contralateral in monkeys than in humans. Proc Natl Acad Sci. 2010; 107:7933–7938. [PubMed: 20385808]

62. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. Nat Neurosci. 2007; 10:1214–1221. [PubMed: 17676057]

63. Chau BKH, Kolling N, Hunt LT, Walton ME, Rushworth MFS. A neural mechanism underlying failure of optimal choice with multiple alternatives. Nat Neurosci. 2014; 17:463–470. [PubMed: 24509428]

64. Constans C, Deffieux T, Pouget P, Tanter M, Aubry J. F Erratum to #x201C;A 200 #x2013;1380 kHz Quadrifrequency Focused Ultrasound Transducer for Neurostimulation in Rodents and Primates: Transcranial In Vitro Calibration and Numerical Study of the Influence of Skull Cavity #x201D; IEEE Trans Ultrason Ferroelectr Freq Control. 2017; 64:1417–1417. [PubMed: 28816660]

65. Wattiez N, et al. Transcranial ultrasonic stimulation modulates single-neuron discharge in macaques performing an antisaccade task. Brain Stimul Basic Transl Clin Res Neuromodulation. 2017; 10:1024–1031.

66. Vanduffel W, Zhu Q, Orban GA. Monkey cortex through fMRI glasses. Neuron. 2014; 83:533–550. [PubMed: 25102559]

67. Velleman PF, Welsch RE. Efficient Computing of Regression Diagnostics. Am Stat. 1981; 35:234–242.

68. Noonan MP, et al. A Neural Circuit Covarying with Social Hierarchy in Macaques. PLOS Biol. 2014; 12:e1001940. [PubMed: 25180883]

69. Neubert F-X, Mars RB, Sallet J, Rushworth MFS. Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. Proc Natl Acad Sci U S A. 2015; 112:E2695–2704. [PubMed: 25947150]

Reporting Summary

Further information on research design is available in the Life Science Reporting Summary linked to this article.
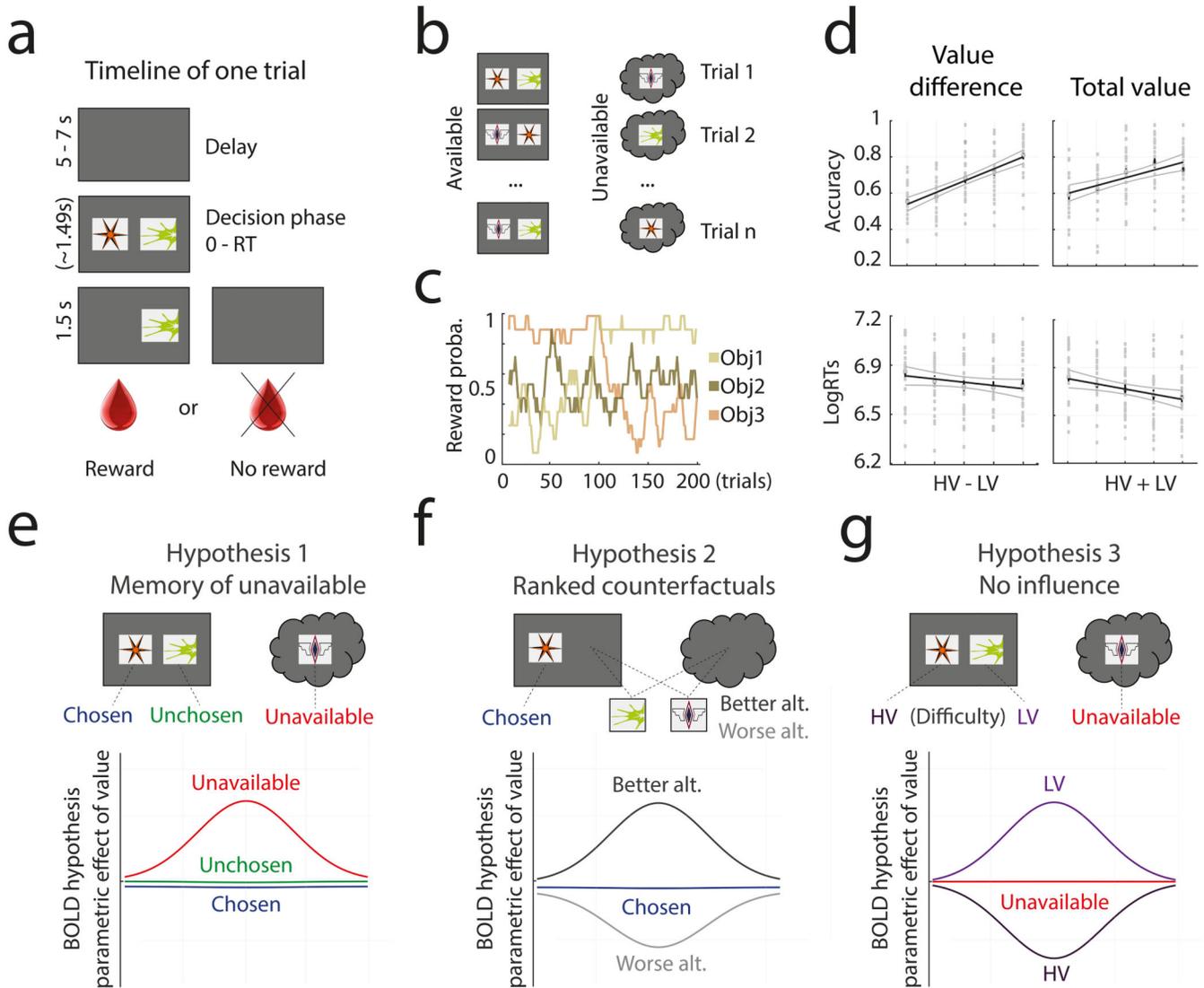
Figure 1. Schematic view of the task, behavioural results and hypothesized neural schemes.
(a) On each trial, animals could choose between two symbols presented on the screen and had to keep in mind a third option, unavailable to them. The position of each symbol on the left/right part of the screen and the combination of available/unavailable options was fully and pseudo-randomized respectively. **(b)** Each trial began with a random delay followed by the presentation of two abstract symbols for a period ending when the animals made a choice. During this time, monkeys pressed one of two touch-sensors to indicate, which of the two symbols (right or left) they believed was more likely to lead to a reward. Finally, the decision outcome was revealed for 1.5 sec. The selected symbol was kept on the screen (or not) to inform the monkeys of a reward delivery (or no reward). **(c)** The plots show the probability of receiving a reward for choosing either options 1 (pink), 2 (blue), or 3 (red) on each trial in the 200-trial sessions. **(d)** The top graphs show the proportion of correct choices (selecting the option with the highest reward probability) plotted as a function of difficulty (distance between the better high value [HV] and the worse low value [LV] presented options: **left panel**) and context value (sum of both HV's and LV's expected values: **right**

panel). Decision accuracy improved with higher value difference between available options and higher total value. The bottom graphs show log-transformed mean RTs for each session plotted as a function of difficulty and context. LogRTs decreased for easier decisions and higher trial value. Red lines are linear fits to the data and the grey lines are the 95% confidence interval, n=25 sessions. **(e)** Because each of the three options' values were uncorrelated with one another it was possible to look for neural activity according to three main coding schemes. If activity in a brain area covaries only with the value of the unavailable option then this suggests the area is concerned with representing the value of an option held in memory on the current trial and which should not interfere with decisions taken on the current trial. **(f)** If instead activity covaries with the ranked value of both the unchosen available option and the option held in memory then it reflects the value of any currently counterfactual choice that might be taken in the future. It is important, however, to distinguish such a pattern from a third possibility **(g)** in which neural activity is only reflecting the currently available options without representing the counterfactual or unavailable option. Thus, the activity would be negatively related to the HV available option value and positively related to the LV option value. This third pattern indicates that the brain area's activity reflects the difficulty or uncertainty of the current decision because the difficulty of selecting an option becomes harder as the LV option increases and as the HV option decreases but it is unaffected by the value of the choice that cannot currently be taken (see discussion by Kolling and colleagues[15]). Note that we also analysed a fourth pattern representing the value of each option separately on supplementary figure 3.
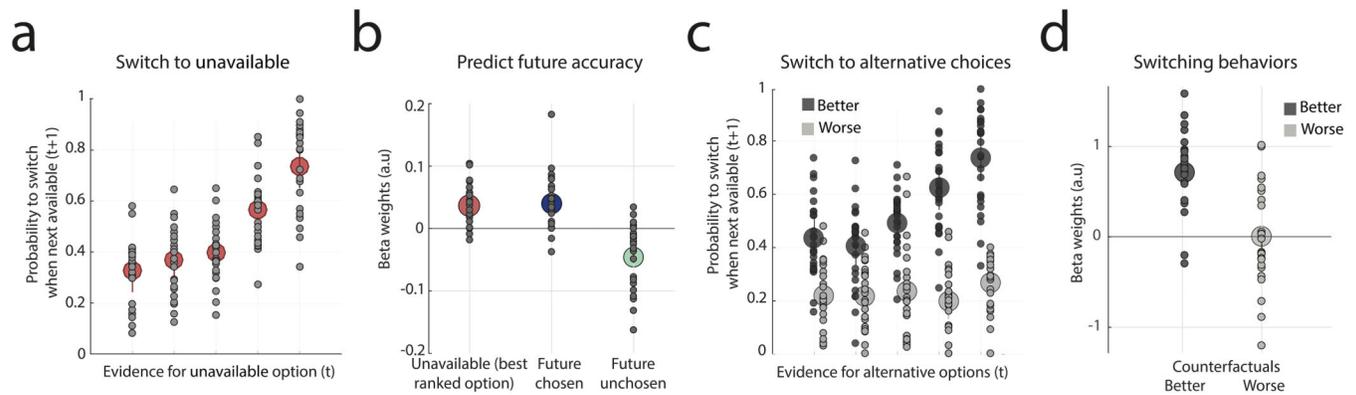
Figure 2. Future switches are explained by the expected value associated with counterfactual options.
(a) Estimated expected values associated with the unavailable option on the current trial predict whether animals switch to it when it reappears on the screen on subsequent trials (y-axis: probability of switching to the currently unavailable option. x-axis: reward probability associated with the unavailable option estimated from the Maintain model). Each bin contains 20% of averaged data across trials (individual sessions in grey dots; average across sessions in red dot). **(b)** A logistic regression confirms that accuracy is explained by the currently unavailable option's value (higher accuracy for trials in which it is the best of the three options vs. when it is not), in addition to the value of the future chosen and unchosen options (each session's beta coefficient is represented as a grey dot and the mean beta coefficients is represented as a coloured dot). **(c)** A similar analysis to the one shown in panel **(a)** is performed but on the basis of a new coding scheme where the counterfactual options (current unchosen option and current unavailable option) are ranked according to their associated reward probabilities as the better and the worse counterfactual choices. **(d)** A logistic regression confirms that the value of the better counterfactual option significantly influenced the frequency with which monkeys subsequently switched to it but this was not the case for the worse counterfactual option. One sample t-tests were used across session on the resulting beta coefficients, n=25, for all analyses.
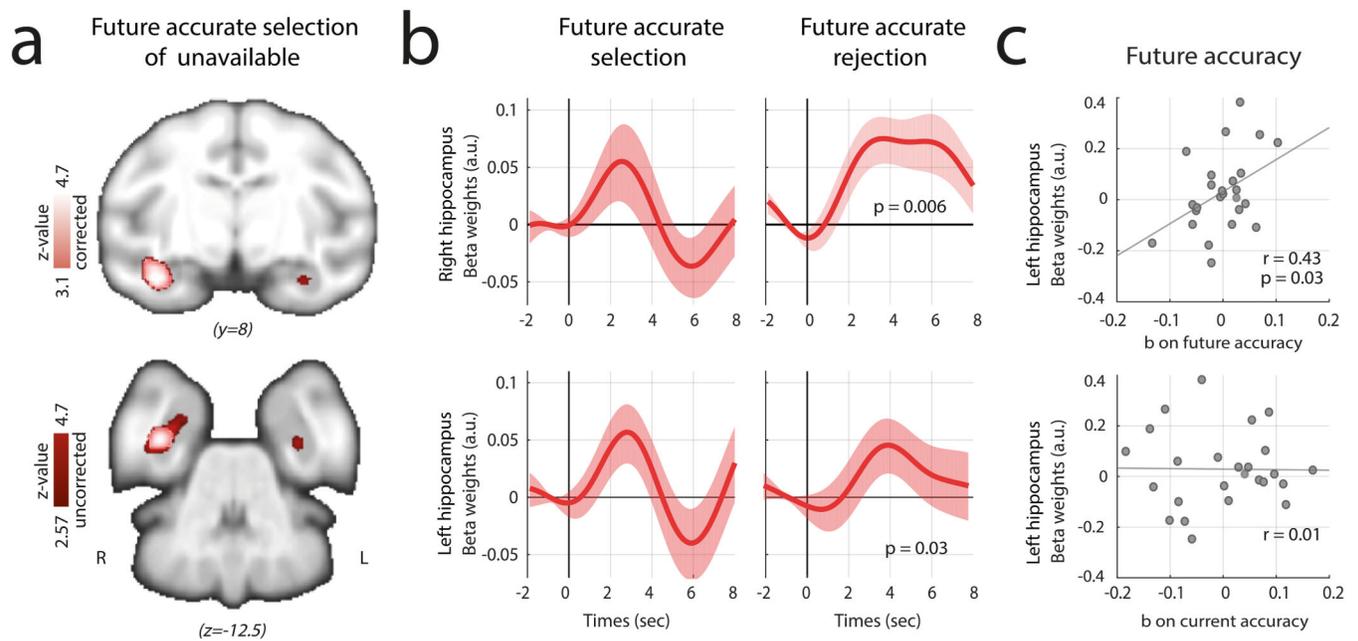
Figure 3. Unavailable option value signal in hippocampus favors accurate future planning.
(a) A whole-brain analysis tested for voxels where activity correlated with the trial-by-trial estimates of the unavailable option binned according to successful future selection. The fMRI analysis was time-locked to the decision phase on trial t and binned according to accurate vs. inaccurate selection of the unavailable option on trial t+1 (in light pink: cluster-corrected, Z > 3.1, P < 0.001; in red: uncorrected, n=25 sessions) **(b)** ROI analyses (multiple regression analysis on the BOLD signal of the ROI) of the right (**top panels**) and left (**bottom panels**) hippocampus illustrate the time course of the aforementioned contrast. BOLD fluctuations reflect the value of the unavailable option on the current trial when it is accurately versus inaccurately selected on the next trial (**left panels** illustrate the contrast show in **(a)**). A leave-one-out procedure (for spatial and temporal peak selections) to assess statistical significance revealed that a similar activity change occurs when contrasting the value of the unavailable option for accurate versus inaccurate future *rejections* of the unavailable option (**right panels**). SEM are presented in the red shaded area across sessions, n=25. **(c)** In the left hippocampus, the beta weights for the contrast used in **(a)** and illustrated in **(b, left panel)** were predictive of how much the unavailable option's reward probability influenced animals' future choice accuracy **(top panel)** but this was not true for current choice accuracy **(bottom panel)**. Scatter plot at the time of the peak effect, n=25 sessions, Pearson R is reported (Results are normalised).
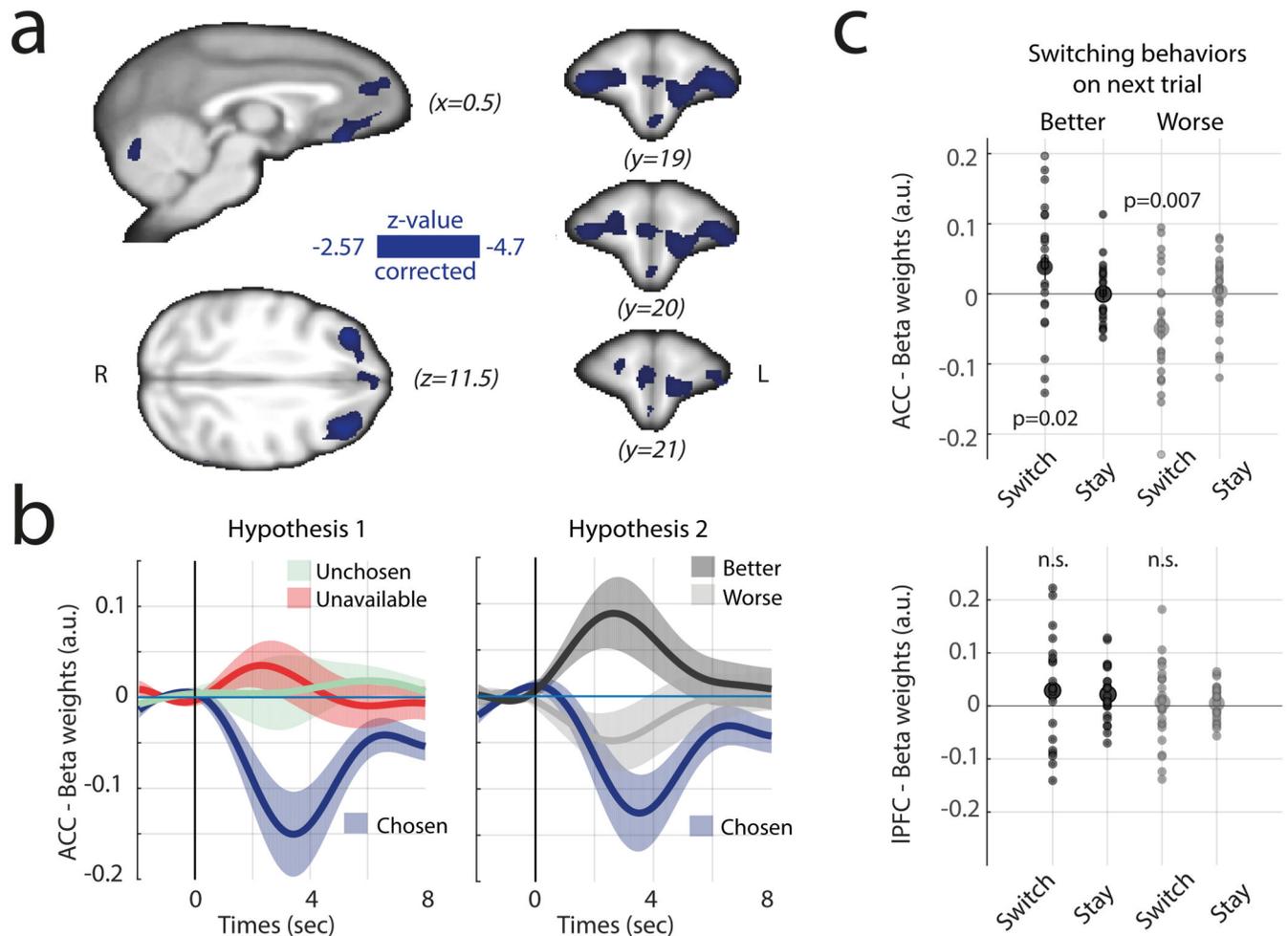
Figure 4. The anterior cingulate ranks expected reward probabilities of counterfactual options.
(a) Whole-brain analysis shows a significant negative relationship between BOLD activity and the difference between the expected value associated with the currently chosen and unchosen options in a distributed brain network, including ACC, bilateral lPFC, and vmPFC/mOFC (cluster corrected, $|Z| > 3.1$, $P < 0.001$, n=25 sessions) **(b)** ROI analysis of the ACC illustrates the relationship between BOLD and the fully parametric representation of the currently chosen, unchosen, and unavailable options (left panel) and shows that a distinct model in which the counterfactual options are ranked according to their associated reward probabilities explains the data better. Note that we avoid double dipping in favour of the hypothesis that we want to support (hypothesis 2) since the ROI has been defined on the basis of hypothesis 1. All shaded areas represent SEM across sessions, n=25. For hypothesis 2, the grey shading represent the Better (dark grey) and Worse (light grey) alternatives. See supplementary figure 3 for a full Bayesian Model Selection across all hypotheses. **(c)** The parametric representation of the better and worse counterfactual values in ACC was further explained by whether a future switch in behavior will occur as opposed to the continued maintenance of behavior ("stay") (leave-one-out procedures for peak selection on time series analyses: **top panel**). This was not true in the lPFC (**bottom panel**). Each session is

represented as a grey dot (bar represents the average beta coefficient across sessions, n=25, one sample t-tests are performed).
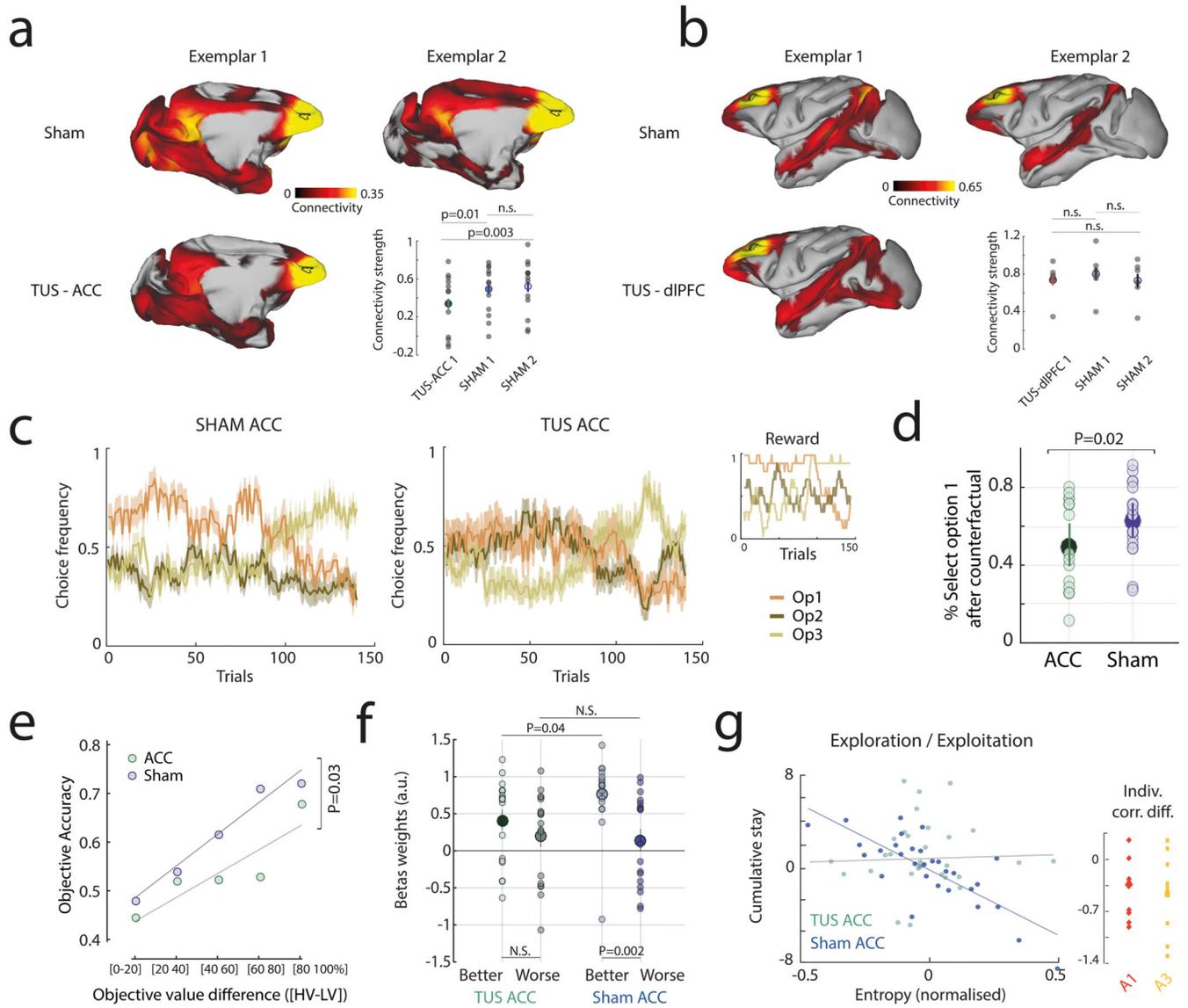
Figure 5. Transcranial Focused Ultrasound Stimulation (TUS) of ACC had a profound and selective effect on resting state connectivity.

(a) Whole-brain functional connectivity between the ACC and the rest of the brain. Left and right top panels show activity coupling between ACC (far-right ROI, black circle) and the rest of the brain in the no stimulation sham condition in two exemplar animals. After ACC TUS in exemplar animal 1, there are strong changes in connectivity (right bottom panel), reflected in changes in a connectivity analysis seeded from ACC with 13 other regions (ROI represented in black circle, for the full details, see supplementary fig.4; table 1) (within subject: two sample t-tests: Cohen's d=-0.84; t_{12}=-3.03; P=0.01, Cohen's d=-1.01; t_{12}=-3.65; P=0.003, n=13 ROIs, between-subject control: non-significant, n=6 ROIs). **(b)** However, while ACC TUS affected ACC connectivity, the effect was selective; ACC TUS did not affect connectivity seeded from lPFC (n.s: non-significant). **(c)** Running average choice frequency for the three options in the control/sham ACC **(left)** and the TUS ACC condition **(middle)** across sessions (the shaded areas represent SEM across session, n= 18

sessions for each group). Predetermined reward schedules used in the sham and in the TUS ACC task for three options, similar to the task used in the fMRI experiment **(right)**. **(d)** The rate of choosing option 1 was significantly reduced on trials that followed those on which it had previously been a counterfactual option – on trials on which it was unavailable in TUS session compared to SHAM sessions, n = 18 sessions for each group. **(e)** Decision accuracy is plotted as a function of the difficulty of the decision – the difference between the objective values of the HV and LV options. Values of HV and LV are objective values (reward probability over the last 10 trials). Each bin contains data binned according to percentile, with each point corresponding to the [0-20%], [20-40%], [40-60%], [60-80%], [80%-100%] of the value difference amplitudes. Accuracy is the rate at which the participant picked the objectively better option. Supplementary fig.5d illustrates accuracy as a function of subjective value differences. Performance differences between TUS and sham conditions do not increase with difficulty (small HV-LV differences on the left); if anything the opposite is true. **(f)** The influence of the better counterfactual option value on future switching behavior (in blue, as per fig.2f) was significantly reduced after TUS ACC (in green), n=18 sessions for each group. **(g)** While entropy (summed entropy of reward probability for all options) is strongly and negatively predictive of a change in exploratory behavior in the sham condition (indexed by the cumulative number of "stay" choices: choices of the same option on one trial after another), this relationship is disrupted in the TUS ACC condition. Each point in the figure illustrates a running average analysis, where each bin contains the derivative of entropy over five trials (thus 30 points). The small panel on the right depicts the difference in regression coefficients – linear fit – between the TUS ACC and the sham conditions (Animals 1 [S1] and 2 [S2] are individually represented as red diamond and yellow square, respectively in all plots, n=9 sessions per animal).
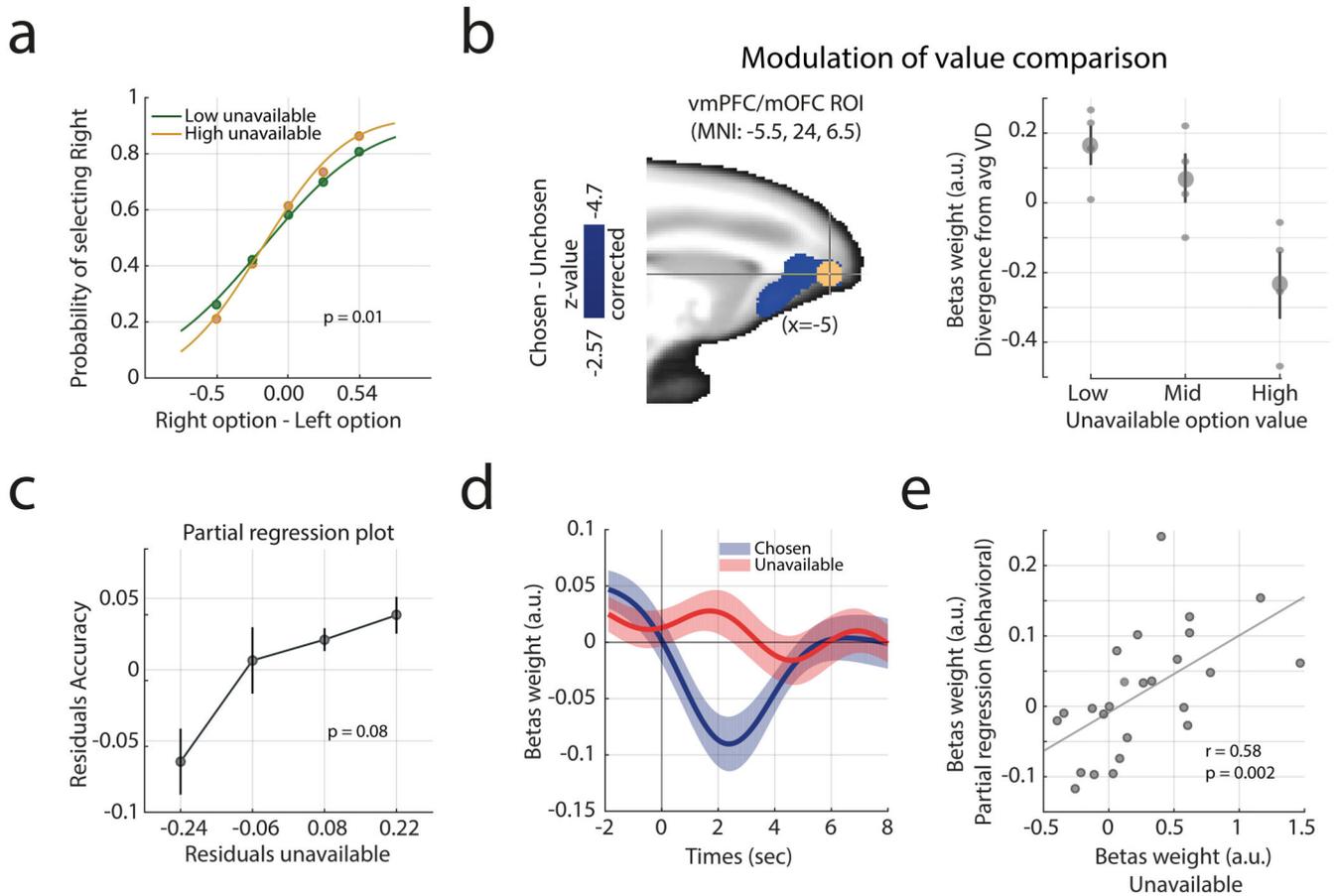
a

b

Modulation of value comparison

c

d

e

Figure 6. Contextual modulation of value-guided choice.
(a) Average choice behavior when choosing between the Left and Right options plotted as a function of the value of the unavailable option (low: green; high: yellow). Decisions were less accurate when they were made in the context of a low value unavailable option. Curves plot logistic functions fit to the choice data, n=25 sessions. **(b)** ROI analysis of the vmPFC/ mOFC (left panel: ROI sphere) illustrates the relationship between the BOLD value-comparison signal and the expected value associated with the unavailable option (binned in Low/Mid/High) (right panel). The greater the value of the unavailable option, the more negative the value difference; a more negative pattern is normally associated with decisions that are easier to take (see panel d). Data for individual animals are indicated by red dots (±SEM in grey, n=4 animals). **(c)** A partial regression plot shows the uncontaminated effect of the unavailable option's value on accuracy (y-axis: accuracy residuals; x-axis: residuals of the unavailable option's value). Each bin contains 20% of averaged data across sessions (±SEM). One sample t-test on betas of regression analysis, n=25 sessions. **(d)** ROI time course analysis of the vmPFC/mOFC illustrates the relationship between BOLD and the fully parametric representation of the currently chosen and unavailable options. The shaded areas represent SEM across session, n=25 sessions. **(e)** While there was not a main effect of the unavailable option value, vmPFC/mOFC variation in activity related to the currently unavailable option's value explains between-session variation in the currently unavailable option's influence on decision making. Scatter plot at the time of the peak effect of

unavailable option value in the vmPFC/mOFC (leave-one-out peak selection, n=25 sessions, Pearson R is reported).
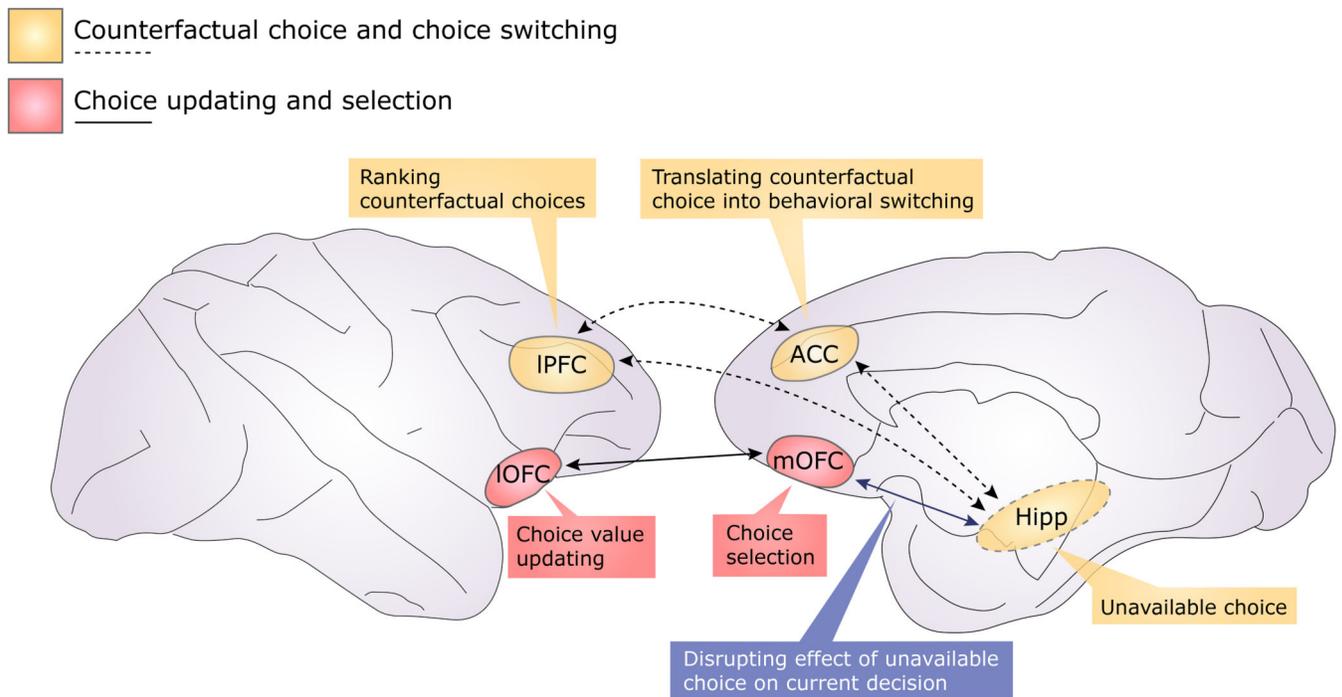
Figure 7. Schematic view of brain regions hypothesized to encode counterfactual choice.
Schematic view of some of the brain regions hypothesized to be involved in encoding counterfactual choice (in yellow and dashed lines, including the anterior cingulate cortex - ACC, lateral prefrontal cortex - lPFC, and the hippocampus - Hippo), and choice updating and selection (in red and continuous lines, including the lateral and the medial orbitofrontal cortex – lOFC and mOFC/vmPFC, respectively). A blue line represents the hypothesized effect exerted by the hippocampus, via mOFC/vmPFC, on the current choice.

Table 1

ROIs for rs-fMRI connectivity analyses.

The XYZ coordinates of the ROIs used in the rs-fMRI connectivity analysis are listed. For the ACC seed analyses, we excluded the ROI "A" (ACC itself) and thus used B, C, D, E, F, G, H, I, J, K, L, M, N. For the lPFC seed analyses, we excluded the two ROIs too close to the seed to avoid circular analyses (namely "L" and "M"). In addition, we excluded the ACC and neighbouring and thus used F, G, H, I, J, and N since TUS over ACC seems to have an influence on the connectivity of lPFC.

ROI	A (ACC)	B	C	D	E	F (MCC)	G (PCC)
X	-2.6	-1.8	-1.5	-1.8	-1.3	-0.9	-1.5
Y	20.4	13.8	6.5	-2.0	-8.9	-15.7	-21.0
Z	10.3	12.8	14.2	15.6	14.3	15.2	11.9
ROI	H (PCC)	I (PCC)	J (PCC)	K (lPFC)	L (dlPFC)	M (dlPFC)	N (lPFC)
X	-1.3	-1.1	-2.0	-6.7	-9.5	-14.8	-8.0
Y	-25.7	-30.7	-24.7	20.4	14.8	14.7	19.4
Z	8.0	8.6	2.5	15.9	18.9	15.8	11.0