Common neural code for reward and information value

Kenji Kobayashi[a,b,1,2] and Ming Hsu[b,c,1]

[a]The Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, NY 10027; [b]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720; and [c]Haas School of Business, University of California, Berkeley, CA 94720

Edited by Terrence J. Sejnowski, Salk Institute for Biological Studies, La Jolla, CA, and approved May 20, 2019 (received for review November 28, 2018)

Adaptive information seeking is critical for goal-directed behavior. Growing evidence suggests the importance of intrinsic motives such as curiosity or need for novelty, mediated through dopaminergic valuation systems, in driving information-seeking behavior. However, valuing information for its own sake can be highly suboptimal when agents need to evaluate instrumental benefit of information in a forward-looking manner. Here we show that information-seeking behavior in humans is driven by subjective value that is shaped by both instrumental and noninstrumental motives, and that this subjective value of information (SVOI) shares a common neural code with more basic reward value. Specifically, using a task where subjects could purchase information to reduce uncertainty about outcomes of a monetary lottery, we found information purchase decisions could be captured by a computational model of SVOI incorporating utility of anticipation, a form of noninstrumental motive for information seeking, in addition to instrumental benefits. Neurally, trial-by-trial variation in SVOI was correlated with activity in striatum and ventromedial prefrontal cortex. Furthermore, cross-categorical decoding revealed that, within these regions, SVOI and expected utility of lotteries were represented using a common code. These findings provide support for the common currency hypothesis and shed insight on neurocognitive mechanisms underlying information-seeking behavior.

value of information | information seeking | reward | decoding | fMRI

Adaptive information seeking is critical in goal-directed behavior in humans. Collecting too little information, paying too much for information, not discriminating relevant information from irrelevant ones, or acting on unreliable or false information, can all result in failure to achieve desired goals. Understanding neurocognitive mechanisms of adaptive information seeking is not only important in neuroscience, psychology, and economics, but also has wide real-world applications, such as policymaking, public health, and artificial intelligence.

Information-seeking behavior is frequently viewed as reflecting agents' curiosity, i.e., motive to collect information for its own sake (1–3). This, however, poses a challenge for decision-making models such as reinforcement learning (RL) because information seeking by itself is not directly reinforced by explicit, tangible rewards. To incorporate curiosity-driven information seeking, decision-making models often postulate that information is intrinsically rewarding, and more specifically, exploratory actions are encouraged by some forms of bonus utility (4–6). Various forms of utility bonus have been proposed, such as surprise (7), novelty (8–10), perceived information gap (2), and anticipatory utility from savoring and dread (11–14). At the neural level, dopaminergic reward system may multiplex utility bonus with signals on extrinsic reward (14–20). Multiplexing extrinsic reward signals and utility bonus would help otherwise myopic agents to achieve appropriate balance of exploration (seeking more information) and exploitation (acting on available information).

Relying solely on curiosity, however, can be detrimental to adaptive goal-directed information seeking. Most importantly, motivation to acquire information should be sensitive to instrumental benefits that can be gained from accruing said information. For instance, our interest in weather forecast would likely be greater if we are trying to decide whether to go hiking or read indoors, compared with if we have already decided to stay indoors. Such goal-driven information seeking is particularly challenging when agents need to acquire information that they have never acquired before (e.g., a morning TV show in a foreign country we have never seen), where the bonus utility may not be adaptively formed based on the reward history.

Maximizing the instrumental benefits of information acquisition instead requires forward-looking simulation of agents' own actions and outcomes under different possible informational states ("I'll go hiking if it will be sunny, but read indoors if rainy.") If agents are driven solely by curiosity but do not explicitly evaluate instrumental benefits, they may fail to discern relevant and useful pieces of information from irrelevant ones. At the neural level, aforementioned curiosity-related dopaminergic activity may not be sufficient for maximization of instrumental benefits, and little is known regarding how dopaminergic reward system represents and integrates information's instrumental benefits and noninstrumental curiosity signals.

The importance to evaluate forward-looking instrumental benefit has long been recognized in economic and ethological studies of decision-making, owing to abundance of information seeking in problems ranging from comparison shopping to job/mate search (21, 22). Normative economic accounts presume that agents acquire information only as a consequence of utility maximization. Specifically, instrumental benefit of information is measured as value of information (VOI), i.e., how much it would improve choices and expected utility (EU); agents acquire the information only if its VOI outweighs its cost. Although normative VOI calculation may be computationally more complex than basic rewards (e.g., food or money), subsequent processes of cost-benefit analysis and action selection can be similar to other types of value-driven choices.

That the motivation to acquire information may be indexed by a single value measure, such as VOI, opens up a number of interesting

Significance

It is more important than ever to seek information adaptively. While it is optimal to acquire information based solely on its instrumental benefit, humans also often acquire useless information because of psychological motives, such as curiosity and pleasure of anticipation. Here we show that instrumental and noninstrumental motives are multiplexed in subjective value of information (SVOI) signals in human brains. Subjects' information seeking in an economic decision-making task was captured by a model of SVOI, which reflects not only information's instrumental benefit but also utility of anticipation it provides. SVOI was represented in traditional value regions, sharing a common code with more basic reward value. This demonstrates that valuation system combines multiple motives to drive information-seeking behavior.

Author contributions: K.K. and M.H. designed research; K.K. performed research; K.K. analyzed data; and K.K. and M.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]To whom correspondence may be addressed. Email: kenji.kobayashi@berkeley.edu or mhsu@haas.berkeley.edu.

[2]Present address: Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820145116/-/DCSupplemental.

Published online June 11, 2019.

Downloaded by guest on August 23, 2021

hypotheses. First, dopaminergic reward system may drive information seeking not only by encoding noninstrumental utility bonus but also instrumental benefits. While it is yet to be established whether normative VOI alone is represented or is multiplexed with noninstrumental motives to constitute subjective value of information (SVOI), the reward system may represent that informational value in the same way as conventional reward signals; for example, monetary reward. Second, individual neurons may encode both informational value and conventional rewards in the same way—the neural common currency hypothesis—which is advantageous for computing trade-offs guiding choice (23, 24). Common currency may particularly provide an elegant solution to the exploration-exploitation dilemma by allowing agents to directly compare action value of respective options (4, 25). Although common currency across reward categories has been observed in humans and monkeys (24, 26–28), it has not been tested with instrumental information value.

To address these questions, we conducted an fMRI study where subjects made choices on costly, but directly actionable, information. Subjects were presented with a lottery with two monetary outcomes (a gain and a loss) and asked to choose whether to accept or reject it. The outcome probability was initially hidden and described as fair, but subjects could purchase the information to reveal the true probability. This information has direct instrumental benefit because subjects could change their choice flexibly based on the revealed probability. For instance, a subject may play a fair lottery with a large gain and a small loss, but reject it if the loss turns out to be more likely. Although there is a chance that the loss probability turns out to be smaller and she retains her original choice, she may purchase the information if the benefit of avoiding the loss is large enough to justify the cost.

We observed that subjects' information-seeking behavior was indeed largely driven by instrumental benefit. Subjects' information purchase choice was systematically sensitive to lotteries' outcomes and possible probabilities, consistently with the normative VOI prediction. We further examined the contribution of additional noninstrumental motives. While we found no evidence for simplistic utility bonus, information-seeking choices were better explained by a SVOI model that involves anticipatory utility in addition to instrumental benefit. Next, using support vector regression (SVR) on voxel-wise BOLD signals, we tested a key prediction of the common currency hypothesis—common code between SVOI and reward values at the level of voxel-wise BOLD signals. We found that SVOI was represented in striatum and ventromedial prefrontal cortex (VMPFC), traditional valuation regions. Lastly, cross-categorical decoding revealed that these representations shared a common coding scheme with more basic values, consistent with the neural common currency hypothesis.

Results

Information Seeking Is Sensitive to Instrumental Benefits. To characterize the extent to which human information seeking is sensitive to instrumental benefits, we used a two-stage task (Fig. 1*A*). Subjects were first asked whether to accept or reject a lottery with two outcomes (one positive outcome x_1 and one negative outcome x_2), assuming they would not receive further information (under the initial informational state s_0: $P(x_1) = 0.5$). Next, two possible probability distributions were presented, one where the positive outcome is more likely (s_+; $P(x_1) = \pi$) and the other where the positive outcome is less likely (s_-; $P(x_1) = 1 - \pi$). One of them would be true but revealed only if subjects purchased the information (Fig. 1*B*). Thus, π determines information's diagnosticity (predictability of outcome) and was randomly chosen on each trial from {2/3, 5/6, 1}. Subjects were then presented with the monetary cost of the information and indicated whether they would purchase it. Even though the true probability (s_+ or s_-) was not revealed during the scanning to prevent learning, subjects were instructed beforehand that they would receive the information and could change their original choices after the scanning. We verified using model-free logistic regression that subjects made information purchase decisions based on all

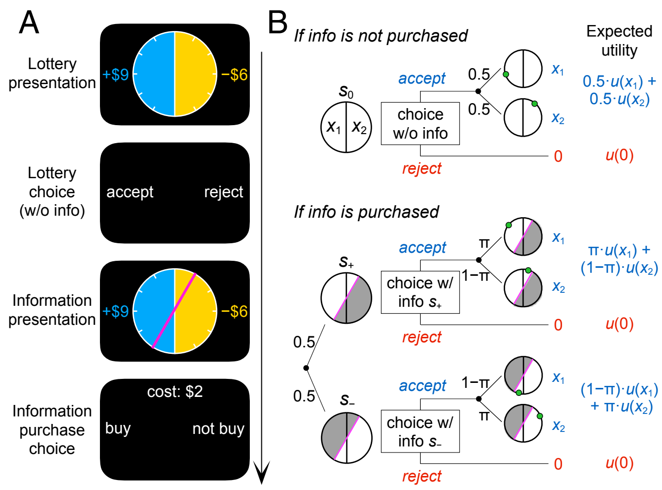

Fig. 1. Experimental task. (*A*) Subjects were presented with lotteries with two possible monetary outcomes, gain and loss, shown as a roulette wheel. When played, a green dot appeared at a random location on the perimeter, and its position determined the outcome (left or right). Outcomes were initially described as equally likely. Subjects indicated whether to accept or reject it, assuming they would not receive any further information. Potential information was then presented as a magenta partition line; if purchased, it would reveal which side of the partition the green dot would appear. Information diagnosticity, π, was determined by the partition angle. Subjects indicated whether to purchase it given the information cost. (*B*) If subjects did not purchase information, they made a choice under the initial informational state (s_0). If subjects purchased information, it revealed which of the two possible probability distributions, s_+ or s_-, was true, based on which subjects made a choice. Because subjects could not predict the true probability (s_+ or s_-) in advance, they need to stimulate their actions and EU in both states to compute instrumental benefit (VOI). Note that costless information is assumed here for illustration purposes; see *SI Appendix*, *SI Methods* for details on how our models deal with sunk cost of information.

experimental variables (x_1, x_2, π, and cost; all Ps < 0.01; *SI Appendix*, Figs. S1–S3).

Under normative economic accounts, agents accept the lottery if its EU under the current informational state (s_0, s_+, or s_-) is higher than the utility of status quo $u(0)$, and reject otherwise (Fig. 1*B*). Furthermore, they purchase the information if its instrumental benefit is higher than the cost, and forgo otherwise. The information provides the instrumental benefit by improving the overall EU, which happens only if EU-maximizing choices differ between informational states. Specifically, instrumental benefit is present if the lottery is preferable under s_0 but turns not to be preferable after unfavorable information (s_-), or if it is not preferable initially but changes to be preferable after favorable information (s_+). Instrumental benefit is nonexistent if the EU-maximizing choice would be the same irrespective of informational state (e.g., if the potential loss is extremely huge and gain is extremely small).

Normative VOI captures this marginal improvement of action utility, i.e., the difference in the expected utilities between the decision with the information and the decision without (Fig. 2*A*). Note that, because agents cannot predict the true probability a priori, they need to simulate their own choices under s_+ and s_-, average their EUs (i.e., overall EU of the informed choice), and compare it against EU under s_0. VOI computed as such is strongly sensitive to size of potential gains and losses; VOI is large if both the potential gain and loss are large, and small if the potential gain is very large and the loss is trivial (or vice versa), because the agent would not change its choice irrespective of the true probability in the latter cases.

We numerically derived instrumental normative VOI predictions based on outcomes (x_1, x_2) and diagnosticity (π) (*SI Appendix*, *SI Methods*). We then compared the predicted VOI against

PNAS PNAS PNAS
Downloaded by guest on August 23, 2021

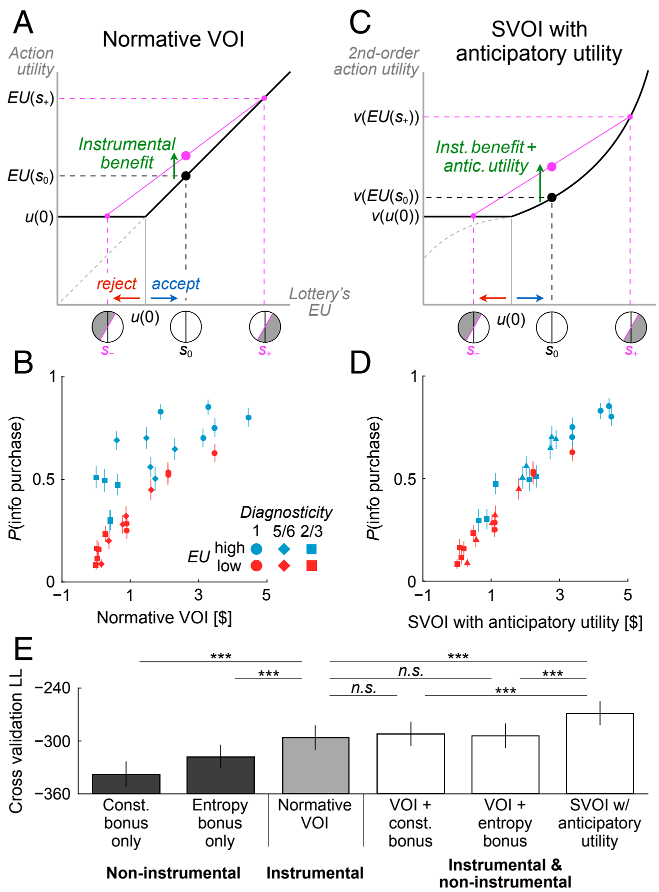

Fig. 2. Behavioral results. (*A*) Under normative VOI prediction, instrumental benefit is captured by the difference between the average EU of informed choices (s_+, s_-) and EU of uninformed choice (s_0). Note that costless information is assumed here for illustration purposes. (*B*) Information purchase probability was correlated with VOI predictions, but subjects over-purchased information in high-EU lotteries. Each dot represents a combination of lottery and diagnosticity, averaged over subjects and cost levels. (*C*) SVOI with anticipatory utility deviates from VOI due to nonlinearity of the second-order utility's aggregator function. (*D*) SVOI model outperformed normative VOI and was able to account for over-purchase of information in high-EU lotteries. (*E*) Normative VOI model achieved better behavioral fit than alternative models of utility bonuses. SVOI with anticipatory utility achieved even better fit, while alternative accounts that combine VOI with utility bonuses did not. ***: $P < 0.001$, *n.s.*: $P > 05$.

the probability that subjects purchased the information. Subjects purchased the information more often when the predicted normative VOI was higher (Kendall's tau = 0.62; Fig. 2*B*), and such relationship was driven by both lotteries (x_1, x_2) and diagnosticity (π) (*SI Appendix*, Fig. S2). More formally, using a softmax choice rule that relates choices to the difference between normative VOI and the information cost, we found the VOI was able to explain a substantial portion of the variation in information purchase choices ($P < 0.0005$; *SI Appendix*, Fig. S4*A*), consistent with the normative prediction that information acquisition should be sensitive to its instrumental benefit, and in particular the magnitudes of future possible outcomes.

To further evaluate the goodness-of-fit of normative VOI model, we compared it against two popular nonforward-looking, noninstrumental motives: a constant utility bonus (fixed utility for information) (4), and a utility bonus scaled by entropy reduction, which is sensitive to π but still not to outcomes (2, 5, 6). We found that normative VOI provided drastically better model fit than both accounts ($P < 10^{-3}$; Fig. 2*E*).

Coexistence of Instrumental and Noninstrumental Motives. Although we found that the instrumental benefit is a critical driver of information purchase in our task, it is possible that noninstrumental motives also contribute to the behavior. Accordingly, we tested the extent to which subjective value of information (SVOI), which would consist of instrumental VOI and noninstrumental motives, improves model fit. Specifically, we tested a prominent model of noninstrumental motives from the literature: anticipatory utility. Anticipatory utility, often referred to as savoring and dread, has been used in economics to explain people's nonnormative preference for information, and in particular timing of information delivery (e.g., many prefer to know if they win a raffle prize earlier because of savoring, while they prefer not to know the results of their cancer diagnosis because of dread) (11, 13, 29–32).

We constructed a model of SVOI that integrates anticipatory utility VOI using a recursive utility approach (11, 33). Recursive utility, as the normative VOI theory, assumes forward-looking, utility-maximizing agents. However, it relaxes the VOI theory by allowing utility functions to change depending on the availability of information; the mere presence of information may increase or decrease the overall utility, which cannot be explained by conventional expected utility theories that assume consistency of utility functions. Specifically, the theory evaluates the lotteries in our task based on expected second-order utility, which aggregates first-order EUs under the possible informational states (s_0, s_+, and s_-) in a nonlinear manner (Fig. 2*C*). A convex (concave) aggregator function amplifies (diminishes) the difference in the overall expected second-order utility between informed and uninformed choices compared with the standard prediction. Therefore, subjects are more (less) information seeking under SVOI than normative VOI if the aggregator function is convex (concave).

We found that our subjects' behavior was consistent with this SVOI composed of instrumental benefit and anticipatory utility. The correlation between information purchase probability and SVOI with anticipatory utility is significantly higher than the normative VOI (Kendall's tau = 0.87 vs 0.62, $P < 0.001$; Fig. 2*D*). Furthermore, trial-by-trial information purchase choices were predicted better by SVOI with anticipatory utility than normative VOI ($P < 0.001$; Fig. 2*E* and *SI Appendix*, Fig. S4*B*).

One important feature of anticipatory utility is its outcome dependence. Since the contribution of anticipatory utility depends on the convexity of the aggregator function, it is naturally allowed to be dependent on the possible outcomes, nicely echoing the widespread notions savoring on reward and dread on punishment. As a direct support for its outcome dependence, we noticed that subjects over-purchased information, compared with the normative VOI prediction, more often in higher valued lotteries than in lower valued ones (median split according to EU, $P < 10^{-3}$; Fig. 2*B* and *SI Appendix*, Fig. S4*A*). Since VOI computation already incorporates subjects' nonlinear utility function, this outcome-dependent over-purchase cannot be explained by factors such as risk preference (*SI Appendix*, *SI Methods*). The outcome-dependent over-purchase disappeared when the behavior was compared against the prediction of SVOI with anticipatory utility ($P > 0.05$; Fig. 2*D* and *SI Appendix*, Figs. S4*B* and S5).

We also tested two alternative models of noninstrumental motives, in which VOI is combined with aforementioned utility bonus term (constant bonus or entropy reduction bonus), which lacks sensitivity to possible outcomes. Neither of the utility bonus models improved the normative VOI model ($P > 0.30$), and both were outperformed by SVOI with anticipatory utility ($P < 0.001$; Fig. 2*E*). These further support that noninstrumental motive is sensitive to possible outcomes, consistent with anticipatory utility.

Neural Representation of SVOI. The above results suggest that subjects acquired information based on SVOI, which consists of forward-looking instrumental benefit and anticipatory utility. We next sought to investigate whether SVOI shapes subjective value function at the neural level. In particular, we asked whether SVOI was represented in valuation regions, and if so, whether that representation employs a common code with more conventional

Downloaded by guest on August 23, 2021

reward values. To this end, we asked subjects to make two value-based choices: whether to gamble on a lottery, and whether to purchase information regarding the said lottery, allowing us to compare encoding of the lottery EU and SVOI.

We first looked for SVOI representation during the presentation of the information's diagnosticity. Combined with potential outcomes, which had been already presented, the diagnosticity is sufficient for subjects to compute subjective benefit of the information. We asked if we could decode trial-by-trial SVOI from voxel-wise BOLD signals in a searchlight (10-mm radius) using one-run-leave-out five-hold cross-validation and SVR (Fig. 3; see *SI Appendix, SI Methods* for details). Prediction accuracy was measured as a partial correlation between the predicted and actual SVOIs controlling for diagnosticity π. This is to ensure that we detect regions engaged in valuation, rather than information-theoretic processing (e.g., entropy reduction) or visual processing (π was visually presented by the partition's angle).

Consistent with the hypothesis that dopaminergic reward system is involved in value-driven information seeking, we found that SVOI was decodable from striatum and VMPFC ($P < 0.05$, voxel-wise FWE corrected). SVOI representation was additionally found in lateral prefrontal cortex (middle frontal gyrus; MFG), right superior frontal gyrus, posterior cingulate cortex, right angular gyrus, and cerebellum (Fig. 4*A* and *SI Appendix*, Figs. S6 and S7 and Table S1). Since we evaluated decoding accuracy while controlling for diagnosticity, this successful decoding cannot be attributed to mere representation of diagnosticity (*SI Appendix*, Figs. S6 and S7). Striatum and VMPFC receive dopaminergic inputs and are the two regions that are the most associated with valuation in fMRI literature. Indeed, we found that lottery's EU was represented in striatum during the presentation of lottery ($P < 0.05$; Fig. 4*A* and *SI Appendix*, Fig. S8 and Table S1), suggesting the involvement of traditional valuation processing in SVOI.

Since SVOI is correlated with the lotteries' EU ($r = 0.62$), some of our SVOI decoding performance might have been attributable to signals related to EU rather than SVOI. This issue is particularly important because our EU cluster and SVOI cluster overlapped in striatum (Fig. 4*A*). However, SVOI decoding could not be explained by the possible presence of EU signals; SVOI decoding accuracy in all clusters was above chance even when EU was controlled for ($P < 0.05$, Bonferroni corrected; Fig. 4*B* and *SI Appendix*, Figs. S6 and S7). This supports that these regions use both outcomes and information diagnosticity to calculate SVOI, as normatively predicted.

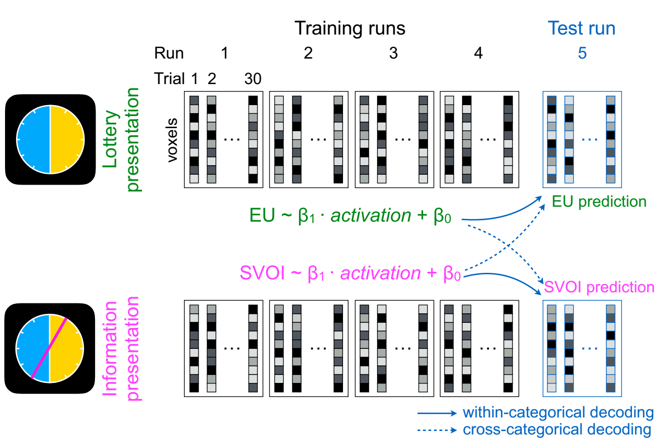

Fig. 3. Schematic illustration of decoding analysis. Values were decoded from activation during the lottery presentation (EU) or information presentation (SVOI) using support vector regression. In addition to the conventional within-categorical decoding (solid arrows), cross-categorical decoding (dotted arrows) was conducted to test the common code.

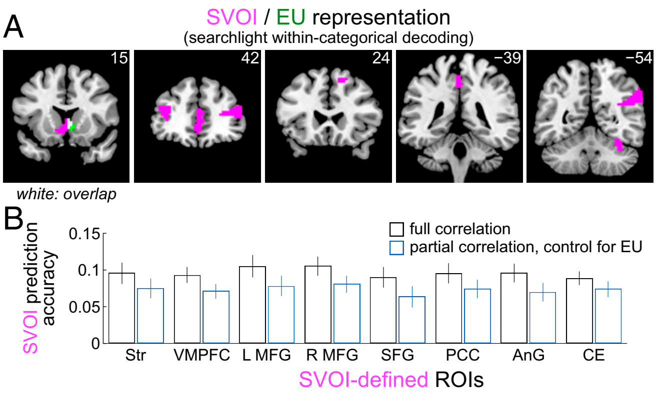

Fig. 4. Neural representation of SVOI and EU. (*A*) SVOI representation was revealed in regions including striatum and VMPFC by within-categorical decoding (magenta). EU representation was also found in striatum (green), overlapping with SVOI (white). (*B*) SVOI representation found in *A* cannot be explained by reinstantiation of EU representation. Prediction accuracy was significantly higher than zero even when EU was controlled for (blue; all $P < 0.05$, Bonferroni corrected). Black: accuracy without controlling for EU. Str: striatum; VMPFC: ventromedial prefrontal cortex; MFG: middle frontal gyrus; SFG: superior frontal gyrus; PCC: posterior cingulate cortex; AnG: angular gyrus; CE: cerebellum.

Common Code of SVOI and EU Representations. Having characterized representations of SVOI and EU, we next investigated their relationship, and in particular whether they are represented using a common code. Although we observed overlap of SVOI and EU clusters, this is not a strong evidence for a common code, because these representations could be distinct at a more fine-grained level. As a more direct test, we adopted cross-categorical decoding approach.

Specifically, if EU and SVOI are indeed represented on a common code in striatum at the voxel level, SVR trained based on EU in striatum should be able to predict SVOI (Fig. 3). We found that the decoder trained by EU could indeed predict SVOI above the chance level, compared with the permutation-based null-hypothesis distribution ($P < 0.05$; Fig. 5*A* and *SI Appendix*, Figs. S6 and S8). This holds when information diagnosticity was controlled for, and more critically, even when EU was controlled for. This provides a clear evidence that striatum did not just maintain or reactivate EU representation; rather, it flexibly switched the content of representation within each trial from EU and SVOI, presumably in preparation for the upcoming choices.

Lastly, to seek for further evidence for common neural code, we examined if decoders trained by SVOI could be used to decode EU. To control for FWE over eight SVOI clusters, we constructed null-hypothesis distribution based on the highest accuracy (t-statistics) over ROIs in each permutation iteration. EU prediction accuracy was above chance in striatum, VMPFC, and right MFG ($P < 0.05$, Fig. 5*B* and *SI Appendix*, Fig. S9). Although EU was not decodable from VMPFC and right MFG in the within-categorical decoding analysis above, it may be because we had used a more stringent statistical threshold. Together, these results show that human brains use a common code to represent SVOI and EU.

Discussion

A substantial portion of our daily actions pertains to information seeking. Particularly in the digital age where a tremendous amount of information is available at our fingertips, acquiring relevant information to an appropriate degree is as important as making use of acquired information. Going back at least to Berlyne (3), psychologists studying functions, causes, and consequences of motivation and interests have hypothesized the relationship between exploratory and information-seeking behavior and reward system. More recently, since Kakade and

Downloaded by guest on August 23, 2021

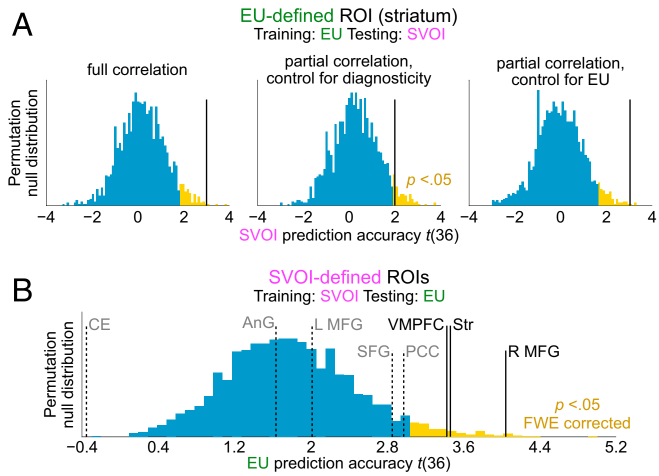

Fig. 5. Evidence for common code. Cross-categorical decoding accuracies (black vertical line) are compared against permutation-based null-hypothesis distributions. (*A*) In striatum (green in Fig. 4*A*), decoders trained on EU-predicted SVOI. (*B*) In striatum, VMPFC, and right MFG (magenta in Fig. 4*A*), decoders trained on SVOI-predicted EU.

Dayan's influential proposal (15), neuroscientists have provided evidence that putative noninstrumental motives are represented in dopaminergic reward system in monkeys (16, 20) and humans (8–10, 14, 17–19), as if they shape subjective value function that favors information seeking. However, because existing studies have largely focused on instances of noninstrumental information seeking, it remains unclear how subjective preference for forward-looking, instrumental information is formed, and to what extent dopaminergic reward system is involved in that process.

Behaviorally, our study provides evidence that subjective value of information (SVOI) consists of (at least) two motives: forward-looking instrumental benefit, consistent with normative economic VOI theories, and anticipatory utility, an example of noninstrumental motives. Other models on noninstrumental motives that are independent of reward value of outcomes, such as constant utility bonus (4), were insufficient in explaining the observed behavior. Particularly, consistent with the notion of savoring, we found outcome-dependent over-purchase of information. Our results extend the findings from the past studies on anticipatory utility, which focused mostly on noninstrumental information and did not quantitatively capture concurrent contribution of instrumental and anticipatory value for information (14, 31, 32). That both motives we identified are strongly sensitive to future possible outcomes highlight the involvement of valuation systems in information-seeking behavior in general, which is sometimes overlooked in curiosity literature.

The possibility that anticipatory utility is an important component of information seeking opens up several important questions. One particular issue concerns the effect of dread, or utility of anticipating negative outcomes (34). The effect of dread may be large enough for some people to avoid potentially negative information even when its instrumental benefit is critical, such as medical conditions, but more studies are needed to empirically quantify its relative contribution in instrumental information settings. Our study could not measure its effect quite reliably because our subjects could reject unfavorable lotteries. Second, anticipatory utility provides a possible explanation for the phenomenon of ambiguity aversion. Intuitively, the desire for information may be causally linked to aversion to the lack thereof (12). It may thus be not a coincidence that nonlinearity of the aggregator function that determines second-order utility, a critical part of recursive utility theory, is also central to some theories on ambiguity and compound lotteries (35). Future studies may be able to use our experimental paradigm to quantify anticipatory utility at the individual level and correlate with ambiguity attitude.

Neurally, if information seeking is driven by subjective value signals in dopaminergic reward system, we should expect such responses to exhibit two features; first, they should be scaled according to subjective preference for information, which would reflect both instrumental and noninstrumental motives, and second, they should be on a common currency with extrinsic reward. Our results that SVOI and EU share the common code in BOLD from striatum and VMPFC are highly consistent with these predictions, because these regions receive massive dopaminergic projection (36) and represent various kinds of values (37, 38), with some evidence for common currency (24, 26–28). In particular, our findings expand existing knowledge by showing that striatum also represents forward-looking instrumental benefits. Our decoding approach is suitable to test common code because it characterizes localized fine-grained representation, while typical brain mapping studies only examine spatially smoothed signals and whole-brain approaches such as elastic net examine representations distributed across the brain. Moreover, our results yield an additional prediction that, when monkeys act on forward-looking instrumental benefit of information, rather than merely receive noninstrumental information (16, 20, 39), it may also be encoded by their midbrain dopamine neurons.

We found SVOI representation in other brain regions as well, but with limited evidence for common code, where cross-categorical decoding was observed only in the right MFG. As SVOI computation requires the simulation of agents' own choices and outcomes under possible informational states, this may reflect the higher need of neurocognitive recourses than basic rewards, particularly working memory and planning. Relatedly, although encoding of noninstrumental information value was reported in orbitofrontal cortex (OFC) in monkeys, it was distinct from reward encoding (39), contrary to midbrain dopaminergic neurons (16). A recent human fMRI study corroborated this distinction, reporting that striatum and midbrain dopaminergic regions represent subjective value of noninstrumental information, which is influenced by possible outcome valence, while OFC merely represents availability of information regardless of valence (14). These suggest that, while OFC may encode signals relevant to information valuation, they seem not to use a common code with other types of values (40, 41). Taken together, information valuation may be supported by multiple neurocognitive processes, and it may converge with other values in striatum and/or VMPFC.

Our evidence for voxel-level common code is consistent with the neural common currency hypothesis. However, due to the nature of fMRI, it still leaves open the possibility that distinct neuron populations represent EU and SVOI but are sampled by overlapping voxels. More direct evidence for common currency at the neural level would come from electrophysiological recording while subjects acquire instrumental information. Our findings also raise an important question regarding the "common scale"; i.e., whether neural responses to SVOI and other reward values are scaled to be in the same range, thereby allowing direct comparison between information and rewards (23). To directly test the common scale, it would be ideal to use experimental paradigms in which subjects choose between information and noninformational goods and examine if cross-categorically decoded values predict such choices (27). Such an approach would also bridge the conceptual gap between one-shot information acquisition and exploration-exploitation dilemma, in which agents choose between myopic reward and information.

Further investigations are also needed on how humans adopt different strategies on information seeking under various goals, from stable to dynamic environments, and from short to long temporal horizons (1, 25). Although we found little support for utility bonus accounts in our experimental paradigm, it is entirely possible that they are responsible for exploratory behavior in more dynamic settings with longer temporal horizon (4, 42). Moreover, other proposed motives we did not study here, such as novelty or surprise (1, 7–10), might be necessary or more suited to ensure the adequate degree of information seeking in certain

NEUROSCIENCE

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Downloaded by guest on August 23, 2021

circumstances, particularly outside value-based decision-making domains. Our results raise an interesting possibility that such difference in motives may be partly caused by whether reliable SVOI signals from dopaminergic system are available, depending on factors such as the difficulty or cognitive load of model-based SVOI computation (43). Potential motives of information seeking have been long studied separately, and the current study marks an important step, both theoretically and empirically, toward integrative understanding.

Methods

All subjects provided informed consent; all protocols were approved by UC Berkeley Committee for the Protection of Human Subjects and Virginia Tech Institutional Review Board. Detailed method descriptions are available in *SI Appendix, SI Methods*.

Task Design. In each trial, a lottery with two outcomes (x_1, x_2) was presented as a roulette wheel, and subjects chose whether to play it assuming no further information (s_0). Then the information was presented as a magenta partition on the wheel, which defined the two possible probability distributions, $P(x_1) = \pi$ (s_+) or $1 - \pi$ (s_-). π, the information's diagnosticity, was determined by the orientation of the magenta partition; π = 1, 5/6, or 2/3 when the partition was vertical, slanted by 30°, or slanted by 60°, respectively. The cost of the information was presented after the delay, and subjects chose whether to purchase it. The purchased information was delivered after the scanning. When the information was delivered, one side of the magenta partition was brightened, indicating the posterior probability (s_+ or s_-), and subjects could change their original lottery choice. Subjects were told that the brighter side would be chosen randomly.

Behavioral Modeling. The predictions of VOI and SVOI with anticipatory utility were obtained as the sunk cost for the information at which agents that maximize EU (or expected second-order utility in the case of SVOI model) would be indifferent between informed and uninformed choices. The aggregator function that maps the first-order to second-order utility in SVOI model was estimated by likelihood maximization of information purchase choices. Models were compared by cross-subject cross validation.

fMRI Decoding Analysis. Voxel-wise activation from the two epochs in each trial, lottery presentation (for EU decoding), and information presentation (for SVOI decoding), were used as features of leave-one-run-out cross-validation SVR. Within-categorical decoding took a whole-brain searchlight approach. SVOI decoding accuracy was evaluated by Pearson partial correlation between predicted and actual SVOI labels while controlling for π. Accuracy of cross-categorical decoding was evaluated within the ROIs defined by the within-categorical decoding. Null-hypothesis distribution was obtained by permuting labels across trials while maintaining the trial-wise pairing of SVOI and EU.

ACKNOWLEDGMENTS. We thank Amanda Savarese, Cassandra Carrin, and Duy Phan for assistance with data collection. This research was funded by National Institute of Mental Health Grant MH098023 and Collaborative Research in Computational Neuroscience/National Institute on Drug Abuse Grant DA043196 (to M.H.).

1. C. Kidd, B. Y. Hayden, The psychology and neuroscience of curiosity. *Neuron* **88**, 449–460 (2015).
2. G. Loewenstein, The psychology of curiosity: A review and reinterpretation. *Psychol. Bull.* **116**, 75–98 (1994).
3. D. E. Berlyne, A theory of human curiosity. *Br. J. Psychol.* **45**, 180–191 (1954).
4. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
5. P. Dayan, T. J. Sejnowski, Exploration bonuses and dual control. *Mach. Learn.* **25**, 5–22 (1996).
6. S. Ishii, W. Yoshida, J. Yoshimoto, Control of exploitation-exploration meta-parameter in reinforcement learning. *Neural Netw.* **15**, 665–687 (2002).
7. A. Barto, M. Mirolli, G. Baldassarre, Novelty or surprise? *Front. Psychol.* **4**, 907 (2013).
8. N. Bunzeck, E. Düzel, Absolute coding of stimulus novelty in the human substantia nigra/VTA. *Neuron* **51**, 369–379 (2006).
9. B. C. Wittmann, N. D. Daw, B. Seymour, R. J. Dolan, Striatal activity underlies novelty-based choice in humans. *Neuron* **58**, 967–973 (2008).
10. R. M. Krebs, B. H. Schott, H. Schütze, E. Düzel, The novelty exploration bonus and its attentional modulation. *Neuropsychologia* **47**, 2272–2281 (2009).
11. D. M. Kreps, E. L. Porteus, Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* **46**, 185–200 (1978).
12. R. Golman, G. Loewenstein, Information gaps: A theory of preferences regarding the presence and absence of information. *Decision* **5**, 143–164 (2016).
13. A. Caplin, J. Leahy, Psychological expected utility theory and anticipatory feelings. *Q. J. Econ.* **116**, 55–79 (2001).
14. C. J. Charpentier, E. S. Bromberg-Martin, T. Sharot, Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7255–E7264 (2018).
15. S. Kakade, P. Dayan, Dopamine: Generalization and bonuses. *Neural Netw.* **15**, 549–559 (2002).
16. E. S. Bromberg-Martin, O. Hikosaka, Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* **63**, 119–126 (2009).
17. M. J. Kang *et al*., The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychol. Sci.* **20**, 963–973 (2009).
18. M. Jepma, R. G. Verdonschot, H. van Steenbergen, S. A. R. B. Rombouts, S. Nieuwenhuis, Neural mechanisms underlying the induction and relief of perceptual curiosity. *Front. Behav. Neurosci.* **6**, 5 (2012).
19. M. J. Gruber, B. D. Gelman, C. Ranganath, States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron* **84**, 486–496 (2014).
20. E. S. Bromberg-Martin, O. Hikosaka, Lateral habenula neurons signal errors in the prediction of reward information. *Nat. Neurosci.* **14**, 1209–1216 (2011).
21. R. A. Howard, Information value theory. *IEEE Trans. Syst. Sci. Cybern.* **2**, 22–26 (1966).
22. W. Edwards Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *J. Math. Psychol.* **2**, 312–329 (1965).
23. F. Grabenhorst, E. T. Rolls, Value, pleasure and choice in the ventral prefrontal cortex. *Trends Cogn. Sci. (Regul. Ed.)* **15**, 56–67 (2011).
24. D. J. Levy, P. W. Glimcher, The root of all value: A neural common currency for choice. *Curr. Opin. Neurobiol.* **22**, 1027–1038 (2012).
25. J. D. Cohen, S. M. McClure, A. J. Yu, Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 933–942 (2007).
26. D. McNamee, A. Rangel, J. P. O'Doherty, Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat. Neurosci.* **16**, 479–485 (2013).
27. J. Gross *et al*., Value signals in the prefrontal cortex predict individual preferences across reward categories. *J. Neurosci.* **34**, 7580–7586 (2014).
28. D. V. Smith *et al*., Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *J. Neurosci.* **30**, 2490–2495 (2010).
29. G. Wu, Anxiety and decision making with delayed resolution of uncertainty. *Theory Decis.* **46**, 159–199 (1999).
30. S. H. Chew, J. L. Ho, Hope: An empirical study of attitude toward the timing of uncertainty resolution. *J. Risk Uncertain.* **8**, 267–288 (1994).
31. K. Eliaz, A. Schotter, Experimental testing of intrinsic preferences for noninstrumental information. *Am. Econ. Rev.* **97**, 166–169 (2007).
32. A. Falk, F. Zimmermann, *Beliefs and Utility: Experimental Evidence on Preferences for Information* (SSRN, 2017).
33. M. Ahlbrecht, M. Weber, The Resolution of uncertainty: An experimental study. *J. Inst. Theor. Econ.* **152**, 593–607 (1996).
34. D. Lovallo, D. Kahneman, Living with uncertainty: Attractiveness and resolution timing. *J. Behav. Decis. Making* **13**, 179–190 (2000).
35. C. Camerer, M. Weber, Recent developments in modeling preferences: Uncertainty and ambiguity. *J. Risk Uncertain.* **5**, 325–370 (1992).
36. S. N. Haber, B. Knutson, The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology* **35**, 4–26 (2010).
37. A. Rangel, T. Hare, Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* **20**, 262–270 (2010).
38. A. Rangel, C. Camerer, P. R. Montague, A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
39. T. C. Blanchard, B. Y. Hayden, E. S. Bromberg-Martin, Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron* **85**, 602–614 (2015).
40. R. C. Wilson, Y. K. Takahashi, G. Schoenbaum, Y. Niv, Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267–279 (2014).
41. P. H. Rudebeck, E. A. Murray, The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron* **84**, 1143–1156 (2014).
42. R. C. Wilson, A. Geana, J. M. White, E. A. Ludvig, J. D. Cohen, Humans use directed and random exploration to solve the explore-exploit dilemma. *J. Exp. Psychol. Gen.* **143**, 2074–2081 (2014).
43. N. D. Daw, Y. Niv, P. Dayan, Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).

Downloaded by guest on August 23, 2021