


The Rostrolateral Prefrontal Cortex Mediates a Preference for High-Agency Environments

Kaitlyn G. Norton¹ and  Mimi Liljeholm^{2,3}

¹Department of Psychological Science, ²Department of Cognitive Sciences, and ³Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, California 92697

The ability to exert flexible instrumental control over one's environment is a defining feature of adaptive decision-making. Here, we investigated neural substrates mediating a preference for environments with greater instrumental divergence, the distance between outcome probability distributions associated with alternative actions. A formal index of agency, instrumental divergence allows an organism to flexibly obtain the currently most desired outcome as preferences change. As such, it may have intrinsic utility, guiding decisions toward environments that maximize instrumental power. Consistent with this notion, we found that a measure of expected value that treats instrumental divergence as a reward surrogate provided a better account of male and female human participants' choice preferences than did a conventional model, sensitive only to monetary reward. Using model-based fMRI, we found that activity in the rostralateral and ventromedial PFC, regions associated with abstract cognitive inferences and subjective value computations, respectively, scaled with the divergence-based account of expected value. Implications for a neural common currency of information theoretic and motivational variables are discussed.

Key words: agency; decision-making; fMRI; instrumental divergence; RL PFC; utility

Significance Statement

Agency is a central concept in philosophy and psychology. While research thus far has focused on cognitive and perceptual measures of agency, recent work demonstrating a strong preference for high-agency environments indicates a salient motivational dimension. Here, using instrumental divergence, the distance between outcome distributions associated with alternative actions, as a formal index of agency, we found that brain regions associated with directed exploration and subjective value computations, respectively, were selectively modulated by a model that treated agency as a reward surrogate, over models that assigned utility only to monetary payoffs. In a subset of regions, such effects were predicted by the influence of instrumental divergence on economic choice preferences. Our results elucidate neural mechanisms mediating the utility of agency.

Introduction

A series of recent studies (Mistry and Liljeholm, 2016; Liljeholm et al., 2018) have demonstrated that individuals prefer environments in which instrumental divergence, the degree to which alternative actions differ with respect to their outcome probability distributions, is relatively high. A high level of instrumental divergence is a necessary feature of flexible instrumental control: If all available action alternatives have identical, or very similar, outcome distributions, such that selecting one action over

another does not significantly alter the probability of any given outcome state, an agent's ability to exert control over its environment is considerably impaired. Conversely, when available action alternatives produce distinct outcomes, discrimination and selection between actions allow an agent to flexibly obtain the currently most desired outcome. Importantly, since subjective outcome utilities often change from one moment to the next, flexible instrumental control is essential for reward maximization and, as such, may have intrinsic value, serving to reinforce and motivate decisions that guide the organism toward high-agency environments (Liljeholm, 2018). In the current study, we investigated neural substrates mediating the apparent preference for high instrumental divergence.

In previous work, Liljeholm et al. (2013) found that activity in the right supramarginal gyrus (rSMG) scaled parametrically with instrumental divergence during performance of a reward-based decision-making task, and increases across training blocks during acquisition of contingencies with relatively high instrumental

Received Oct. 16, 2019; revised Apr. 2, 2020; accepted Apr. 6, 2020.

Author contributions: K.G.N. and M.L. performed research; K.G.N. and M.L. analyzed data; M.L. designed research; M.L. wrote the first draft of the paper; M.L. wrote the paper.

The authors declare no competing financial interests.

This work was supported by National Science Foundation Career Grant 1654187 to M.L. We thank Rongwen Tai and Nidhi Banavar for assistance with data acquisition.

Correspondence should be addressed to Mimi Liljeholm at m.liljeholm@uci.edu.

<https://doi.org/10.1523/JNEUROSCI.2463-19.2020>

Copyright © 2020 the authors

divergence (Liljeholm et al., 2015). Although the designs of these previous studies did not permit investigation of the influence of instrumental divergence on behavioral choice preferences, nor of a common neural value scale for instrumental divergence and conventional reward, the results suggest that the rSMG implements a basic representation of instrumental divergence. With respect to established neural correlates of subjective value, a plethora of research suggests that the ventromedial PFC (vmPFC) retrieves and compares the values of decision outcomes (for review, see O’Doherty, 2011). Intriguingly, activity in the vmPFC scales with the values of a wide variety of goods, including food, money, and clothes, suggesting a common neural value-scale for distinct stimulus categories (Chib et al., 2009; McNamee et al., 2013). It is unknown, however, whether this common value-scale might also extend to more abstract, cognitive, commodities, such as instrumental divergence. Here, using a task in which participants choose between gambling environments based on differences in both instrumental divergence and monetary payoffs, we combine computational cognitive modeling with fMRI to investigate neural representations of the utility of agency.

Materials and Methods

Participants

Thirty undergraduates at the University of California at Irvine (19 females; mean age = 21.6 ± 3.9 years) participated in the study for monetary compensation. The target sample size was based on our previous work assessing a neural implementation of instrumental divergence (Liljeholm et al., 2013). Three participants were excluded due to excessive head movement (>6 mm), and 1 participant due to severe banding artifacts, leaving a sample size of 26. All participants gave informed consent, and the Institutional Review Board of the University of California, Irvine, approved the study.

Experimental design and statistical analysis

Task and procedure. Participants were scanned with fMRI while performing a simple gambling task, illustrated in Figure 1 and described in detail by Mistry and Liljeholm (2016). At the start of the experiment, participants were instructed that they would assume the role of a gambler in a casino, playing a set of four slot machines (labeled A1, A2, A3, and A4, respectively) that yielded three different colored tokens (blue, green, and red), each worth a particular amount of money, with different probabilities. They were further told that, in each of several rounds, they would be required to first select a room in which only two slot machines were available, that they could only gamble on the two machines in the selected room on several subsequent trials in that round, and that, in some rooms, they would be forced to accept the gambling choices of a computer algorithm. The critical measure was the decision at the beginning of each gambling round, between rooms that could differ in terms of divergence, monetary payoffs, and free versus forced choice.

While the machine-token probabilities remained constant throughout the study, the monetary values of the tokens changed intermittently throughout the task, and these changes always occurred after the participant had already committed to a particular room in a given round.

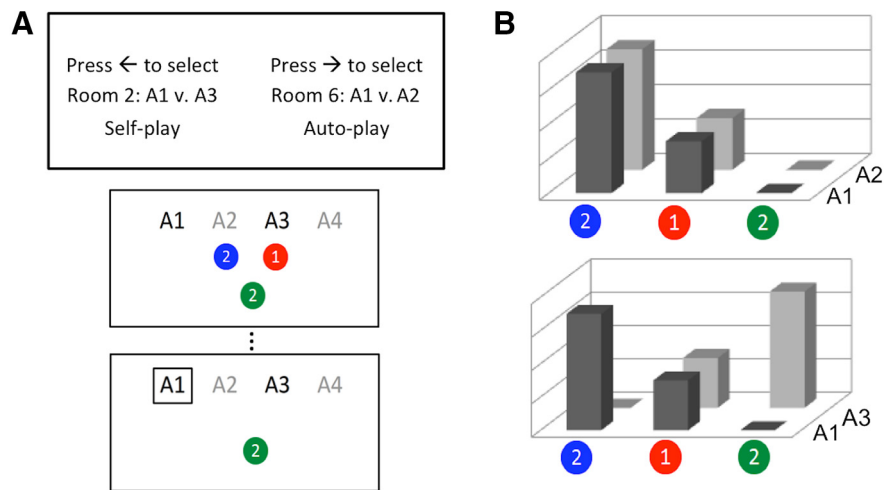


Figure 1. Task description. **A**, Trial illustration, showing the room-choice screen at the beginning of a block (top), and the choice (middle) and feedback (bottom) screens on a trial in the selected room. Choice screens remained until a response was performed and feedback screens had a 2 s duration. **B**, Example of zero- (top) and high-divergence (bottom) gambling rooms, where the bars represent the probability of each color token for each slot machine, and with current monetary values printed on the token surfaces. All variables, except instrumental divergence (e.g., mean and maximum expected payoffs and outcome entropies), are the same for both rooms.

Consequently, although changes in value were explicitly announced, and the current values of tokens were always printed on their surface once a “room” had been entered, participants regularly found themselves in a room in which the expected monetary payoffs of the two available slot machines had suddenly been altered. Three token-reward distributions were intermittently alternated across rounds (changing on average every third round), such that expected monetary payoffs were either the same across rooms or differed across rooms in either the same or opposite direction of instrumental divergence. In addition to mimicking dynamic changes in the utilities of natural rewards, the sporadic changes in token payoffs allowed us to pit the value of instrumental divergence against that of monetary reward. Participants were instructed that, at the end of the study, a single gambling round would be selected randomly, and they would receive any monetary gain earned in that round, up to \$15.

Two distinct probability distributions over the three possible token outcomes were used, and the assignment of outcome distributions to slot machines was such that two of the machines (either A1 and A2 or A1 and A3, counterbalanced across subjects) always shared one distribution, whereas the other two machines shared the other distribution. This yielded a low (zero) divergence for rooms in which the two available slot machines shared the same probability distribution, and a relatively high divergence for rooms in which slot machines had different outcome probability distributions (see Fig. 1B). Before starting the gambling task, participants were given a practice session to learn the probabilities with which each slot machine produced the different colored tokens, and were explicitly instructed that these probabilities would remain the same throughout the task. If a participant’s estimate of any given probability deviated by >0.2 from the programmed probability, the participant was returned to the beginning of the practice phase, and this continued until all rated probabilities were within 0.2 points of programmed probabilities. At the end of the study, participants again provided estimates of the machine-token probabilities.

Given a constant outcome entropy level, increases in instrumental divergence are accompanied by increases in the perceptual diversity of obtainable outcomes, a variable previously shown to elicit preferences in economic tasks (Ayal and Zakay, 2009). To rule out perceptual diversity as an explanation for any effects of instrumental divergence, gambling rooms differed in terms of whether the participant was allowed to choose freely between slot machines in the room (self-play) or a computer

algorithm alternated between machines across trials in that room (auto-play). In auto-play rooms, participants were still required to press a key corresponding to the slot machine indicated by the computer, to control for movement execution. Critically, in the absence of voluntary choice, high divergence no longer yields flexible instrumental control. However, the computer algorithm still yields greater perceptual diversity in high- than in low-divergence rooms (i.e., an algorithm alternating across slot machines in each panel of Fig. 1B, while still yielding greater perceptual diversity in the bottom panel, would categorically eliminate instrumental control, and thus instrumental divergence). Consequently, if choices were driven by a desire to maximize perceptual diversity, rather than instrumental divergence, they should not differ depending on whether the participant or an alternating computer algorithm chose between the slot machines in a room. The self-play versus auto-play manipulation also relates the preference for high instrumental divergence to a well-established preference for free over forced choice (e.g., Leotti and Delgado, 2011, 2014).

There were a total of 44 gambling rounds, with participants choosing between two gambling rooms (the decision of interest) at the start of each round, followed by 3–5 gambling trials within the selected room. For all participants, there were 12 room-choice scenarios in which divergence differed across the two rooms while monetary payoffs were the same, 16 room-choice scenarios in which both divergence and monetary payoffs differed across rooms, in the same (8 scenarios) or opposite (8 scenarios) direction, and 16 choice scenarios in which divergence was the same (high or low) across room options while monetary payoffs were either the same (8 or 4 scenarios, balanced across subjects) or differed (8 or 12 scenarios, balanced across subjects). For the last case, in which divergence was the same while payoffs differed, the divergence was necessarily low for both room options, given the probability and reward distributions. Room-choice scenarios were further split into cases where both room options were self-play, one room option was self-play and the other was auto-play, or both room options were auto-play. The order of room choice scenarios, and of the different reward distributions, was counterbalanced across subjects. Henceforth, we distinguish between choice scenarios with High Instrumental Divergence (Hi_ID), in which at least one room option was both high divergence and self-play, and choice scenarios with No Instrumental Divergence (No_ID), which included scenarios in which high-divergence options were auto-play, as well as scenarios in which both options had zero divergence. Of the total 44 room-choice trials, 18 or 14 (balanced across subjects) were Hi_ID.

Computational models. Instrumental divergence is formalized as the Jensen-Shannon (JS) divergence of sensory-specific outcome probability distributions (Liljeholm et al., 2013). Let P_1 and P_2 be the respective token outcome probability distributions for two available slot machines, let O be the set of possible token outcomes, and let $P(o)$ be the probability of a particular token outcome, o . The instrumental divergence, ID , is as follows:

$$ID = \frac{1}{2} \sum_{o \in O} \log \left(\frac{P_1(o)}{P_*(o)} \right) P_1(o) + \frac{1}{2} \sum_{o \in O} \log \left(\frac{P_2(o)}{P_*(o)} \right) P_2(o), \quad (1)$$

where

$$P_* = \frac{1}{2} (P_1 + P_2)$$

JS divergence is intimately related to Shannon entropy, a decision variable frequently shown to influence economic choice (e.g., Abler et al., 2009), which is greatest when the distribution over outcomes is uniform. Despite the close relationship between the two measures (JS divergence is the symmetrized relative entropy), they have dramatically different implications: While Shannon entropy reflects uncertainty about the state of the outcome variable given performance of a particular action, JS divergence, as applied here, reflects the degree to which discrimination and selection between available actions increase the controllability of the outcome (Liljeholm, 2018). For example, in Figure 1B,

the mean and maximum Shannon entropy is the same across the top and bottom panels, whereas instrumental (IS) divergence is zero in the top panel and relatively high in the bottom panel. In our previous work, we have demonstrated that these closely related information theoretic variables elicit neural activity in distinct brain regions (Liljeholm et al., 2013). In the current study, the Shannon entropy (i.e., unpredictability) of outcomes given a particular slot machine was held constant across slot machines.

Also, instrumental divergence is defined with respect to the sensory rather than motivational features of outcome states. Since subjective outcome utilities often change from one moment to the next (e.g., due to sensory satiety, instantiated in our task as changes in monetary token values), a measure of divergence based on outcome utilities would be inherently unstable, and thus poorly suited for the proposed role of instrumental divergence as an organizing guide of behavior toward high-agency environments (Liljeholm, 2018). That is not to say, of course, that outcome utilities do not critically influence motivated choice. We defined the expected value of a gambling room, EV_R , as follows:

$$EV_R = \sum_{m \in M} \sum_{o \in O} (p(o|m) * \$ (o)) + w * ID_R \quad (2)$$

where M is the set of slot machines available in a given room, $p(o|m)$ is the probability of a particular token outcome, o , given selection of a particular slot machine, m , and $\$(o)$ is the monetary value of that particular token outcome. ID_R is the instrumental divergence of the gambling room, and w is a parameter indicating the subjective utility of instrumental divergence. In a conventional utility model ($\$EV$), w is set to zero for all participants, so that value is always defined solely in terms of expected monetary payoffs. However, in an alternative model ($IDEV$), instrumental divergence is treated as a potential reward surrogate, with w being fit to each individual's room choices. Thus, in the latter variant, the expected values of gambling rooms reflect both the expected monetary payoff and the instrumental divergence associated with that room.

While there are many instances for which both the maximum and mean slot machine payoff is the same across rooms while divergence differs, and yet others for which divergence is the same while both the mean and maximum payoff differ, there are also some for which the room with the greater divergence has a greater maximum, but not mean, monetary payoff. To address the possibility that the influence of instrumental divergence on choice preferences reflects expectations of a greater maximum monetary payoff in the selected room, we specify a third, policy, model ($polEV$) such that:

$$polEV_R = \sum_{m \in M} \sum_{o \in O} (p(o|m) * \$ (o) * p(choose_m)) + w * ID_R \quad (3)$$

where w is again set to zero, as in $\$EV$, and where $p(choose_m)$ is equal to the degree to which a given participant selected the slot machine with the greater monetary payoff when payoffs differed across machines in self-play rooms. In auto-play rooms, $p(choose_m)$ is fixed at 0.5, reflecting the alternating response strategy of the computer algorithm, and reducing $polEV$ to $\$EV$.

For each model, a softmax rule with a noise parameter, τ , was used to translate the expected values of gambling rooms into choice probabilities, and free parameters were fit to behavioral data by minimizing the negative log likelihood of observed choices, separately for each subject. Choice scenarios in which at least one room option was both high-divergence and self-play (Hi_ID), yielding high instrumental divergence, and those in which the high-divergence room option was auto-play or both rooms had zero divergence (No_ID), were modeled separately and contrasted. When choosing between two available room options, the participant does not know whether monetary token values will change once a room has been selected, or what those new values might be; accordingly, all expected room values are computed based on the last experienced token values (i.e., those from the previous gambling round).

Model performance was evaluated in several ways. First, the corrected Akaike Information Criterion (AICc) was used to compare the fit of each EV model to behavior. Second, to validate the models, best-fit parameters were used to simulate gambling-room decisions in our task, based on each EV model (using Eqs. 1–3), to assess whether the models would generate the basic qualitative prediction that, mean and maximum monetary payoffs being equal, a room with high divergence would be significantly preferred over a room with zero divergence, when the high-divergence room was self-play but not when it was auto-play. Third, to assess parameter and model recovery, 1000 parameter values, for each parameter (w and τ), were drawn from a uniform distribution with the same bounds as those used for behavioral model fitting ($0 \leq w \leq 1$ and $0 \leq \tau \leq 10$), and gambling-room decisions were simulated for each set of parameter values and for each EV model.

Neuroimaging acquisition and analyses.

All MR images were obtained in a 3T Siemens Prisma Scanner, fitted with a 32-channel RF receiver head coil, padded to minimize head motion, at the Facility for Imaging and Brain Research (FIBRE) at the University of California, Irvine. Functional images covered the whole brain with 48 continuous 3-mm-thick axial slices with T2*-weighted gradient EPI (TR = 2.65 s, TE = 28 ms, 3-mm² in-plane voxel size, 64 × 64 matrix). All participants had a high-resolution structural image taken before functional scanning commenced (T1-weighted FSPGR sequence: 208 continuous 0.8 mm axial slices 0.4-mm² in-plane voxel size; 640 × 640 matrix). All stimulus materials were presented, and all responses recorded, using MATLAB. All imaging data were preprocessed and analyzed with MATLAB and SPM. Functional images were preprocessed with standard parameters, including slice timing correction, spatial realignment, coregistration of the high-resolution structural image to functional images, segmentation of the structural image into tissue types, spatial normalization of functional images into MNI space, and spatial smoothing with an 8 mm FWHM kernel.

At the first level, for each participant, a GLM was specified with two regressors, respectively indicating the onsets of Hi_ID room-choice screens, in which at least one room option was both high divergence and self-play, and No_ID room-choice screens, in which the high-divergence options were auto-play, or both options had zero divergence, parametrically modulated by the absolute difference between rooms in IDEVs, as well as by response times. We used the absolute difference in IDEVs, rather than signed difference measures, such as chosen-unchosen (Hunt et al., 2012) or unchosen-chosen (Wunderlich et al., 2009), because we were looking for a predecision signal, reflecting a contrast of available options rather than of decisions and their counterfactuals. Additional regressors modeled the onsets of choice trials within a selected room (i.e., between available slot machines), separately for self-play and auto-play rooms, with response times specifying durations and with two parametric modulators, respectively, specifying the room divergence and expected monetary payoffs on each trial. Finally, two onset regressors modeled the outcome period of each slot machine trial, modulated by the monetary value of the obtained token, for self-play and auto-play rooms, respectively, and two motor regressors, respectively, modeled the onsets of key presses for room selections and slot machine selections. Regressors of no interest indicated separate scanning runs and accounted for the residual effects of head motion. For comparison, two

Table 1. Mean ± SD probability ratings^a

	0.7	0.0	0.3
Before	0.69 ± 0.04	0.00 ± 0.00	0.31 ± 0.03
After	0.66 ± 0.11	0.02 ± 0.05	0.32 ± 0.04

^aProgrammed probabilities are shown in the top row. Mean ratings, obtained before and after the gambling task, are averaged across identical objective probabilities, yielding three unique values.

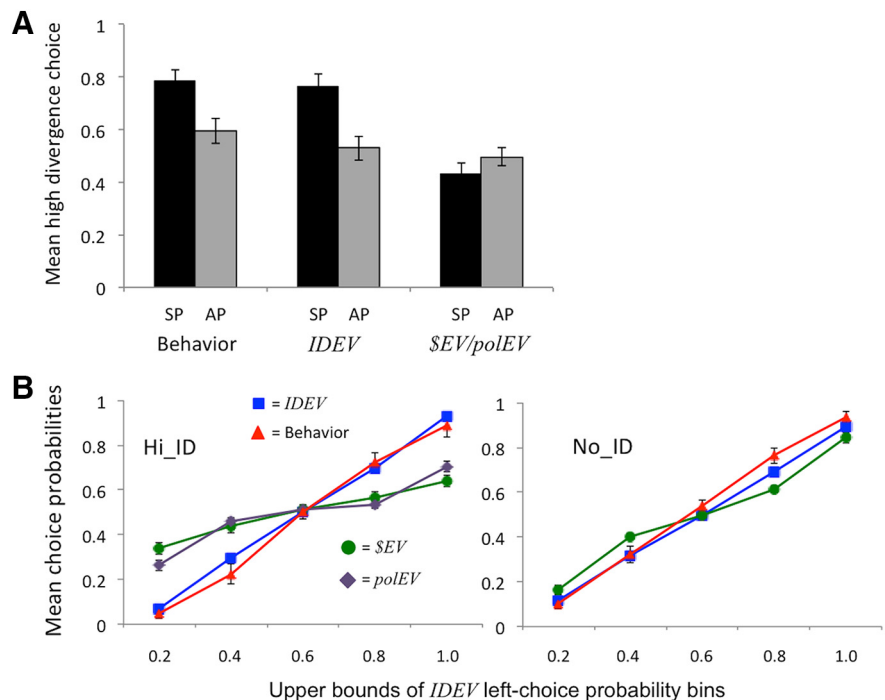


Figure 2. Behavioral results. **A**, Mean proportions of choosing a high-divergence room over a zero-divergence room when the high-divergence room was self-play (SP) versus when it was auto-play (AP), for behavioral choice data, and for data simulated, using best-fit parameters, with the IDEV and \$EV/poEV models, respectively. **B**, Mean probability of choosing the left room option derived using the IDEV model (blue), poEV model (purple), and \$EV model (green), respectively, together with actual behavioral left-choice proportions (red), plotted as a function of binned IDEV left-choice probabilities, for Hi_ID (left) and No_ID (right) choice scenarios. The poEV model converges with the \$EV model in the No_ID condition. Error bars indicate SEM.

separate GLMs were specified for each participant, identical to the first, except that room choice scenarios were modulated by the absolute difference between rooms in \$EVs and poEVs, respectively. Fixed-effects models were estimated using restricted maximum likelihood and an AR (1) model for temporal autocorrelation. Group-level statistics were generated by entering contrasts of first-level parameter estimates into between-subject analyses.

We specified two ROIs. First, based on previous work implicating the rSMG in instrumental divergence (Liljeholm et al., 2013, 2015), we used an anatomical mask of this region from the WFU PickAtlas (https://www.nitrc.org/projects/wfu_pickatlas/). Second, Chib et al. (2009) identified a subregion of the vmPFC encoding a common-currency value scale across stimulus categories. Here, we created a vmPFC mask by centering a 20 mm sphere on their peak coordinates (−6, 41, −6; averaged across replication experiments). While these ROIs are good candidates for a neural implementation of the motivating and reinforcing properties of agency, the notion of instrumental divergence as a decision variable is nascent, and the brain basis of its impact on choice virtually unknown; it is likely, therefore, that exploratory analyses will be particularly informative. We report exploratory effects at a whole-brain FWE cluster-corrected threshold of $p < 0.05$, calculated using the Statistical non-Parametric Mapping toolbox (SnPM13; <http://warwick.ac.uk/snpm>) (Nichols and Holmes, 2002), with 5000

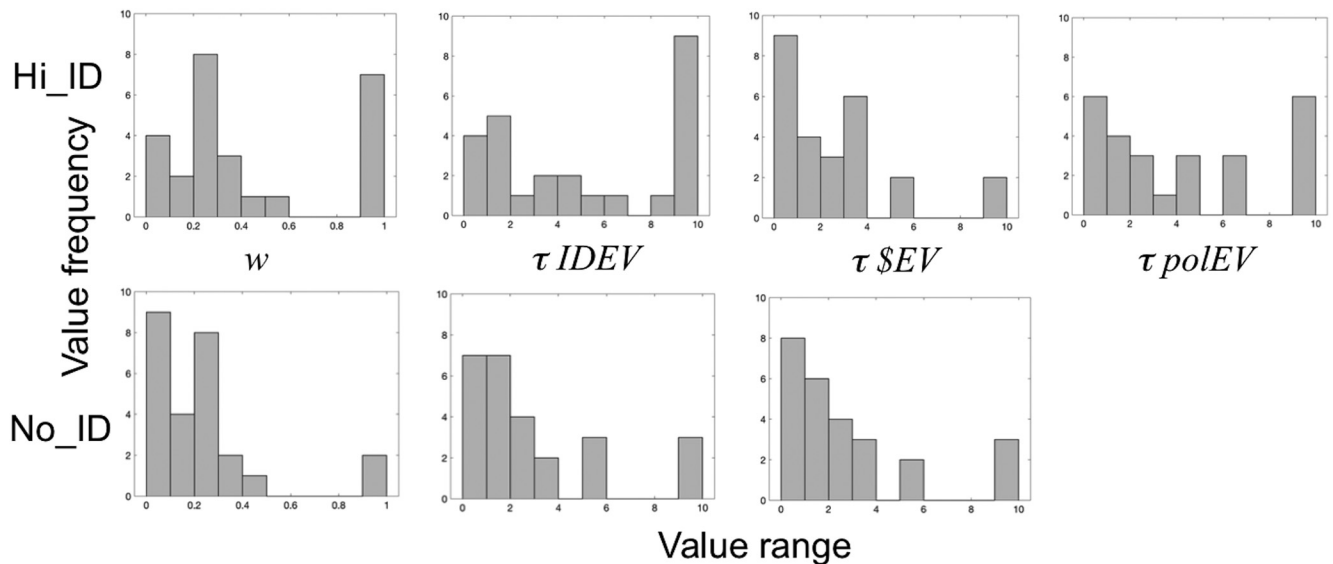


Figure 3. Histograms of best-fitting parameters. Distribution of the w parameter, indicating how much an individual values high divergence, and of the decision noise parameter, τ , for the *IDEV*, *\$EV* and *poleV* models, respectively, in the Hi_ID (top) and No_ID (bottom) conditions.

permutations, 8 mm variance smoothing, and an uncorrected height threshold of $p < 0.005$. Finally, to follow-up on initial exploratory effects, an additional ROI was specified, covering the right rostralateral PFC (RLPFC), by averaging across the coordinates of three ROIs tested by Badre et al. (2012) as follows: $x, y, z = 27, 50, 28$ (Badre et al., 2012); $x, y, z = 27, 57, 6$ (Daw et al., 2006); and $x, y, z = 35, 54, 0$ (Boorman et al., 2009). All ROI analyses were performed using MarsBar (<http://marsbar.sourceforge.net>).

Results

Behavioral results

All t tests performed on behavioral data were two-tailed paired comparisons. Participants required on average 1.93 (SD = 1.07) cycles of practice on the action-token probabilities. Mean probability ratings, obtained right before and right after the gambling phase, and averaged across identical programmed probabilities, are shown in Table 1.

The decision of interest was that at the beginning of each gambling round, as participants chose between rooms that differed in terms of their divergence, expected monetary payoffs and self-play versus auto-play. Mean choice proportions and model-derived choice probabilities are illustrated in Figure 2. The behavioral results closely replicate those of Mistry and Liljeholm (2016) and Liljeholm et al. (2018), revealing a clear preference for rooms with greater instrumental divergence: For choice scenarios in which both *poleV* and *\$EV* were the same across room options while divergence differed, participants were significantly more likely to select a room with high divergence over a room with zero divergence when the high-divergence room was self-play than when it was auto-play ($t_{(25)} = 2.78$, $p = 0.01$). Model validation, using the best-fit parameter values (τ and w) to simulate choice data, revealed that this qualitative result was uniquely captured by the *IDEV* model (Fig. 2A).

Moreover, as illustrated in Figure 2B, across all choice scenarios, choice probabilities derived using the *IDEV* model provided a much closer fit to behavioral choices than did choice probabilities derived using the conventional *\$EV* or *poleV* models, for Hi_ID choice scenarios, in which high divergence yielded instrumental divergence, but not for No_ID choice scenarios, in which

high divergence yielded perceptual diversity but no instrumental control. Consistent with this pattern of results, a repeated-measures ANOVA performed on the AICc scores revealed a significant interaction ($F_{(1,25)} = 13.25$, $p < 0.005$), such that scores were significantly lower, indicating a better fit, for the *IDEV* model (18.72 ± 5.92) than the *\$EV* model (22.30 ± 3.65) for Hi_ID choice scenarios ($t_{(25)} = 2.85$, $p = 0.009$), while being significantly lower for the *\$EV* (35.21 ± 7.99) than the *IDEV* (36.30 ± 8.95) model for No_ID choice scenarios ($t_{(25)} = 3.12$, $p = 0.005$). Recall that the *poleV* model reduces to the *\$EV* model in No_ID choice scenarios, precluding its inclusion in a balanced ANOVA. Nevertheless, planned comparisons revealed analogous results when comparing the *IDEV* and *poleV* models, with the *IDEV* model yielding significantly lower AICc scores (18.72 ± 5.92) than the *poleV* model (21.10 ± 3.05) in the Hi_ID condition ($t_{(25)} = 2.37$, $p = 0.020$).

Histograms of fit parameter values are provided in Figure 3. The w parameter is greater in the Hi_ID condition than in the No_ID condition ($t_{(25)} = 2.64$, $p = 0.014$), reflecting a greater utility of high divergence in the presence of instrumental control; likewise, for the *IDEV* model, τ is significantly greater in the Hi_ID condition than in the No_ID condition ($t_{(25)} = 3.88$, $p < 0.0001$), reflecting a greater reliance on model derived utilities when they reflect agency, rather than just perceptual diversity. In contrast, for the *\$EV* model, which categorically fails to capture agency, τ does not differ across Hi_ID and No_ID conditions ($p = 0.73$). (Only the w parameter impacts the model-derived values regressed against the BOLD signal; the τ parameter simply modulates the influence of those values on softmax-derived choice probabilities.) Model simulations across a large parameter space yielded significant recovery of all free parameters ($0.66 < r < 0.8$) and significant recovery of each model (with all p values < 0.0001). Specifically, in both the Hi_ID and No_ID condition, AICc scores were significantly lower for the *IDEV* model when fit to choice data generated by the *IDEV* model than when fit to data generated by either the *\$EV* or *poleV* model. Likewise, scores were significantly lower for the *\$EV* model when fit to data generated by the *\$EV* model than when fit to data generated by either the *IDEV* or *poleV* models, and significantly lower for the *poleV* model when

Table 2. Significant effects of the different EV models, across hi_ID and no_ID conditions, during the room-choice period^a

Contrast	Region	Peak MNI	Cluster size		
<i>IDEV</i> (Hi_ID>No_ID)	Mid-cingulate	0, 2, 28	1232		
	Anterior cingulate	2, 38, 18			
	RLPFC	26, 50, 14			
	Left premotor	-26, 8, 52			
Correlation between above contrast and behavioral choice preferences	Right STG	62, -4, 4	2907		
	Right LOFC	46, 52, 0			
	Right insula	40, 12, 2	1938		
	Right IFG	52, 36, 0			
	Left STG	-44, -12, 0			
	Left insula	-38, -4, 6			
	Left LOFC	-44, 34, -14			
	Left IFG	-54, 22, 6			
	\$EV (Hi_ID-No_ID) (-)	Right precentral/postcentral gyrus		40, 2, 38	2134
		Right IFG		34, 12, 32	
Right MTG		64, -46, 12	1523		
Right ITG		52, -62, 4			
Left lingual gyrus		-28, -92, -14	1340		
Left cerebellum		-12, -58, -16			
<i>poEV</i> (Hi_ID>No_ID) (-)		Left precentral/postcentral gyrus		-34, -30, 70	1157

^aSnPM-corrected cluster sizes are shown in the fourth column, with empty rows indicating that a cluster is continuous with that listed above. (-), Negative correlation; STG, superior temporal gyrus; LOFC, lateral orbitofrontal gyrus; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; ITG, inferior temporal gyrus.

fit to data generated by the *poEV* model than when fit to data generated by either the *IDEV* or *\$EV* models, in both Hi_ID and No_ID conditions.

Neuroimaging results

All significant results of whole-brain corrected exploratory analyses are listed in Tables 2 and 3. Significant effects of ROI analyses are reported in the text, together with the relevant *t*-statistic and corrected (for number of ROIs) *p* values. Maps of *t* statistics in figures are uncorrected for display purposes only. All figure plots of neural effects are unbiased, showing mean betas or contrast values extracted from entire, independently specified, ROIs.

An interaction contrast assessing greater parametric modulation by the difference in *IDEVs* across room options for Hi_ID but not for No_ID room-choice scenarios revealed significant whole-brain corrected effects in the RLPFC, the mid-cingulate and anterior cingulate cortex (ACC), and left premotor cortex, as well as effects in the vmPFC ROI ($t = 2.18, p = 0.040$). To probe the relevance of neural activity to behavior, a subsequent test assessed whether neural effects of the interaction contrast depended on the influence of instrumental divergence on economic choice performance; specifically, how much more likely a participant was to choose a room option with greater *\$EV* when that room had high instrumental divergence versus when it had zero instrumental divergence. Effects of the difference in *IDEVs* across room options, specific to Hi_ID choice scenarios, were significantly predicted by this behavioral measure in both the rSMG ($t = 2.20, p = 0.038$) and vmPFC ($t = 2.10, p = 0.046$) ROIs, as were significant whole-brain corrected exploratory effects throughout the bilateral insula, superior temporal and inferior frontal gyri, and lateral orbitofrontal cortex. A scatter plot of the results in the rSMG is presented in Figure 4. We caution the reader that, although significant, the correlation coefficient ($r = 0.41$) yields a relatively low power of 0.53 at $p < 0.05$ with our sample size.

Table 3. Significant effects of the expected values (EV) of slot machines, and of the monetary token outcomes, in self-play and auto-play rooms, during the machine-choice and machine-outcome periods^a

Contrast	Region	Peak MNI	Cluster size		
<i>EV</i> self-play (-)	Left caudate	-16, 30, 2	2569		
	Left MFG				
	Left IFG				
	Left insula	-30, 18, 6			
	Anterior cingulate	-4, 26, 28			
	Cuneus	10, -98, 12			
	Precuneus	12, -78, 50			
	Right occipital	32, -84, 28			
	Left cerebellum	-44, -58, -26			
	Left calcarine	-6, -100, 6			
<i>EV</i> (self-play-auto-play)	Right cerebellum	34, -52, -22	1218		
	Left cerebellum	-44, -58, -26	2101		
\$ outcome self-play	SMA	12, -8, 56	22,797		
	Right postcentral gyrus	62, -4, 18			
	Right caudate	18, -6, 28			
	Left caudate	-6, 12, -4			
	Mid-cingulate	6, -18, 36			
	Left MFG	-42, 20, 50			
	Anterior cingulate	0, 38, 6			
	DMPFC	-4, 46, 50			
	vmPFC	-4, 32, -12			
	vlPFC	-38, 40, -14			
	\$ outcome auto-play	Left amygdala		-18, -4, 12	1124
		Left putamen		-16, 10, -6	
		Right caudate		20, 22, 14	
		Right MTG		46, -56, -2	
		Right precentral/postcentral gyrus		14, -20, 68	
Left postcentral gyrus		-16, -32, 68			
SMA		6, 0, 56			
Right occipital		28, -78, 40			

^aSnPM-corrected cluster sizes are shown in the fourth column, with empty rows indicating that a cluster is continuous with that listed above. (-), Negative correlation; DMPFC, dorsomedial PFC; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; pTG, posterior temporal gyrus; SMA, supplementary motor area; vlPFC, ventrolateral PFC; vmPFC, ventromedial PFC.

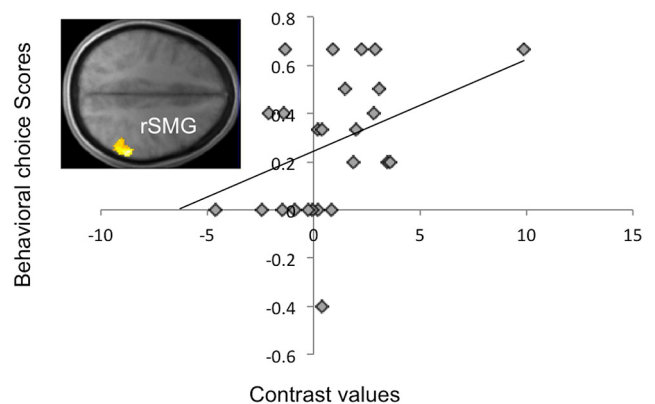


Figure 4. Neuroimaging results. Map of the *t* statistics for a test of the correlation between effects of the *IDEV* (Hi_ID-No_ID) contrast and the influence of instrumental divergence on behavioral choices, showing results in the rSMG ROI, thresholded at $p < 0.005$ uncorrected for display purposes. Scatter plot represents behavioral choice scores as a function of unbiased contrast values, extracted from the entire rSMG ROI.

When applied to the *\$EV* model, the same interaction contrast revealed significant negative whole-brain corrected effects (i.e., activity decreased as the difference between rooms in *\$EV* increased, for Hi_ID but not for No_ID choice scenarios), extending throughout the right primary and premotor cortex,

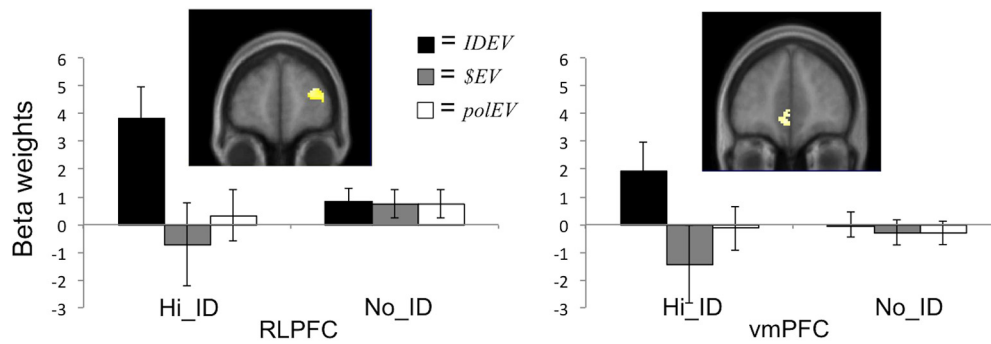


Figure 5. Neuroimaging results. Map of the t statistics for an interaction test of greater parametric modulation by the difference across rooms in *IDEV* than *SEV*, for Hi_ID but not for No_ID choice scenarios, showing effects in the right RLPFC ROI (left) and the vmPFC ROI (right), thresholded at $p < 0.01$ uncorrected for display purposes. Bar plots represent unbiased β weights extracted from entire ROIs, for differences across rooms in *IDEV*, *SEV* and *polEV*, in Hi_ID and No_ID conditions respectively. Error bars indicate SEM.

the inferior and middle temporal gyri, and the lingual gyrus and cerebellum. Likewise, for the *polEV* model, only negative effects emerged, throughout the left precentral and postcentral gyrus. No effects of these contrasts emerged in *a priori* ROIs (all p values > 0.67), and none was significantly predicted by the influence of instrumental divergence on behavioral choice preferences, for either the *SEV* or *polEV* model.

To directly compare the *IDEV* model with the two conventional *EV* models, we entered contrasts of β weights estimated in the first-level analysis into two group-level 2 (*EV* model) \times 2 (Play Type) ANOVAs, assessing significantly greater parametric modulation by the *IDEV* model than the *SEV* model and *polEV* model, respectively, in the Hi_ID condition, but not in the No_ID condition, using our independently specified ROIs in the vmPFC and right RLPFC (see Materials and Methods). As illustrated in Figure 5, neural activity in the right RLPFC ROI ($t = 2.45$, $p = 0.016$), as well as the vmPFC ROI ($t = 2.33$, $p = 0.020$), was significantly better accounted for by the *IDEV* than the *SEV* model, in Hi_ID but not in No_ID choice scenarios. The same comparison of the *IDEV* model with the *polEV* model again yielded significant effects in the right RLPFC ROI ($t = 2.50$, $p = 0.014$), as well as a marginally significant effect in the vmPFC ROI ($t = 1.97$, $p = 0.051$).

Additional contrasts assessed neural responses on trials in which participants chose between slot machines available in a particular room. During the choice period of each slot machine trial, for self-play rooms, activity in the ACC, left anterior caudate, precuneus, cuneus, occipital cortex, left middle and inferior frontal gyrus, and left insula increased with the sum of expected monetary payoffs for slot machines available in the current gambling room: While no effects of expected monetary payoffs reached significance for auto-play rooms, there was a significant interaction of expected monetary payoffs with self-play versus auto-play rooms, such that activity throughout the bilateral cerebellum decreased with an increase in expected monetary payoffs on gambling trials in auto-play rooms, but not in self-play rooms. No significant effects emerged for the divergence of slot machines in either self-play or auto-play rooms during the choice period of each slot machine trial, nor was there a significant interaction of divergence-modulation for self-play versus auto-play rooms.

Finally, during the outcome period of each slot machine trial, in self-play rooms, activity increased with an increase in the monetary value of the delivered token throughout the

following: ventral and dorsal striatum; vmPFC, ventrolateral PFC, dorsomedial PFC, and dorsolateral PFC; mid- to posterior cingulate; supplementary motor area; and postcentral gyri. In auto-play rooms, significant effects of this contrast emerged in the amygdala, striatum, precentral and postcentral gyri, supplementary motor area, right posterior middle temporal gyrus, and occipital gyrus, with no significant outcome-by-play interaction.

Discussion

Countless studies on motivated behavior have investigated neural representations of primary and monetary rewards (e.g., Cador et al., 1989; Belova et al., 2007; Abler et al., 2009; Abe and Lee, 2011). Here, having previously demonstrated a behavioral preference for instrumental divergence, a formal index of flexible instrumental control, we explored neural substrates mediating the influence of this information theoretic variable on economic choice. Specifically, participants were scanned with fMRI as they chose between gambling rooms that differed with respect to divergence, expected monetary payoffs, and free versus forced choice. Using a model-based analysis, we found that activity in the RLPFC and vmPFC scaled with a divergence-based measure of expected value (*IDEV*) that reflected both the level of divergence and monetary payoffs, but only for (Hi_ID) choice scenarios in which differences in divergence across rooms reflected differences in instrumental control.

It is worth reflecting on why neural activity would be selectively modulated by the *IDEV* variable in Hi_ID choice scenarios, particularly since, as shown in Figure 2B, behavioral choice preferences were well predicted by the *IDEV* model across Hi_ID and No_ID choice scenarios, and by both the *IDEV* and *SEV* model in No_ID choice scenarios. First, recall that the free parameter indicating how much a given participant values high divergence, w , was fit separately across Hi_ID and No_ID conditions, and was significantly lower for the No_ID condition, reducing *IDEV* to *SEV*. In such cases, behavioral sensitivity to variations in *IDEV* in the No_ID condition simply reflects sensitivity to *SEV*. More broadly speaking, whereas in the Hi_ID condition, high divergence reflects the potential consequences of intentional choices, in the No_ID condition, high divergence reflects the diversity of perceptual outcomes, a variable that is completely divorced from the participant's slot machine selection. Thus, while these distinct constructs may each impact choice preferences, they should nonetheless be expected to have distinct neural signatures.

We found a strong modulation of activity in the right RLPFC by the difference in IDEVs across room options, specific to Hi_ID choice scenarios. Previous work has implicated the RLPFC in exploratory behavior (Daw et al., 2006; Boorman et al., 2009; Badre et al., 2012; Zajkowski et al., 2017). For example, Zajkowski et al. (2017) found that inhibition of the right RLPFC by theta-burst TMS significantly impaired directed exploration, reflecting active information seeking, but not random exploration, driven by decision noise. Specifically, following a set of forced-choice trials designed to provide partial information about the probabilistic payoffs of two one-armed bandits, RLPFC inhibition (relative to vertex) decreased free-choice exploration of a less sampled, and thus high-information, bandit, but not of an equally sampled bandit. As with the selection of a high-divergence self-play gambling room in the current study, such directed exploration reflects a preference for options associated with consequential voluntary choice.

However, the RLPFC has also been implicated in a range of higher-level cognitive processes less obviously related to those addressed by the current study, including subgoal management and task sequencing (Braver and Bongiolatti, 2002; Desrochers et al., 2015, 2019), relational reasoning and rule induction (Strange et al., 2001; Bunge et al., 2009; Davis et al., 2017), and episodic memory encoding and retrieval (Grasby et al., 1993; Shallice et al., 1994). Christoff et al. (2003) argued that a common thread among these apparently disparate tasks is the processing of self-generated information that is not apparent in the external environment, but must be inferred or otherwise internally generated. Using a task in which the task-relevant dimension in a target sample had to be inferred or was explicitly stated, they found that RLPFC activity selectively increased during the evaluation of inferred stimulus dimensions, regardless of cognitive load, and concluded that the RLPFC specifically implements the evaluation of self-generated products of reasoning, planning, and long-term memory retrieval. Further work is needed to determine whether the currently observed modulation of RLPFC activity, by a utility signal that incorporates both instrumental divergence and monetary payoffs, reflects a general involvement in higher-order cognitive integration, or a more specific contribution to the representation of agency.

As with the RLPFC, we found that the difference in IDEVs across room options modulated activity in the vmPFC. Considerable evidence from neurophysiological and neuroimaging studies suggests that the vmPFC encodes the subjective values of primary rewards, such as tastes and odors (Anderson et al., 2003; Rolls et al., 2003; Small et al., 2003), as well as pleasant visual stimuli, including the attractiveness of faces or pictorial scenes (O'Doherty et al., 2003; Kirk et al., 2009), and more abstract goods, such as social praise (Elliott et al., 1997) and monetary gain (O'Doherty et al., 2001). Two notable features of the vmPFC shed important light on the current results: First, value encoding in the vmPFC appears to be relative, such that the value signal for a particular stimulus depends on the values of other, proximal, stimuli (O'Doherty, 2011). One might expect, thus, that the vmPFC signal would respond most clearly to a difference in value between concurrently available stimuli. Second, recent findings suggest that the vmPFC encodes stimulus values that are independent of the particular stimulus category, essentially implementing a common neural value scale for different types of goods (Chib et al., 2009; McNamee et al., 2013). The currently demonstrated value signal in the vmPFC, corresponding to a difference between options in divergence-based utility, suggest that this common value scale can be extended to a relative analysis of exceedingly abstract concepts.

Our previous work has implicated the rSMG of the inferior parietal lobule in encoding instrumental divergence. Specifically, using a simple value-based decision-making task, Liljeholm et al. (2013) found that activity in the rSMG scaled parametrically with trial-by-trial estimates of instrumental divergence, and that this signal was dissociable from other information theoretic and motivational variables, including outcome entropy and expected utility. In a subsequent task, aimed at assessing neural substrates mediating the acquisition of goal-directed versus habitual instrumental behavior, Liljeholm et al. (2015) found that activity in the rSMG increased across blocks of instrumental acquisition in a high-divergence, but not in a zero-divergence, condition. Moreover, in a subsequent test, the degree to which rSMG activity discriminated between high- and zero-divergence conditions predicted the degree to which those conditions generated different levels of outcome devaluation sensitivity, a standard measure of goal-directedness. In the current study, while we did not replicate a main effect of instrumental divergence in the rSMG, we did find that modulation of rSMG activity by a divergence-based utility measure predicted the degree to which instrumental divergence influenced participants' preferences for greater monetary payoffs. Together, these results suggest that the rSMG mediates an influence of instrumental divergence on goal-directed behavior.

In conclusion, we have used model-based fMRI to investigate the neural computations mediating a behavioral preference for instrumental divergence. We found that activity in the RLPFC and vmPFC was significantly modulated by a variant of expected value that reflected both instrumental divergence and monetary payoffs, but not by a conventional model of expected value based solely on monetary gain, and that activity in the rSMG and vmPFC predicted the degree to which instrumental divergence influenced participants' preferences for greater monetary payoffs. The recently demonstrated influence of instrumental divergence on economic choice behavior suggests that this variable has affective properties, whether acquired through experience or conferred by a selective advantage. The current work breaks new ground by extending the neuroscientific study of agency to its potential role as a motivational decision variable. Our results contribute to a growing literature on the neural integration of cognitive and affective processes.

References

- Abe H, Lee D (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70:731–741.
- Abler B, Herrnberger B, Gron G, Spitzer M (2009) From uncertainty to reward: BOLD characteristics differentiate signaling pathways. *BMC Neurosci* 10:154.
- Anderson AK, Christoff K, Stappen I, Panitz D, Ghahremani DG, Glover G, Gabrieli JD, Sobel N (2003) Dissociated neural representations of intensity and valence in human olfaction. *Nat Neurosci* 6:196–202.
- Ayal S, Zakay D (2009) The perceived diversity heuristic: the case of pseudo-diversity. *J Pers Soc Psychol* 96:559–573.
- Badre D, Doll BB, Long NM, Frank MJ (2012) Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron* 73:595–607.
- Belova MA, Paton JJ, Morrison SE, Salzman CD (2007) Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* 55:970–984.
- Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62:733–743.
- Braver TS, Bongiolatti SR (2002) The role of frontopolar cortex in subgoal processing during working memory. *Neuroimage* 15:523–536.

- Bunge SA, Helskog EH, Wendelken C (2009) Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46:338–342.
- Cador M, Robbins TW, Everitt BJ (1989) Involvement of the amygdala in stimulus-reward associations: interaction with the ventral striatum. *Neuroscience* 30:77–86.
- Chib VS, Rangel A, Shimojo S, O'Doherty JP (2009) Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J Neurosci* 29:12315–12320.
- Christoff K, Ream JM, Geddes L, Gabrieli JD (2003) Evaluating self-generated information: anterior prefrontal contributions to human cognition. *Behav Neurosci* 117:1161–1168.
- Davis T, Goldwater M, Giron J (2017) From concrete examples to abstract relations: the rostrolateral prefrontal cortex integrates novel examples into relational categories. *Cereb Cortex* 27:2652–2670.
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876–879.
- Desrochers TM, Chatham CH, Badre D (2015) The necessity of rostrolateral prefrontal cortex for higher-level sequential behavior. *Neuron* 87:1357–1368.
- Desrochers TM, Collins AG, Badre D (2019) Sequential control underlies robust ramping dynamics in the rostrolateral prefrontal cortex. *J Neurosci* 39:1471–1483.
- Elliott R, Frith CD, Dolan RJ (1997) Differential neural response to positive and negative feedback in planning and guessing tasks. *Neuropsychologia* 35:1395–1404.
- Kirk U, Skov M, Hulme O, Christensen MS, Zeki S (2009) Modulation of aesthetic value by semantic context: an fMRI study. *Neuroimage* 44:1125–1132.
- Grasby PM, Frith CD, Friston KJ, Bench CR, Frackowiak RS, Dolan RJ (1993) Functional mapping of brain areas implicated in auditory-verbal memory function. *Brain* 116:1–20.
- Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MF, Behrens TE (2012) Mechanisms underlying cortical activity during value-guided choice. *Nat Neurosci* 15:470.
- Leotti LA, Delgado MR (2011) The inherent reward of choice. *Psychol Sci* 22:1310–1318.
- Leotti LA, Delgado MR (2014) The value of exercising control over monetary gains and losses. *Psychol Sci* 25:596–604.
- Liljeholm M, Wang S, Zhang J, O'Doherty JP (2013) Neural correlates of the divergence of instrumental probability distributions. *J Neurosci* 33:12519–12527.
- Liljeholm M, Dunne S, O'Doherty JP (2015) Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control. *Eur J Neurosci* 41:1358–1371.
- Liljeholm M, Mistry P, Koh S (2018) The influence of schizotypal traits on the preference for high instrumental divergence. Proceedings of the 40th Annual Meeting of the Cognitive Science Society, pp 2053–2058. Madison, WI: Cognitive Science Society.
- Liljeholm M (2018) Instrumental divergence and goal-directed choice. *Goal Directed Decision Making* (Morris R, Bornstein A, Shenhav A eds), Ed 1, pp 27–48. San Diego: Academic Press.
- McNamee D, Rangel A, O'Doherty JP (2013) Category-dependent and category-independent goal value codes in human ventromedial prefrontal cortex. *Nat Neurosci* 16:479–485.
- Mistry P, Liljeholm M (2016) Instrumental divergence and the value of control. *Sci Rep* 6:36295.
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping* 15:1–25.
- O'Doherty JP (2011) Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Ann NY Acad Sci* 1239:118–129.
- O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* 4:95–102.
- O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ (2003) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147–155.
- Shallice T, Fletcher P, Frith CD, Grasby P, Frackowiak RS, Dolan RJ (1994) Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature* 368:633–635.
- Strange BA, Henson RN, Friston KJ, Dolan RJ (2001) Anterior prefrontal cortex mediates rule learning in humans. *Cereb Cortex* 11:1040–1046.
- Rolls ET, Kringelbach ML, De Araujo IE (2003) Different representations of pleasant and unpleasant odours in the human brain. *Eur J Neurosci* 18:695–703.
- Small DM, Gregory MD, Mak YE, Gitelman D, Mesulam MM, Parrish T (2003) Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39:701–711.
- Wunderlich K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. *Proc Natl Acad Sci USA* 106:17199–17204.
- Zajkowski WK, Kossut M, Wilson RC (2017) A causal role for right frontopolar cortex in directed, but not random, exploration. *Elife* 6:e27430.