Behavioral/Cognitive

Dynamic Representation of the Subjective Value of Information

Kenji Kobayashi,[1] Sangil Lee,[1] Alexandre L. S. Filipowicz,[1,2,3] Kara D. McGaughey,[2,3] Joseph W. Kable,[1] and Matthew R. Nassar[4,5]

[1]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, [2]Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania 19104, [3]Computational Neuroscience Initiative, University of Pennsylvania, Philadelphia, Pennsylvania 19104, [4]Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, Rhode Island 02912, and [5]Department of Neuroscience, Brown University, Providence, Rhode Island 02912

To improve future decisions, people should seek information based on the value of information (VOI), which depends on the current evidence and the reward structure of the upcoming decision. When additional evidence is supplied, people should update the VOI to adjust subsequent information seeking, but the neurocognitive mechanisms of this updating process remain unknown. We used a modified beads task to examine how the VOI is represented and updated in the human brain of both sexes. We theoretically derived, and empirically verified, a normative prediction that the VOI depends on decision evidence and is biased by reward asymmetry. Using fMRI, we found that the subjective VOI is represented in the right dorsolateral prefrontal cortex (DLPFC). Critically, this VOI representation was updated when additional evidence was supplied, showing that the DLPFC dynamically tracks the up-to-date VOI over time. These results provide new insights into how humans adaptively seek information in the service of decision-making.

Key words: **decision-making; fMRI; information seeking; valuation**

Significance Statement

For adaptive decision-making, people should seek information based on what they currently know and the extent to which additional information could improve the decision outcome, formalized as the VOI. Doing so requires dynamic updating of VOI according to outcome values and newly arriving evidence. We formalize these principles using a normative model and show that information seeking in people adheres to them. Using fMRI, we show that the underlying subjective VOI is represented in the dorsolateral prefrontal cortex and, critically, that it is updated in real time according to newly arriving evidence. Our results reveal the computational and neural dynamics through which evidence and values are combined to inform constantly evolving information-seeking decisions.

Introduction

Information seeking is critical for adaptive decision-making. To improve future decisions, we collect information that would help us predict decision outcomes. For instance, we check the weather forecast to decide whether to go out for a hike, and we read about the policies and characters of candidates to decide how to vote. Recent work raises the possibility that deficits in information seeking underlie some psychiatric diseases such as schizophrenia and obsessive-compulsive disorder (Ross et al., 2015; Dudley et al., 2016; Hauser et al., 2017; Baker et al., 2019).

In economic theories, information seeking should be primarily driven by information instrumentality, or how much it would help the agent acquire rewards in an upcoming decision. Instrumentality is formally characterized as the value of information (VOI), defined as the improvement in the expected value that the agent can achieve by making the decision based on the information (Edwards, 1965; Howard, 1966). Although this normative VOI theory does not incorporate psychological motives of curiosity (Kreps and Porteus, 1978; Caplin and Leahy, 2001; Kidd and Hayden, 2015; Kobayashi et al., 2019), it successfully predicts how humans acquire costly information that can be used to maximize rewards (Edwards and Slovic, 1965; Wendt, 1969; Wilson et al., 2014; Kobayashi and Hsu, 2019). The theory is further supported by evidence that the VOI is encoded in reward-related regions, such as

Received Mar. 1, 2021; revised Aug. 4, 2021; accepted Aug. 5, 2021.

Author contributions: J.W.K. and M.R.N. designed research; S.L. and M.R.N. performed research; K.K., A.L.S.F., and K.D.M. analyzed data; K.K., S.L., A.L.S.F., K.D.M., J.W.K., and M.R.N. wrote the paper.

This work was supported by National Institute of Mental Health Grant R01 MH098899 (J.W.K.) and National Institute of Aging Grant R00 AG054732 (M.R.N.). Computing resources used for MRI analysis were supported by National Institutes of Health Grant 1S10 OD023495-01.

A.L.S. Filipowicz's present address is Machine Assisted Cognition, Toyota Research Institute, Los Altos, CA.

The authors declare no competing financial interests.

Correspondence should be addressed to Kenji Kobayashi at kenjik@sas.upenn.edu.

https://doi.org/10.1523/JNEUROSCI.0423-21.2021

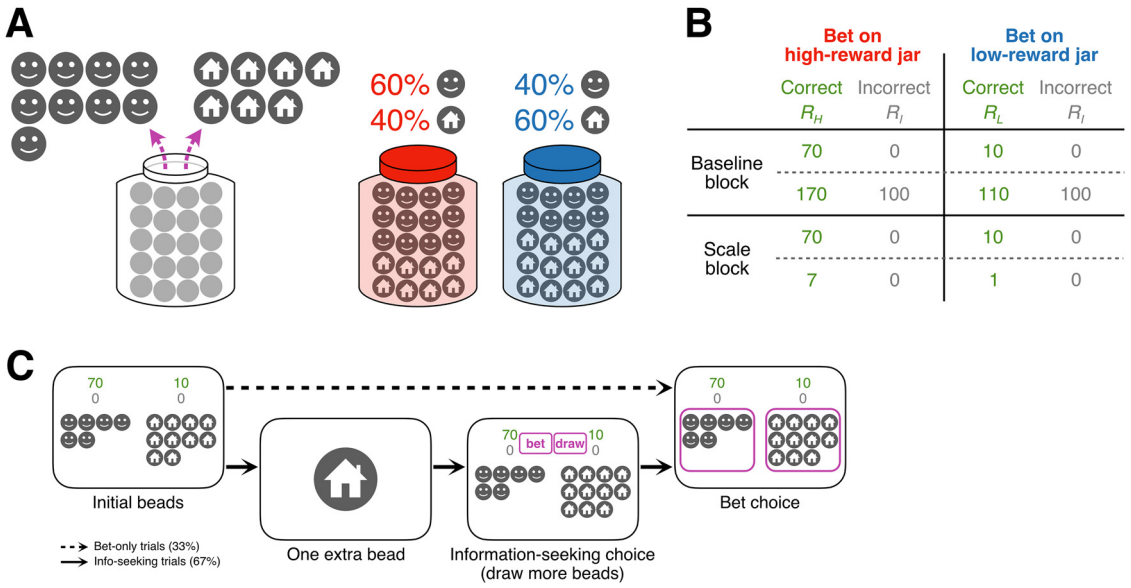

Figure 1. Experimental paradigm. We adopted the beads task with three key modifications: asymmetry in the reward structure, initial evidence before information seeking, and an updating event (one extra bead). ***A***, Participants observed a number of beads drawn from a jar and made a bet on its composition. Each bead was marked with a face or a house. There were two possible jar compositions: 60% face beads and 40% house beads, or 40% face beads and 60% house beads. The jars are colored here only for illustrative purposes. ***B***, Reward structure. Participants earned more reward points by correctly betting on one of the two jar types. The experiment consisted of two blocks, and in each block, one of the two reward structures was presented in each trial. The first block involved a baseline shift, and the second block involved a scale manipulation. ***C***, Trial sequence. In a third of the trials (bet-only trials), participants were presented with a number of beads from the jar, and they immediately made a bet on its type. In the remaining trials (information-seeking trials), they were presented with the initial beads and an extra bead, and were then allowed to seek further information by drawing more beads from the jar before making a bet on one of the two jars. Participants could draw as many beads as they needed within 5 s, but each additional draw incurred a cost (0.1 point). The extra bead was presented to evoke updating in the value of information.

the nucleus accumbens, ventromedial prefrontal cortex, anterior cingulate cortex, and dorsolateral prefrontal cortex (DLPFC) (Bromberg-Martin and Hikosaka, 2009, 2011; Kang et al., 2009; Krebs et al., 2009; Gruber et al., 2014; Charpentier et al., 2018; Kobayashi and Hsu, 2019; White et al., 2019; Kaanders et al., 2020; Lau et al., 2020).

The notion that the VOI is based on its instrumentality has two important implications. First, the VOI should not be determined by how much the information would contribute to accurate predictions of the state of the world but rather how much it would help the agent maximize rewards. Therefore, the VOI depends on the reward structure of the upcoming decision (e.g., the value of a weather forecast depends on how much the hiker prefers different weather conditions; those who don't mind hiking in the rain may not value the forecast as much as those who do). Second, the VOI depends on decision evidence that the agent already possesses before information seeking. The VOI tends to be smaller when the agent already has more evidence because the agent may already know what to do, and additional information is less likely to influence the agent (e.g., hikers may not need to check the weather forecast if other hikers have already informed them that it is going to snow). Thus, the agent needs to combine the available decision evidence with the reward structure to assess the VOI and seek information adaptively.

Crucially, when the decision evidence available to the agent changes, the agent should update the VOI. Situations requiring such updates are ubiquitous, either because the environment gradually supplies evidence over time (e.g., a recent weather forecast is more accurate than an old one) or because the agent sequentially samples multiple pieces of information (the hiker can check multiple sources of weather forecasts). Despite the importance, to the best of our knowledge, no study has examined how the human brain updates the VOI based on the most recent decision evidence.

We conducted an fMRI study to examine whether human information-seeking behavior is sensitive to reward structure and current decision evidence, and how human brains track the up-to-date VOI after acquiring additional evidence. First, we theoretically derive, and empirically demonstrate, a simple and generalizable prediction for how information seeking should be biased by asymmetry in a reward structure. Second, we show that the right DLPFC tracks the up-to-date VOI over time as a new piece of evidence is supplied. These results suggest that the right DLPFC plays a critical role in information seeking in dynamic decision-making contexts.

Materials and Methods

All procedures were approved by the Institutional Review Board at the University of Pennsylvania. Our experiment was not preregistered.

Participants. Fifteen people (11 female, 4 male, 18–28 years old, mean = 21.27, SD = 2.79) participated in the experiment. The *a priori* target sample size for this study was 16, determined according to effect sizes in previous studies that looked for updating signals using fMRI (McGuire et al., 2014). However, when the scanner used to collect data from 15 participants went down for an extensive period of repairs (because of a quench event during a flood), we opted to proceed to our planned analyses to avoid the need to pool data across different scanning conditions. All participants provided informed consent in accordance with the Declaration of Helsinki.

Experimental Design. We adopted a variant of the beads task (Phillips and Edwards, 1966; Huq et al., 1988; Furl and Averbeck, 2011); the participant was presented with a jar containing two types of beads and asked to guess its composition (i.e., which type made up the majority of the beads) by drawing some beads from the jar (Fig. 1*A*). Our variant had three important features. First, the participant was rewarded for identifying the correct jar composition, but the reward structure was asymmetric; the participant could earn more rewards by correctly betting on one jar type than the other (Fig. 1*B*). Second, a variable number of beads was drawn from the jar and presented to the participant at the beginning of each trial, empirically manipulating the evidence available to participants before they seek information. Third, an extra bead was

presented on a subset of trials to update the initial evidence. These features allowed us to examine how the brain represents and updates the VOI based on evidence that changes over time.

The experiment consisted of two interleaved trial types, bet-only trials and information-seeking trials (Fig. 1*C*). In the bet-only trials, the participant was first presented with a number of beads drawn from the jar. Each bead was marked with a rounded picture of a face or a house (one picture for the face or the house was used throughout the experiment). Beads marked with a face were presented to the left and those marked with a house to the right. The participant was told that these beads were drawn from one of two jars: a face-majority jar, which consisted of 60% face beads and 40% house beads, and a house-majority jar, which consisted of 60% house beads and 40% face beads. Rewards for correct and incorrect bets (in points) were also presented in green and gray, respectively. Rewards for a bet on the face-majority jar were shown above the face beads, and rewards for a bet on the house-majority jar was above the house beads. Rewards for a correct bet on one jar were numerically larger than rewards for a correct bet on the other jar (reward asymmetry), whereas an incorrect bet on either jar yielded the same rewards (Fig. 1*B*). After the presentation of the initial beads for 3 s, the participant was asked to make a bet. During the bet phase of the task, face and house beads were separately outlined by magenta boxes, and the participant could press the left or right button on a response box to bet on the face- or house-majority jar. Trials in which the participant did not make a bet within 3 s were terminated and discarded from the analysis.

In the information-seeking trials, the participant was first presented with the initial beads screen (same as the bet-only trials), followed by a blank screen (0–2 s). Next, an extra bead drawn from the jar was presented, either marked with a face or a house (1 s), which was added to the corresponding group of beads on the initial screen (0–2 s). The participant was then asked to decide whether to draw more beads from the jar before making a bet on its composition (information-seeking phase). Two choices appeared on the screen, draw and bet, and the participant pressed one button to draw one more bead and another button to terminate the information-seeking phase and proceed to the bet (the sides of the options were randomized across trials). The participant was allowed to draw as many beads as desired within 5 s, and a face or house bead was added to the screen every time the participant pressed the draw button. The participant was told that each draw incurred a constant small cost (0.1 point). Once they pressed the bet button (or when 5 s have passed), participants were presented with the bet screen (same as the bet-only trials).

The task was programmed in MATLAB (MathWorks) using MGL (http://justingardner.net/mgl/) and SnowDots (http://code.google.com/p/snow-dots/) extensions.

Procedure. In a separate task session before scanning, participants received extensive training on the task, in which various aspects of the task were gradually introduced (betting on the jar composition, asymmetric rewards, costly draws, and multiple reward structures). During the subsequent session, participants completed the task inside the scanner. Participants gave responses using an MRI-compatible button box. They were compensated based on the total points they acquired in the scanning session (500 points = $1.00).

The scanning experiment consisted of two blocks, which differed in reward structure (Fig. 1*B*). In the first block (the baseline block), one of the two reward structures, $(R_H, R_L, R_I) = (70, 10, 0)$ or $(170, 110, 100)$, was randomly presented in each trial, where R_H is the reward for a correct bet on the high-reward jar, R_L is the reward for a correct bet on the low-reward jar, and R_I is the reward for an incorrect bet; thus, participants earned a baseline reward of 100 points regardless of their bet in half of the trials. In the second block (the scale block), one of the two reward structures, $(R_H, R_L, R_I) = (70, 10, 0)$ or $(7, 1, 0)$, was randomly presented in each trial; thus, the participant earned a tenth of the rewards in half of the trials. Each block consisted of two scanning runs, one in which the high-reward jar was the face-majority jar and one in which the high-reward jar was the house-majority jar; the order was counterbalanced across participants.

On each trial, the participant was presented with 20 or 30 initial beads from the jar. The difference in the number of initial beads marked with a face or house was uniformly sampled from a discrete set of values ranging from −10 to 10 in increments of 2. Unbeknown to the participant, the true jar type was stochastically determined following the Bayesian posterior conditional on the initial bead difference (Eq. 4). In the information-seeking trials, the type of the extra bead presented and all additional beads drawn by the participant (face or house) were stochastically determined based on the hidden jar type (Eq. 1). The participant was not provided with feedback on bet accuracy or rewards on a trial-by-trial basis. Participants were, however, informed of the total number of points they had accumulated at the end of each run.

Theory. Normative predictions about the VOI, or how much an optimal agent should pay for information, were derived under assumptions that the agent conducts full-Bayesian inference on the jar type, deterministically makes an optimal choice to maximize the expected value (EV), is risk neutral, and optimally seeks information based on instrumentality, or how much it would improve the EV of the subsequent bet choice. Our theoretical framework is consistent with previous Bayesian models (Furl and Averbeck, 2011; Moutoussis et al., 2011), except that it explicitly examines the effects of the asymmetric reward structure and current decision evidence on the VOI. Our theoretical framework did not consider any additional information-seeking motives, such as curiosity, savoring, dread, or uncertainty reduction.

Let s_H be the state where the true jar is the high-reward jar and s_L the state where it is the low-reward jar. Let a_H be the action to bet on s_H and a_L the action to bet on s_L. Let us further refer to the majority beads in the high-reward jar as high-reward beads and the majority beads in the low-reward jar as low-reward beads (e.g., if the high-reward jar is the house-majority jar, a house bead is a high-reward bead, and a face bead is a low-reward bead; note that the beads were not directly associated with rewards per se). The goal for the agent is to choose between a_H and a_L to maximize the EV given the current evidence (i.e., the number of high-reward beads n_H and low-reward beads n_L drawn from the jar so far) and the reward structure (R_H, R_L, R_I).

The likelihood of drawing a high-reward bead b_H or a low-reward bead b_L conditional on the jar type is known to the agent:

$$P(b_H|s_H) = P(b_L|s_L) = q$$

$$P(b_L|s_H) = P(b_H|s_L) = 1 - q \tag{1}$$

where $q = 0.6$. Thus, the likelihood of observing n_H and n_L conditional on the jar type is:

$$\begin{aligned} P(n_H, n_L|s_H) &= \binom{n_H+n_L}{n_H} P(b_H|s_H)^{n_H} P(b_L|s_H)^{n_L} \\ &= \binom{n_H+n_L}{n_H} q^{n_H}(1-q)^{n_L} \end{aligned}$$

$$\begin{aligned} P(n_H, n_L|s_L) &= \binom{n_H+n_L}{n_H} P(b_H|s_L)^{n_H} P(b_L|s_L)^{n_L} \\ &= \binom{n_H+n_L}{n_H} (1-q)^{n_H} q^{n_L}. \end{aligned} \tag{2}$$

Using these likelihood functions and Bayes theorem, the posterior probability of the jar type after observing n_H and n_L follows:

$$\begin{aligned} \frac{P(s_H|n_H, n_L)}{P(s_L|n_H, n_L)} &= \frac{P(n_H, n_L|s_H)P(s_H)}{P(n_H, n_L|s_L)P(s_L)} = \frac{\binom{n_H+n_L}{n_H} q^{n_H}(1-q)^{n_L}}{\binom{n_H+n_L}{n_H} (1-q)^{n_H} q^{n_L}} \\ &= \left(\frac{q}{1-q}\right)^{n_H - n_L} \end{aligned} \tag{3}$$

assuming that the agent has a flat prior on the jar type ($P(s_H) = P(s_L) = 0.5$). Because $P(s_H|n_H, n_L) + P(s_L|n_H, n_L) = 1$, we obtain

$$P(s_H|n_H, n_L) = \frac{\left(\frac{q}{1-q}\right)^{n_H - n_L}}{\left(\frac{q}{1-q}\right)^{n_H - n_L} + 1}, \quad (4)$$

which is a function of the bead difference, $n_H - n_L$ [e.g., the posterior is the same when $(n_H, n_L) = (5, 2)$ or $(15, 12)$].

Given the posterior, the agent makes a choice among the following three options: to bet on s_H, to bet on s_L, or to seek information and draw an additional bead from the jar, which incurs a cost c_{draw} (0.1 point). The agent should decide whether to draw an additional bead based on the VOI, or the improvement in the EV of the bet because of the next bead:

$$VOI(n_H, n_L) = EV_{\text{draw}}(n_H, n_L) - EV_{\text{bet}}(n_H, n_L), \quad (5)$$

where EV_{draw} is the highest EV the agent could achieve after drawing the next bead (without considering the information-seeking cost), and EV_{bet} is the highest EV the agent could achieve by making a bet without any further information. The agent should draw a bead if and only if the VOI is higher than the drawing cost c_{draw}. EV_{bet} is the higher of the two bet EVs based on the current evidence, namely,

$$EV_{\text{bet}}(n_H, n_L) = \max_a EV(a|n_H, n_L), \quad (6)$$

where

$$a \in \{a_H, a_L\}$$

and

$$EV(a_H|n_H, n_L) = R_H \cdot P(s_H|n_H, n_L) + R_I \cdot P(s_L|n_H, n_L)$$

$$EV(a_L|n_H, n_L) = R_L \cdot P(s_L|n_H, n_L) + R_I \cdot P(s_H|n_H, n_L).$$

Because the posterior is determined by the bead difference (Eq. 4), EV_{bet} is also determined by the bead difference.

To evaluate EV_{draw}, we have to take into account two important facets of our information-seeking paradigm. First, the content of information (the type of the next bead, b_H or b_L) is stochastic, and second, the agent can decide whether to draw yet another bead or not after observing the next bead. Therefore, we have to evaluate the likelihood of the next bead type and combine it with the EV of an optimal choice conditional on each bead type. The likelihood of the next bead type based on the current evidence is evaluated according to the posterior on the jar type:

$$P(b_H|n_H, n_L) = P(b_H|s_H) \cdot P(s_H|n_H, n_L) + P(b_H|s_L) \cdot P(s_L|n_H, n_L)$$

$$P(b_L|n_H, n_L) = P(b_L|s_H) \cdot P(s_H|n_H, n_L) + P(b_L|s_L) \cdot P(s_L|n_H, n_L). \quad (7)$$

If the next bead is b_H, it would update the evidence from (n_H, n_L) to (n_H+1, n_L). Then the agent can either make an optimal bet and achieve $EV_{\text{bet}}(n_H+1, n_L)$ or pay the cost to draw another bead and achieve $EV_{\text{draw}}(n_H+1, n_L) - c_{\text{draw}}$. Similarly, if the next bead is b_L, it would update the evidence to (n_H, n_L+1), based on which the agent can either make an optimal bet and achieve $EV_{\text{bet}}(n_H, n_L+1)$ or draw another bead and achieve $EV_{\text{draw}}(n_H, n_L+1) - c_{\text{draw}}$. Therefore, the highest EV the agent can achieve after drawing an additional bead is

$$EV_{\text{draw}}(n_H, n_L) = P(b_H|n_H, n_L) \cdot \max[EV_{\text{bet}}(n_H + 1, n_L), EV_{\text{draw}}(n_H + 1, n_L) - c_{\text{draw}}] + P(b_L|n_H, n_L) \cdot \max[EV_{\text{bet}}(n_H, n_L + 1), EV_{\text{draw}}(n_H, n_L + 1) - c_{\text{draw}}]. \quad (8)$$

In Equation 8, $EV_{\text{draw}}(n_H, n_L)$ in the left-hand side depends on $EV_{\text{draw}}(n_H+1, n_L)$ and $EV_{\text{draw}}(n_H, n_L+1)$ in the right-hand side because of the aforementioned sequentiality of information seeking. We thus solved Equation 8 by backward recursion. Specifically, we arbitrarily assumed that the agent cannot draw more than 200 beads, set $EV_{\text{draw}}(n_H, n_L) = 0$ where $n_H+n_L = 200$, and we used Equation 8 to obtain $EV_{\text{draw}}(n_H, n_L)$ where $n_H+n_L = 199$. We then used Equation 8 recursively to obtain $EV_{\text{draw}}(n_H, n_L)$ for all cases where $0 < n_H+n_L < 200$. Although the obtained $EV_{\text{draw}}(n_H, n_L)$ depends on n_H+n_L, it reaches an asymptote over the course of recursion quickly. We substituted the asymptotic EV_{draw} in Equation 5 and obtained the theoretical VOI as a function of the bead difference.

The VOI was obtained for each of the three reward structures, $(R_H, R_L, R_I) = (70, 10, 0)$, $(170, 110, 100)$, and $(7, 1, 0)$. The baseline shift affects both EV_{draw} and EV_{bet} by the same amount, which is canceled out in Equation 5 and does not affect the VOI. On the other hand, as c_{draw} was not scaled along with rewards and remained the same across conditions (0.1 point), the scale manipulation affects not only the magnitude but also the shape of EV_{draw} (Eq. 8) and thus the VOI (Eq. 5).

The most important prediction of this theoretical framework is that information seeking should be biased because of the reward asymmetry. The VOI takes an inverted U shape as a function of the bead difference, and its peak is at a moderate negative bead difference ($n_H - n_L = -5$). This is because the information would directly improve the subsequent bet choice; when $n_H - n_L = -5$, $EV(a_H|n_H, n_L)$ is close to $EV(a_L|n_H, n_L)$, but the next bead would increase the difference in either direction [if a high-reward bead b_H is observed, $EV(a_H|n_H+1, n_L) > EV(a_L|n_H+1, n_L)$; if a low-reward bead b_L is observed, $EV(a_H|n_H, n_L+1) < EV(a_L|n_H, n_L+1)$]. Therefore, the agent can bet on s_H after b_H and bet on s_L after b_L, and such flexibility improves the overall EV. In contrast, the VOI is effectively zero when the bead difference is positive ($n_H - n_L > 0$) because the agent would bet on s_H regardless of the next draw. The VOI is also effectively zero when low-reward beads outnumber high-reward beads by a large enough margin ($n_H - n_L < -7$), because the agent would bet on s_L regardless of the next draw.

This qualitative prediction, a bias in information seeking toward a negative bead difference, does not depend on most of our assumptions (e.g., choice optimality, risk neutrality). Information seeking would be biased as far as the agent is sensitive to the rank order of rewards and the bead difference. On the other hand, if an agent is not motivated to maximize rewards but to maximize the accuracy of the prediction (i.e., utility function U follows $U(R_H) = U(R_L) > U(R_I)$), the agent would exhibit unbiased information seeking; the uncertainty about the jar type is determined by $|n_H - n_L|$ and is highest when $n_H = n_L$, which is when the agent would draw beads most frequently. Therefore, a bias in information seeking would suggest that information seeking is motivated by instrumentality of the information for future reward seeking.

Statistical analyses—behavioral data. To examine information-seeking behavior, we analyzed the frequency at which participants drew at least one bead as a function of the bead difference. We specifically focused on whether they drew the first bead as a function of the current evidence and examined if the choice was biased by the reward asymmetry as theoretically predicted. The relationship between information-seeking behavior and the bead difference was analyzed using Gaussian process (GP) logistic regression (Rasmussen and Williams, 2006). GP logistic regression estimates a latent function that smoothly varies with the independent variable (the bead difference) and yields likelihoods of binary choices (whether participants drew a bead in each trial). The estimated latent function can be interpreted as the subjective VOI function (the higher the VOI is, the more likely participants draw a bead). The latent function with isotropic squared exponential covariance was estimated using the variational Bayes approximation, as implemented in Gaussian processes for the Machine Learning Toolbox, version 4.2 (https://github.com/alshedivat/gpml; Rasmussen and Nickisch, 2010).

To test whether information-seeking behavior systematically differed across blocks and reward conditions within each block, we compared four models. Model 1 implemented the theoretical prescription that information seeking is sensitive to the scale manipulation but not to the baseline manipulation. It thus consisted of three separate latent value functions, one used in all trials in the baseline block, one used in trials where $(R_H, R_L, R_I) = (70, 10, 0)$ in the scale block, and one used in trials where $(R_H, R_L, R_I) = (7, 1, 0)$ in the scale block. We constructed several alternative models. Model 2 postulated different value functions for

reward conditions not only in the scale block but also in the baseline block, one for trials where $(R_H, R_L, R_I) = (70, 10, 0)$, and another for trials where $(R_H, R_L, R_I) = (170, 110, 0)$ (i.e., four value functions in total). Model 3 postulated the lack of sensitivity to reward conditions in both blocks but a separate value function for each block (i.e., two value functions in total), and Model 4 postulated one common value function for all trials in both blocks. These models were compared based on log likelihood (LL) in leave-one-participant-out cross-validation (LOPO CV) and leave-one-trial-out cross-validation (LOTO CV). We also adopted the same analytic approach to the bet choices, comparing the performance of Models 1–4.

We found that Model 3 outperformed other models for both information-seeking and bet choices (see below, Results). To test whether information-seeking behavior was biased by the reward asymmetry, we next compared Model 3 with another model (Model 5), which assumed value functions that are symmetric with respect to the bead difference (i.e., value functions that only vary with the absolute value of bead difference). We found that Model 3 fit information-seeking behavior better than Model 5, supporting a bias in information seeking (see below, Results).

The fact that Model 3 performed better than Models 1, 2, and 4 suggests that, although participants did not change their behavioral strategies based on the trial-by-trial reward manipulation, they adapted to the different reward statistics across blocks. However, such changes across blocks could potentially reflect time-induced behavioral changes as well, such as boredom or fatigue, as all participants completed the baseline block first and the scale block second. To examine the possibility that the population-level behavioral pattern was not stationary over time, we tested another model (Model 6), which assumed distinct value functions between the first and second scanning runs within each block (one value function for each run, four functions in total). Model 6 performed worse than Model 3 (information-seeking choices: LOPO CV LL = −1222.05 vs −1214.73, LOTO CV LL = −1145.39 vs −1142.25, bet choices: LOPO CV = −288.55 vs −283.56, LOTO CV LL = −266.00 vs −265.26), suggesting that changes in participants' behavior were systematically driven by reward statistics, such as the average reward rate over trials rather than time.

MRI data acquisition. MRI data were collected using a Siemens (Erlangen) Trio 3T scanner with a 32-channel head coil at the University of Pennsylvania. A 3D high-resolution anatomic image was acquired using a T1-weighted MPRAGE sequence [voxel size = 0.9375 × 0.9375 × 1 mm, matrix size = 192 × 256, 160 axial slices, inversion time (TI) = 1100 ms, repetition time (TR) = 1810 ms, echo time (TE) = 3.51 ms, flip angle = 9 degrees]. Functional images were acquired using a T2*-weighted multiband gradient echoplanar imaging (EPI) sequence (voxel size = 2 × 2 × 2 mm, matrix size = 98 × 98, 72 axial slices with no interslice gap, 400 volumes, TR = 1500 ms, TE = 30 ms, flip angle = 45 degrees, multiband factor = 4), followed by field map images (TR = 1270 ms, TE = 5 ms and 7.46 ms, flip angle = 60 degrees).

Statistical analyses—MRI data. MRI data were analyzed using Functional MRI of the Brain (FMRIB) Software Library (FSL) version 6.0; (Smith et al., 2004; Jenkinson et al., 2012). MPRAGE anatomic images were skull stripped using FSL BET. EPI functional images were slice time corrected, motion corrected, high-pass filtered (cutoff = 90 s), geometrically undistorted using field map images, registered to the MPRAGE anatomic image, normalized to the Montreal Neurological Institute (MNI) space, and spatially smoothed (Gaussian kernel FWHM = 6 mm).

To look for regions that represent the subjective VOI on the initial beads presentation, we ran a GLM analysis (GLM 1). The regressor of interest modeled the initial beads presentation (3 s boxcar) and was parametrically modulated by the trial-by-trial subjective VOI, which was the latent function estimated in the winning model (Model 3) of GP logistic regression on the information-seeking behavior. GLM 1 also included nuisance regressors that modeled the initial beads presentation (unmodulated), the extra bead presentation, and button presses. The regressors were convolved with the canonical double-gamma hemodynamic response function (HRF). GLM 1 additionally incorporated six head motion parameters (three translations and three rotations, estimated by MCFLIRT) as confound regressors. GLM 1 was run following the standard approach of FSL FEAT, the GLM was first fit to BOLD signals in each run (first level), and the estimated coefficients of interest were combined across runs (second level). Individual-level T statistics were entered into the population-level inference using FSL randomize, in which clusters that showed positive response to the subjective VOI were defined at the voxelwise cluster-forming threshold of $p < 0.001$ and evaluated by sign-flipping permutation on cluster mass controlling for whole-brain familywise error (FWE).

To illustrate how the cluster's activation varied as a function of the bead difference, we ran another GLM (GLM 2) using FSL FEAT, which included a regressor for each level of bead difference separately, along with the same nuisance regressors as GLM 1. Then T statistics for each regressor of interest were averaged across runs within each block and then averaged across all voxels in the right DLPFC cluster defined as above.

To examine how the DLPFC responds to the updating of the VOI, we ran another GLM (GLM 3) using FSL FEAT to estimate the time course of signals related to the initial VOI and the VOI updating, which were derived from Model 3 of GP logistic regression. The VOI updating was calculated as the signed difference between the posterior VOI, which depends both on the initial beads and the extra bead, and the prior VOI, which depends only on the initial beads. GLM 3 included three sets of the finite impulse response (FIR) function, one unmodulated (intercept), one parametrically modulated by the initial VOI, and one parametrically modulated by the VOI updating. These FIRs were aligned to the onset of the extra bead and sampled every 1.5 s (equal to TR) for the total duration of 21 s. GLM 3 also included nuisance regressors that modeled the initial beads presentation and button presses, convolved with the canonical HRF, along with head motion parameters. T statistics of parametrically modulated FIR sets were averaged across all voxels in the right DLPFC cluster for each participant. Population-level inference on the updating signal was conducted at the cluster level across time; clusters were defined at the eventwise cluster-forming threshold of $p < 0.05$ and evaluated by sign-flipping permutation on cluster mass, correcting for FWE across time.

Finally, we tested whether different regions represent the VOI as the actual information-seeking choice approaches. To do so, we ran additional whole-brain analyses using four GLMs (4–7). GLMs 4 and 5 modeled parametric effects of either the VOI based on the initially presented beads alone (GLM 4) or the updated VOI, which also took into account the extra bead (GLM 5) on the extra bead presentation (1 s boxcar). Similarly, GLMs 6 and 7 modeled parametric effects of the initial (GLM 6) or updated (GLM 7) VOI on the onset of the information-seeking screen (stick function). These GLMs included the same nuisance regressors and were subject to the same statistical inference procedure as GLM 1. For clusters identified in GLM 4 (the initial VOI on the extra bead), we tested for the VOI updating using GLM 3. For clusters identified in GLM 6 (the initial VOI on the information-seeking screen), we tested for updating by running another GLM (GLM 8). GLM 8 was constructed similarly to GLM 3, except that the FIRs in GLM 8 were aligned to 6 s before the information-seeking screen onset and covered the duration of 27 s. In both analyses, statistical significance of the updating signals was evaluated at the cluster level, as described in the previous paragraph.

Data Availability. Behavioral data and custom code for behavioral analysis are available at https://gitlab.com/kenji.k/beadsVOI/, raw MRI data are available at https://openneuro.org/datasets/ds003758/, and unthresholded population-level statistics images are available at https://neurovault.org/collections/MORTKKAM/.

Results

Experimental paradigm

To examine neural representations of the VOI and its updating, we adopted a variant of the beads task, an experimental paradigm widely used to study probability judgment and information seeking (Phillips and Edwards, 1966; Huq et al., 1988; Furl and Averbeck, 2011). As in the conventional version of the beads task, participants were presented with a jar containing two

types of beads, one marked with a face and the other marked with a house, and were asked to make a bet on the bead composition of the jar by observing some of the beads drawn from the jar. There were two possible compositions of the jar: one that consists of 60% face beads and 40% house beads, and the other that consists of 40% face beads and 60% house beads (Fig. 1*A*).

Our variant of the beads task had three key features. First, we introduced reward asymmetry; participants could earn more rewards by correctly betting on one jar type (e.g., the face-majority jar) than the other (e.g., the house-majority jar; Fig. 1*B*). If participants were motivated to seek information to maximize rewards in the bet, their information-seeking strategy should be sensitive not only to the current evidence (the number of observed beads so far) but also to the reward asymmetry (the jar type they should bet on to maximize rewards). On the other hand, if participants were motivated to accurately guess the jar type, their information seeking should not be sensitive to the reward asymmetry. Therefore, the reward asymmetry allowed us to test whether information seeking was driven by the instrumentality of information for future reward seeking as normatively prescribed in economic theories.

Second, we provided initial evidence in the form of 20 or 30 bead draws from the jar. On a subset of trials, participants could then seek more information about the jar composition by drawing additional beads or elect to make a bet on the jar type (Fig. 1*C*). The difference in the numbers of face beads and house beads was parametrically manipulated to range from strong evidence favoring the low-reward jar to strong evidence favoring the high-reward jar. Additional draws incurred a small constant cost (0.1 point per draw) to monetarily incentivize participants to seek information only when necessary. This design allowed us to empirically measure the subjective VOI, or how much participants were willing to seek costly information, as a function of the current evidence.

Third, on the trials that allowed for information seeking, participants were presented with one extra bead draw from the jar before the information-seeking phase (Fig. 1*C*). The extra bead complemented the initial beads, shifting the evidence on the jar compositions, and thus updated the VOI originally evaluated based on the initial beads. We analyzed neural responses on this extra bead event to examine how the neural representation of the VOI is dynamically updated based on the up-to-date evidence over time.

Participants completed the task inside the scanner. In each trial, after the presentation of initial beads and an extra bead, participants were allowed to draw as many additional beads as they wanted within 5 s and then made a binary bet on the jar type. Additionally, to empirically elucidate participants' reward-seeking behavior in a way that is not contaminated by information seeking, participants were asked to make a bet on the jar type without information seeking in a subset of trials (bet-only trials). Finally, to explore how information seeking is sensitive to rewards, we introduced trialwise manipulation of the reward structures. Specifically, participants earned a baseline reward of 100 points, regardless of their bet, in half of the trials in one block (henceforth the baseline block), and they earned a tenth of the rewards in half of the trials in the other block (henceforth the scale block). Importantly, the reward of a correct bet was asymmetric across all trials and blocks (Fig. 1*B*).

Theory

We first derived a theoretical prediction on how agents should seek information to optimize their bet and maximize rewards. We obtained a theoretical VOI under the assumption that the agents aim to maximize the expected value (EV) of the upcoming bet, which they evaluate based on the posterior probability of the jar type inferred in a perfectly Bayesian manner.

The posterior of the jar type is determined by the numbers of high-reward beads (the majority bead in the high-reward jar, e.g., face) and low-reward beads (the majority bead in the low-reward jar, e.g., house) observed from the jar so far (Fig. 2*A*). The more high-reward beads that have been drawn, the more likely the jar is the high-reward jar, and vice versa. More specifically, the posterior is determined by the difference in the numbers of observed beads (high-reward beads minus low-reward beads; Fig. 2*B*; Eq. 4). When more high-reward beads have been observed than low-reward beads (bead difference > 0), the probability of the high-reward jar is higher than the probability of the low-reward jar, and it increases with the bead difference. Conversely, when more low-reward beads have been observed (bead difference < 0), evidence favors the low-reward jar.

To evaluate the EV of a bet, the agent needs to combine the posterior on the jar type with the reward structure (Fig. 2*C*). Because of the reward asymmetry, when the current evidence does not favor either jar (the bead difference = 0; Fig. 2*C*, diagonal), the EV to bet on the high-reward jar is higher than the EV to bet on the low-reward jar. The EVs to bet on the two jars are closest to each other when more low-reward beads have been observed (bead difference = −5; Fig. 2*C*, white region). This prediction holds across all our reward structures (Fig. 2*D*); a baseline shift in rewards does not affect the EV difference, and a scale manipulation in rewards multiplicatively affects both EVs without changing the relative magnitudes of the EVs. Therefore, if forced to bet on one of the two possible jars, the EV-maximizing agent would experience the highest choice uncertainty not when equal numbers of beads have been observed but when more low-reward beads have been observed than high-reward beads.

Under economic theories, the VOI, or the value of drawing an additional bead, is evaluated based on how much the next bead would improve the upcoming bet on average (Eq. 5). Qualitatively, the theoretical VOI tends to increase with the uncertainty about which jar type to bet on because an additional bead would provide more evidence for either jar type and resolve the uncertainty over possible actions (Fig. 2*E*). For instance, when the agent is under high uncertainty on the bet (bead difference = −5; Fig. 2*E*, black region), an additional bead would help the agent make a bet regardless of its type; if the next bead is a high-reward bead, it provides additional evidence in favor of the high-reward jar, whereas if it is a low-reward bead, it favors the low-reward jar. The agent can thus improve the EV by making a bet conditional on the next bead type. On the other hand, when the agent has observed more high-reward beads than low-reward beads (e.g., bead difference = +10), or when the agent has observed many more low-reward than high-reward beads (e.g., bead difference = −10), an additional bead would not affect the subsequent bet; the agent would bet on the high-reward jar or low-reward jar no matter what the next bead would be. Therefore, the theoretical VOI takes an inverted U shape as a function of the bead difference, with its peak at a negative bead difference (−5; Fig. 2*F*).

Therefore, our theoretical framework yields an important prediction that the information-seeking strategy should be biased by the reward asymmetry; participants should draw additional

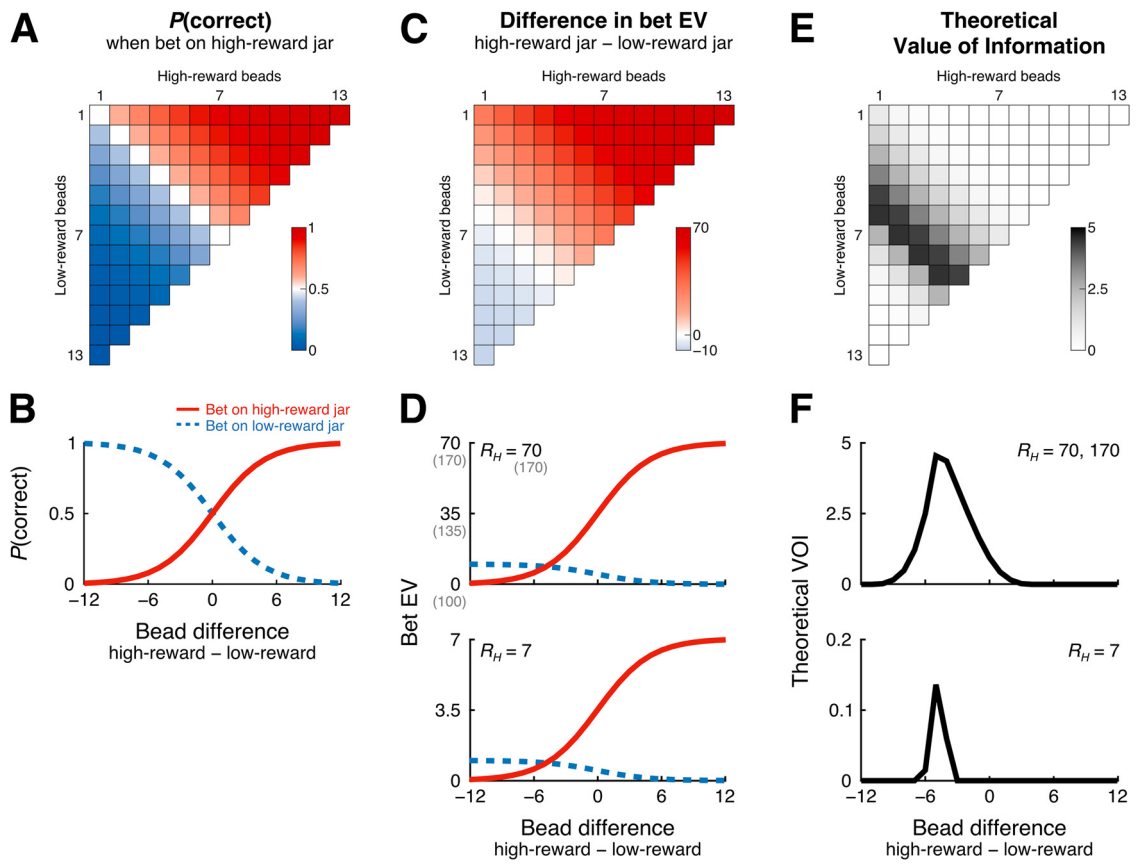

Figure 2. Theoretical predictions. ***A***, The probability of the jar type (specifically, the probability that the true jar is the high-reward jar) increases with the number of observed high-reward beads and decreases with the number of observed low-reward beads. ***B***, The probability of the jar type is determined by the bead difference. ***C***, Because of the reward asymmetry, when equal numbers of high-reward and low-reward beads have been observed (the diagonal), the EV to bet on the high-reward jar is higher than the low-reward jar. The agent would experience the smallest EV difference and, hence the highest uncertainty on the bet, when more low-reward beads have been drawn (the white region). ***D***, The EV difference is smallest when the bead difference is −5 across all reward structures. Top, Bet EVs are not affected by a baseline shift in rewards. Bottom, The relative magnitudes of EVs remain the same when rewards are scaled down overall. ***E***, The theoretical VOI is evaluated based on how much an additional bead would help the upcoming bet and increase the EV. It is highest when the uncertainty on the bet is highest (bead difference = −5, the black region) because the next bead would affect the bet regardless of the type of the bead; an additional high-reward bead would provide evidence in favor of betting on the high-reward jar, and an additional low-reward bead would provide evidence in favor of betting on the low-reward jar. ***F***, The theoretical VOI is highest at a negative bead difference across all reward structures. Top, The VOI is unaffected by a baseline shift in rewards. Bottom, When the rewards are scaled down, the magnitude of the VOI becomes smaller as well, but the peak location remains the same.

beads more frequently when more low-reward beads have been observed than high-reward beads (bead difference < 0). The predicted bias holds across reward structures (Fig. 2*F*); manipulation of the reward baseline (in the baseline block) does not affect the VOI, and manipulation of the reward scaling (in the scale block) affects the overall magnitude of the VOI but does not drastically alter its inverted U shape. This prediction might be somewhat counterintuitive, as the motivation for information seeking is expected to be higher when the current evidence favors the less desirable state (the low-reward jar). However, it is consistent with the widespread notion of confirmation bias that an agent needs less evidence to bet on a desirable state than an undesirable state (Gesiarz et al., 2019). More generally, the prediction echoes the general assumption that information seeking should be driven not by the motivation to predict the state (Which jar is the true jar?) but to maximize rewards (Which jar to bet on?). If, in contrast to our theoretical assumption, an agent is solely motivated to accurately predict the state, the agent would seek information the most when the bead difference is zero. Therefore, a bias in information seeking would suggest that participants seek information based on the instrumentality of the information for future reward seeking as normatively prescribed. To our knowledge, this bias in information seeking under the reward asymmetry is a novel theoretical prediction, which has not yet been directly tested.

Behavior

We examined participants' information-seeking behavior, and in particular, whether it was biased by the reward asymmetry as predicted. If participants sought to improve their subsequent bet choice and maximize rewards, the frequency of information seeking (i.e., how often they drew at least one bead) should be biased toward a negative bead difference (i.e., when more low-reward beads have been drawn than high-reward beads).

Observed information-seeking behavior was biased in the predicted direction (Fig. 3*A*). In both baseline and scale blocks, the frequency of drawing an additional bead was highest when more low-reward beads had been drawn than high-reward beads. Sensitivity to the reward asymmetry was also confirmed by the bet on the jar type in the bet-only trials (Fig. 3*B*). The frequency of betting on the high-reward jar increased with the bead difference, and the indifference point (the point at which participants were equally likely to bet on either jar) was shifted toward a negative bead difference. These results show that participants incorporated both the current evidence and reward asymmetry in reward-seeking and information-seeking choices.

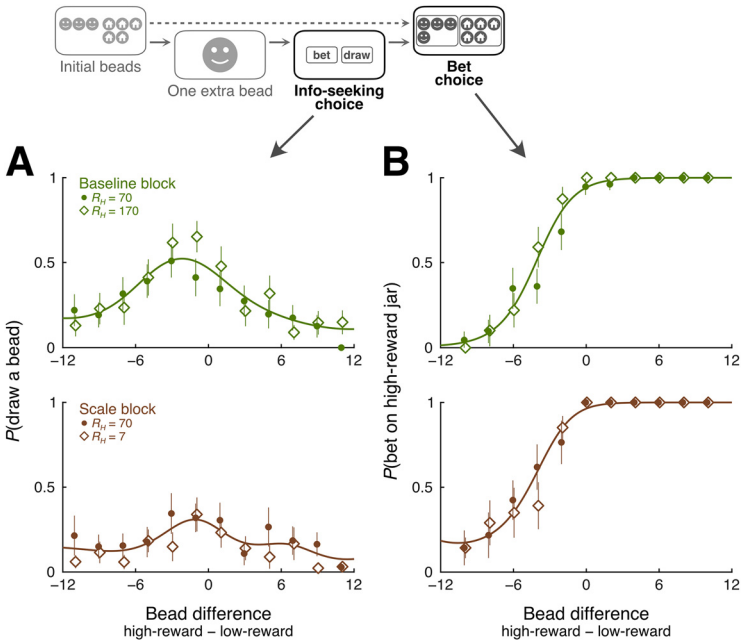

Figure 3. Behavior. Participants' information-seeking and reward-seeking behavior was biased by the reward asymmetry as predicted. ***A***, Participants' information seeking, or the frequency at which they drew at least one bead, peaked when more low-reward beads had been drawn than high-reward beads. ***B***, In the bet-only trials, the frequency with which they bet on the high-reward jar increased with the difference in the beads and was biased by the reward asymmetry. Lines indicate the best-fit model, which assumed sensitivity to blocks but not to reward manipulations within blocks. Error bars indicate SEM based on bootstrap resampling participants.

A notable deviation from the theoretical prediction is that participants' information seeking was not sensitive to the reward scale manipulation. In our framework, the theoretical VOI is smaller when the rewards are scaled down (although the peak location remains the same), but it is unaffected by a reward baseline shift (Fig. 2*F*). Thus, if our participants were perfectly sensitive to the reward structure on a trial-by-trial basis, their information seeking should be affected by trialwise reward manipulation in the scale block but not in the baseline block. To test this, we examined how information-seeking behavior differed across reward conditions and blocks. To characterize the relationship between information seeking and the bead difference without assuming a functional form, we used GP logistic regression (Rasmussen and Williams, 2006). We fit four models to participants' behavior; Model 1 assumed sensitivity to the scale manipulation but not to the baseline manipulation as normatively prescribed, Model 2 assumed sensitivity to both manipulations, Model 3 assumed a difference between blocks but no sensitivity to the manipulation in either block, and Model 4 assumed no difference between blocks or reward conditions. We found that Model 3 outperformed other models, including Model 1, according to both leave-one-participant-out cross-validation (LOPO CV; LL = −1216.93, −1216.15, −1214.73, and −1232.15) and leave-one-trial-out cross-validation (LOTO CV; LL = −1143.61, −1143.76, −1142.25, and −1166.17). Therefore, although participants did not change their information-seeking strategy based on the reward structure on a trial-by-trial basis, their behavior was systematically different between blocks, likely because of differences in reward statistics between blocks (e.g., the average reward rate over trials).

We speculate that shifting information-seeking strategies on a trial-by-trial basis was too cognitively taxing for our participants because we also manipulated the bead difference and the trial type (information-seeking or bet-only). Despite this limitation, we observed that participants' information seeking exhibited a clear bias in both blocks. Indeed, we observed that Model 3, which allowed asymmetry in information seeking, performed better than another model (Model 5), which assumed symmetric information seeking (baseline block LOPO CV LL = −666.82 vs −679.52; LOTO CV LL = −630.87 vs −645.10; scale block LOPO CV LL = −547.92 vs −548.13; LOTO CV LL = −511.68 vs −512.53). Furthermore, analysis on betting choices also preferred Model 3 to Models 1 and 2 (comparison between Models 3 and 4 is equivocal; LOPO CV LL = −287.01, −285.67, −283.56, and −281.19; LOTO CV LL = −267.77, −266.25, −265.26, and −268.46), showing that participants were insensitive to trialwise reward manipulation not only in information seeking but also in reward seeking. These results are qualitatively consistent with our theoretical prediction and lend support to the general notion that people seek information to improve their subsequent choices and maximize rewards.

Neural representation of VOI in the DLPFC on initial decision evidence

Next, we examined how the VOI was represented in the brain. Although previous fMRI studies reported VOI representations in a set of regions including the DLPFC, ventromedial prefrontal cortex, and striatum, most of these studies focused on situations where participants obtained information that would not be useful for future decisions (i.e., information seeking for its own sake), and one study that examined instrumentality-driven information seeking used a one-shot paradigm that did not involve any updating (Kobayashi and Hsu, 2019). Thus, it remains unknown to what extent the neural representation of VOI is generalizable across tasks and decision contexts, and whether previously reported regions also represent and update the VOI in our experimental paradigm.

To look for brain regions that represent the VOI, we empirically estimated subjective VOI from the information-seeking behavior. We used the winning model of our GP logistic regression analysis (Model 3) to obtain the latent value function, which varied smoothly with the bead difference and differed between blocks (Fig. 4*A*). We then looked for regions where neural responses at the presentation of initial beads covaried with the subjective VOI.

We found a cluster in the right DLPFC representing subjective VOI (Fig. 4*B*; cluster-forming threshold $p < 0.001$, cluster mass $p < 0.05$, whole-brain FWE corrected; Table 1). Activation in this cluster peaked when more low-reward beads had been drawn in both blocks, consistent with the prediction (Fig. 4*C*). Interestingly, the DLPFC cluster overlaps with a VOI cluster reported in a previous study, which examined one-shot instrumentality-driven information seeking (Kobayashi and Hsu,

2019), providing converging evidence that the right DLPFC represents the VOI across decision contexts, at least when information is primarily acquired based on its instrumentality for future value-guided decisions.

Updating of VOI representation in the DLPFC on additional decision evidence

We then examined how the VOI representation was updated on the arrival of additional evidence. When the evidence available to agents changes, they need to track the up-to-date VOI to seek information adaptively over time. Specifically, we examined how the right DLPFC responds to the extra bead presented after the initial beads but before the information-seeking choice (Fig. 5*A*). We derived the VOI updating, or the difference between the posterior and prior VOI, as a function of the initial bead difference (the prior evidence) and the type of the extra bead (the evidence that causes updating). For instance, if participants have observed many more low-reward beads than high-reward beads (bead difference < -5), an extra high-reward bead would positively update the VOI as it slightly increases the uncertainty on the bet, whereas an extra low-reward bead would negatively update the VOI as it further decreases the uncertainty on the bet. The directionality of updating is the opposite when more high-reward beads have been observed (bead difference > 0).

We hypothesized that the right DLPFC tracks the up-to-date VOI over time, and it responds not only to the VOI based on the initial beads but is dynamically updated to the appropriate updated VOI after observation of the extra bead. To test this, we estimated the effects of the initial VOI and VOI updating on BOLD signals from the region of interest (ROI) defined above (Fig. 4*B*). To avoid a strong assumption about the time course of the updating process, we estimated the effects of initial VOI and VOI updating across time using FIR functions aligned to the presentation of the extra bead (Fig. 5, top). We included three FIRs in a GLM, one parametrically modulated with the initial VOI, one modulated with the VOI updating, and one without parametric modulation (intercept). Because the ROI was originally defined based on its response to the initial VOI (albeit in an earlier time window), the estimated effect of the initial VOI is biased, but the estimated effect of the VOI updating depends critically on the exact bead that was drawn and thus is independent of our ROI selection process (Fig. 5*A*).

The estimated time courses are shown in Figure 5*B*. As expected, the right DLPFC represents the initial VOI early on. Importantly, the right DLFPC also positively responded to the VOI updating (cluster-forming threshold $p < 0.05$, cluster mass $p < 0.05$, FWE corrected across time). The rise of the VOI updating signal lags behind the initial VOI signal in time but the signals go back to baseline in parallel. The estimated time courses look somewhat sluggish, which presumably reflects the temporal nature of our experimental paradigm. After the extra bead, the participants were presented with another screen that incorporated both the initial beads and the extra bead, and they could start making the information-seeking choices only after a jittered delay. Therefore, the VOI representation did not have to be immediately updated to drive information seeking adaptively; indeed, a separate whole-brain analysis that used the canonical double-gamma HRF did not reveal any significant clusters for VOI updating.

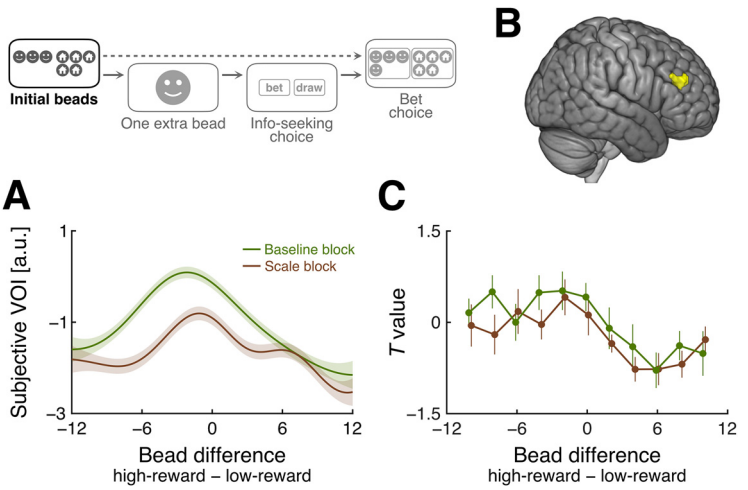

Figure 4. Neural representation of the VOI. ***A***, The subjective VOI was estimated for each block based on information-seeking behavior (Fig. 3*A*). ***B***, The right DLPFC represented the subjective VOI (cluster mass $p < 0.05$, whole-brain FWE corrected). ***C***, As predicted, the right DLPFC activation peaks at a negative bead difference in both blocks. Error bars indicate SEM.

This evidence demonstrates that neural representations in the right DLPFC shift from the initial (*a priori*) VOI to the updated (*a posteriori*) VOI, suggesting that this brain region dynamically tracks the VOI based on the up-to-date evidence in service of adaptive information seeking over time.

Widespread neural representations of the VOI closer to information seeking

Finally, we examined how the VOI was represented in the brain later in the trials. Although we identified the VOI cluster in the right DLPFC on the initial beads presentation, it is possible that other regions would represent the VOI as the participants started preparing for actual information-seeking choices. We conducted additional whole-brain analyses related to two events, (1) the presentation of the extra bead and (2) onset of the information-seeking screen, and looked for brain regions that represented either the initial VOI (only based on the initially presented beads) or the updated VOI (based both on the initial and extra beads).

These analyses revealed widespread cortical representations of the VOI above and beyond the right DLPFC (Fig. 6*A*,*B*; Table 1). The set of regions identified in these analyses was largely consistent regardless of the regressor (initial vs updated VOI) and the timing (extra bead vs information-seeking screen). Most notably, a cluster was found in the dorsomedial prefrontal cortex (DMPFC; largely in the superior frontal gyrus), consistent with previous reports on the representation of the VOI in this region (White et al., 2019; Kaanders et al., 2020).

We tested whether these regions show evidence of the VOI updating by conducting the same FIR modeling described above. To maintain the independence of the updating analysis from the ROI definition, we focused on the regions identified by the initial VOI (Fig. 6*C*,*D*). Although the only region that exhibited statistically significant updating ($p < 0.05$) was the right DLPFC, all regions exhibited numerically positive updating signals that are temporally consistent with each other (starting ~9 s from the

Table 1. Clusters with significant VOI signals

Region	Voxels	Cluster-level *p*	Peak T statistics	Peak MNI coordinate
Initial VOI at initial bead presentation				
Right dorsolateral prefrontal cortex (middle frontal gyrus)	137	0.0395	4.96	48, 40, 24
Initial VOI at extra bead presentation				
Right dorsolateral prefrontal cortex	504	0.00585	5.50	56, 14, 36
			4.94	46, 44, 28
Right inferior parietal lobule (supramarginal gyrus)	484	0.00669	5.26	54, −38, 56
Dorsomedial prefrontal cortex (superior frontal gyrus, paracingulate gyrus)	476	0.00669	5.01	4, 28, 40
Right anterior insula	278	0.00690	5.59	32, 22, −6
Right superior frontal gyrus	231	0.0136	5.12	16, 18, 66
Updated VOI at extra bead presentation				
Dorsomedial prefrontal cortex	400	0.0111	4.86	2, 20, 56
Right inferior parietal lobule	318	0.0148	4.79	54, −38, 56
Right anterior insula	252	0.0171	5.59	38, 20, −8
Right superior frontal gyrus	233	0.0188	5.27	16, 18, 66
Initial VOI at information-seeking screen				
Right dorsolateral prefrontal cortex	1520	0.000627	6.38	48, 38, 22
			5.32	44, 4, 38
Right inferior parietal lobule	893	0.00313	6.38	54, −38, 56
Dorsomedial prefrontal cortex	476	0.00899	5.18	2, 28, 46
Right central orbitofrontal cortex	260	0.0192	5.94	24, 54, −10
Right lateral occipital cortex	256	0.0219	5.22	24, −66, 50
Right anterior insula	245	0.0219	5.50	32, 26, 0
Updated VOI at information-seeking screen				
Right inferior parietal lobule	643	0.00522	5.73	54, −40, 56
Right dorsolateral prefrontal cortex	637	0.00564	5.61	48, 38, 22
			5.34	44, 28, 42
Right premotor cortex (precentral gyrus)	425	0.0117	5.15	44, 4, 34
Dorsomedial prefrontal cortex	404	0.0121	4.99	0, 20, 52
Right anterior insula	223	0.0242	5.31	32, 26, 0
Right central orbitofrontal cortex	209	0.0247	5.50	24, 54, −10
Right lateral occipital cortex	178	0.0322	4.98	24, −66, 50

Shown are clusters that were formed at voxelwise $p < 0.001$ and survived cluster mass $p < 0.05$, whole-brain FWE corrected.

extra bead onset or 10.5 s from the information-seeking screen onset), and statistical trends ($p < 0.10$) were observed in the right supramarginal gyrus, right central orbitofrontal cortex, and right lateral occipital cortex. These suggest that, although the right DLPFC starts to represent the VOI earliest and dynamically tracks the VOI over time, the resultant up-to-date VOI signals are also represented in a network of regions, perhaps to support cognitive or motor processes related to actual information-seeking choices.

Discussion

To make better decisions, we need to seek information adaptively based on what we already know (up-to-date decision evidence) and what is at stake (reward structure). When our knowledge is updated, we need to update the VOI accordingly to decide whether to seek further information. In this study, we used a variant of the beads task to examine how information seeking is shaped by current evidence and asymmetric reward structure and how the VOI is represented and updated in the brain.

We theoretically derived, and empirically verified, the normative prediction that information seeking should be biased by reward asymmetry. To maximize rewards in the upcoming decision, participants sought information more frequently when the current evidence preferred the less rewarding state. This finding is related to, but distinct from, the widespread observation of confirmation biases. Confirmation biases are commonly framed as biases in updating processes and/or decision criteria because of reward asymmetry or other factors such as precommitment (Luu and Stocker, 2018; Talluri et al., 2018; Gesiarz et al., 2019; Leong et al., 2019). We showed that even without such confirmation biases, information-seeking behavior is biased because of reward asymmetry, which is normative from the perspective of reward maximization. Our findings thus raise an important question regarding the extent to which phenomena attributed to confirmation biases could be accounted for by instrumentality-driven information seeking. Furthermore, although the current study used monetary rewards, our theoretical framework can be generalized based on the notion that people assign intrinsic values to beliefs they can hold (Kunda, 1990; Sharot and Garrett, 2016); if people are incentivized to hold certain beliefs, their information seeking would be biased depending on the extent to which the current evidence supports the desirable belief. It is worth noting, however, that the current study only examined reward structures where a correct bet yields asymmetric rewards but an incorrect bet does not. Our theoretical framework needs to be expanded to different reward structures to achieve a more comprehensive understanding of information-seeking biases across domains.

Our theoretical framework derived a quantitative prediction on how the VOI is sensitive to decision evidence and asymmetric reward structure (Fig. 2*E*,*F*). Although our participants' actual information-seeking behavior (Fig. 3*A*) did not precisely match the theoretical VOI, it is important to note that the prediction was derived under specific assumptions about decision processes, such as risk neutrality and perfectly Bayesian probability

estimation. In reality, our participants exhibited risk aversion in the betting choices (Fig. 3*B*), and their internal probability estimation could have deviated from the Bayesian predictions (e.g., as assumed in the prospect theory). Relaxing these assumptions quantitatively affects the shape of the VOI function (e.g., risk aversion would make the peak closer to the bead difference of zero, consistent with the observed information-seeking behavior in Fig. 3*A*), and the current study is not ideally designed to precisely characterize these decision processes. Nonetheless, the most important prediction of our theoretical framework is the existence of biases in information seeking as this prediction holds even when some assumptions are relaxed (e.g., nonlinear utility or probability weighting functions). To further test the validity of the VOI theory, future studies need to empirically characterize how information-seeking choices are related to individual differences in decision preferences. This approach is also a critical step toward characterizing the extent to which information seeking is driven by noninstrumental, psychological motives along with an instrumental benefit (Hunt et al., 2016; Kobayashi and Hsu, 2019).

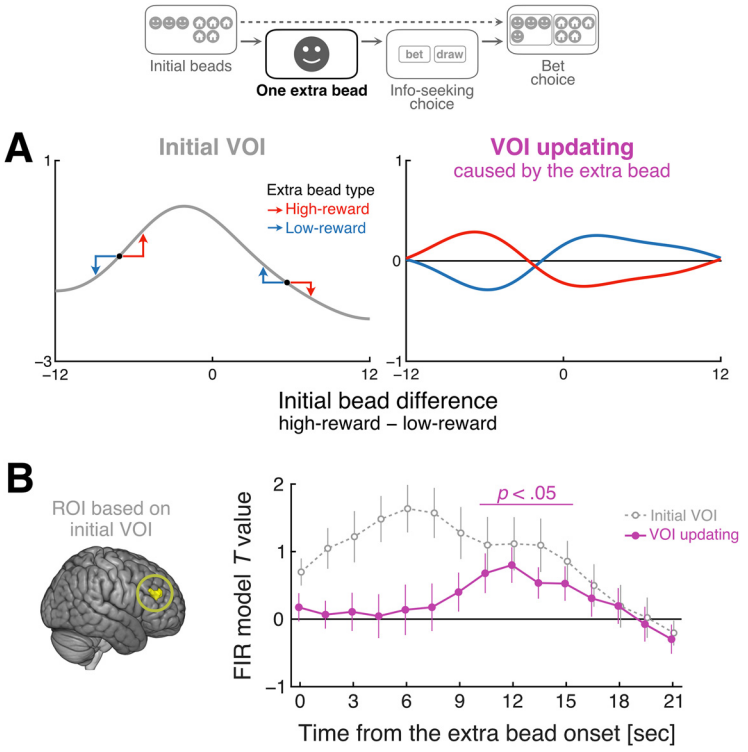

Figure 5. Updating of the VOI representation. The right DLPFC tracks VOI as it is updated by an extra bead, presented after the initial beads but before information seeking. ***A***, The VOI update was calculated as the signed difference between the VOI after the extra bead and the VOI before the extra bead. ***B***, Time courses of the initial VOI signal (gray) and the VOI updating signal (purple) in the right DLPFC. The right DLPFC responds not only to the initial VOI but also to the updating of the VOI (temporal cluster mass $p < 0.05$, FWE corrected). Because the region of interest was defined based on the initial VOI signal, estimation of the initial VOI signal is biased, but estimation of the updating signal is unbiased. Error bars indicate SEM.

Our finding that the VOI is represented in the DLPFC is consistent with a previous fMRI study on instrumentality-driven information seeking (Kobayashi and Hsu, 2019), despite a few key differences in task design. For instance, our paradigm required probabilistic inference on the hidden jar composition based on observable evidence, whereas Kobayashi and Hsu (2019) provided explicit visual presentation of outcome probability. Furthermore, our paradigm manipulated decision evidence available to the participant on each trial and examined its effect on information-seeking behavior and underlying neural signals. Thus, the current study not only replicates but also critically extends the previous finding by showing that the DLPFC is sensitive to the current evidence and biased by reward asymmetry. Along with neuroimaging evidence that the DLPFC is also activated on information seeking driven by factors other than instrumentality (Kang et al., 2009; Jepma et al., 2012; Gruber et al., 2014), these results suggest that the DLPFC is critical for adaptive information seeking across decision contexts.

Our theoretical and empirical results indicate that the VOI is tightly coupled with choice difficulty, or the difficulty of deciding which jar to bet on to maximize rewards, in the current paradigm. Although the close relationship between these two variables may be observed across a wide range of real-world information seeking, it raises the possibility that the right DLPFC cluster that we identified tracks choice difficulty as opposed to the VOI. We think this is unlikely for two reasons. First, we observed the VOI signals in the DLPFC on the initial evidence presentation, which is temporally distant from actual betting choices (in the majority of the trials, the participants observed an extra bead and then had the opportunity to draw even more beads before making a bet). Second, previous work has shown that the VOI signals in the right DLPFC could not be entirely accounted for by choice difficulty alone (Kobayashi and Hsu, 2019). This previous study manipulated the diagnosticity of information (i.e., to what extent the information would affect outcome probability) independently from the choice difficulty and observed that the diagnosticity systematically affected information-seeking behavior and the underlying VOI signals in the right DLPFC. Nonetheless, this previous study differed from our current design in several ways as discussed above, and future studies need to experimentally decouple the VOI from choice difficulty in belief-updating tasks such as ours.

Importantly, we showed that the DLPFC not only represents the VOI based on the initial evidence but also updates it when additional evidence is supplied. In other words, the DLPFC tracks the up-to-date VOI based on the most recent evidence. Such DLPFC signals may be critical for adaptive information seeking in situations where people accumulate decision evidence over time, either because information is gradually supplied from the environment or because people sequentially acquire multiple pieces of information. The DLPFC may be well suited for sustained and dynamically updated representation of the VOI as DLPFC neurons are known to exhibit sustained activity for working memory retention (Fuster and Alexander, 1971; Funahashi et al., 1989; Sreenivasan and D'Esposito, 2019).

In addition to the DLPFC, we observed that the VOI was also represented in several additional regions later in the trial, closer to the time of the decision whether to seek information or not. Among these regions, the most notable is the DMPFC, which past studies have also suggested is involved in information

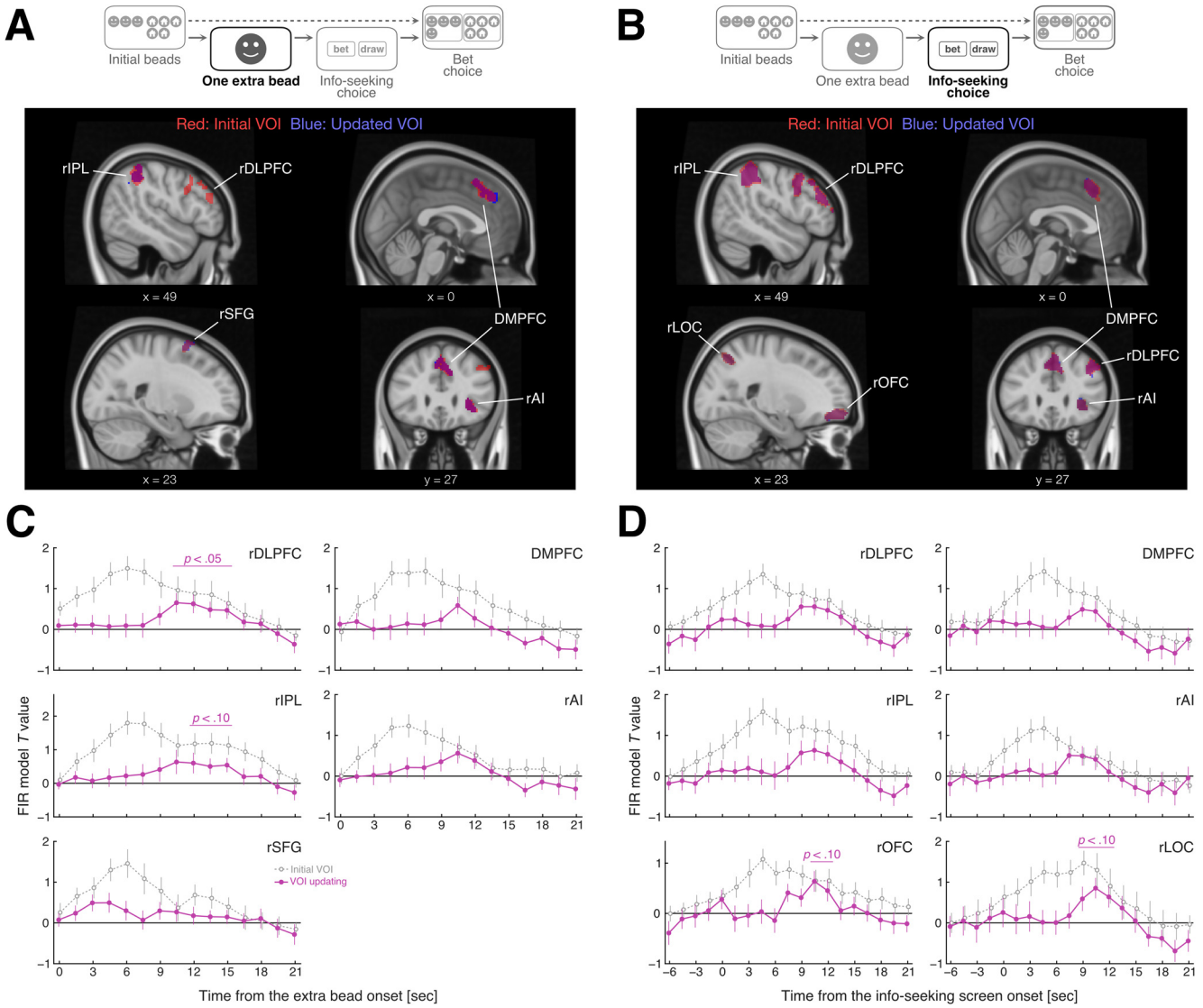

Figure 6. The VOI is widely represented in time epochs closer to the information-seeking choices. ***A***, ***B***, A number of regions were identified by parametric effects of the initial VOI (red, evaluated based on the initially presented beads only) or the updated VOI (blue, incorporating both the initial and extra beads) on the extra bead presentation (***A***) or information-seeking screen (***B***; cluster mass $p < 0.05$, whole-brain FWE corrected). ***C***, ***D***, Time courses of the initial VOI signal (gray) and the VOI updating signal (purple) in the regions identified by the initial VOI signals in ***A*** and ***B***, respectively. Although the clearest evidence of updating was observed in the right DLPFC, there were statistical trends in a few other regions (temporal cluster mass $p < 0.10$, FWE corrected). Error bars indicate SEM. IPL, Inferior parietal lobule; SFG, superior frontal gyrus; AI, anterior insula; OFC, orbitofrontal cortex; LOC, lateral occipital cortex.

seeking (White et al., 2019; Kaanders et al., 2020). One interpretation is that, while the DLPFC starts representing the VOI as early as some decision evidence is presented and keeps track of it over time, the VOI is also represented in other regions to support information-seeking decisions in an on-demand manner. In line with this interpretation, our analysis provided some evidence, although not statistically significant, that the regions outside the DLPFC represent the updated, most recent VOI (as opposed to the initial VOI), possibly by reading out dynamic representations in the DLPFC. Future studies should examine this possibility further, for instance by testing for causal relationships between computations in the DLPFC and other regions. Another interpretation, however, is that some of these regions do not represent the VOI but represent choice difficulty; in particular, the DMPFC could be involved in evaluating the uncertainty or conflict in which action to take (Rudebeck et al., 2008; Rushworth and Behrens, 2008; Kennerley et al., 2011; Shenhav et al., 2016). One reason that we did not observe the VOI signals in the DMPFC on the initial beads presentation could be that action uncertainty (regarding which jar to bet on) is evaluated later in time. This possibility could be tested by experimentally decoupling choice difficulty from the VOI as discussed above. Finally, it is worth noting that our current study included a modest sample size ($n = 15$) and thus may have lacked statistical power to detect signals related to the VOI reliably across time in regions outside the DLPFC.

Our results may have important implications for information-seeking deficits in clinical populations. For instance, schizophrenia has been associated with the tendency to make premature decisions without enough information seeking (Ross et al., 2015; Dudley et al., 2016; but see Baker et al., 2019), which could be accompanied by DLPFC hypoactivity (Barch and Ceaser, 2012) and/or lack of sensitivity in the DLPFC to decision evidence and reward asymmetry. Similarly, individuals with obsessive-compulsive disorder exhibit excessive information seeking (Hauser et al., 2017), which could be caused by

hyperactivity in the DLPFC (Eng et al., 2015) and/or lack of VOI updating in the DLPFC. Our experimental and theoretical framework provides a novel approach to characterize instrumentality-driven information seeking, which can be readily applied in future research with clinical populations.

References

Baker SC, Konova AB, Daw ND, Horga G (2019) A distinct inferential mechanism for delusions in schizophrenia. Brain 142:1797–1812.

Barch DM, Ceaser A (2012) Cognition in schizophrenia: core psychological and neural mechanisms. Trends Cogn Sci 16:27–34.

Bromberg-Martin ES, Hikosaka O (2009) Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron 63:119–126.

Bromberg-Martin ES, Hikosaka O (2011) Lateral habenula neurons signal errors in the prediction of reward information. Nat Neurosci 14:1209–1216.

Caplin A, Leahy J (2001) Psychological expected utility theory and anticipatory feelings. Q J Econ 116:55–79.

Charpentier CJ, Bromberg-Martin ES, Sharot T (2018) Valuation of knowledge and ignorance in mesolimbic reward circuitry. Proc Natl Acad Sci U S A 115:E7255–E7264.

Dudley R, Taylor P, Wickham S, Hutton P (2016) Psychosis, delusions and the "jumping to conclusions" reasoning bias: a systematic review and meta-analysis. Schizophr Bull 42:652–665.

Edwards W (1965) Optimal strategies for seeking information: models for statistics, choice reaction times, and human information processing. J Math Psychol 2:312–329.

Edwards W, Slovic P (1965) Seeking information to reduce the risk of decisions. Am J Psychol 78:188–197.

Eng GK, Sim K, Chen S-HA (2015) Meta-analytic investigations of structural grey matter, executive domain-related functional activations, and white matter diffusivity in obsessive compulsive disorder: an integrative review. Neurosci Biobehav Rev 52:233–257.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.

Furl N, Averbeck BB (2011) Parietal cortex and insula relate to evidence seeking relevant to reward-related decisions. J Neurosci 31:17572–17582.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Gesiarz F, Cahill D, Sharot T (2019) Evidence accumulation is biased by motivation: A computational account. PLOS Comput Biol 15:e1007089.

Gruber MJ, Gelman BD, Ranganath C (2014) States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. Neuron 84:486–496.

Hauser TU, Moutoussis M, Consortium N, Dayan P, Dolan RJ (2017) Increased decision thresholds trigger extended information gathering across the compulsivity spectrum. Transl Psychiatry 7:1296.

Howard RA (1966) Information value theory. IEEE Trans Syst Sci Cyber 2:22–26.

Hunt LT, Rutledge RB, Malalasekera WMN, Kennerley SW, Dolan RJ (2016) Approach-induced biases in human information sampling. PLoS Biol 14: e2000638.

Huq SF, Garety PA, Hemsley DR (1988) Probabilistic judgements in deluded and non-deluded subjects. Q J Exp Psychol A 40:801–812.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790.

Jepma M, Verdonschot RG, Steenbergen HV, Rombouts SARB, Nieuwenhuis S (2012) Neural mechanisms underlying the induction and relief of perceptual curiosity. Frontiers in Behavioral Neuroscience 6:5.

Kaanders P, Nili H, O'Reilly JX, Hunt LT (2020) Medial frontal cortex activity predicts information sampling in economic choice. BioRxiv 2020.11.24.395814.

Kang MJ, Hsu M, Krajbich IM, Loewenstein G, McClure SM, Wang JT, Camerer CF (2009) The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. Psychol Sci 20:963–973.

Kennerley SW, Behrens TEJ, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci 14:1581–1589.

Kidd C, Hayden BY (2015) The psychology and neuroscience of curiosity. Neuron 88:449–460.

Kobayashi K, Hsu M (2019) Common neural code for reward and information value. Proc Natl Acad Sci U S A 116:13061–13066.

Kobayashi K, Ravaioli S, Baranès A, Woodford M, Gottlieb J (2019) Diverse motives for human curiosity. Nat Hum Behav 3:587–595.

Krebs RM, Schott BH, Schütze H, Düzel E (2009) The novelty exploration bonus and its attentional modulation. Neuropsychologia 47:2272–2281.

Kreps DM, Porteus EL (1978) Temporal resolution of uncertainty and dynamic choice theory. Econometrica 46:185–200.

Kunda Z (1990) The case for motivated reasoning. Psychol Bull 108:480–498.

Lau JKL, Ozono H, Kuratomi K, Komiya A, Murayama K (2020) Shared striatal activity in decisions to satisfy curiosity and hunger at the risk of electric shocks. Nat Hum Behav 4:531–543.

Leong YC, Hughes BL, Wang Y, Zaki J (2019) Neurocomputational mechanisms underlying motivated seeing. Nat Hum Behav 3:962–973.

Luu L, Stocker AA (2018) Post-decision biases reveal a self-consistency principle in perceptual inference. ELife 7:E3548.

McGuire JT, Nassar MR, Gold JI, Kable JW (2014) Functionally dissociable influences on learning rate in a dynamic environment. Neuron 84:870–881.

Moutoussis M, Bentall RP, El-Deredy W, Dayan P (2011) Bayesian modelling of jumping-to-conclusions bias in delusional patients. Cogn Neuropsychiatry 16:422–447.

Phillips LD, Edwards W (1966) Conservatism in a simple probability inference task. J Exp Psychol 72:346–354.

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. Cambridge, MA: MIT.

Rasmussen CE, Nickisch H (2010) Gaussian processes for machine learning (GPML) toolbox. J Mach Learn Res 11:3011−−3015.

Ross RM, McKay R, Coltheart M, Langdon R (2015) Jumping to conclusions about the beads task? a meta-analysis of delusional ideation and data-gathering. Schizophr Bull 41:1183–1191.

Rudebeck PH, Behrens TE, Kennerley SW, Baxter MG, Buckley MJ, Walton ME, Rushworth MFS (2008) Frontal cortex subregions play distinct roles in choices between actions and stimuli. J Neurosci 28:13775–13785.

Rushworth MFS, Behrens TEJ (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci 11:389– 397.

Sharot T, Garrett N (2016) Forming beliefs: why valence matters. Trends Cogn Sci 20:25–33.

Shenhav A, Cohen JD, Botvinick MM (2016) Dorsal anterior cingulate cortex and the value of control. Nat Neurosci 19:1286–1291.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, Luca MD, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, Stefano ND, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1:S208–S219.

Sreenivasan KK, D'Esposito M (2019) The what, where and how of delay activity. Nat Rev Neurosci 20:466–481.

Talluri BC, Urai AE, Tsetsos K, Usher M, Donner TH (2018) Confirmation bias through selective overweighting of choice-consistent evidence. Curr Biol 28:3128–3135.

Wendt D (1969) Value of information for decisions. J Math Psychol 6:430–443.

White JK, Bromberg-Martin ES, Heilbronner SR, Zhang K, Pai J, Haber SN, Monosov IE (2019) A neural network for information seeking. Nat Commun 10:5168.

Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD (2014) Humans use directed and random exploration to solve the explore-exploit dilemma. J Exp Psychol Gen 143:2074–2081.