Independent and Interacting Value Systems for Reward and Information in the Human Prefrontal Cortex

Authors: I. Cogliati Dezza[1,2]*, A. Cleeremens[1], W. Alexander[3,4]

Affiliations:

[1]Center for Research in Cognition & Neurosciences, ULB Neuroscience Institute, Université Libre de Bruxelles, Brussels, Belgium

[2]Department of Experimental Psychology, Faculty of Brain Sciences, University College London, London, UK

[3]Department of Experimental Psychology, Ghent University, Ghent, Belgium

[4]Center for Complex Systems and Brain Sciences, Florida Atlantic University, USA

*Correspondence to: irene.cogliatidezza@gmail.com.

Abstract:

Theories of Prefrontal Cortex (PFC) as optimizing *reward* value have been widely deployed to explain its activity in a diverse range of contexts, and appear to have substantial empirical support in neuroeconomics and decision neuroscience. Theoretical frameworks of brain function, however, suggest the existence of a second, independent value system for optimizing *information* during decision-making. To date, however, there has been little direct empirical evidence in favor of such frameworks. Here, by using computational modeling, model-based fMRI analysis, and a novel experimental paradigm, we aim at establishing whether independent value systems exist in human PFC. We identify two regions in the human PFC which independently encode distinct value signals. These value signals are then combined in subcortical regions in order to implement choices. Our results provide empirical evidence for PFC as an optimizer of independent value signals during decision-making. And, it suggests a new perspective on how to look at decision-making processes in the human brain under realistic scenarios, with clear implications for the interpretation of PFC activity in both healthy and clinical population.

One Sentence Summary: Distinct Value Systems for Reward and Information in the Human PFC

Introduction

A general organizational principle of *reward* value computation and comparison in PFC has accrued widespread empirical support in neuroeconomics and decision neuroscience [1-3]. According to this account, the relative reward value of immediate, easily-obtained, or certain outcomes positively contribute to the net-value of a choice [4 5] (Figure 1A), while delay, difficulty, cost or uncertainty in realizing prospective outcomes negatively contribute to it [6-8]. Although substantial empirical evidence supports the interpretation of PFC function as a single distributed system that performs a cost-benefit analysis in order to optimize the net value of rewards [1-3 9], other perspectives have suggested the existence of a second, independent value system for optimizing *information* within PFC. To date, however, direct empirical evidence for such a system is currently lacking. Using computational modeling, model-based fMRI analysis, and a novel experimental paradigm, we aim at establishing whether independent value systems exist in human PFC.

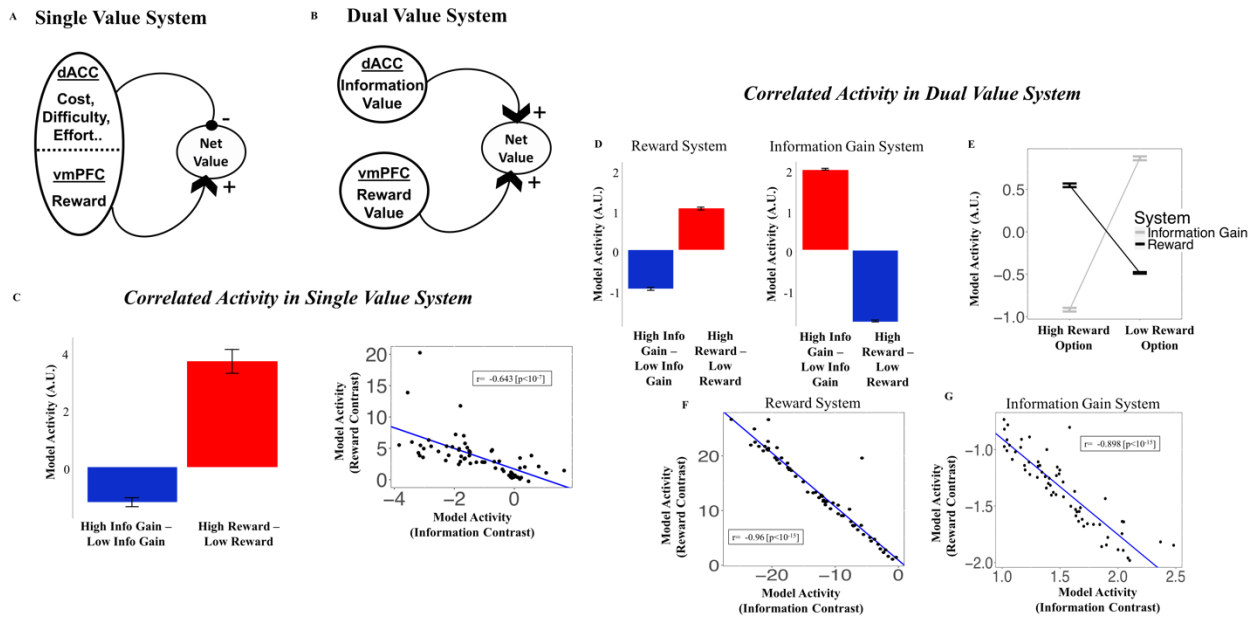
Within PFC, two regions, ventromedial PFC (vmPFC) and dorsal Anterior Cingulate Cortex (dACC), are frequently identified as calculating the positive (vmPFC) and negative (dACC) components of a cost-benefit analysis. In general, vmPFC activity appears to reflect the relative reward value of immediate, easily-obtained, or certain outcomes, while dACC activity signals delay, difficulty, or uncertainty in realizing prospective outcomes. Activity observed in vmPFC and dACC frequently exhibits a pattern of symmetric opposition: as dACC activity increases, vmPFC activity decreases- a pattern that holds across a wide range of value-based decision-making contexts, including foraging [10,11], risk [12], intertemporal [13,14] and effort-based choice [15,16] (see supplementary text for additional discussion). The variety of contexts in which this pattern is observed suggests a general role for these regions in contributing to the net-value associated with a choice [1-3], with vmPFC positively and dACC negatively contributing to the net-value computation (Figure 1A). While evidence reporting this symmetrically-opposed activity is common in the neuroeconomics and decision neuroscience literature, other studies have reported dissociations between dACC and vmPFC during value-based decision-making [17 18 19 9 20 21 22]. However, even when activity in dACC and vmPFC is dissociated, activity in vmPFC is generally linked to reward value, while activity in dACC is often interpreted as indexing negative or non-rewarding attributes of a choice (including ambiguity [23] difficulty [22], negative reward value [16], cost and effort [24]; see supplementary text for additional discussion). The interpretation of dACC and vmPFC as opposing one another therefore includes both symmetrically-opposed activity, as well as a more general functional opposition in value-based choice.

Despite activity in PFC exhibiting characteristics of a net-value computation[9], theoretical frameworks of brain function suggest the existence of a second, independent value system for optimizing *information* during decision-making. Unifying theories of brain organization and function propose that information gain plays a similar role as does reward in jointly minimizing surprise [25-27], allowing a behaving agent to better

anticipate environmental contingencies. Some reinforcement learning (RL) frameworks distinguish extrinsically-motivated (reward-based) behavior from intrinsically-motivated behavior to explain phenomena such as curiosity [28], directed exploration [29], and play [30] in the absence of explicit reward. Computational models of PFC [18,31,32] and neural recordings in monkeys [33] suggest that dACC activity can be primarily explained as indexing prospective information about an option independent of reward value, although these findings do not explain why dACC sometimes appears to encode quantities related to reward value such as cost, effort [23] and difficulty [22]. More broadly, results from machine learning demonstrate that explicitly incorporating information optimization in choice behavior can dramatically improve performance on complex tasks [34,35]. Altogether these perspectives suggest that information is intrinsically valuable [36] and positively contributes to the net-value computation [37] [26] during decision-making (Figure 1B). Although these results are suggestive of *why* and *where* a dedicated and independent value system for information in the human brain *might* exist, direct empirical evidence for such a system is currently lacking.

Simulations of an RL model which consists of independent value systems [38] independently optimizing information and reward demonstrate *how* reward-focused fMRI analysis [10,11,22,39,40] may be unsuccessful in identifying an independent information value system as a consequence of correlated activity (Figure 1). In such systems, independently optimizing reward and information entails a tradeoff: optimizing reward means *not* optimizing information, and vice-versa. In other words, even if reward and information systems are independent, they are nonetheless (negatively) correlated through behavior [41]. This correlation is consistent across different decision-making tasks [10,11,41] (Fig.1; Supplementary Material, Figure S1), and is also observed in single-value system models (Figure 1 C). Model simulations further demonstrate that by interpreting the function of an information system as contributing negatively to net value (e.g., as indexing effort level), it is possible to dissociate reward and information value systems while still observing a functional opposition (Supplementary Material, Figure S1). Crucially, the results of our simulations imply that reward-focused univariate fMRI analyses [10,11,22,39,40] (which uniquely focus on the reward dimension) may misattribute information value to a system computing costs (diminishing reward value), rather than to an independent information value system. Here, we adopt a novel experimental paradigm in which the relative contribution of reward and information as motivating factors in choice behavior can be dissociated. And, by using model-based fMRI analysis we identify their subjective representation in the human PFC.

Figure. 1. *Correlation between reward and information value in single-value and dual-value RL models*



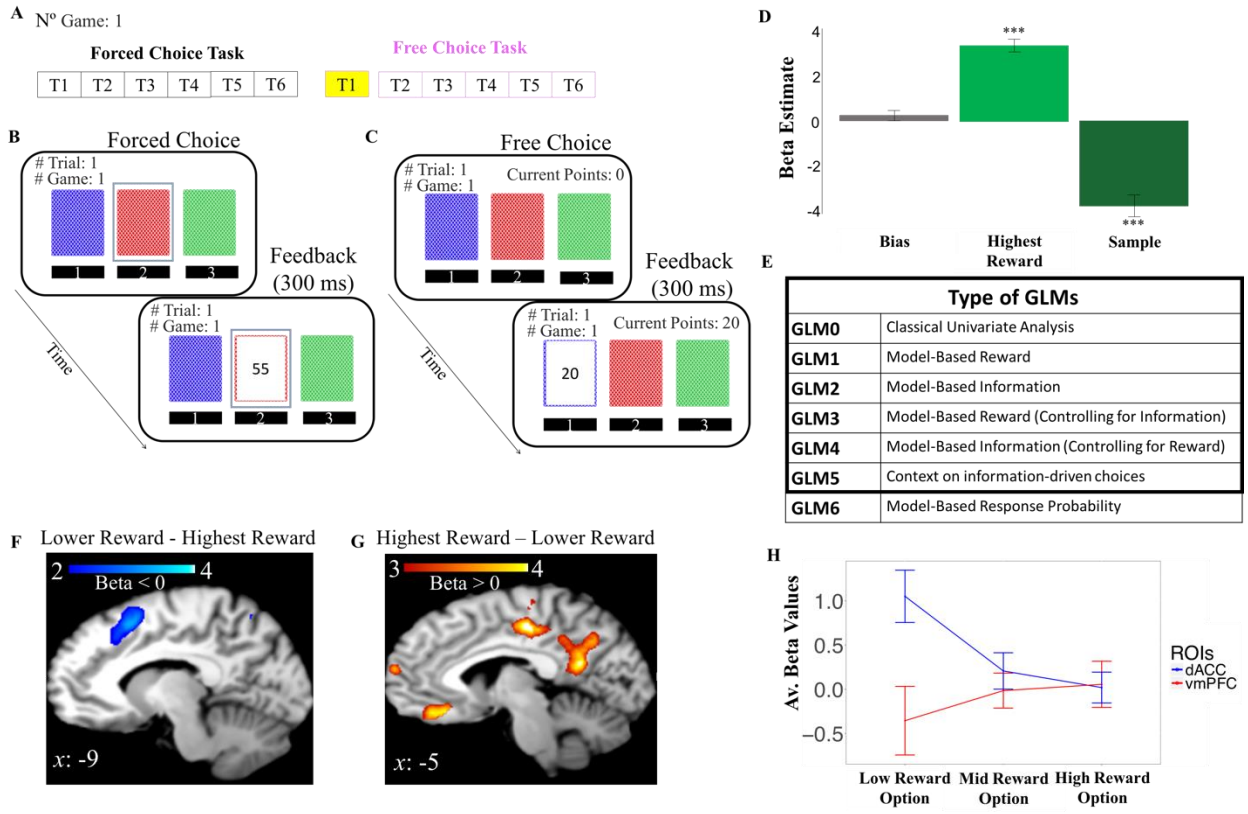
(A) In the single value system framework costs, effort etc. and rewards interact to produce a net-value estimate, (**B**) while in the dual value system framework information value and reward value are estimated independently. **(C)** In the single value system components representing costs/effort/difficulty and rewards negatively correlate. This correlation is also observed in the dual value system, despite the independence of information and reward systems in this framework: **(D)** optimizing either reward or information gain is associated with decreased activity in the alternate value system, leading to **(E)** symmetrically opposed activity between the systems. The values in C and D are standardized. This negative correlation holds across different model parameterizations – activity in the **(F)** reward system and **(G)** information system is negatively correlated across independent model simulations. Note: 'HighReward'- trials in which the models choose the deck associated with highest experienced reward; LowReward (otherwise). HighInfoGain – trials in which the models choose the never-sampled deck during the forced-choice task; and LowInfoGain (otherwise). The "model activity" was computed by running a first level analysis over the average of reward values (information values) in the HighReward trials minus the average of reward values in the LowReward trials (Reward Contrast); and reward values (information values) in the HighInfoGain trials minus the average of reward values (information values) in LowInfoGain trials (Information Contrast). For the single value system only reward values were entered into the analysis.

Results

Reward and information jointly influence choices

Human participants made sequential choices among 3 decks of cards over 128 games, receiving between 1 and 100 points after each choice (Figure 2). The task consisted of two phases (Figure 2A): a learning phase (forced-choice task) in which participants were instructed which deck to select on each trial (Figure 2B), and a decision phase (free-choice task) in which participants made their own choices with the goal of maximizing the total number of points obtained at the end of the experiment (Figure 2C). By controlling the levels of reward (i.e. points received) and information (i.e. # of samples per deck) experienced during the learning phase, it is possible, using appropriate analyses, to decorrelate both reward and information values in the first free-choice trial of each game (Fiure 2A) [41]. Logistic regression of subjects' behavior on that trial shows that, overall, choices were driven both by the reward (3.22, t(1,19) = 12.4, $p < 10^{-9}$) and information levels (-3.68, t(1,19) =-7.84, , $p < 10^{-6}$) experienced during the learning phase (Figure 2D). Reward-focused univariate fMRI analyses support previous findings: dACC activity is negatively correlated with the reward value of the selected option (FDR p = 0.076, voxel extent = 87, peak voxel coordinates (-2, 12, 58), t (19) = 4.66; FDR p = 0.076, voxel extent = 92, peak voxel coordinates (26, 6, 52), t (19) = 4.52; Figure 2F), while vmPFC activity is positively correlated with reward value (FEW p = 0.009, voxel extent = 203, peak voxel coordinates (-6, 30, -14), t (19) = 5.48; Fig 2G), following a symmetrically opposite pattern (Figure 2H) [11].

Figure 2. *Behavioral Task and Behavior*



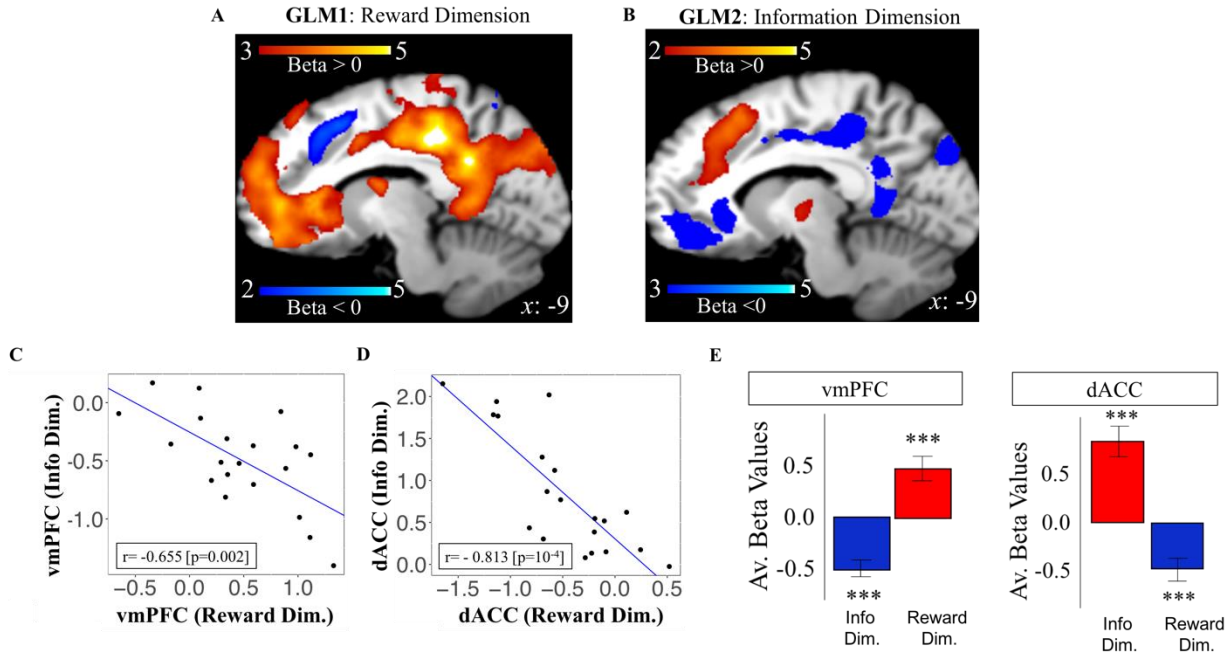
(**A**) One game of the behavioral task consisted of 6 consecutive forced choice trials and from 1 to 6 free-choice trials. FMRI analyses focused on the first free-choice trial (shown in yellow) in which reward and information were decorrelated. (**B**) In the *forced-choice task* participants chose a pre-selected deck of cards (outlined in blue), and were free to choose a deck of cards during the (**C**) *free-choice task* in order to maximize the total number of points obtained. (**D**) Participants' behavior was driven by both experienced reward and number of times the options was chosen in previous trials (beta weights from logistic regression; dependent variable is participants' exploitative choices). (**E**) Types of GLMs adopted in the fMRI analysis. Activity related to selecting the lower reward (**F**) and highest reward option options (**G**) was observed in vmPFC/dACC. Activity scale represents z-score. (**H**) DACC and vmPFC BOLD beta weights negatively correlated over the relative reward of subjects' choices.

Symmetrical activity in dACC and vmPFC as consequence of correlated variables

To carry out our model-based analyses of fMRI data, we obtained trial-by-trial estimates of subjects' expected reward and level of prospective information gain for selecting decks by fitting a reinforcement learning (RL) model with information integration [38] to participants' behavior (Methods). Being the first free-choice trial of each game where information and reward are orthogonalized by the experiment design, we focus the fMRI analysis to the time window preceding the first free choices. The relative reward value (or Reward) and the information that could be gained from sampling each deck (or Information Gain) derived from the RL model were regressed on the BOLD signal recorded on the first free-choice trial of

each game. In the first set of analyses, we ignore potential correlations between reward and information in order to replicate classical reward-focused analysis [10,40]. Reward and Information Gain were used as the only parametric modulators in separate GLMs (Figure 2E) to identify BOLD activity related to reward (GLM1) and to information (GLM2) respectively, on the first free-choice trial. Reward and Information Gain refers to the value associated with the chosen option in the first trial before the feedback is delivered. Unless otherwise specified, all results for these and subsequent analyses are cluster-corrected with a voxel-wise threshold of 0.001. Activity in vmPFC on the first free-choice trial correlated positively with relative reward (FWE $p < 0.001$, voxel extent = 1698, peak voxel coordinates (8, 28, -6), t (19) = 6.62) (Figure 3A) and negatively with information gain (FWE $p < 0.001$, voxel extent = 720, peak voxel coordinates (-10, 28, -2), t (19) = 5.36) (Figure 3B), while activity in dACC was negatively correlated with reward value (FWE $p = 0.001$, voxel extent = 321, peak voxel coordinates (6, 24, 40), t (19) = 4.59) (Figure 3A) and positively with information gain (FWE $p < 0.001$, voxel extent = 1441, peak voxel coordinates (8, 30, 50), t (19) = 7.13) (Figure 3B). These results support the findings from our univariate analyses (Figure 1E & 1F), and replicate the frequently-reported opposition effect: vmPFC activity positively correlates with the reward of the selected option, while dACC activity is negatively correlated. We additionally observed this symmetric opposition along the information dimension. Therefore, these results suggest the existence of a single value system for information and reward along the human PFC as suggested by reward-maximization theories [10,11,22,39,40]. Because of the confound between reward and information[41], however, only considering one choice dimension at a time may mislead the interpretation of value computation in the PFC (Figure 1). In support of our hypothesis that reward and information are confounded in our analyses, we observed that the beta values for GLM1 and GLM2 were positively correlated across subjects for both the vmPFC cluster (Figure 3C) and the dACC cluster (Figure 3D). Directly contrasting the beta estimates for Information Gain and Reward in both clusters revealed a symmetrically-opposed pattern of activity in both dimensions (Figure 3E).

Figure 3. *Symmetrical opposition between dACC and vmPFC as consequence of correlated variables*



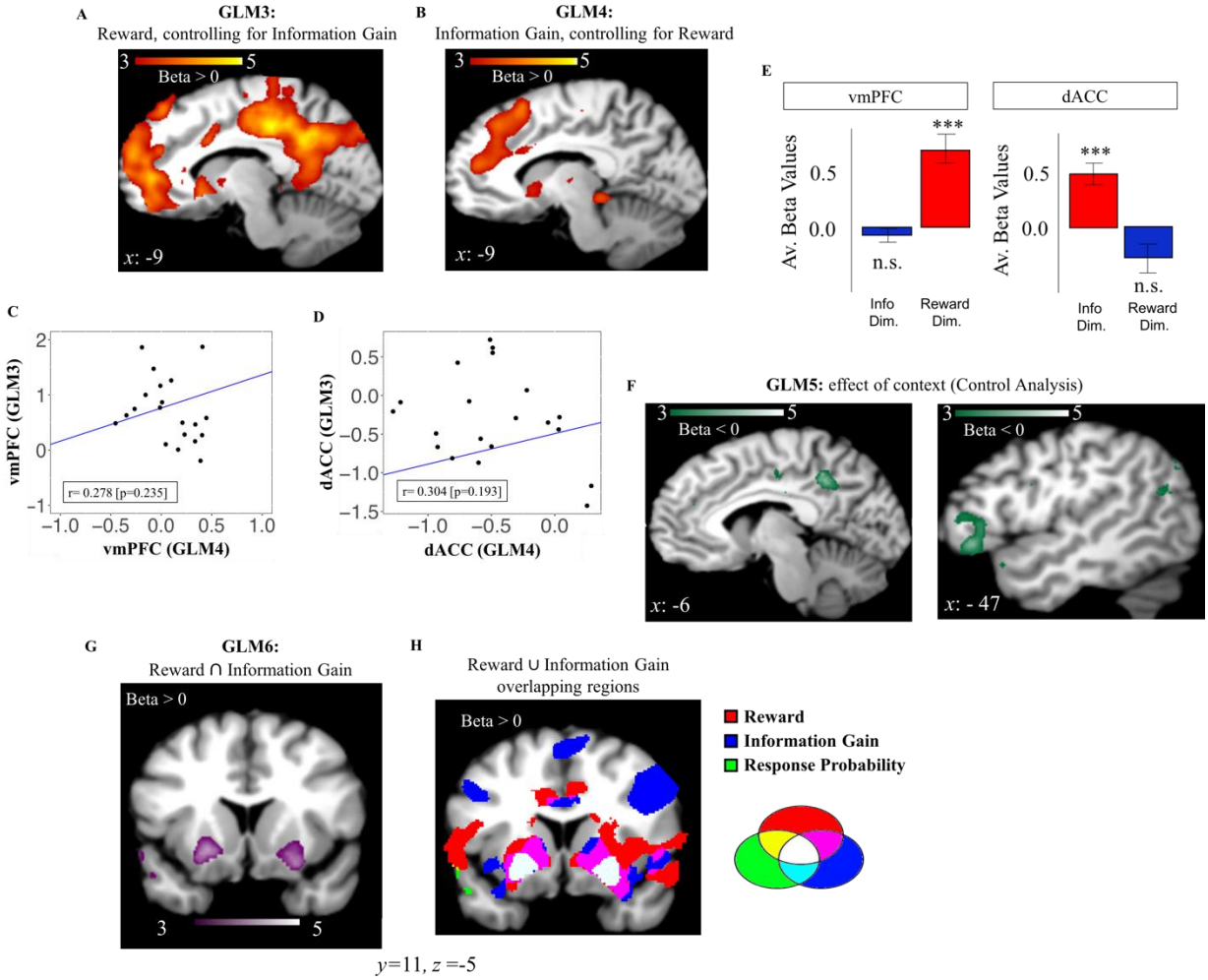
A) VMPFC correlated positively with model-based reward value for the selected option (in red) while dACC was negatively correlated (in blue). **B**) DACC (in red) positively correlated with model-based information gain, while vmPFC was negatively correlated (in blue). Activity scale represents z-score. BOLD signal estimates for Information Gain and Reward Value were negatively correlated across subjects for both **(C)** vmPFC and **(D)** dACC ROIs, and average BOLD beta estimates **(E)** for each ROI were dissociated along the Information and Reward dimensions, in line with model predictions (Figure 1).

Independent value systems for reward and information

In order to control for possible correlations between information and reward that may underlie our results for GLMs 1 & 2, a second set of analyses was conducted in which we investigated the effects of Reward after controlling for Information Gain (GLM3), and, conversely, the effects of Information Gain after controlling for Reward (GLM4; Methods). Activity in vmPFC remained positively correlated with relative reward of the chosen deck (Figure 4A; FWE $p < 0.001$, voxel extent = 1655, peak voxel coordinates (6, 46, -2), t(19) = 6.56) after controlling for Information Gain in GLM3. In contrast, whereas Reward was negatively correlated with dACC activity in GLM2, no significant cluster was observed after the removing variance associated with Information Gain in GLM3. Similarly, after controlling for the effects of Reward in GLM4, we observed significant activity in dACC positively correlated with Information Gain (Figure 4B; FWE $p < 0.001$, voxel extent = 764, peak voxel coordinates (10, 24, 58), t(19) = 5.89), while we found no correlated activity in vmPFC as observed in GLM1. Correlations across subjects between the beta estimates for Information Gain (after controlling for Reward) and Reward (after controlling for Information

Gain) from GLMs 3 & 4 demonstrates that activity in vmPFC is specifically related to the relative reward value of the chosen deck (Figure 4C) while activity in dACC is specifically related to information to be gained from the chosen deck (Figure 4D) [42]. Directly contrasting the beta estimates for Information and Reward in both clusters reveal an asymmetrical pattern of activity in the two dimensions (Figuer 4E). These results were replicated after contrasting GLM3 and GLM4 using a paired-t-test (GLM3>GLM4: vmPFC (FWE $p < 0.001$, voxel extent = 467, peak voxel coordinates (-4, 52, 16), t(19) = 5.59); GLM4>GLM3: dACC (FWE $p < 0.001$, voxel extent = 833, peak voxel coordinates (10, 24, 46), t(19) = 5.70); Supplementary Text). These findings thus reveal the coexistence of two independent value systems for reward and information in human PFC.

Figure 4. *Independent value systems for reward and information in PFC and their interaction in subcortical regions*



A) After controlling for information effects (GLM3), vmPFC activity (in red) positively correlated with model-based reward value, while no correlations were observed for dACC. **(B**) After controlling for reward effects (GLM4), dACC activity (in red) positively

correlated with model-based information gain, while no correlation was observed for vmPFC. The correlation of BOLD signal estimates between Reward Value and Information Gain was no longer observed in either **(C)** vmPFC nor **(D)** dACC, and comparison of average BOLD beta values **(E)** confirms that effects of Information Gain are only observed in dACC, while Reward Value is observed in vmPFC. "Info Dim" corresponds to the ROIs extracted from GLM4, while "Reward Dim" to the ROIs extracted from GLM3. **F**) Activity in the ventral putamen (striatum region) correlated with response probabilities derived from the GKRL model and **G**) Both Reward Value and Information Gain overlap in the striatum region (in white). Activity scale represents z-score.

Activity in dACC signals information and not long-term reward maximization

In our task, information-seeking behaviors may be driven by two motives: information-seeking for the sake of information, or information-seeking for long-term reward maximization. In other words, when subjects make choices that maximize information gain (when choosing an option in the first free-choice trial that was not observed during the learning phase) a tension between obtaining information *per se* vs. assisting long-term reward optimization may occur. In previous research, dACC activity has been interpreted as reflecting long-term reward maximization [14], suggesting that the primary purpose of information-seeking ultimately concerns reward. In order to rule out the possibility that dACC activity observed in our study reflects long-term reward maximization, we conducted an additional set of analyses (GLM5) with an additional reward context modulator included for trials in which subjects selected the most-informative option. The reward context modulator was calculated as the average reward obtained from the two decks sampled during the learning phase. If dACC is involved in *long-term* reward maximization, a modulation of its activity should be observed as a function of the context in which information-driven choices were made: dACC activity should be lower in richer reward contexts, while poorer reward contexts should increase its activity. That is, selecting an unknown option when known options are highly-rewarding is less beneficial for long-term reward maximization than selecting an unknown option when known options offer only small rewards. We observed no activity in dACC that correlates either positively or negatively with reward context (p unc. > 0.05). However, we did observe a negative correlation between the reward context modulator on the first free-choice trial with activity in ventrolateral PFC (FWE p = 0.001, voxel extent = 263, peak voxel coordinates (-46, 30, -4), t (19) = 5.65) and posterior cingulate cortex (FWE p = 0.028, voxel extent = 132, peak voxel coordinates (-8, -46, 36), t (19) =5.62). No activity was detected for the positive contrast (p unc. > 0.05). To better link the context modulator to the overall behavior, we ran the same analysis but on all choices (both motivated by reward and information-driven behaviors). The context modulator negative correlated with activity in ventrolateral region (FWE p = 0.038, voxel extent = 167, peak voxel coordinates (-12, 52, 4), t (19) = 4.92) and posterior cingulate cortex (FWE p = 0.075 p unc.= 0.08, voxel extent = 137, peak voxel coordinates (-6, -50, 38), t (19) = 4.72). As previously, activity in

dACC was not observed for either the negative or the positive contrast (p unc. > 0.05). Although it is difficult to interpret a null result, our finding that activity in ventrolateral PFC correlates with long-term reward maximization, in line with previous studies [43], suggests that our design was sufficiently-powered to detect long-term reward effects (if any) in dACC as well.

Information value and choice difficulty

Activity in dACC has been often associated with task difficulty [22] and conflict [44]. Trials with greater levels of choice difficulty or conflict may lead to prolonged reaction times, and dACC activity may index time on task [45] rather than task-related decision variables. In order to rule out the possibility that dACC activity, associated with information value in our task, might instead be driven by time on task, we correlated the standardized estimates of information value with choice reaction times on the first free choice trials. The correlation was run for each subject and correlation coefficients were tested against zeros using a Wilcoxon Singed Test. Overall, correlation coefficients were not significantly different from zero (Z= 164; p = 0.0958) suggesting that pursuing an option with higher or lower information value was not associated with higher or lower choice reaction times as predicted by a choice difficulty or conflict account of dACC function.

Reward and information signal combine in the striatum region

While distinct brain regions independently encode values across different dimensions of the chosen option, these values appear to converge at the level of the basal ganglia. In a final analysis (GLM6), we entered choice probabilities derived from the RL model (where Reward and Information Gain combine into a common option value; eq. 4) as a single parametric modulator, and we observed positively-correlated activity in bilateral ventral putamen (striatum region; right: FWE $p < 0.01$, voxel extent = 238, peak voxel coordinates (22, 16, -6), t(19) = 5.59); left: FWE $p < 0.01$, voxel extent = 583, peak voxel coordinates (-26, 8, -10), t(19) = 5.89) (Figure 4F). Additionally, ventral putamen overlaps with voxels passing a threshold of $p < 0.001$ for effects of both relative reward and information gain (Figure 4G) from GLMs 3 & 4.

Discussion

Decision-making outcomes are influenced by both reward and information about available options in the environment [38]. Here, we present evidence for dedicated and independent value systems for such decision variables in the human PFC. When correlations between reward and information were taken into account, we observed that dACC and vmPFC distinctly encode information value and relative reward value of the chosen option, respectively. These value signals were then combined in subcortical regions in order to implement choices. These findings are direct empirical evidence for a dedicated information value system in human PFC, independent of reward value. Our finding is in line with a view of human PFC as an optimizer of independent value signals [25,27,46].

Our main finding that dACC and vmPFC distinctly encode information gain and relative reward supports theoretical accounts such as active inference and certain RL models (e.g., upper confidence bound) which predict independent computations in the brain for information value (epistemic value) and reward value (extrinsic value) [26,47] [37]. Consistent with our findings, the activity of single neurons in the monkey orbitofrontal cortex independently and orthogonally reflects the output of the two value systems [48]. Therefore, our results may highlight a general coding scheme that the brain adopts during decision-making evaluation.

Our finding that activity in dACC positively correlates with the information value of the chosen option suggests the existence of a dedicated system for information in the human PFC independent of the reward value system. This result is in line with recent findings in monkey literature that identified a population of neurons in dACC which selectively encodes the information signal [33]. Additionally, our results are in line with computational models of PFC which predict that dACC activity can be primarily explained as indexing prospective information about an option independent of reward value [18,31,32] . DACC has often been associated with conflict [44] and uncertainty [23], and recent findings suggest that activity in the region corresponds to unsigned prediction errors, or "surprise" [49]. Our results enhance this perspective by showing that the activity observed in dACC during decision-making can be explained as representing the subjective representation of decision variables (i.e., information value signal) elicited in uncertain or novel environments. It is worth highlighting that other regions might be involved in processing information-relate components of the value signal not elicited by our task. In particular, orbitofrontal cortex signals the opportunity to receive knowledge vs. ignorance [36] and, rostrolateral PFC signals the changes in relative uncertainty associated to the exploration of novel and uncertain environments [50]. Neural recordings in monkeys also showed an interconnected cortico-basal ganglia network which resolve uncertainty during information seeking [33]. Taken together, these findings highlight an intricate and dedicate network for

processing information signals, independent of reward. Further research is therefore necessary to map the information network in the human brain.

Our finding that vmpFC positively correlates with the relative reward value of the chosen option agrees with previous research that identifies vmPFC as a region involved in value computation and reward processing [51]. VmPFC appears not only to code reward-related signals [52 53,54] but to specifically encode the relative reward value of the chosen option [55], in line with the results of our study. We also observed clusters in posterior cingulate cortex which were positively correlated with the relative reward value of the chosen option in a similar fashion as observed for vmPFC, suggesting a role of posterior cingulate in reward processing and exploitative behaviors as previously reported in monkey studies [56 57].

These independent value systems interact in the striatum, consistent with its hypothesized role in representing expected policies [47]. The convergence of reward and information signal in the striatum region is also consistent with the identification of basal ganglia as a core mechanism that supports stimulus-response associations in guiding actions [58] as well as recent findings demonstrating distinct corticostriatal connectivity for affective and informative properties of a reward signal [59]. Furthermore, our results are in line with recent evidence on multidimensional value encoding, as opposed to "pure" value encoding, in the striatum [60 61 62]. Moreover, activity in this region was computed from the softmax probability derived from our RL model, consistent with previous modeling work that identified the basal ganglia as the output of the probability distribution expressed by the softmax [63].

In addition to dACC, we observe activity in additional regions of the cognitive control network which correlated with the information value signal, including bilateral anterior insula cortex and dorsolateral PFC (dlPFC). Activity in these regions is frequently observed in conjunction with dACC activity, and this result is in line with a wide literature that associates anterior insula and dlPFC with behavioral control [64 65] and suppressing default behavior [66 67]. Although activity in these additional regions correlates with information value, it is unclear whether they, like dACC, represent information value per se, or instead may represent variables that correlate with information value but were not controlled for in this experiment, e.g., context uncertainty [32]. Additional work is needed to determine the unique contributions of these regions in signaling information value.

At the same time, our results question emerging views regarding the symmetrically opposing roles of dACC and vmPFC in value-based choice [11,68] and the role of PFC in explicitly calculating cost-benefit tradeoffs [10,39], and instead suggest the two regions encode distinct decision variables that are frequently confounded in studies of sequential decision-making [41]. The results of our study are in line with [69] who warned on the possibility that in most of neuroeconomics and decision neuroscience studies activity identified as a value signal might instead capture informational signaling of an outcome or particular hidden structure of a

decision problem. Recent work has emphasized ecologically-valid tasks for investigating behavior and brain function; while it is critical to characterize the function of brain structures in terms of the behaviors they evolved to support, increased task realism frequently entails a loss of control over experimental variables. While the present study focuses on reward and information, our results suggest that other decision dimensions (e.g., effort and motivation, cost, affective valence, or social interaction) may also be confounded in the same manner. Indeed, symmetrical opposition between dACC and vmPFC has been reported for a wide range of contexts involving decision variables such as effort, delay, and affective valence (Table S1). Our findings therefore suggest caution is needed when interpreting findings from such tasks.

Taken together, by showing the existence of independent value systems in the human PFC, this study provides the first empirical evidence in support of theoretical work aimed at developing a unifying framework for interpreting brain functions. Additionally, this study individuates a dedicated value system for information, independent of reward value. And, it suggests a new perspective on how to look at decision-making processes in the human brain under realistic scenarios, with clear implications for the interpretation of PFC activity in both healthy and clinical conditions.

Methods

Participants

Twenty-one right-handed, neurologically healthy young adults were recruited for this study (12 women; aged 19 - 29 years, mean age = 23.24). Of these, one participant was excluded from the analysis due to problems in the registration of the structural T_1 weighted MPRAGE sequence. The sample size was based on previous studies e.g., [10 14 22]. Participants also presented normal color vision and absence of psychoactive treatment. The entire group belonged to the Belgian Flemish-speaking community. The experiment was approved by the Ethical Committee of the Ghent University Hospital and conducted according to the Declaration of Helsinki. Informed consent was obtained from all participants prior to the experiment.

Procedure

Participants performed a gambling-task where on each trial choices needed to be made among three decks of cards [38] (Figure 2). The gambling-task consisted of 128 games. Each game contains two phases: a *forced-choice task* where participants selected options highlighted by the computer for 6 consecutive trials, and a *free-choice task* where participants produced their own choices in order to maximize the total gain obtained at the end of the experiment (from 1 to 6 trials). In the forced-choice task, participants were forced to either choose each deck 2 times (*equal information condition*), or to choose one deck 4 times, another deck 2 times, and 0 times for the remaining deck (*unequal information condition*). By using this two phase-task, Wilson *et al.* showed that the difference in the number of time each option is sampled and the differences in the mean reward is orthogonalized [41] (i.e., options associated with the lowest amount of information were least associated with experienced reward values [38]). In other words, the use of the forced-choice task allows to orthogonalize available information and reward delivered to participants in the first free choice trial. For this reason, the focus of our fMRI analyses is on the first free-choice of each game (resulting in 128 trials for the fMRI analysis). However, we adopted trial-by-trial fMRI analyses to have a better estimate of neural activity over the overall performance. Therefore, we treated *equal information condition* and *unequal information condition* altogether. This introduces information-reward confound in our analysis (Figure 1).

On each trial, the payoff was generated from a Gaussian distribution with a generative mean between 10 and 70 points and standard deviation of 8 points. Participants' payoff on each trial ranged between 1 and 100 points and the total number of points was summed and converted into a monetary payoff at the end of the experimental session (0.01 euros every 60 points). Participants underwent a training session outside the scanner in order to make the task structure familiar to them.

The forced-choice task lasted about 8 sec and was followed by a blank screen, for a variable jittered time window (1 sec - 7 sec). The temporal jitter allows to obtain neuroimaging data at the onset of the first-free choice trial and right before the option was selected (decision window). After participants performed

the first free-choice trial, a blank screen was again presented for a variable jittered time window (1 sec - 6 sec) before the feedback, indicating the number of points earned, was given for 0.5 sec and another blank screen was shown to them for a variable jittered time window. As the first free-choice trial was the main trial of interest for the fMRI analysis, subsequent free-choice trials were not jittered.

Image acquisition

Data were acquired using a 3T Magnetom Trio MRI scanner (Siemens), with a 32-channel radio-frequency head coil. In an initial scanning sequence, a structural T_1 weighted MPRAGE sequence was collected (176 high-resolution slices, TR = 1550 ms, TE = 2.39, slice thickness = 0.9 mm, voxel size = 0.9 x 0.9 x 0.9 mm, FoV = 220 mm, flip angle = 9°). During the behavioral task, functional images were acquired using a T_2* weighted EPI sequence (33 slices per volume, TR = 2000 ms, TE = 30 ms, no inter-slice gap, voxel size = 3 x 3 x 3mm, FoV = 192 mm, flip angle = 80°). On average 1500 volumes per participants were collected during the entire task. The task lasted approximately 1h split in 4 runs of about 15 minutes each.

Behavioral Analysis

To estimate participants' expected reward value and information value, we adopted a previously implemented version of a reinforcement learning model that learns reward values and information gained about each deck during previous experience - the gamma-knowledge Reinforcement Learning model (gkRL; [38,70]). This model was already validated for this task and it was better able to explain participants' behavior compared to other RL models [4].

Expected reward values were learned by gkRL adopting on each trial a simple δ learning rule [71]:

$$Q_{t+1,j}(c) = Q_{t,j}(c) + \alpha \times \delta_{t,j} \quad (1)$$

where $Q_{t,j}(c)$ is the expected reward value for deck *c* (= Left, Central or Right) at trial *t* and game *j* and $\delta_{t,j} = R_{t,j}(c) - Q_{t,j}(c)$ is the *prediction error*, which quantifies the discrepancy between the previous predicted reward values and the actual outcome obtained at trial *t* and game *j*.

Information was computed as follows:

$$I_{t,j}(c) = \left(\sum_{1}^{t} i_{t,j}(c)\right)^{\gamma}$$

$$\text{where, } i_{t,j}(c) = \begin{cases} 0, & choice \neq c \\ 1, & choice = c \end{cases} \quad (2)$$

$I_{t,j}(c)$, is the amount of information associated with the deck *c* at trial t and game j. $I_{t,j}(c)$, is computed by including an exponential term γ that defines the degree of non-linearity in the amount of observations obtained from options after each observation. γ is constrained to be > 0. Each time deck c is selected, $i_{t,j}(c)$

takes value of 1, and 0 otherwise. On each trial, the new value of $i_{t,j}(c)$ is summed to the previous $i_{t-1,1:j}(c)$ estimate and the resulting value is elevated to γ, resulting in $I_{t,j}(c)$.

Before selecting the appropriate option, gkRL subtracts the information gained $I_{t,j}(c)$ from the expected reward value $Q_{t+1,j}(c)$:

$$V_{t,j}(c) = Q_{t+1,j}(c) - I_{t,j}(c) * \omega \quad (3)$$

$V_{t,j}(c)$ is the final value associated with deck *c*. Here, information accumulated during the past trials scales values $V_{t,j}(c)$ so that increasing the number of observations of one option decreases its final value.

In order to generate choice probabilities based on expected reward values, the model uses a softmax choice function [72]. The softmax rule is expressed as:

$$P(c/V_{t,j}(c_i)) = \frac{\exp(\beta \times V_{t,j}(c))}{\sum_i \exp(\exp \beta \times V_{t,j}(c_i))} \quad (4)$$

where β is the inverse temperature that determines the degree to which choices are directed toward the highest rewarded option. By minimizing the negative log likelihood of $P(c/V_{t,j}(c_i))$ model parameters α, β, and ω, γ, were estimated for participants' choices made during the first free-choice trials. The fitting procedure was performed using MATLAB and Statistics Toolbox Release 2015b function *fminsearch*. Model parameters were then used to compute the value of $Q_{t+1,j}(c)$ and $I_{t,j}(c)$ for each participant. The results of this fit are reported in the table S2. To identify regions that tracked reward value and information value during the gambling-task, we entered them as regressors in a model-based fMRI analysis as explained below.

fMRI analysis

The first 4 volumes of each functional run were discarded to allow for steady-state magnetization. The data were preprocessed with SPM12 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Functional images were motion corrected (by realigning to the first image of the run). The structural T_1 image was coregistered to the functional mean image for normalization purposes. Functional images normalized to a standardized (MNI) template (Montreal Neurological Institute) and spatially smoothed with a Gaussian kernel of 8 mm full width half maximum (FWHM).

All the fMRI analyses focus on the time window associated to the onset of the first free-trials prior the choice was actually made (see *Procedure*). The rationale for our model-based analysis of fMRI data is as follows (Table S3). First, in order to link participants' behavior with neural activity, GLM0 was created with a regressor modelling choice onset associated with highest rewarded options (Highest Reward), and another regressor modelling choice onset associated with lower rewarded options (Lower Reward). Activity related to Highest Reward was then subtracted from the activity associated with Lower Reward (giving a value 1 and -1 respectively) at the second level. Next, in order to identify regions with activity

related to reward and information, two GLMs were created with a single regressor modeling the onset of the first free-choice trial as a 0 duration stick function. In GLM1, a single parametric modulator was included using the relative reward value of the chosen deck c computed by subtracting the average expected reward values for the unchosen decks from the expected reward values of the chosen deck c from the gkrl model ($Q^{R}_{t+1,j}(c=1) = Q_{t+1,j}(c=1) - mean\,(\,Q_{t+1,j}(c=2)\,, Q_{t+1,j}(c=3)\,)$). We adopted a standard computation of relative reward values [22]. It has already been shown that vmPFC represent reward values following the above computation. We refer to this regressor as Reward. In GLM2, a single parametric modulator was included using the negative value of gkrl model-derived information gained from the chosen option ($-\,I_{t,j}(c)$). We have already shown that humans represent information value as computed by our model compared to alternative computations when performing the behavioral task adopted in this study [38]. The negative value $I_{t,j}(c)$ relates to the information to be gained about each deck by participants. We refer to this regressor as Information Gain. Second, in order to identify regions with activity related to Reward (Information Gain) independent of effects due to Information Gain (Reward), two additional GLMs were created, also with a single regressor modeling the onset of the first free-choice trial. In GLM3, two parametric modulators were included in the order: Information Gain, Reward. In GLM4, the same two parametric modulators were included, with the order reversed (Reward, Information Gain). Because information and reward are expected to be partially correlated, the intent of GLMs 3&4 was to allow us to investigate the effects of the 2nd parametric modulator after accounting for variance that can be explained by the 1st parametric modulator. In SPM12, this is accomplished by enabling modulator orthogonalization (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Finally, to determine whether regions with activity related to reward (independent of information) and information (independent of reward) were *specific* to either quantity, beta weights for Reward (GLM3, parametric modulator 2) and Information Gain (GLM4, parametric modulator 2) were entered into a 2nd level (random effects) paired-sample t-test. In order to determine activity related to the context in which information-driven choices were made we created GLM5 with a context modulator modelling the onsets of information-driven choices. The context modulator constitutes the averaged rewards obtained from the two decks during the forced-choice task (e.g., if outcome from deck3 is observed in the forced-choice task, *Context*= mean (R_{deck1}, R_{deck2})). In order to determine activity related to the combination of information and reward value, GLM6 was created with the softmax probability of the chosen option ($P(c/V_{t,j}(c_i))$ modelling the onsets of first free-choices.

In order to denoise the fMRI signal, 24 nuisance motion regressors were added to the GLMs where the standard realignment parameters were non-linearly expanded incorporating their temporal derivatives and the corresponding squared regressors [73]. Furthermore, in GLM3 & 4 regressors were standardized to

avoid the possibility that parameter estimates were affected by different scaling of the models' regressors alongside with the variance they might explain [74]. During the second level analyses, we corrected for multiple comparison in order to avoid the false positive risk [75]. We corrected at cluster level using both FDR and FEW. Both corrections gave similar statistical results therefore we reported only FEW correction.

References:

1 Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci* **9**, 545-556, doi:10.1038/nrn2357 (2008).

2 Doya, K. Modulators of decision making. *Nat Neurosci* **11**, 410-416, doi:10.1038/nn2077 (2008).

3 Montague, P. R., King-Casas, B. & Cohen, J. D. Imaging valuation models in human choice. *Annu Rev Neurosci* **29**, 417-448, doi:10.1146/annurev.neuro.29.051605.112903 (2006).

4 Glimcher, P. W., Camerer, C., Fehr, E. & Poldrack, R. A. *Neuroeconimcs- decision-making and the brain*. (Academic Press, 2009).

5 Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An introduction*. (MIT Press, 1998).

6 Ellsberg, D. Risk, ambiguity, and the Savage axioms. *Q. J. Econ* **75**, 643–669, doi:doi: 10.2307/1884324 (1961).

7 Rosati, A. G., Stevens, J. R., Hare, B. & Hauser, M. D. The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. *Curr Biol* **17**, 1663-1668, doi:10.1016/j.cub.2007.08.033 (2007).

8 Botvinick, M. M., Huffstetler, S. & McGuire, J. T. Effort discounting in human nucleus accumbens. *Cogn Affect Behav Neurosci* **9**, 16-27, doi:10.3758/CABN.9.1.16 (2009).

9 Rushworth, M. F., 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, doi:10.1016/j.conb.2012.04.011 (2012).

10 Kolling, N., Behrens, T. E., Mars, R. B. & Rushworth, M. F. Neural mechanisms of foraging. *Science* **336**, 95-98, doi:10.1126/science.1216930 (2012).

11 Shenhav, A., Straccia, M. A., Botvinick, M. M. & Cohen, J. D. Dorsal anterior cingulate and ventromedial prefrontal cortex have inverse roles in both foraging and economic choice. *Cogn Affect Behav Neurosci* **16**, 1127-1139, doi:10.3758/s13415-016-0458-8 (2016).

12 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, doi:10.1016/j.neuron.2014.01.033 (2014).

13 Wittmann, M. K. *et al.* Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nat Commun* **7**, 12327, doi:10.1038/ncomms12327 (2016).

14 Boorman, E. D., Rushworth, M. F. & Behrens, T. E. Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. *J Neurosci* **33**, 2242-2253, doi:10.1523/JNEUROSCI.3022-12.2013 (2013).

15 Arulpragasam, A. R., Cooper, J. A., Nuutinen, M. R. & Treadway, M. T. Corticoinsular circuits encode subjective value expectation and violation for effortful goal-directed behavior. *Proc Natl Acad Sci U S A* **115**, E5233-E5242, doi:10.1073/pnas.1800444115 (2018).

16 Skvortsova, V., Palminteri, S. & Pessiglione, M. Learning to minimize efforts versus maximizing rewards: computational principles and neural correlates. *J Neurosci* **34**, 15621-15630, doi:10.1523/JNEUROSCI.1350-14.2014 (2014).

17 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, doi:10.1038/nature04766 (2006).

18 Behrens, T. E., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. Learning the value of information in an uncertain world. *Nat Neurosci* **10**, 1214-1221, doi:10.1038/nn1954 (2007).

19 Hogan, P. S., Galaro, J. K. & Chib, V. S. Roles of Ventromedial Prefrontal Cortex and Anterior Cingulate in Subjective Valuation of Prospective Effort. *Cereb Cortex* **29**, 4277-4290, doi:10.1093/cercor/bhy310 (2019).

20 Marsh, A. A., Blair, K. S., Vythilingam, M., Busis, S. & Blair, R. J. Response options and expectations of reward in decision-making: the differential roles of dorsal and rostral anterior cingulate cortex. *Neuroimage* **35**, 979-988, doi:10.1016/j.neuroimage.2006.11.044 (2007).

21 Kim, H. Y., Shin, Y. & Han, S. The reconstruction of choice value in the brain: a look into the size of consideration sets and their affective consequences. *J Cogn Neurosci* **26**, 810-824, doi:10.1162/jocn_a_00507 (2014).

22 Shenhav, A., Straccia, M. A., Cohen, J. D. & Botvinick, M. M. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nat Neurosci* **17**, 1249-1254, doi:10.1038/nn.3771 (2014).

23 Silvetti, M., Seurinck, R. & Verguts, T. Value and prediction error estimation account for volatility effects in ACC: a model-based fMRI study. *Cortex* **49**, 1627-1635, doi:10.1016/j.cortex.2012.05.008 (2013).

24 Hillman, K. L. & Bikey, D. K. Neural encoding of competitive effort in the anterior cingulate cortex. *Nature Neuroscience*, 1290-1297 (2012).

25 Friston, K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* **11**, 127-138, doi:10.1038/nrn2787 (2010).

26 FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J. & Friston, K. Active inference, evidence accumulation, and the urn task. *Neural Comput* **27**, 306-328, doi:10.1162/NECO_a_00699 (2015).

27 Friston, K. Learning and inference in the brain. *Neural Netw* **16**, 1325-1352, doi:10.1016/j.neunet.2003.06.005 (2003).

28 Kidd, C. & Hayden, B. Y. The Psychology and Neuroscience of Curiosity. *Neuron* **88**, 449-460, doi:10.1016/j.neuron.2015.09.010 (2015).

29 Bellemare, M. G. *et al.* Unifying count-based exploration and intrinsic motivation. *Advances in Neural Information Processing Systems* (2016).

30 Singh, S., Barto, A. G. & Chentanez, N. Intrinsically motivated reinforcement learning. *Adv. Neural Inform. Process. Syst. 17* (2005).

31 Alexander, W. H. & Brown, J. W. Medial prefrontal cortex as an action-outcome predictor. *Nat Neurosci* **14**, 1338-1344, doi:10.1038/nn.2921 (2011).

32 Alexander, W. H. & Brown, J. W. Frontal cortex function as derived from hierarchical predictive coding. *Sci Rep* **8**, 3843, doi:10.1038/s41598-018-21407-9 (2018).

33 White, J. K. *et al.* A neural network for information seeking. *Nat Commun* **10**, 5168, doi:10.1038/s41467-019-13135-z (2019).

34 Chung, J. J., Lawrance, N. R. J. & Sukkarieh, S. Learning to soar: Resource-constrained exploration in reinforcement learning. *The international journal of robotics research* **34**, 158-172 (2015).

35 Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O. & Clune, J. Go-Explore: a New Approach for Hard-Exploration Problems. *arXiv:1901.10995* (2019).

36 Charpentier, C. J., Bromberg-Martin, E. S. & Sharot, T. Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proc Natl Acad Sci U S A* **115**, E7255-E7264, doi:10.1073/pnas.1800547115 (2018).

37 Auer, P., Cesa-Bianchi, N. & Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning* **47**, 235-256 (2002).

38 Cogliati Dezza, I., Yu, A. J., Cleeremans, A. & Alexander, W. Learning the value of information and reward over time when solving exploration-exploitation problems. *Sci Rep* **7**, 16919, doi:10.1038/s41598-017-17237-w (2017).

39 Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217-240, doi:10.1016/j.neuron.2013.07.007 (2013).

40 Shenhav, A., Cohen, J. D. & Botvinick, M. M. Dorsal anterior cingulate cortex and the value of control. *Nat Neurosci* **19**, 1286-1291, doi:10.1038/nn.4384 (2016).

41 Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of experimental psychology. General* **143**, 2074-2081, doi:10.1037/a0038199 (2014).

42 Blanchard, T. C. & Gershman, S. J. Pure correlates of exploration and exploitation in the human brain. *Cogn Affect Behav Neurosci* **18**, 117-126, doi:10.3758/s13415-017-0556-2 (2018).

43 Tanaka, S. C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* **7**, 887-893, doi:10.1038/nn1279 (2004).

44 Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict monitoring and cognitive control. *Psychol Rev* **108**, 624-652 (2001).

45 Grinband, J. *et al.* The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage* **57**, 303-311, doi:10.1016/j.neuroimage.2010.12.027 (2011).

46 Friston, K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* **360**, 815-836, doi:10.1098/rstb.2005.1622 (2005).

47 Friston, K. *et al.* Active inference and epistemic value. *Cogn Neurosci* **6**, 187-214, doi:10.1080/17588928.2015.1020053 (2015).

48 Blanchard, T. C., Hayden, B. Y. & Bromberg-Martin, E. S. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron* **85**, 602-614, doi:10.1016/j.neuron.2014.12.050 (2015).

49 Vassena, E., Deraeve, J. & Alexander, W. H. Surprise, value and control in anterior cingulate cortex during speeded decision-making. *Nat Hum Behav*, doi:10.1038/s41562-019-0801-5 (2020).

50 Badre, D., Doll, B. B., Long, N. M. & Frank, M. J. Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron* **73**, 595-607, doi:10.1016/j.neuron.2011.12.025 (2012).

51 Smith, D. V. & Delgado, M. R. in *Brain, Mapping: An Encyclopedic Reference* Vol. 3 361-366 (Academic Press, 2015).

52 Chib, V. S., Rangel, A., Shimojo, S. & O'Doherty, J. P. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J Neurosci* **29**, 12315-12320, doi:10.1523/JNEUROSCI.2575-09.2009 (2009).

53 Kim, H., Shimojo, S. & O'Doherty, J. P. Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb Cortex* **21**, 769-776, doi:10.1093/cercor/bhq145 (2011).
54 Hampton, A. N., Bossaerts, P. & O'Doherty, J. P. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* **26**, 8360-8367, doi:10.1523/JNEUROSCI.1010-06.2006 (2006).
55 Boorman, E. D., Behrens, T. E., Woolrich, M. W. & Rushworth, M. F. 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, doi:10.1016/j.neuron.2009.05.014 (2009).
56 McCoy, A. N., Crowley, J. C., Haghighian, G., Dean, H. L. & Platt, M. L. Saccade reward signals in posterior cingulate cortex. *Neuron* **40**, 1031-1040 (2003).
57 Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y. & Platt, M. L. Posterior cingulate cortex: adapting behavior to a changing world. *Trends Cogn Sci* **15**, 143-151, doi:10.1016/j.tics.2011.02.002 (2011).
58 Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337-1340, doi:10.1126/science.1115270 (2005).
59 Smith, V. D., Rigney, A. E. & Delgado, M. R. Distinct Reward Properties are Encoded via Corticostriatal Interactions. *Scientific Reports*, doi: DOI: 10.1038/srep20093 (2016).
60 Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898-1902, doi:10.1126/science.1077349 (2003).
61 Cai, X., Kim, S. & Lee, D. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* **69**, 170-182, doi:10.1016/j.neuron.2010.11.041 (2011).
62 Costa, V. D., Mitz, A. R. & Averbeck, B. B. Subcortical Substrates of Explore-Exploit Decisions in Primates. *Neuron* **103**, 533-545 e535, doi:10.1016/j.neuron.2019.05.017 (2019).
63 Humphries, M. D., Khamassi, M. & Gurney, K. Dopaminergic Control of the Exploration-Exploitation Trade-Off via the Basal Ganglia. *Front Neurosci* **6**, 9, doi:10.3389/fnins.2012.00009 (2012).
64 Singer, T., Critchley, H. D. & Preuschoff, K. A common role of insula in feelings, empathy and uncertainty. *Trends Cogn Sci* **13**, 334-340, doi:10.1016/j.tics.2009.05.001 (2009).
65 Morris, R. W., Dezfouli, A., Griffiths, K. R. & Balleine, B. W. Action-value comparisons in the dorsolateral prefrontal cortex control choice between goal-directed actions. *Nat Commun* **5**, 4390, doi:10.1038/ncomms5390 (2014).
66 Lerner, A. *et al.* Involvement of insula and cingulate cortices in control and suppression of natural urges. *Cereb Cortex* **19**, 218-223, doi:10.1093/cercor/bhn074 (2009).
67 Sridharan, D., Levitin, D. J. & Menon, V. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A* **105**, 12569-12574, doi:10.1073/pnas.0800005105 (2008).
68 Shenhav, A., Straccia, M. A., Musslick, S., Cohen, J. D. & Botvinick, M. M. Dissociable neural mechanisms track evidence accumulation for selection of attention versus action. *Nat Commun* **9**, 2485, doi:10.1038/s41467-018-04841-1 (2018).

69 O'Doherty, J. P. The problem with value. *Neurosci Biobehav Rev* **43**, 259-268, doi:10.1016/j.neubiorev.2014.03.027 (2014).
70 Cogliati Dezza, I., Cleeremans, A. & Alexander, W. Should we control? The interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma. *Journal of experimental psychology. General*, doi:10.1037/xge0000546 (2019).
71 Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning: Current research and theory*, 64-99 (1972).
72 Daw, N. D. & Doya, K. The computational neurobiology of learning and reward. *Curr Opin Neurobiol* **16**, 199-204, doi:10.1016/j.conb.2006.03.006 (2006).
73 Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. & Turner, R. Movement-related effects in fMRI time-series. *Magn Reson Med* **35**, 346-355 (1996).
74 Erdeniz, B., Rohe, T., Done, J. & Seidler, R. D. A simple solution for model comparison in bold imaging: the special case of reward prediction error and reward outcomes. *Front Neurosci* **7**, 116, doi:10.3389/fnins.2013.00116 (2013).
75 Chumbley, J. R. & Friston, K. J. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage* **44**, 62-70, doi:10.1016/j.neuroimage.2008.05.021 (2009).

Acknowledgments: funded by F.R.S.-fNRS (I.C.D.), FWO-Flanders Odysseus II Award #G.OC44.13N (W.A.) and A.C. was partly supported by an Advanced Grant (RADICAL) from the European Research Council.

Author Contribution: I.C.D. and W.A. designed and carried out the experiment and discussed the computational modelling and fmri analysis. I.C.D. performed the fmri analysis and the model analysis. I.C.D. and W.A. discussed and interpreted the data. I.C.D, A.C. and W.A. wrote the manuscript.

Supplementary Material: Supplementary text and Materials and Methods, Figures S1-3, Tables S1-S3, References (*1-16*) accompanies this paper (bottom of this document).

Competing Interests: The authors declare that they have no competing interests.

Supplementary Materials for

Distinct Value Systems for Reward and Information in Human Prefrontal Cortex

I. Cogliati Dezza, A. Cleeremens, W. Alexander

Correspondence to: irene.cogliatidezza@gmailcom

This PDF file includes:

Supplementary Text and Results
Figure S1 to S3
Tables S1 to S3

Supplementary Text and Results

VmPFC and dACC symmetrical opposition as evidence for single distributed value system in PFC

In the main text, we argue that the symmetric opposition between vmPFC and dACC in value-based decision-making is extensively documented in the neuroscientific literature, and identify several recent studies that make opposition claims (Table S1). Here, we discuss some relevant papers in more details.

Using a sequential decision-making task which alternates engage choices (engaging with choices that are offered to participants) and forage choices (explore alternatives options presented in the environment), Kolling et al. showed that vmPFC activity reflects decision to engage, and it negatively correlated with the value of foraging. In contrast, dACC positively correlates with the value of foraging and negatively correlates with the value of engaging [1]. In an additional study from the same group, Kolling et al. reported opposing effects in vmPFC and dACC as a function of risk: vmPFC activity decreased with increased risk, while dACC activity increased with increased risk during riskier choices [12]. Using an implementation of Kolling et al. 2012 sequential decision-making task, Shenhav et al. showed that foraging value is encoded in an opposite fashion in vmPFC (going from negative to positive) and dACC (going from positive to negative) as choice difficulty decreases [13]. Boorman et al. reported the same opposition effect using an alternative sequential decision-making paradigm. Participants have to make repeated choices among 3 options based on reward expectations learned throughout the task. Their results showed that vmPFC activity reflects the value of the chosen option, while dACC activity reflects the value of the long-term best option [2]. Furthermore, this symmetrical opposition is also observed in effort-based choices: vmPFC activity positively correlated with expected subjective value of the chosen option, while dACC negatively correlate with it [14]; vmPFC activity positively correlated with the expected reward of the chosen option, while dACC negatively correlated with it [15]. Overall, this empirical evidence suggests for a single distributed system along the human PFC that performs a cost/benefit analysis across a wide range of value-based decision-making contexts.

vmPFC and dACC opposition in value-based choice in absence of symmetrical opposition

In order to show how functional opposition between vmPFC and dACC in value-based choices may be observed even in absence of clear symmetric opposition of activity, we simulated an effort-based environment where rewards could be obtained only after exerting effort. In many effort-based paradigms[16], subjects must choose between a small, default reward that requires little effort to obtain, or a larger reward that requires greater effort, and consequently a chance of failing to perform the task adequately and not receiving a reward. We adapted our RL model in order to simulate choices made by an agent performing this task. In this implementation, the information value is equal to the entropy (-p*log(p); where p is the probability of successfully performing the task) resulting in the following value function:

$$V_{t,j}(c) = Q_{t+1,j}(c) * p + p * \log(p(c)) * \omega$$

We simulated the model across different ranges of effort and rewards. While the probability of the model selecting the non-default option decreased with effort level (Figure S1A) and increased with relative reward value (Figure S1B), consistent with research in this area, there was no correlation between the relative reward value and effort level (Figure S1C). Finally, for the range of effort levels included in this simulation, the level of effort correlated with the information value signal (Figure S1D). This result suggests that even when activity in dACC (frequently interpreted as indicating effort) and vmPFC (relative reward) can be dissociated in value-based decision-tasks, the interpretation that the regions serve functionally opposed roles may be misguided.

Information and reward confound in decision-making tasks

In this section, we first report the results of simulations of a dual value system RL model on our gambling task as well as the sequential decision-making task adopted by [1 13]. Next, we report the simulations of a single value system.

We ran 63 simulations of the gkRL model on our gambling-task. The model parameters were selected in the range of those estimated in our sample. Next, we classified model's choices as HighReward (when choosing the deck associated with highest experienced reward), LowReward (otherwise), HighInfoGain (when choosing the never-sampled deck during the forced-choice task) and LowInfoGain (otherwise). Subsequently, we computed Information Gain and Reward (as explained in *fMRI analysis* section) in order to simulate the activity associated with the reward system and the information system. We then compute the "model activity" by running a first level analysis over the average of Reward (Information Gain) in the HighReward trials minus the average of reward values in the LowReward trials (Reward Contrast); and the average of Reward (Information Gain) in the HighInfoGain trials minus the average of Reward (Information Gain) in LowInfoGain trials (Information Contrast). This analysis was repeated for all model simulations. As reported in Figure 1, activity associated with the Reward and the Information Contrast are correlated in both value systems. Moreover, along both the reward dimension and information dimension, the two systems are represented in a symmetrically opposing manner. These results suggest that, even if reward and information are represented by distinct and independent systems, reward and information signals are nevertheless correlated within the same system. Next, we extended this analysis to other decision-making tasks already published in the literature (e.g., [1,13]). As in previous versions[1,13], two cards are displayed on each game and their reward magnitude is visible to the agent. The model has to decide either to engage, which will lead to an economic decision between the two cards (engage cards), or to forage, which will lead to sample alternatives options from the back-up cards. The model has access to the reward magnitude of the back-up cards. Choosing to forage is associated with a cost (ranging between 0 and 3 points). We presented the agent with two

conditions: High Information and Low Information. In High Information, half of the back-up cards had lower values than those of the engaged options while the other half had lower values. Therefore, this condition has maximal uncertainty, i.e., the mean value of new cards obtained through foraging was equally likely to be higher or lower than the mean value of the current cards. Therefore, if the agent decides to forage it has no information on the actual value of the card that will be selected. In the Low Information condition, all back-up cards could have higher or lower values with respect of the engage options. Therefore, this condition has minimum uncertainty, since the mean value of new cards was guaranteed to be either higher or lower than the current mean card value. The task lasts 135 trials. On each trial, the model computes the value of foraging (i.e., the mean reward of the back-up cards minus the cost of foraging plus the uncertainty associated with the back-up cards: mean (Reward back-up cards)-cost + sd (Reward back-up cards)) and the value of engaging (i.e., the mean reward of engage cards). The model computes decision policies by entering both values into a softmax function. As in our previous set of analyses, we classified model's choices as HighReward (when choosing the option- forage or engage-associated with the highest mean reward), LowReward (otherwise); HighInfoGain (when choosing forage in High Information condition) and LowInfoGain (otherwise). Subsequently, we computed Information Gain as the value of foraging and Reward as the value of engaging in order to simulate neural activity associated with the reward system and information system. We then ran a first level analysis over the averaged value of engaging (value of foraging) in the HighReward choices minus the average of reward values in the LowReward choices (Reward Contrast); and the averaged value of engaging (value of foraging) in the HighInfoGain trials minus the averaged value of engage (value of foraging) in LowInfoGain trials (Information Contrast). As already shown with the simulation of our gambling task, activity associated with Reward and Information Contrast are correlated in both value systems and both the reward dimension and information dimension are represented in a symmetrically opposing manner within the two systems (Figure S2). These results suggest that the confound in the representation of reward and information with value systems can be generalize to other sequential decision-making tasks.

Next, we simulated a single value system. In this simulation, the single value system is a standard RL model where the expected reward value in eq 2 enters directly in eq 4 without integrating any information. We ran this model on our gambling task and we conducted the same analyses reported above (i.e., interpreting model choices as HighInfoGain, LowInfoGain, HighReward, LowReward; compute information and reward contrast associated with the Reward system). We observed the same prediction of a dual value system: reward and information correlates within the reward system. This suggests that predictions of neural activity made by a single and a dual value system are indistinguishable, if the confound between reward and information is not taken into account.

Behavioral results

In order to investigate participants' behavior during the scanner session, we performed a logistic regression for each participant over exploitative choices against the following normalized variables: highest experienced reward (Highest Reward) and number of samples for the highest rewarded option (Nº samples). In particular, the dependent variable had binary output {exploitative choices =1; non-exploitative choices – or exploration = 0 otherwise}. Exploitative choice trials were classified as those trials in which participants chose the option in the first free-choice trial associated with the highest average of points collected during the forced-choice task of the same game. Beta coefficients were collected for the entire group and a one sample t-test was conducted as shown in the main text to test whether coefficients differed from 0 (Figure 2D).

Univariate Analysis

To investigate the neural correlates of participants' behavior during the task, we conducted one sample t-test on the beta weights estimated for GLM0. For the positive t-test (Highest Reward – Lower Reward), we observed significant activity in vmPFC (FEW p = 0.009, voxel extent = 203, peak voxel coordinates (-6, 30, -14), t (19) = 5.48), in posterior cingulate (FEW p < 0.001, voxel extent = 732, peak voxel coordinates (-6, -22, 42), t (19) = 5.88; FEW p < 0.001, voxel extent = 732, peak voxel coordinates (8, -44, 28), t (19) = 6.77) and in medial orbitofrontal cortex (FEW p =0.037, voxel extent = 146, peak voxel coordinates (-10, 60, 16), t (19) = 5.68; Figure 2F). For the negative t-test (Lower Reward – Highest Reward), we observed significant activity in dACC at p uncorr < 0.001 (FDR p = 0.076, voxel extent = 87, peak voxel coordinates (-2, 12, 58), t (19) = 4.66; FDR p = 0.076, voxel extent = 92, peak voxel coordinates (26, 6, 52), t (19) = 4.52; Figure 2G).

Reward & Information Value under correlated activity

To investigate regions involved in processing the relative reward value associated with the chosen options during the first free-choice trials of the gambling-task, we conducted a one sample t-test on the beta weights estimated for the parametric modulator (Reward) for GLM1. For the positive t-test (beta > 0), indicating activity correlated with the relative reward value of the chosen deck, we observed significant activity in vmPFC (as reported in the main text) and in posterior cingulate (FWE $p < 0.001$, voxel extent = 560, peak voxel coordinates (-8, -50, 28), t (19) = 8.61) (Figure 3A). For the negative t-test (beta < 0), results showed significant activity in dACC (as reported in the main text) (Figure 3A).

To identify brain regions involved in processing information gain of the selected options during the first free-choice trials of the gambling-task, we conducted a one sample t-test on the beta weights estimated for the parametric modulator (Information Gain) for GLM2. For the positive t-test (beta < 0), indicating activity correlated with choosing options about which the participant had gained the most information from, regions commonly associated with Reward were observed, including a cluster in vmPFC (as reported in the main text) and posterior cingulate (FWE $p < 0.001$, voxel extent = 362, peak voxel coordinates (-10, -34,

46), t(19) = 6.62) (Figure 3B). For the negative t-test (beta > 0), indicating activity associated with choosing options about which the participant had the least amount of information, significant activity was observed in regions commonly associated with cognitive control, including dACC (as reported in the main text), bilateral anterior insula (right: FWE $p < 0.01$, voxel extent = 300, peak voxel coordinates (34, 22, -8), t(1, 19) = 5.06); left: FWE $p < 0.05$, voxel extent = 214, peak voxel coordinates (-34, 20, 6), t(1, 19) = 5.04) and right dlPFC (FWE $p = 0.001$, voxel extent = 300, peak voxel coordinates (42, 16, 40), t(1, 19) = 4.93) (Figure 3B).

Additionally, for each subject we computed the average beta estimates for vmPFC-cluster and dACC-cluster in both GLM1 and GLM2 and we correlated those estimated between the two GLMs. VMPFC in GLM1 positive correlated with vmPFC in GLM2 (Figure 3C) and dACC in GLM1 positive correlated with dACC in GLM2 (Figure 3D).

Dissociable regions for Reward and Information

In the previous analyses, we observed regions with overlapping activity for reward and information. Regions frequently associated with reward, including vmPFC and posterior cingulate, also appeared to correlate with the information already known about the chosen option, while cognitive control regions such as dACC and anterior insula, implicated in overriding default or prepotent value-based responses, were more active for trials in which participants selected lower-value options as well as options for which more information could be gained. As noted previously, gained information and experienced reward are frequently correlated in studies of value-based decision-making. Therefore, in order to determine whether the activity in regions observed in our previous analysis was specific to either reward value or the amount of information that could be gained about the chosen option, we turn to GLMs 3 & 4. In GLM 3, we investigate the effects of Reward after accounting for variance that can be explained by Information Gain, while in GLM 4, we investigate effects of Information Gain after accounting for variance that can be explained by Reward. If the activity of regions observed in our previous analyses is due only to the variance shared by Information Gain and Reward, then no activity should be observed after removing that component of the variance. On the other hand, if activity is best explained by variance unique either Reward (GLM3) or Information Gain (GLM4), the regions observed in the previous analyses should also be observed here. In GLM3, we first account for variance explained by Information Gain, after which we conduct a one sample t-test on the beta weights estimated for the effects of Reward. In GLM3 Reward still explains a significant proportion of variance in regions typically associated with reward value, including vmPFC (as reported in the main text), and posterior cingulate (posterior cingulate: FWE $p < 0.001$, voxel extent = 603, peak voxel coordinates (-2, -50, 26), t(19) = 7.30, (Figure 4A). Conversely, no significant cluster was observed for negative beta. In GLM4, we first accounted for variance explained by Reward, after which we conducted a one sample t-test on beta

weights estimated for the effects of Information Gain. Here, for the effect of Information Gain (beta > 0), we find significant activity in dACC (as reported in the main text) and bilateral insula (left: FWE $p < 0.05$, voxel extent = 229, peak voxel coordinates (-38, 18, -10), t(19) = 5.31); right: FWE $p < 0.05$, voxel extent = 220, peak voxel coordinates (34, 22, -8), t(19) = 4.44) (Figure 4B). Conversely, no significant cluster was observed for negative betas. Additionally, we correlate average beta estimates for vmPFC-cluster and d-ACC cluster in both GLMs. Results did not show any correlation between vmPFC in GLM3 and GLM4 (Figure 4C) and dACC in GLM3 and GLM4 (Figure 4D).

While our results from GLMs 3 & 4 demonstrate that activity in vmPFC & posterior cingulate is explained by Reward after controlling for Information Gain, and activity in dACC & anterior insula is explained by Information Gain after controlling for Reward, these analyses do not allow us to conclude that one set of regions is specific to reward while the other is specific to information (i.e., while we can say the, for example, Reward is different than 0, and Information Gain is not different than 0, we cannot say Reward is different than Information Gain). In order to do so, we directly compare the beta weights estimated for Reward (after orthogonalizing with respect to Information Gain) from GLM3 and the beta weights estimated for Information Gain (orthogonalized with respect to Reward) from GLM4 using a paired-sample t-test. We find clusters of activity in vmPFC (as reported in the main text), posterior cingulate (FWE $p < 0.001$, voxel extent = 1493, peak voxel coordinates (-14, -48, 36), t(19) = 6.02) and putamen (FWE $p < 0.001$, voxel extent = 920, peak voxel coordinates (24, 10, -8), t(19) = 6.13) in which Reward > Information Gain (Figure S3A), indicating that these regions are specifically involved in reward processing, while a significant cluster is observed in dACC (as reported in the main text), right insula (FWE $p < 0.05$, voxel extent = 157, peak voxel coordinates (34, 24, -6), t(19) = 4.89)) and dlPFC (FWE $p < 0.05$, voxel extent = 158, peak voxel coordinates (48, 32, 32), t(19) = 4.89) for (negative) Information Gain > Reward (Figure S3B), indicating that dACC is specifically involved in representing uncertainty.

References

1 Kolling, N., Behrens, T. E., Mars, R. B. & Rushworth, M. F. Neural mechanisms of foraging. *Science* **336**, 95-98, doi:10.1126/science.1216930 (2012).

2 Boorman, E. D., Rushworth, M. F. & Behrens, T. E. Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. *J Neurosci* **33**, 2242-2253, doi:10.1523/JNEUROSCI.3022-12.2013 (2013).

3 Shenhav, A., Straccia, M. A., Cohen, J. D. & Botvinick, M. M. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nat Neurosci* **17**, 1249-1254, doi:10.1038/nn.3771 (2014).

4 Cogliati Dezza, I., Yu, A. J., Cleeremans, A. & Alexander, W. Learning the value of information and reward over time when solving exploration-exploitation problems. *Sci Rep* **7**, 16919, doi:10.1038/s41598-017-17237-w (2017).

5 Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of experimental psychology. General* **143**, 2074-2081, doi:10.1037/a0038199 (2014).

6 Cogliati Dezza, I., Cleeremans, A. & Alexander, W. Should we control? The interplay between cognitive control and information integration in the resolution of the exploration-exploitation dilemma. *Journal of experimental psychology. General*, doi:10.1037/xge0000546 (2019).

7 Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning: Current research and theory*, 64-99 (1972).

8 Daw, N. D. & Doya, K. The computational neurobiology of learning and reward. *Curr Opin Neurobiol* **16**, 199-204, doi:10.1016/j.conb.2006.03.006 (2006).

9 Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. & Turner, R. Movement-related effects in fMRI time-series. *Magn Reson Med* **35**, 346-355 (1996).

10 Erdeniz, B., Rohe, T., Done, J. & Seidler, R. D. A simple solution for model comparison in bold imaging: the special case of reward prediction error and reward outcomes. *Front Neurosci* **7**, 116, doi:10.3389/fnins.2013.00116 (2013).

11 Chumbley, J. R. & Friston, K. J. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage* **44**, 62-70, doi:10.1016/j.neuroimage.2008.05.021 (2009).

12 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, doi:10.1016/j.neuron.2014.01.033 (2014).

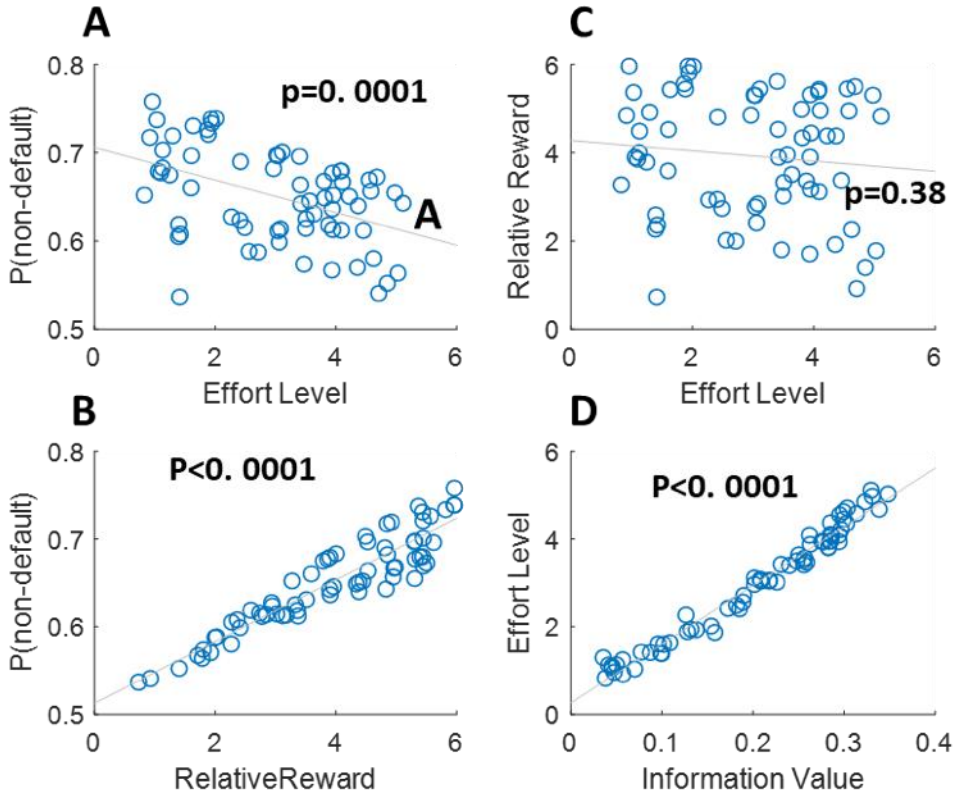
13 Shenhav, A., Straccia, M. A., Botvinick, M. M. & Cohen, J. D. Dorsal anterior cingulate and ventromedial prefrontal cortex have inverse roles in both foraging and economic choice. *Cogn Affect Behav Neurosci* **16**, 1127-1139, doi:10.3758/s13415-016-0458-8 (2016).

14 Arulpragasam, A. R., Cooper, J. A., Nuutinen, M. R. & Treadway, M. T. Corticoinsular circuits encode subjective value expectation and violation for effortful goal-directed behavior. *Proc Natl Acad Sci U S A* **115**, E5233-E5242, doi:10.1073/pnas.1800444115 (2018).

15 Skvortsova, V., Palminteri, S. & Pessiglione, M. Learning to minimize efforts versus maximizing rewards: computational principles and neural correlates. *J Neurosci* **34**, 15621-15630, doi:10.1523/JNEUROSCI.1350-14.2014 (2014).

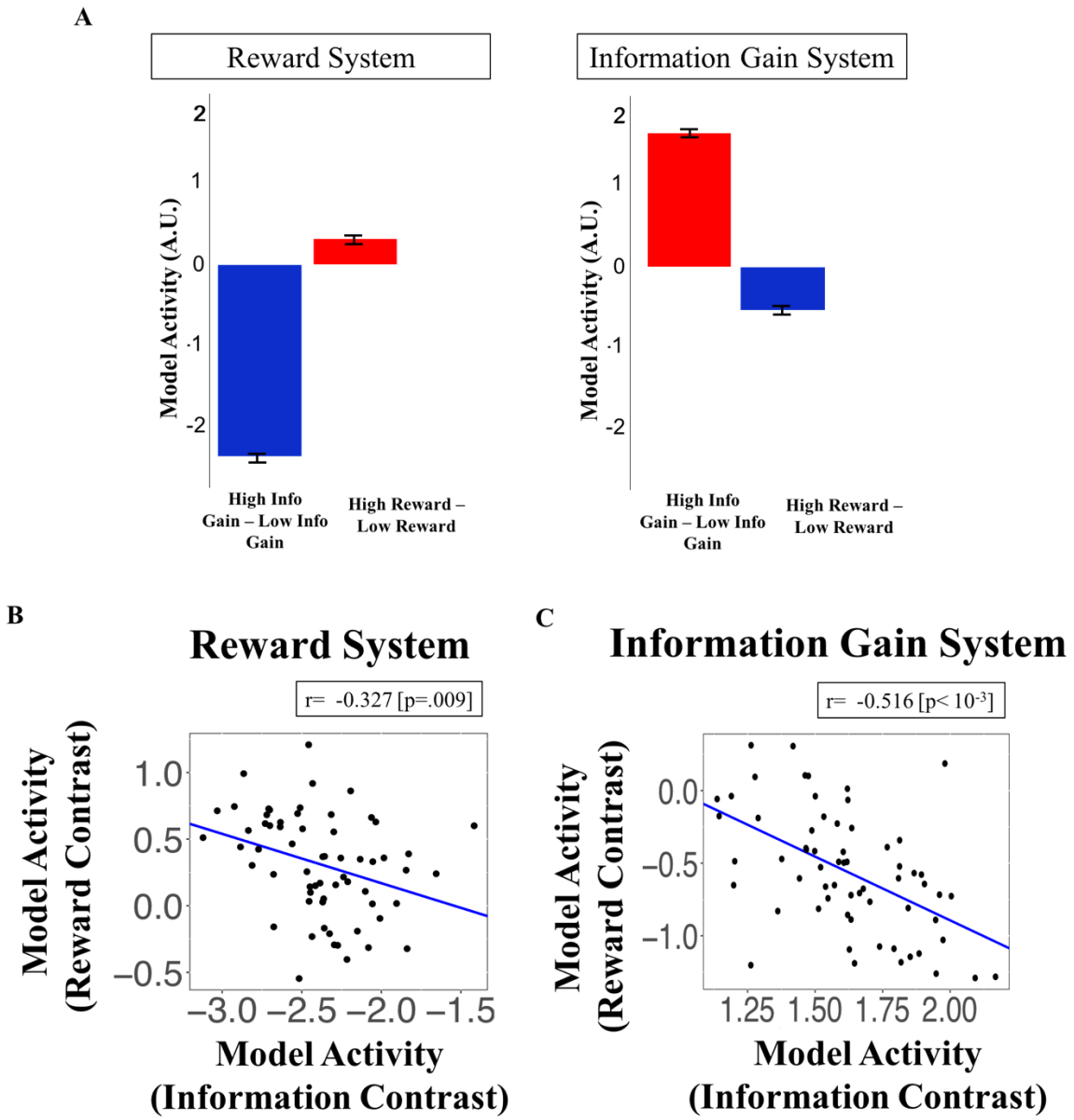
16 Hogan, P. S., Galaro, J. K. & Chib, V. S. Roles of Ventromedial Prefrontal Cortex and Anterior Cingulate in Subjective Valuation of Prospective Effort. *Cereb Cortex* **29**, 4277-4290, doi:10.1093/cercor/bhy310 (2019).

Figure S1. *Functional opposition between vmPFC and dACC in absence of symmetrical opposition*



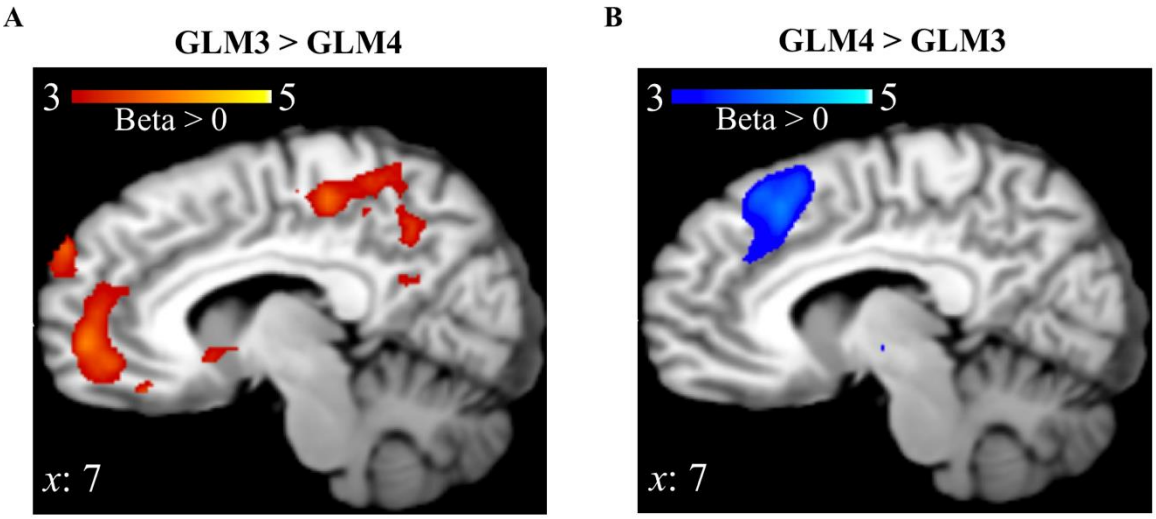
Probability of the model selecting the non-default option across effort levels (**A**) and its relative reward values (**B**). Correlation between relative reward value and effort levels (C). Correlation between information value and effort levels (D).

Figure S2. *Correlated activity in foraging task*



A) Simulating a dual value system on the sequential decision-making task adopted by [1 13]. Despite the independence of information and reward systems, the systems' activity are correlated: optimizing information is associated with decreased activity in the reward value system, and optimizing reward is associated with decreased activity in the information value system. Activity within the **(B)** reward system and **(C)** the information system is negatively correlated across independent model simulations.

Figure S3. *Domain specificity in vmPFC and dACC.*



A paired t-test between GLM3 and GLM4 shows **A**) specificity for reward (and not for information) in vmPFC, and **B**) for information (and not for reward in dACC).

Tables:

Table S1. *vmPFC and dACC opposition across different decision-making contexts.*

STUDIES	VMPFC	dACC
Blair *et al.* 2006	Greater reward	Lesser punishment
Kolling *et al.* 2012	Value of engaging	Value of foraging
Boorman *et al.* 2013	Immediate reward	Long-term reward
Chen et al. 2014	Positive correlation with creativity questionnaire	Negative correlation with creativity questionnaire
Kolling *et al.* 2014	Reduced risk	Increased risk
Skvortsova *et al.* 2014	Expected reward	Negative Expected reward
Becker *et. al.* 2016	Exploit reward	Switch to alternatives
Sinha *et al.* 2016	Early decrease, later increase during stress manipulation	Early increase, later decrease during stress manipulation
Shenhav *et al.* 2016	Choice ease	Choice difficulty
Delaveau *et al.* 2017	Default Mode Network	Task Positive Network
Arulpragasam *et al*. 2018	Subjective value	Negative Subjective value
Suarez-Jimenez *et al.* 2018	Learning danger	Approaching danger

The table shows a selection of studies that report the symmetric opposition between vmPFC and dACC in value-based decision-making.

Table S2. *Model estimated parameters from participants' behavior*

	gkRL			
Participants	α	β	ω	γ
Number1	0.004	7.916	0.457	0.150
Number2	0.151	0.443	2.723	1.103
Number3	0.010	2.281	0.961	0.195
Number4	0.495	0.358	33.333	0.200
Number5	0.518	0.201	12.980	0.232
Number6	0.459	0.168	48.663	0.000
Number7	0.528	0.147	9.181	0.473
Number8	0.593	0.158	24.743	0.158
Number9	0.658	0.214	33.661	0.158
Number10	0.054	1.819	1.714	1.149
Number11	0.761	0.155	31.631	0.000
Number12	0.687	0.142	1.114	0.997
Number13	0.660	0.125	29.913	0.072
Number14	0.393	0.321	17.714	0.000
Number15	0.23	0.338	4.808	0.875
Number16	0.269	0.791	7.528	0.653
Number17	0.046	0.244	3.032	0.088
Number18	0.269	0.208	9.174	0.837
Number19	0.011	1.562	0.508	0.526
Number20	0.012	2.084	0.26	0.923
Total	0.374 (0.264)	0.879 (0.176)	14.53 (1.76)	0.43 (0.414)

The table shows parameter estimates after fitting the model to participants' data. Mean and standard deviation estimates are also reported for each parameter.

Table S3. *GLMs for fMRI data.*

NAME	***REGRESSORS***
GLM0	[Highest Reward choice; Lower Reward choice]
GLM1	[First free choice; $RQ_{t+1,j}(c)$; 24 motion regressors]
GLM2	[First free choice; $I_{t,j}(c)$; 24 motion regressors]
GLM3	[First free choice; $RQ_{t+1,j}(c)$; $I_{t,j}(c)$; 24 motion regressors]
GLM4	[First free choice; $I_{t,j}(c)$; $RQ_{t+1,j}(c)$; 24 motion regressors]
GLM5	[First free info choice; $Context$; 24 motion regressors]
GLM6	[First free choice; $P(c/V_{t,j}(c_i))$; 24 motion regressors]

The table shows the 7 GLMs adopted in the fmri data analysis all referring to activity associated with the onset of the first-free choice trial. GLM0 is the univariate analysis, whereas GLMs 1-6 relates with the model-based analysis.