nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-019-0375-6

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

Elsa F. Fouragnan[1,2,7*], Bolton K. H. Chau[2,3,7], Davide Folloni[2,7], Nils Kolling[2], Lennart Verhagen[2], Miriam Klein-Flügge[2], Lev Tankelevitch[2], Georgios K. Papageorgiou[2,4], Jean-Francois Aubry[5], Jerome Sallet[2,8] and Matthew F. S. Rushworth[2,6,8]

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 the hippocampus, the anterior lateral prefrontal cortex and the anterior cingulate cortex. Anterior cingulate cortex activity also reflected whether the internal value representations would be translated into actual behavioral change. To establish the causal importance of the anterior cingulate cortex for this translation process, we used a novel technique, transcranial focused ultrasound stimulation, to reversibly disrupt anterior cingulate cortex activity.

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 toward 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, the 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 the anterior cingulate cortex (ACC)[3,5,6]. Other studies have emphasized a related role for the 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: the hippocampus. Within the hippocampal formation, the subiculum projects monosynaptically to the 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 the 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 the human lFPC. The human frontal polar cortex can be subdivided into the lateral and medial sub-regions, lFPC and mFPC[11,12]. While resting state connectivity patterns exhibited by the human mFPC and the macaque FPC are similar, human lFPCs more closely resemble the 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 the 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 and 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 cortex (vmPFC) and/or medial orbitofrontal cortex (mOFC).

[1]School of Psychology, University of Plymouth, Plymouth, UK. [2]Wellcome Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Oxford, UK. [3]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China. [4]McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Physics for Medicine Paris, Inserm, ESPCI Paris, CNRS, PSL Research University, Paris, France. [6]Wellcome Integrative Neuroimaging, Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, UK. [7]These authors contributed equally: Elsa F. Fouragnan, Bolton K. H. Chau, Davide Folloni. [8]These authors jointly supervised this work: Jerome Sallet, Matthew F. S. Rushworth. *e-mail: elsa.fouragnan@plymouth.ac.uk

Four macaques chose between pairs of abstract visual stimuli while in the magnetic resonance imaging (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 toward 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 functional MRI (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). Thus, an option not actually available does not affect the difficulty of the current decision and therefore does not influence the 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, TUS can alter neural activity[17] but unlike transcranial magnetic stimulation it can even do so in relatively deep structures such as the 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 10–12 mm distant[19]. Importantly, the minimally invasive nature of the stimulation made it possible to examine not just a region of interest such as the 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 and 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 $d = 1.42$, $t_{24} = 7.12$, $P = 2.3 \times 10^{-7}$; total value: Cohen's $d = 0.82$, $t_{24} = 4.10$, $P = 4.04 \times 10^{-4}$) and reaction times (value difference: Cohen's $d = -0.74$, $t_{24} = -3.68$, $P = 0.001$; total value: Cohen's $d = -1.11$, $t_{24} = -5.54$, $P = 1.07 \times 10^{-5}$; 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 $d = 1.59$, $t_{24} = 7.95$, $P = 3.5 \times 10^{-8}$; Fig. 2a). A complementary analysis confirmed these results and showed that the 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 $d = 1.06$, $t_{24} = 5.32$, $P = 1.87 \times 10^{-5}$; Fig. 2b) beyond the effect of the current chosen and unchosen options (chosen: Cohen's $d = 0.98$, $P = 5.04 \times 10^{-5}$; unchosen: Cohen's $d = -0.87$, $P = 2.92 \times 10^{-5}$).

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 that animals' decisions to switch to the better counterfactual choice were influenced by its expected value (Cohen's $d = 1.23$, $t_{24} = 6.16$, $P = 2.32 \times 10^{-6}$)

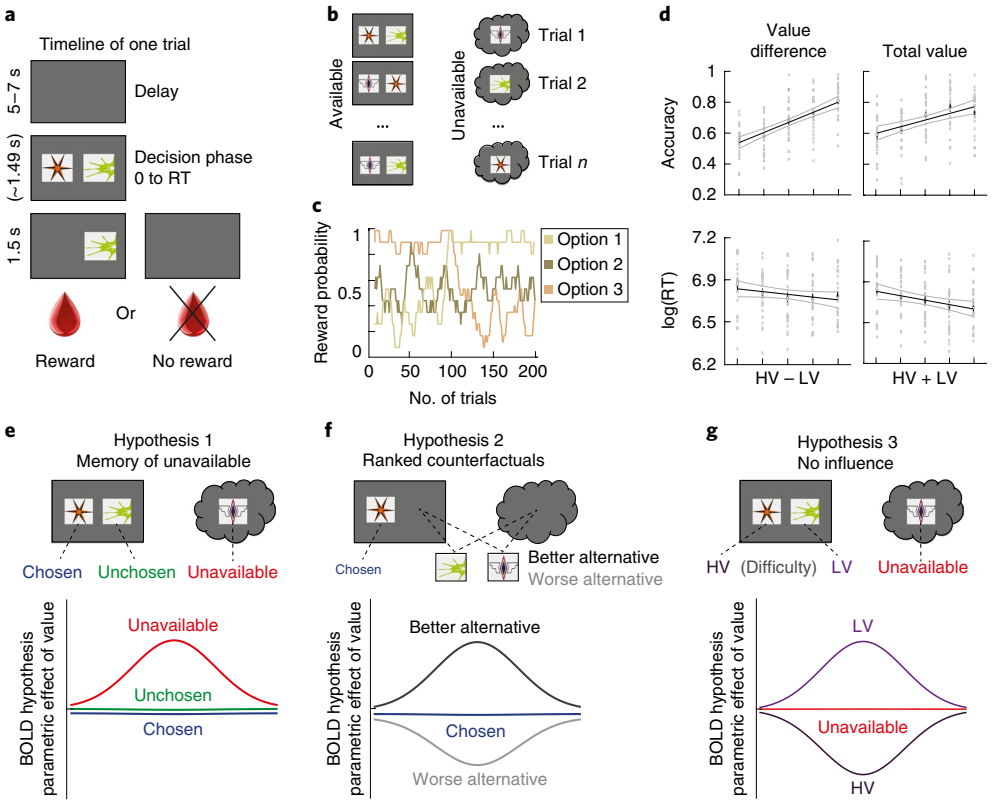

Fig. 1 | Schematic view of the task, behavioral results and hypothesized neural schemes. a, On each trial, animals could choose between two symbols presented on the screen and had to keep in mind a third option, unavailable to them. The position of each symbol on the left or right part of the screen and the combination of available or unavailable options were fully and pseudorandomized, 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 s. 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 option 1 (light green), 2 (dark green) or 3 (brown) 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 expected values of both HV and LV values, right panel). Decision accuracy improved with the higher value difference between available options and the higher total value. The bottom graphs show log-transformed mean reaction times (RT) for each session plotted as a function of difficulty and context. The log(RT) values decreased for easier decisions and higher trial values. Dark gray lines are linear fits to the data, and the lighter gray lines are the 95% confidence interval; $n = 25$ sessions. **e**, Because each of the values of the three options 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, this suggests that 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**,**g**, If instead activity covaries with the ranked value of both the unchosen available option and the option held in memory, it reflects the value of any currently counterfactual choice that might be taken in the future (**f**). 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 is unaffected by the value of the choice that cannot currently be taken (see the discussion by Kolling and colleagues[15]). Note that we also analyzed a fourth pattern, representing the value of each option separately, in Supplementary Fig. 3.

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 general linear model (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

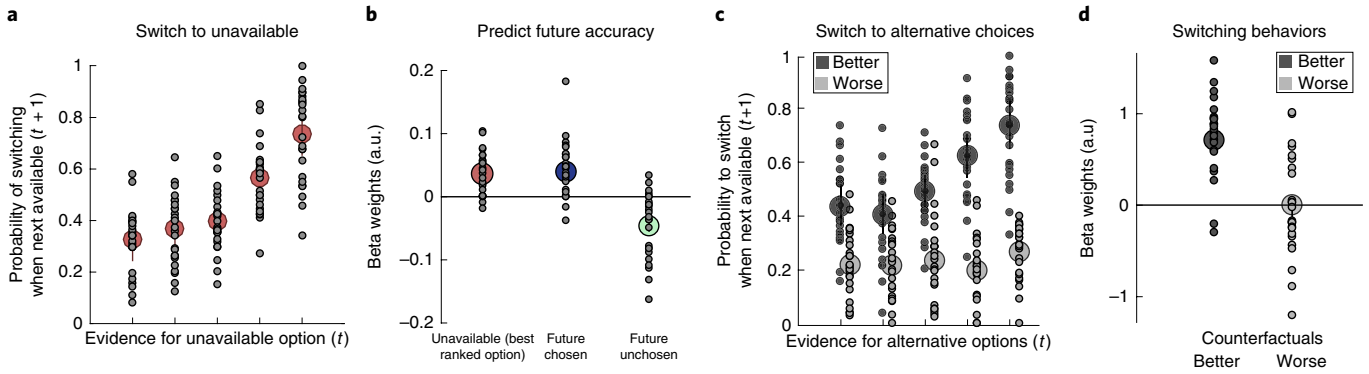

Fig. 2 | Future switches are explained by the expected value associated with counterfactual options. a, Estimated expected values associated with the unavailable option on the current trial predict whether animals switch to it when it reappears on the screen on subsequent trials (*y* axis, probability 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 gray dots; average across sessions in red dots). **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 versus when it is not), in addition to the value of the future chosen and unchosen options (each session's β coefficient is represented as a gray dot and the mean β coefficients is represented as a colored dot); a.u., arbitrary units. **c**, A similar analysis to the one shown in **a**, but on the basis of a new coding scheme in which 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 regression coefficient β; $n = 25$, for all analyses.

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 panel). 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 the 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 role of the hippocampus 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 (for example, 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 the memory of unavailable options, we hypothesized that the variation in such activity (at trial *t*) across sessions might predict the 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 versus 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 vmPFC and adjacent 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 (for example, 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);

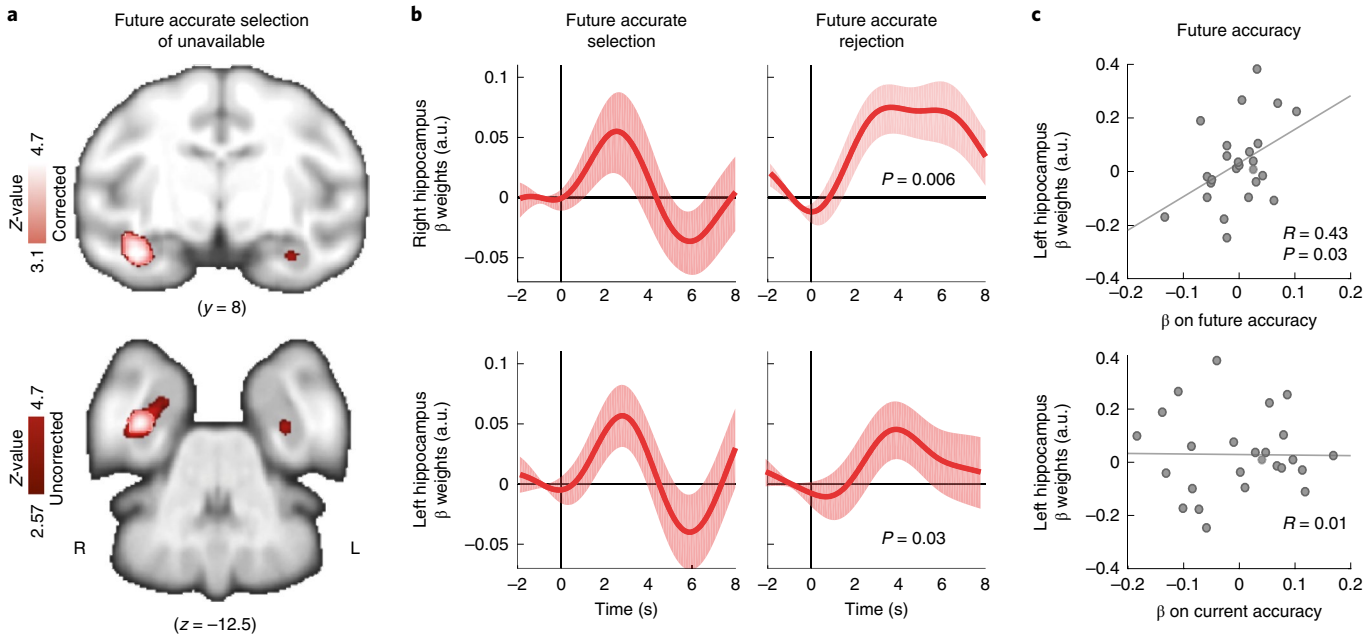

Fig. 3 | Unavailable option value signal in hippocampus favors accurate future planning. a, A whole-brain analysis tested for voxels where activity correlated with the trial-by-trial estimates of the unavailable option, binned according to successful future selection. The fMRI analysis was time-locked to the decision phase on trial *t* and binned according to accurate versus 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) and left (bottom) 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 shown 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). The s.e.m. are presented in the red shaded area across sessions, $n = 25$. **c**, In the left hippocampus, the β weights for the contrast used in **a** and illustrated in **b** (left) were predictive of how much the unavailable option's reward probability influenced animals' future choice accuracy (top), but this was not true for current choice accuracy (bottom). Scatter plot at the time of the peak effect, $n = 25$ sessions; Pearson's *R* is reported (results are normalized).

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 the 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 versus 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 the 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 exceedance probability $\varphi s > 0.95$ (Fig. 4b; see Supplementary Fig. 3 and the Methods for full Bayesian model comparison[29]). Thus, while the 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 the unchosen/unavailable (first scheme or Hypothesis 1) to focusing on the best alternative (second scheme or Hypothesis 2). 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 the 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 (1) the expected value of the better alternative if the future trial is a switch to that option; (2) the expected value of the better alternative if the future trial is a stay (that is, a repetition of the same choice as on the current trial); (3) the expected value of the worse alternative if the future trial is a switch to that option; (4) 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 $d < 0.23$, $P > 0.02$). Such a pattern is consistent with a role for the ACC in evaluating future strategies before execution[3,30–32]. By contrast, the 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 value, the decision is easy. However, 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

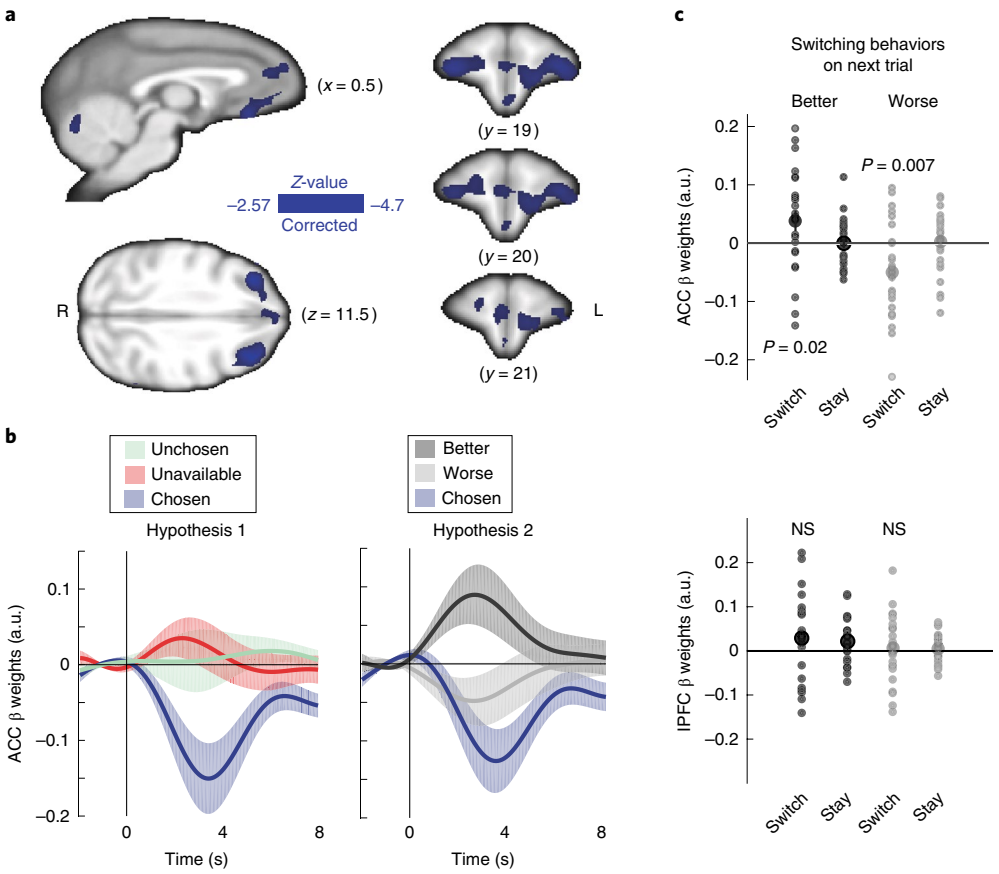

Fig. 4 | The anterior cingulate ranks expected reward probabilities of counterfactual options. a, Whole-brain analysis shows a significant negative relationship between BOLD activity and the difference between the expected value associated with the currently chosen and unchosen options in a distributed brain network, including the 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) 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 favor of the hypothesis that we want to support (Hypothesis 2), as the ROI has been defined on the basis of Hypothesis 1. All shaded areas represent s.e.m. across sessions, $n = 25$. For Hypothesis 2, the gray shadings represent the better (dark gray) and worse (light gray) alternatives. See Supplementary Fig. 3 for a full Bayesian model selection across all hypotheses. **c**, The parametric representation of the better and worse counterfactual values in the 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). This was not true in the lPFC (bottom). Each session is represented as a gray dot (the bar represents the average β coefficient across sessions; $n = 25$; one-sample t-tests).

time (RT). We controlled for this in all our analyses by parametrically modulating the duration of the boxcar regressor locked at the time of the decision by RT (regressor DEC in GLMs 1, 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 40 s sonification at 250 kHz reaches the ACC and does so in a relatively focal manner having a lesser 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). The rs-fMRI scans were collected for two healthy animals (the rs-fMRI scans from the two animals were acquired under no stimulation; rs-fMRI scans from one animal were 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 the ACC using MRI-guided frameless stereotaxy[19,33] immediately before nine testing sessions that were interleaved, across days, with nine control sessions in which no TUS was applied (Fig. 5a and 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 (the 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 the moni-

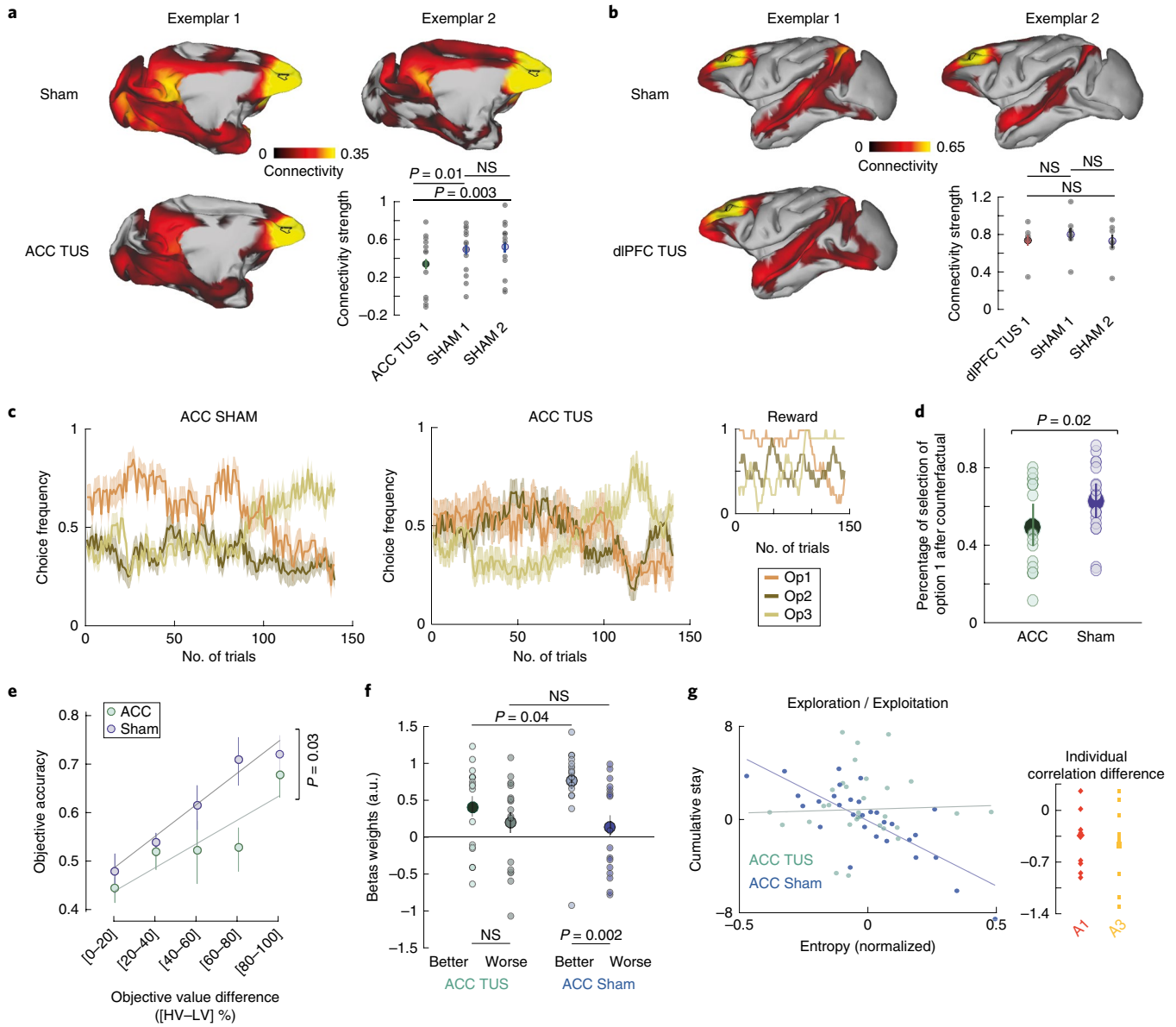

Fig. 5 | TUS of ACC had a profound and selective effect on resting state connectivity. a, Whole-brain functional connectivity between the ACC and the rest of the brain. Top: left and right panels show activity coupling between ACC (far-right ROI, black circle) and the rest of the brain in the no-stimulation sham condition in two exemplar animals. Bottom: after ACC-TUS in exemplar animal 1, there are strong changes in connectivity (right), reflected in changes in a connectivity analysis seeded from ACC with 13 other regions (ROI represented by the black circle, for the full details see Supplementary Fig. 4 and 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 dlPFC (NS, non-significant). **c**, Running average choice frequency for the three options (Op1, Op2 and Op3) in the control (sham) ACC (left) and the ACC-TUS condition (middle) across sessions (the shaded areas represent s.e.m. across sessions; $n = 18$ sessions for each group). Predetermined reward schedules used in the sham and in the ACC-TUS 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, i.e., on trials on which it was unavailable in the TUS sessions compared to sham TUS (SHAM) sessions; $n = 18$ sessions for each group. **e**, Decision accuracy is plotted as a function of the difficulty of the decision, i.e., 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 ten trials). Each bin contains data binned according to percentile, with each point corresponding to the [0–20%], [20–40%], [40–60%], [60–80%] or [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 ACC TUS (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, i.e., choices of the same option on one trial after another), this relationship is disrupted in the ACC-TUS 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 ACC-TUS and the sham conditions (animals 1 (A1) and 2 (A2) are individually represented as red diamonds and yellow squares, respectively in all plots, $n = 9$ sessions per animal).

Table 1 | ROIs for rs-fMRI connectivity analyses

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

The *XYZ* coordinates of the ROIs used in the rs-fMRI connectivity analysis are listed. For the ACC seed analyses, we excluded ROI A (the ACC itself) and thus used B–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 neighboring ROIs and thus used F–J and N, since TUS over ACC seems to have an influence on the connectivity of the lPFC. MCC, middle cingulate cortex; PCC, posterior cingulate cortex; dlPFC; dorsolateral prefrontal cortex.

toring of 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 reinforcement learning (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 that ACC activity encodes the better counterfactual alternative but not the worse counterfactual alternative (Figs. 2f and 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 $d=1.57$, $t_{17}=6.7$, $P=3.62\times10^{-6}$) 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 a 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 (that is, 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 $d=-1.20$, $t_{28}=-6.59$, $P=3.77\times10^{-7}$) 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 (ACC TUS: Cohen's $d=-0.67$, $t_{28}=-3.65$, $P=0.001$; ACC-SHAM: Cohen's $d=-0.90$, $t_{28}=-4.95$, $P=3.17\times10^{-5}$; 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 of the 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 in the effect of the best counterfactual on switching behaviors between the lOFC-TUS and lOFC-SHAM (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 ACC-TUS experiment: Cohen's $d=0.21$, $t_{9}=0.46$, $P=0.66$). Further direct comparisons between lOFC-TUS and ACC-TUS showing significant differences between the two types of TUS are reported in Supplementary Fig. 6. Additionally, there was no difference between the strength of the relationship between entropy and cumulative stay in lOFC-TUS and lOFC-SHAM conditions (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, the vmPFC/mOFC, also carried a choice value-comparison signal (Figs. 4a and 6b). This pattern of decision-related fMRI activity in the vmPFC/mOFC has been reported previously in macaques[38]. Given the 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 the results presented in Fig. 4c). Instead, the 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 the 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 obtained 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

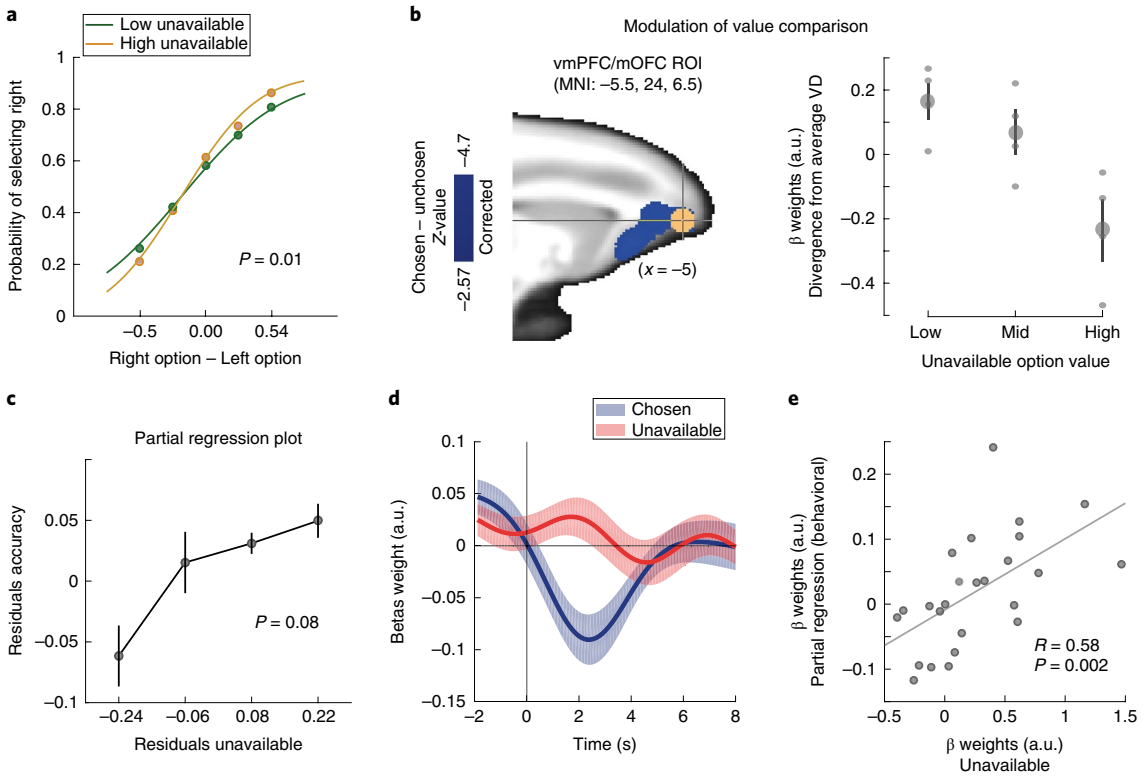

Fig. 6 | Contextual modulation of value-guided choice. a, Average choice behavior when choosing between the left and right options plotted as a function of the value of the unavailable option (low, green; high, yellow). Decisions were less accurate when they were made in the context of a low-value unavailable option. Curves plot logistic functions fitted to the choice data; $n = 25$ sessions. **b**, ROI analysis of the vmPFC/mOFC (left, ROI sphere) illustrates the relationship between the BOLD value-comparison signal and the expected value associated with the unavailable option (binned by low, mid and high values) (right). 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 make (see **d**). Data for individual animals are indicated by red dots (±s.e.m. in gray; $n = 4$ animals). MNI, Montreal Neurological Institute; VD, value difference. **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 (± s.e.m.). 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 s.e.m. across sessions; $n = 25$ 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's *R* is reported).

right option) as a function of the value difference between the two available options (right and left option values) for two different levels of the unavailable option values (high versus 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), and 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 (Figs. 4b and 6d); as the chosen option's value falls and choosing it becomes more difficult, there is more activity in the 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 do not represent such modulation of value comparison by the unavailable option (all $P > 0.25$).

To further test the strength of the link between the contextual factor's impact on the current decision and its neural impact in the 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 the 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

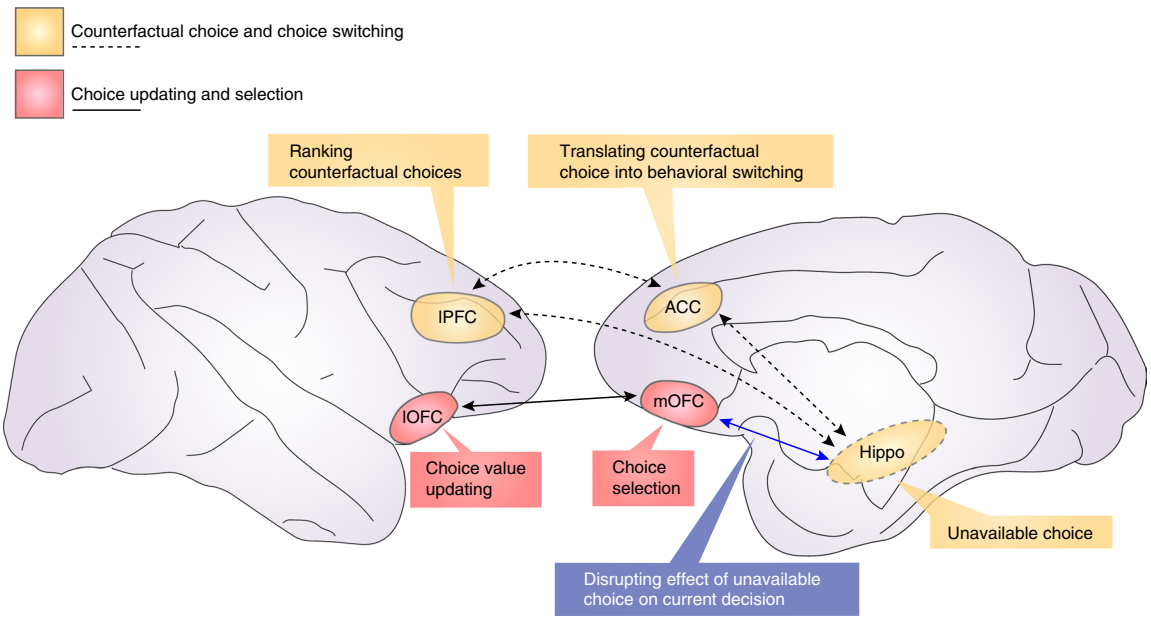

Fig. 7 | Schematic view of brain regions hypothesized to encode counterfactual choice. Encoding counterfactual choice (in yellow, dashed lines, including the ACC, lPFC and the hippocampus), and choice updating and selection (in red, continuous lines, including the lOFC and mOFC/vmPFC, respectively). A blue line represents the hypothesized effect exerted by the hippocampus, via the mOFC/vmPFC, on the current choice. Hippo, hippocampus.

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. 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 the 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 the mOFC/vmPFC (Fig. 6), then this distracts animals from the current choice to be taken. In contrast, translating the currently unavailable choice's value into a counterfactual plan that can be executed in the future depends on the ACC (Fig. 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 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 the mOFC/vmPFC and the ACC on the basis of different criteria[42,43]. The ACC is especially concerned with signaling the value of behavioral change and alternative courses of action[7,44,45].

Like the ACC, the lPFC holds counterfactual choice values. In this respect, lPFC activity resembles that seen in or near the human lateral frontal pole (FPl). 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 the anterior prefrontal cortex that is distinct to the 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 medial frontal pole (FPm) share cytoarchitectonic features, it is possible that some of the circuit level interactions and functions of macaque 46 are associated with the 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 in another, and connections previously associated with another area may become associated with the new area[48].

Notably, while the 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 the ACC (Fig. 4). By contrast, both the fMRI and TUS results suggest that the ACC is concerned with the translation of counterfactual information into a change of behavior.

The 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 the hippocampus, dlPFC and ACC hold information about currently unavailable choices to guide future behavioral change, other mechanisms associated with the vmPFC/mOFC have been linked to comparison of the values of specific choice options on the current trial (Fig. 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 the 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 the vmPFC/mOFC, in the absence of the vmPFC/mOFC, the opposite effect prevails and decisions are particularly vulnerable to disruption by high-value alternatives[34,50].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0375-6.

Received: 7 November 2018; Accepted: 6 March 2019;
Published online: 15 April 2019

References

1. Noser, R. & Byrne, R. W. Mental maps in chacma baboons (*Papio ursinus*): using inter-group encounters as a natural experiment. *Anim. Cogn.* **10**, 331–340 (2007).
2. Shadlen, M. N. & Shohamy, D. Decision making and sequential sampling from memory. *Neuron* **90**, 927–939 (2016).
3. Boorman, E. D., Behrens, T. E. & Rushworth, M. F. Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biol.* **9**, e1001093 (2011).
4. Boorman, E. D., Behrens, T. E. J., Woolrich, M. W. & Rushworth, M. F. S. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* **62**, 733–743 (2009).
5. Scholl, J. et al. The good, the bad, and the irrelevant: neural mechanisms of learning real and hypothetical rewards and effort. *J. Neurosci.* **35**, 11233–11251 (2015).
6. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
7. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural mechanisms of foraging. *Science* **336**, 95–98 (2012).
8. Kolling, N., Behrens, T., Wittmann, M. K. & Rushworth, M. Multiple signals in anterior cingulate cortex. *Curr. Opin. Neurobiol.* **37**, 36–43 (2016).
9. Aggleton, J. P., Wright, N. F., Rosene, D. L. & Saunders, R. C. Complementary patterns of direct amygdala and hippocampal projections to the macaque prefrontal cortex. *Cereb. Cortex* **25**, 4351–4373 (2015).
10. Jang, A. I. et al. The role of frontal cortical and medial-temporal lobe brain areas in learning a Bayesian prior belief on reversals. *J. Neurosci.* **35**, 11751–11760 (2015).
11. Neubert, F.-X., Mars, R. B., Thomas, A. G., Sallet, J. & Rushworth, M. F. S. Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* **81**, 700–713 (2014).
12. Bludau, S. et al. Cytoarchitecture, probability maps and functions of the human frontal pole. *Neuroimage* **93**, 260–275 (2014).
13. Abe, H. & Lee, D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* **70**, 731–741 (2011).
14. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Fictive reward signals in the anterior cingulate cortex. *Science* **324**, 948–950 (2009).
15. Kolling, N. et al. Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* **19**, 1280–1285 (2016).
16. Miyamoto, K. et al. Causal neural network of metamemory for retrospection in primates. *Science* **355**, 188–193 (2017).
17. Verhagen, L. et al. Offline impact of transcranial focused ultrasound on cortical activation in primates. *eLife* **8**, e40541 (2019).
18. Folloni, D. et al. Manipulation of subcortical and deep cortical activity in the primate brain using transcranial focused ultrasound stimulation. *Neuron* https://doi.org/10.1016/j.neuron.2019.01.019 (2019).
19. Deffieux, T. et al. Low-intensity focused ultrasound modulates monkey visuomotor behavior. *Curr. Biol.* **23**, 2430–2433 (2013).
20. Fouragnan, E., Queirazza, F., Retzler, C., Mullinger, K. J. & Philiastides, M. G. Spatiotemporal neural characterization of prediction error valence and surprise during reward learning in humans. *Sci. Rep.* **7**, 4762 (2017).
21. Fouragnan, E., Retzler, C., Mullinger, K. & Philiastides, M. G. Two spatiotemporally distinct value systems shape reward-based learning in the human brain. *Nat. Commun.* **6**, 8107 (2015).
22. Niv, Y. et al. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* **35**, 8145–8157 (2015).
23. Klein-Flügge, M. C. & Bestmann, S. Time-dependent changes in human corticospinal excitability reveal value-based competition for action during decision processing. *J. Neurosci.* **32**, 8373–8382 (2012).
24. Hunt, L. T. et al. Mechanisms underlying cortical activity during value-guided choice. *Nat. Neurosci.* **15**, S1–S3 (2012).
25. Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
26. Kolling, N., Wittmann, M. & Rushworth, M. F. S. Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* **81**, 1190–1202 (2014).
27. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
28. Martin, V. C., Schacter, D. L., Corballis, M. C. & Addis, D. R. A role for the hippocampus in encoding simulations of future events. *Proc. Natl Acad. Sci. USA* **108**, 13858–13863 (2011).
29. Palminteri, S., Khamassi, M., Joffily, M. & Coricelli, G. Contextual modulation of value signals in reward and punishment learning. *Nat. Commun.* **6**, 8096 (2015).
30. Wittmann, M. K. et al. Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nat. Commun.* **7**, 12327 (2016).
31. Schuck, N. W. et al. Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron* **86**, 331–340 (2015).
32. Procyk, E., Tanaka, Y. L. & Joseph, J. P. Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nat. Neurosci.* **3**, 502–508 (2000).
33. Paus, T. Imaging the brain before, during, and after transcranial magnetic stimulation. *Neuropsychologia* **37**, 219–224 (1999).
34. Noonan, M. P. et al. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl Acad. Sci. USA* **107**, 20547–20552 (2010).
35. O'Reilly, J. X. et al. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc. Natl Acad. Sci. USA* **110**, E3660–E3669 (2013).
36. Gallistel, C. R., Mark, T. A., King, A. P. & Latham, P. E. The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. *J. Exp. Psychol. Anim. Behav. Process.* **27**, 354–372 (2001).
37. Rudebeck, P. H. & Murray, E. A. The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron* **84**, 1143–1156 (2014).
38. Papageorgiou, G. K. et al. Inverted activity patterns in ventromedial prefrontal cortex during value-guided decision-making in a less-is-more task. *Nat. Commun.* **8**, 1886 (2017).
39. Chau, B. K. H., Kolling, N., Hunt, L. T., Walton, M. E. & Rushworth, M. F. S. A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nat. Neurosci.* **17**, 463–470 (2014).
40. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nat. Neurosci.* **19**, 973–980 (2016).
41. Strait, C. E., Blanchard, T. C. & Hayden, B. Y. Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron* **82**, 1357–1366 (2014).
42. Hunt, L. T. & Hayden, B. Y. A distributed, hierarchical and recurrent framework for reward-based choice. *Nat. Rev. Neurosci.* **18**, 172–182 (2017).
43. Rushworth, M. F. S., Kolling, N., Sallet, J. & Mars, R. B. Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Curr. Opin. Neurobiol.* **22**, 946–955 (2012).
44. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Neuronal basis of sequential foraging decisions in a patchy environment. *Nat. Neurosci.* **14**, 933–939 (2011).
45. Kolling, N., Scholl, J., Chekroud, A., Trier, H. A. & Rushworth, M. F. S. Prospection, perseverance, and insight in sequential behavior. *Neuron* **99**, 1069–1082.e7 (2018).
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.* **32**, 1940–1950 (2010).

47. Sallet, J. et al. The organization of dorsal frontal cortex in humans and macaques. *J. Neurosci.* **33**, 12255–12274 (2013).
48. Krubitzer, L. The magnificent compromise: cortical field evolution in mammals. *Neuron* **56**, 201–208 (2007).
49. Louie, K., Khaw, M. W. & Glimcher, P. W. Normalization is a general neural mechanism for context-dependent decision making. *Proc. Natl Acad. Sci. USA* **110**, 6139–6144 (2013).
50. Noonan, M. P., Chau, B. K. H., Rushworth, M. F. S. & Fellows, L. K. Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. *J. Neurosci.* **37**, 7023–7035 (2017).

Acknowledgements

Funding for this work was provided by the Wellcome Trust (grant nos. 203139/Z/16/Z, WT100973AIA, 103184/Z/13/Z and 105238/Z/14/Z), the Medical Research Council (grant nos. MR/P024955/1 and G0902373), the Bettencourt Schueller Foundation and the Agence Nationale de la Recherche (grant no. ANR-10-EQPX-15) and Christ Church, University of Oxford. We are very grateful for the care afforded to the animals by the veterinary and technical staff at the University of Oxford. We also thank J. Scholl for helpful comments on the manuscript.

Author contributions

E.F.F., B.K.H.C. and M.F.S.R. designed the experiments. B.K.H.C., G.K.P., D.F. and J.S. collected the data. E.F.F. analyzed the behavioral, fMRI and TUS data. L.V. contributed to the rs-fMRI analysis. B.K.H.C., N.K. and M.K.F. contributed to fMRI analysis tools. L.T. contributed preprocessing analysis tools. J.S. and J.F.A. contributed to the ultrasound. E.F.F. and M.F.S.R. wrote the manuscript. All authors discussed the results and implications and commented on the manuscript at all stages.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0375-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to E.F.F.

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 h light dark cycle, with access to water for 12–16 h 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 behavioral task in the MRI scanner (a horizontal 3T MRI scanner with a full-size bore). 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 the ACC. It was, however, possible to apply TUS to the lateral location appropriate for targeting lOFC in all four animals.

Behavioral training. Before 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 infrared 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). 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-hand sides of the screen and monkeys responded with a right or left infrared 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 toward 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 s. 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.).

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 (Figs. 1 and 2): trials with reaction times 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 the animals' responses[52]. For all models, if stimulus A was selected on trial i, its value, v, was updated via a prediction error, δ, as follows: $v_A(i+1) = v_A(i) + \alpha \times \delta(i)$, where α is the learning rate and the prediction error was given by $\delta(i) = r(i) - v_A(i)$, where r is the reward obtained. The values of the unselected stimulus (for example, B) were not updated. The two first models differ in their assumptions of the stimulus that was not shown on that trial (for example, 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: $v_C(i+1) = v_C(i) + \gamma(v_C(1) - v_C(i))$, where γ is the decay parameter. The third and fourth models assumed that a subjective value can be distorted by risk preference. 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[23,24,53], we fitted an additional free parameter η using the following equation, where w represents the new value associated with each option:

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

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 $P_A(i) = \sigma(\beta(v_A(i) - v_B(i)))$ or $P_A(i) = \sigma(\beta(w_A(i) - w_B(i)))$ for the distortion models, where $\sigma(z) = 1/(1 + e^{-z})$ is the logistic function and β is 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 (α, β, γ, η).

Model fitting. To estimate the free parameters (α, β, γ, η), we used a maximum likelihood estimation and a constrained nonlinear optimization procedure (as implemented in fmincon in MATLAB) separately for each session. The associated likelihood function was given by

$$\log L = \frac{\sum B_A \log P_A + \sum B_B \log P_B}{N_A + N_B}$$

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) 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 analysi of variance 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 the mean (mean ± s.e.m.). 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 Supplementary Fig. 1c), 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. (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). Second (right panel in Supplementary Fig. 1c), 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 (Supplementary Fig. 1d), the last feature of the data characterizing the task is the influence of valence (win/loss) on the switch/stay pattern. We thus compared 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). 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 and 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 magnetization-prepared rapid gradient echo (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 artifacts by an offline-SENSE reconstruction method[54] (Offline_SENSE GUI, Windmiller Kolster Scientific). 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[55] (Align_EPI GUI and Align_Anatomy GUI, Windmiller Kolster Scientific). The aligned data were processed with high-

pass temporal filtering (3-dB cutoff of 100 s) and Gaussian spatial smoothing (with full-width half-maximum of 3 mm). The data that were already registered to each subject's structural space were then registered to the CARET macaque F99 template[56] using affine transformation.

fMRI data analysis. We employed a univariate approach within the GLM framework to perform whole-brain statistical analyses of functional data as implemented in the FMRIB Software Library[57,58]:

$$Y = X\beta + \varepsilon = \beta_1 X_1 + \beta_2 X_2 + 1/4 + \beta_N X_N + \varepsilon$$

where Y is a $T\times1$ (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[59,60]. β is a $N\times1$ column vector of regression coefficients and ε a $T\times1$ 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 (FMRIB's local analysis of mixed effects) 1 + 2) treating session as random effects. For all analysis, we performed a cluster inference using a cluster-defining threshold of $|Z| > 3.1$ with a family-wise error (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 versus 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 to account for difficulty effects, as well as two boxcar regressors with a duration of 100 ms that were then convolved with the hemodynamic response function: (1) a modulated regressor indexing the occurrence of a decision (Dec; all event amplitudes set to one and the duration set to the reaction time for that trial); (2–3) two parametric regressors whose event amplitudes were modulated by the expected value of the unavailable option for (a) future correct selection ($unav_{corr}$) and (b) future incorrect selection ($unav_{incorr}$). 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:

$$Y = \beta_1 \mathrm{Dec} + \beta_2 \mathrm{unav}_{\mathrm{cor}} + \beta_3 \mathrm{unav}_{\mathrm{incor}} + \beta_4 \mathrm{Ch} + \beta_5 \mathrm{Unch} + \beta_6 \mathrm{Fbk} + \beta_7 \mathrm{Side} + \varepsilon$$

Finally, to further reduce variance and noise in the BOLD signal, we added two unconvolved regressors locked at the time of feedback and with a duration of TR (2.28 s) 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 versus 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 regressor parametrically modulated by reaction times (to account for difficulty effects), as well as three boxcar regressors with a duration of 100 ms that were then convolved with the hemodynamic response function: (1) a modulated regressor indexing the occurrence of a decision (Dec; all event amplitudes set to one and the duration set to the reaction time 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:

$$Y = \beta_1 \mathrm{Dec} + \beta_2 \mathrm{Ch} + \beta_3 \mathrm{Unch} + \beta_4 \mathrm{Unav} + \beta_6 \mathrm{Fbk} + \beta_7 \mathrm{Side} + \varepsilon$$

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 options 1, 2 and 3 (see Fig. 1 and 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. This 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 indicated which GLM best explained the neural data. We report the results for the 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 ($x = 14.5$, $y = 17.5$, $z = 9.5$) and vmPFC/mOFC ($x = -5$, $y = 14$, $z = 2$). We used procedures now standardly employed in most human and animal neuroimaging studies[39,51,61] in which the mean and s.e.m. (denoted in all figures by lines and shadings, respectively) of all the within-subject β weights were calculated across sessions for plotting the effect size time courses (each animal had a similar number of sessions).

Leave-one-out procedures 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 atidentifying 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 1−6 s 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. A single-element ultrasound transducer (H115-MR, diameter 64 mm, Sonic Concept) 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). Overall, the stimulation lasted for 40 s. A 75-W amplifier (75A250A, Amplifier Research) 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 136 V and corresponded to a peak negative pressure of 1.152 to 1.292 MPa, respectively, measured in water with an in-house heterodyne interferometer (see ref. [62] for more details about the simulation protocol). Based on a mean 66% transmission through the skull[63], 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) so that the focal spot was centered on the targeted brain region, namely, the ACC (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 on 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, neuronavigation, 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 lOFC SHAM and 20 lOFC TUS (4 animals × 5 sessions) using the same experimental design as the ACC-TUS protocol. Two out of the four animals tested were also used in the ACC-TUS protocol. TUS and control days were interleaved in one of two pseudo-random 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[64]. Data collection and analysis were not performed blind to the conditions of the experiments.

Finally, given that the TUS procedure lasts for 40 s 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 $E(i) = \sum_{i=1}^{\text{trials}} p(x_{i,j}) \cdot \log(p(x_{i,j}))$, 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 A on the next trial ($t+1$) 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 the 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^* (ref. [65]) for each session separately (see the average effect in Fig. 6c).

Macaque rs-fMRI data acquisition, preprocessing and analysis. The rs-fMRI and anatomical MRI scans were collected for two healthy animals (rs-fMRI scans from the two animals were acquired under no stimulation; rs-fMRI scans from one animal were acquired post ACC TUS) under inhalational isoflurane anesthesia using a protocol that was previously proven successful[66,67] in preserving whole-brain functional connectivity as measured with BOLD signal. In the case of the TUS conditions, we used the same procedure as that employed in refs. [17,18]. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[64].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

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

Code availability

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

References

51. Chau, B. K. H. et al. Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. *Neuron* **87**, 1106–1118 (2015).
52. Sutton, R. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
53. Farashahi, S., Azab, H., Hayden, B. & Soltani, A. On the flexibility of basic risk attitudes in monkeys. *J. Neurosci.* **38**, 4383–4398 (2018).
54. Kolster, H. et al. Visual field map clusters in macaque extrastriate visual cortex. *J. Neurosci.* **29**, 7031–7039 (2009).
55. Kolster, H., Janssens, T., Orban, G. A. & Vanduffel, W. The retinotopic organization of macaque occipitotemporal cortex anterior to V4 and caudoventral to the middle temporal (MT) cluster. *J. Neurosci.* **34**, 10168–10191 (2014).
56. Van Essen, D. C. et al. Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res.* **41**, 1359–1378 (2001).
57. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**(Suppl 1), S208–S219 (2004).
58. Fouragnan, E. et al. Reputational priors magnify striatal responses to violations of trust. *J. Neurosci.* **33**, 3602–3611 (2013).
59. Nakahara, K., Hayashi, T., Konishi, S. & Miyashita, Y. Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science* **295**, 1532–1536 (2002).
60. Kagan, I., Iyer, A., Lindner, A. & Andersen, R. A. Space representation for eye movements is more contralateral in monkeys than in humans. *Proc. Natl Acad. Sci. USA* **107**, 7933–7938 (2010).
61. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
62. Constans, C., Deffieux, T., Pouget, P., Tanter, M. & Aubry, J. F. Erratum to 'A 200–1380 kHz quadrifrequency focused ultrasound transducer for neurostimulation in rodents and primates: transcranial in vitro calibration and numerical study of the influence of skull cavity'. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**, 1417 (2017).
63. Wattiez, N. et al. Transcranial ultrasonic stimulation modulates single-neuron discharge in macaques performing an antisaccade task. *Brain Stimul.* **10**, 1024–1031 (2017).
64. Vanduffel, W., Zhu, Q. & Orban, G. A. Monkey cortex through fMRI glasses. *Neuron* **83**, 533–550 (2014).
65. Velleman, P. F. & Welsch, R. E. Efficient computing of regression diagnostics. *Am. Stat.* **35**, 234–242 (1981).
66. Noonan, M. P. et al. A neural circuit covarying with social hierarchy in macaques. *PLoS Biol.* **12**, e1001940 (2014).
67. Neubert, F.-X., Mars, R. B., Sallet, J. & Rushworth, M. F. S. Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proc. Natl Acad. Sci. USA* **112**, E2695–E2704 (2015).

Corresponding author(s): Elsa Fouragnan
elsa.fouragnan@psy.ox.ac.uk

Last updated by author(s): Feb 15, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Stimuli and responses were presented and collected using Presentation software (Neurobehavioral Systems Inc., Albany, CA). Functional Resonance Magnetic Imaging (fMRI) and resting state fMRI (rs-fMRI) data were acquired 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). Transcranial Focused Ultrasound Stimulation (TUS) was applied using a single element ultrasound transducer (H115-MR, diameter 64 mm, Sonic Concept, Bothell, WA, USA) with a 51.74 mm focal depth and a coupling cone filled with degassed water and sealed with a latex membrane (Durex).
Data analysis	Matlab R2016a - in house scripts; FSL (FMRIB Software Library v5.0); scripts from HCP Workbench (https://www.humanconnectome.org), and the Magnetic Resonance Comparative Anatomy Toolbox (MrCat; www.neuroecologylab.org). Magnetic Resonance Comparative Anatomy Toolbox (MrCat; http://www.rbmars.dds.nl/lab/toolbox.html); Bash shell - in house scripts; R (3.5.0) with R studio (AGPL v3 licence) - in house scripts; For Bayesian comparison we used the function spm_BMS (Statistical Parametric Mapping 12).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Four healthy male rhesus monkeys (Macaca mulatta) were involved in the experiment. Each animal performed between 5 to 7 fMRI scans. A total of 25 fMRI scans were used. No statistical methods were used to pre-determine sample sizes but our sample sizes are in accordance with gold standards as described in (Friston, Neuroimage, 1999; Desmond, Journal of Neuroscience Methods, 2002) and our previous work (Chau, B. K. H. et al. Neuron, 2015; Papageorgiou, G. K. et al. Nat. Commun., 2017). 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. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications (Vanduffel, W., Zhu, Q. & Orban, G. A. 2014, Neuron). An additional rsfMRI 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). No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications (Vanduffel, W., Zhu, Q. & Orban, G. A. 2014, Neuron).
Data exclusions	No fMRI session were excluded from the fMRI. 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. Trial exclusion criteria for the behavioral analyses was pre-established and was defined as three standard deviation above the median of the absolute log-transformed reaction times (Hoaglin,Mosteller, Tukey, 1983). Indeed, very slow response trials may have been subject to interference in the choice selection process.
Replication	All the effects that are reported are ones that were replicated across the group of individuals tested; they were found across the whole group of individuals in a mixed effects analysis. In addition some of the key effects relating to activity in the anterior cingulate cortex were examined in a second experiment in which the impact of the anterior cingulate cortical disruption was tested. The results of the second experiment confirmed the inferences drawn in the first experiment.
Randomization	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.
Blinding	Data collection and analysis were not performed blind to the conditions of the experiments. It was not possible to use TUS while being blind given our protocol.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Six male rhesus monkeys (Macaca mulatta) were used in the experiment (four for the fMRI part, and two for the rsfMRI part). 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.
Wild animals	The study did not involved wild animals.

Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All procedures were conducted under licenses from the United Kingdom (UK) Home Office in accordance with the UK The Animals (Scientific Procedures) Act 1986 and with the European Union guidelines (EU Directive 2010/63/EU).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Event related fMRI design and resting-state fMRI.
Design specifications	Each animal performed 200 trials. The details of the events in each trial are summarized in figure 1b. Each trial started with an inter-trial interval (ITI) showing a blank screen ranging between 5 and 7 seconds. Two options were then presented on the screen, monkeys chose one of the options by reaching the touch sensor placed in front of it (decision phase). If the stimulus chosen yielded a reward then it remained on the screen for 1.5 s while the reward was delivered. For the resting state collection, rs-fMRI from the two animals were acquired under no stimulation and rsfMRI from one animal was acquired post ACC-TUS under inhalational isoflurane anesthesia using a protocol which was previously proven successful in preserving whole-brain functional connectivity as measured with BOLD signal. In the case of the TUS conditions, fMRI data collection began only after completion of the TUS train (delay between ultrasound stimulation offset and scanning onset: 37.5 minutes; SEM: 2.21 minutes).
Behavioral performance measures	200 responses were recorded in each session as well as reaction times. We fitted a reinforcement learning model to the animal choices, estimating trial by trial subjective values associated with each option. The fitting procedure is described in the Methods. We also used multiple logistic and linear regressions to explain behavioral responses.

Acquisition

Imaging type(s)	Functional
Field strength	3 Teslas
Sequence & imaging parameters	The full sequence and imaging parameters are reported in the Methods in the section Imaging Data Acquisition. FMRI data were acquired using a gradient-echo T2* echo planar imaging (EPI) sequence with 1.5 x 1.5 x 1.5 mm3 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 x 0.5 x 0.5 mm3 resolution, TR = 2,5 ms, TE = 4.01 ms) were acquired in separate anesthetized scanning sessions.
Area of acquisition	Whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	The preprocessing of the fMRI data used tools of FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki) and the Magnetic Resonance Comparative Anatomy Toolbox (MrCat; http://www.rbmars.dds.nl/lab/toolbox.html).
Normalization	Linear and non-linear registration to F99 space was achieved using FLIRT (Jenkinson et al., 2002; Jenkinson and Smith, 2001) and FNIRT (Andersson et al., 2007; Jenkinson et al., 2012) with configurations adjusted to reflect macaque rather than human brain characteristics.
Normalization template	We used the macaca mulatta F99 template in Caret (Van Essen, 2002; Van Essen and Dierker, 2007) as our group template.
Noise and artifact removal	FMRI data were corrected for body motion artefacts by an offline-SENSE reconstruction method 5 (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 6 (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 template7 using affine transformation.
Volume censoring	We did not remove volumes during which significant movement occurred, instead, we used our motion-related artifacts (i.e. regression of motion parameters) as regressors of non interest that were not convolved in our general linear models.

Statistical modeling & inference

Model type and settings	We employed a univariate approach within the general linear model framework to perform whole-brain statistical analyses of functional data as implemented in the FMRIB Software Library. 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 sessions as a random effects.

Effect(s) tested

The full models are described in the methods including the nature of all regressors entered in the analyses. See GLM1, GLM2, GLM3 and GLM4.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s)

We used a leave-one-out procedure to identify ROI peak voxels for the analyses of main effects for areas identified in all fMRI 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). 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.

Statistic type for inference
(See Eklund et al. 2016)

We performed a cluster inference using a cluster-defining threshold with |Z| > 3.1.

Correction

FWE-corrected threshold of P = 0.001.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis