NeuroImage 215 (2020) 116834

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Goal-oriented and habitual decisions: Neural signatures of model-based and model-free learning

Yi Huang [a], Zachary A. Yaple [b], Rongjun Yu [a,b,*]

[a] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore
[b] Department of Psychology, National University of Singapore, Singapore

A B S T R A C T

Human decision-making is mainly driven by two fundamental learning processes: a slow, deliberative, goal-directed model-based process that maps out the potential outcomes of all options and a rapid habitual model-free process that enables reflexive repetition of previously successful choices. Although many model-informed neuroimaging studies have examined the neural correlates of model-based and model-free learning, the concordant activity among these two processes remains unclear. We used quantitative meta-analyses of functional magnetic resonance imaging experiments to identify the concordant activity pertaining to model-based and model-free learning over a range of reward-related paradigms. We found that: 1) both processes yielded concordant ventral striatum activity, 2) model-based learning activated the medial prefrontal cortex and orbital frontal cortex, and 3) model-free learning specifically activated the left globus pallidus and right caudate head. Our findings suggest that model-free and model-based decision making engage overlapping yet distinct neural regions. These stereotaxic maps improve our understanding of how deliberative goal-directed and reflexive habitual learning are implemented in the brain.

1. Introduction

Human decision-making is influenced by both habitual and goal-directed processes. A large number of psychological studies in both animals and humans have provided support for these two distinct strategies (Daw and O'Doherty, 2013). Within the context of reward processing, a slow, deliberative, goal-directed process compares the potential outcomes of each action and identifies the action most likely to generate a desired outcome. In contrast, an automatic and rapid habitual process links reward to associated action and enables reflexive repetition of previously successful choices (Balleine and O'Doherty, 2010; Daw et al., 2005; Dickinson, 1985; Doll et al., 2012). Recent research employing the computational framework has greatly elucidated the underlying process of reward learning.

Traditionally, two classes of reinforcement learning (RL) models have been proposed to capture the key behavior patterns of the two learning strategies (Daw et al., 2005). 'Model-free' models utilize trial-and-error feedback to update an action value associated with a stimulus. This learning scheme promotes the execution of experienced behavior with little effort. On the other hand, 'model-based' models make decisions through a flexible, but computationally demanding process by ascribing a "decision-tree", comprised of associations between state transitions and outcomes. Model-based learning can be used adaptively to compute ideal action by simulating their outcomes, enabling flexible behaviors in dynamic situations.

Functional magnetic resonance imaging (fMRI) studies have shed light on dissociating the neural mechanisms of model-free and model-based value signals, by using learning tasks inspired by the computational reinforcement learning literature (Sutton and Barto, 1998). For example, the most commonly used tasks are sequential decision tasks, such as mazes or more abstract multistep action-outcome tasks (Daw et al., 2011). Regarding the classic two-step task (see Fig. 1), at the first stage (State A), participants are required to choose between two options (A1 vs. A2). This decision determines the transition to one of two subsequent states (State B and State C) via a common (e.g., 70% possibility) or a rare (e.g., 30% possibility) transition. At the second stage, participants are asked to make another decision between two options (B1 vs. B2 in State B or C1 vs. C2 in State C) and are then provided with a reward, determined by stochastic randomization. Another, less commonly used paradigm involves explicit or implicit counterfactual structure, in which the rewards are not actually received but can be inferred or observed. A typical example is the serial reversal contingency task in which the value decreases in one option, implying an increase in value for the alternative option (Doll et al., 2012; Worthy et al., 2016).

Using these types of tasks, model-free and model-based RL can be distinguished by the pattern of staying or switching of the first stage choice following the rewards of the second stage. The RL algorithm would classify a person as using the model-free approach if the learner

* Corresponding author. Department of Psychology, National University of Singapore, Block AS4, #02-17, 9 Arts Link, Singapore, 117570.
E-mail address: psyyr@nus.edu.sg (R. Yu).

https://doi.org/10.1016/j.neuroimage.2020.116834
Received 26 August 2019; Received in revised form 3 March 2020; Accepted 8 April 2020
Available online 10 April 2020

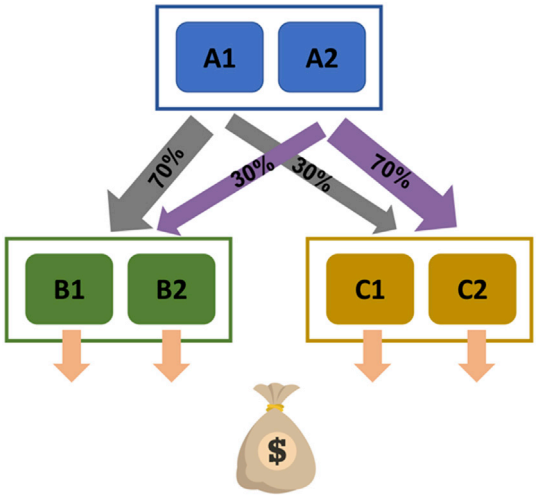

Fig. 1. Illustration of the classic two-step task. At the first stage (State A), participants are required to choose between two options (A1/A2). This decision determines the transition to one of two subsequent states (State B and C) via a common (e.g., 70% possibility) or a rare (e.g., 30% possibility) transition. At the second stage, participants are asked to make another decision between two options (B1/B2 in State B, C1/C2 in State C) and then are given stochastically rewards.

tends to repeat a rewarded action without considering whether the reward occurred after a common or a rare transition. On the other hand, the RL algorithm may classify data as model-based if their actions correspond with a decision tree, such that reward following a rare transition actually increases the value of the unchosen option and thus predicts switching. For the most part, human subjects typically demonstrate a mixture of both model-free and model-based learning strategies in this task (Daw et al., 2011; Decker et al., 2016; Deserno et al., 2015a).

With respect to cognitive neuroscience, studies on instrumental learning have shown that model-free learning depends on the dopamine-rich striatum (Daw et al., 2005; Glascher et al., 2010). Surprisingly, signatures from model-based computations seem to be pervasive in the ventral striatum (VS), which were previously thought to support model-free learning (Daw et al., 2011; Kroemer et al., 2019; Nebe et al., 2018). The most commonly reported brain regions that encode model-based value signals are ventromedial prefrontal cortex (vmPFC) and adjacent orbitofrontal cortex (OFC), which have been demonstrated in goal-directed behavior with human subjects as well as animal specimen (Balleine and O'Doherty, 2010; de Wit et al., 2009; Padoa-Schioppa and Assad, 2006; Valentin et al., 2007). Moreover, in model-based RL, subjects anticipate and memorize the future states of outcomes rather than simply recall the immediate reward. Hippocampus, the key region in memory formation and future planning (Schacter et al., 2007), could be another candidate region involved in model-based learning. Human lesion studies, as well as animal studies, have provided causal evidence directly linking the hippocampus with model-based planning behavior and spatial memory (Miller et al., 2017; Stoianov et al., 2018; van der Meer et al., 2010). Several human fMRI studies also reported the activation of the hippocampus in the model-based learning process (Bornstein and Daw, 2012; Sebold et al., 2017). Taken together, previous research pertaining to the neural signatures of model-free and model-based learning has yielded inconsistent findings.

Here, we performed quantitative whole-brain meta-analyses on neural representations of model-free and model-based learning using the activation likelihood estimation method. Our goal was to assess the concordance of the reward and executive control regions among studies using the model-free vs. model-based task. To this end, we aimed to create and compare stereotaxic maps for model-free and model-based learning to assess which events comprise of striatum and prefrontal cortex activity. We expected to reveal striatum activity for the meta-analysis associated with both types of learning yet more striatum activity for the meta-analysis associated with model-free, confirming the notion that model-free is more striatum based. Therefore, we hypothesize that regions associated with planning and memory, such as the hippocampus, would be involved specifically when participants engage in model-based learning. Together, these comparisons allowed us to confirm the neural networks involving model-free and model-based processing.

2. Material and methods

2.1. Literature search and article selection

Our search was performed in PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and Web of Science (https://www.webofknowledge.com) using the keywords: "model-free" AND "model-based" AND "fMRI" on 6th March 2019. In addition, the references of the included studies, relevant review articles and articles searched on Google Scholar were checked for additional relevant studies. This search yielded 79 unique articles in total, and they were screened for eligibility. Articles were considered eligible if they included whole-brain data with random-effects analysis, reported coordinates (i.e., foci) in Talairach or Montreal Neurology Institute (MNI) space for model-free and pure model-based signals separately (or conjunction) and included human as subjects. The final dataset included 21 eligible articles and 22 experiments (one article includes two different samples; (Worthy et al., 2016). Fig. 2 displays a flowchart representing the steps taken to screen and identify eligible articles. A summary of the studies used in the present meta-analysis is shown in Table 1. All coordinates were transformed into the same MNI space.

2.2. ALE methodology

Two separate meta-analyses for model-free and model-based signals were performed. To perform the meta-analyses, we used GingerALE version 3.0. (http://brainmap.org), a freely available, quantitative meta-analysis method first developed by Turkeltaub et al. (2002), and further developed by Eickhoff et al. (2017) and Turkeltaub et al. (2012). GingerALE relies on activation likelihood estimation (ALE) algorithm, which aims at identifying significant overlapping clusters of activation across studies. The most recent algorithm minimizes within-group effects and provides increased power by allowing for the inclusion of all possible relevant experiments (Eickhoff et al., 2017; Turkeltaub et al., 2012). Statistical maps were thresholded at $p < 0.05$ using a cluster-level correction for multiple comparisons (5000 permutations) and a cluster forming threshold at $p < 0.001$ (Eickhoff et al., 2017). To compare the results of each meta-analysis representing model-free and model-based signals, conjunction and contrast analyses for model-free and model-based were conducted to assess areas consistently activated across both maps and specify activations unique to each map, respectively. Conjunction analyses were conducted according to Eickhoff et al. (2009), whereby the voxel minimum value model-based and model-free thresholded ALE images were used to produce a conjunction image. Similarly, contrast images were produced by subtracting each thresholded ALE map from one another. The ALE conjunction analysis revealed significant clusters of convergence between two conditions. ALE contrast analyses revealed specific activation for model-based > model-free and model--free > model-based learning. Conjunction and contrast analyses clusters were thresholded at uncorrected $P < 0.05$. The conjunction/contrast image is also a thresholded ALE image because conjunction and contrast analyses were based on two thresholded ALE images.

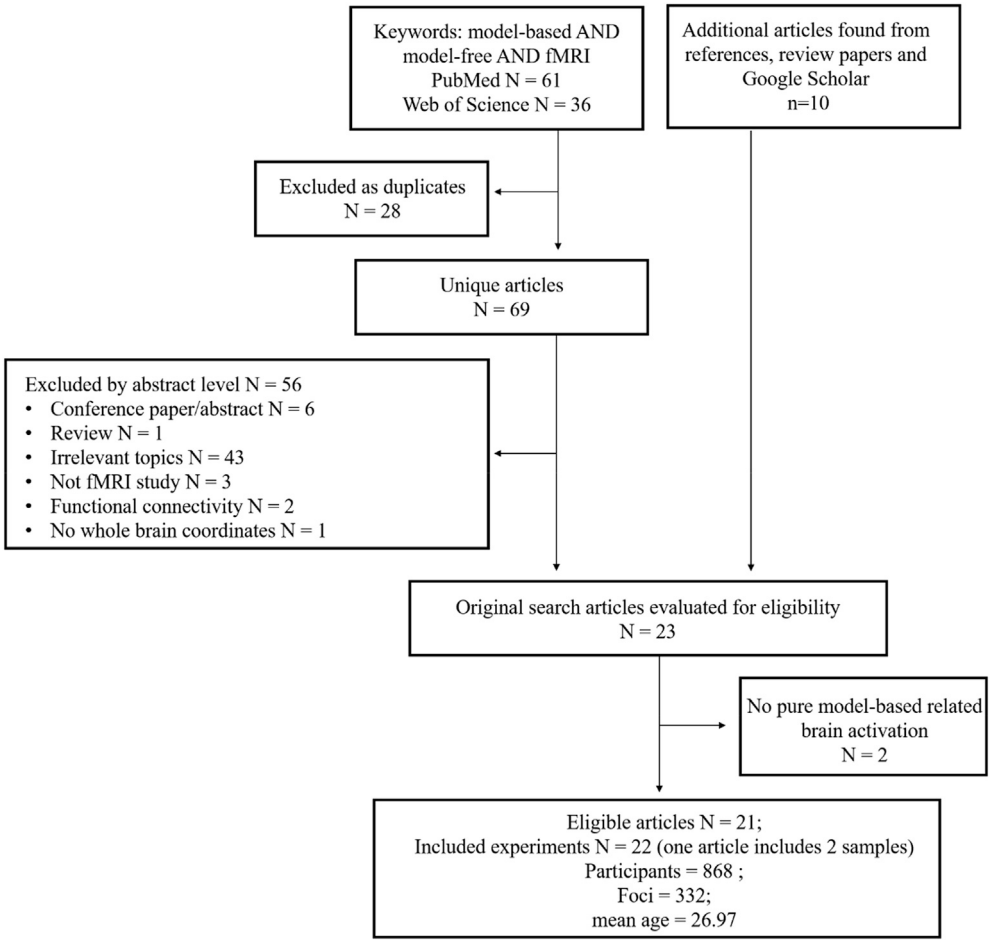

Fig. 2. PRISMA flowchart illustrating exclusion criteria and eligibility for relevant studies.

3. Results

Articles included in the meta-analyses of exploration reported data on 868 participants, yielding a total of 332 foci. Fig. 2 reveals the number of articles, number of experiments, and the number of foci included in the meta-analysis. All the papers used in the meta-analysis are indicated by an asterisk in the reference list. Contributing experiments to the individual clusters for the meta-analysis were listed in the Supplementary material. Significant results for each meta-analysis, conjunction, and contrast analyses are displayed in Fig. 3.

3.1. ALE map

The meta-analysis associated with model-free signals revealed large clusters within bilateral ventral striatum and bilateral globus pallidus (see Fig. 3A), supporting the hypothesis that model-free recruits the striatum. Interestingly, the meta-analysis of model-based learning produced bilateral ventral striatum as well as anterior cingulate (Brodmann area [BA] 24) and anterior/medial prefrontal cortices (BA 9/32), indicating that model-based processing is not limited to the striatum (Fig. 3B). Table 2 shows a list of all regions concordant across studies for model-free and model-based learning. Larger clusters that include several peaks of activation may include several peak activations, listed separately.

3.2. Conjunction and contrast maps

Conjunction analysis had shown significant concordant activation within the bilateral ventral striatum for both model-free and model-based processes (Fig. 3C), confirming the notion that ventral striatum is crucial for both types of learning. Contrasts analyses revealed that relative to model-based, model-free learning significantly activated globus pallidus, superior temporal gyrus (BA 34) and caudate head (Fig. 3C). Contrast analysis of model-based > model-free produced vmPFC (BA 9) and anterior cingulate (BA 25; see Fig. 3D). These findings suggest that model-free strategies selectively recruit caudate, globus pallidus, and superior temporal gyrus, whereas model-based processing is more associated with anterior cingulate and vmPFC, despite having striatum clusters as demonstrated by the single meta-analysis on model-based processing.

3.3. Post-hoc meta-analyses

In addition to the main analyses, an additional meta-analysis was performed for those with only healthy young drug-free subjects. In general, the results are consistent with the findings reported above. A list of regions concordant for model-free and model-based in healthy young drug-free adults are shown in Supplementary Table 1. To test whether different tasks/paradigms influence our results, we limited our analysis to homogenous subgroups. The most commonly used tasks are sequential decision tasks, such as mazes or more abstract multistep action-outcome tasks (e.g., two-step and maze, type 1). Another study involves a task with a counterfactual structure (e.g., state-maximizing and multi-armed bandit, type 2, see Table 1) (Doll et al., 2012). Similar results were found with only type 1 tasks, comprised of 14 studies (see Supplementary Table 2).

Table 1
Information on source datasets for studies on model-free and model-based decisions.

Article	Sample size	Age (SD)	Paradigm	Type[b]	Contrasts	Participants (conditions)	Foci (MF/MB)
Deserno et al. (2015a)	29	28.3 (4.95)	two-step	1	Parametric Model-free/-based PE	Healthy adults	7/4
Daw et al. (2011)	17	25.8	two-step	1	Parametric Model-free/-based PE	Healthy adults	3/3
Deserno et al. (2015b)	50	27.4 (3.7)	two-step	1	Parametric Model-free/-based PE	Across adults with high- and low-impulsive traits	18[#]
Doll et al. (2015)	20	23.8 (4.6)	two-step	1	Parametric Model-free/-based PE	Healthy adults	7/2
Glascher et al. (2010)	18	24 (7.6)	sequential two-choice Markov decision	1	Parametric reward/state PE	Healthy adults	1/4
Kroemer et al. (2019)	61	37 (3.6)	two-step	1	Parametric Model-free/-based PE	Across conditions with L-DOPA and PLC treatments	8/4
Lee et al. (2014)	22	28	sequential two-choice Markov decision	1	Parametric reward/state PE	Healthy adults	11/8
McNamee et al. (2015)	19	22.9 (4.1)	binary decision	2	Multivariate pattern analysis	Healthy adults	3/8
Reiter et al. (2017)	42	28.4 (7.3)	Anti-correlated decision-making	2	Parametric Single-update/double-update PE	Across Binge Eating disorder and healthy control	28/4
Worthy et al. (2016)[a]	18 18	23.61 61	state-maximizing	2	Parametric reward/state PE	Younger adults Older adults	1/7 6/1
Wunderlich et al. (2012)	21	N.A.	sequential two-choice Markov decision	1	Parametric Planned values vs. RPE	Healthy adults	7/20
Dunne et al. (2016)	17	23.3 (3.6)	multi-armed bandit	2	Parametric reward/state PE	Healthy adults	3/4
Fermin et al. (2016)	18	26 (5.0)	grid-sailing	1	The contrast between MF and MB conditions	Healthy adults	2/1
Nebe et al. (2018)	188	18	two-step	1	Parametric Model-free/-based PE	Social drinker	35/11
Simon and Daw (2011)	18	N.A.	continuous spatial navigation task	1	Parametric Temporal difference vs. planned values	Healthy adults	1/7
Sebold et al. (2017)	186	44.5 (10.8)	two-step	1	Parametric Model-free/-based PE	Across alcohol-dependent adults and healthy control	6[#]
Bornstein and Daw (2012)	18	25	serial reaction time (SRT)	2	Parametric Forward entropy vs. reward PE	Healthy adults	6/2
Bornstein and Daw (2013)	17	28	Serial reaction time (SRT)	2	Parametric Forward entropy vs. reward PE	Healthy adults	17/22
Anggraini et al. (2018)	27	N.A.	wayfinding (spatial navigation)	1	Parametric Model-free/-based PE	Healthy adults	16/11
Beierholm et al. (2011)	23	N.A.	door ranking	1	Parametric Reward/Bayesian PE	Healthy adults	11/9
Wimmer et al. (2012)	21	19.3	modified four-armed bandit	2	Parametric Model-free/-based PE	Healthy adults	2/1

Note: SD = Standard deviation; N.A. = not available; L-DOPA: a treatment used to increase dopamine; PE = prediction errors; PLC = placebo; MF = model-free learning; MB = model-based learning; [a] study includes two samples; [b] Type 1: sequential decision-making tasks. Type 2: task involving counterfactual structure; [#]report conjunction coordinates only. All the papers used in the meta-analysis are indicated by an asterisk in the reference list.

4. Discussion

Computational models of reinforcement learning posit that there are two types of learning: model-free learning that updates expectations based on past rewards and the model-based learning that maps possible actions to their potential outcomes (Daw et al., 2005). Although several brain regions are implicated in reward-based learning, the exact mapping between brain regions and functions in reinforcement learning remains unclear. Specifically, it is still debatable whether model-based learning also recruits ventral striatum. It is also unclear whether the hippocampus is engaged in model-based processing across tasks. The present study uses a meta-analytic approach to summarize the results of various studies on model-free and model-based learning processes in fMRI. Our results demonstrated that both model-based and model-free learning activated bilateral ventral striatum. In addition, we showed that the meta-analysis of model-based learning yield executive control regions such as the prefrontal cortex and cingulate cortex. Moreover, we showed that striatal activity differentiates model-free and model-based learning, such that left globus pallidus and right caudate head were more concordant for model-free learning.

4.1. Striatal signals for both model-free and model-based processing

Our results confirmed the important role of the ventral striatum in processing model-free information. The ventral striatum is thought to supply a common-currency reward prediction to midbrain dopamine neurons that compute differences between predicted and received outcomes (Garrison et al., 2013; Schultz, 2001). Not surprising, the ventral striatum may encode reward prediction errors during model-free processing. Critically, we provide confirming evidence that the ventral striatum also signals model-based information. Findings from animal behavior also implicate the ventral striatum in the use of model-based representations (Han et al., 2016). Tasks based on the specific features of rewards, such as reinforcer devaluation and Pavlovian to instrumental transfer, would also require a model-based representation of information (Corbit and Balleine, 2011; Corbit et al., 2001). Interindividual variability in ventral striatal presynaptic dopamine reflects a balance in the behavioral expression and the neural signatures of model-free and model-based control (Daw et al., 2011; Deserno et al., 2015a; Sebold et al., 2017). Other possible explanations relate to ventral striatal presynaptic dopamine levels which are positively associated with the coding

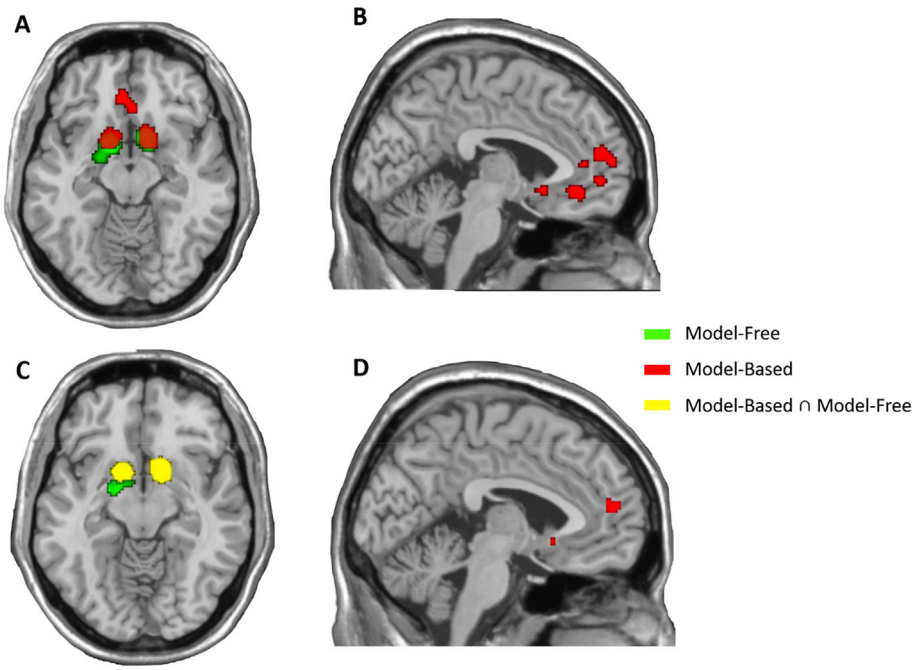

Fig. 3. Concordant activation across studies for model-based and model-free processing. **A&B**: regions concordant across studies for model-free (in green) and model-based learning (in red). **C**: the bilateral ventral striatum for both model-free and model-based processes and the globus pallidus for model-free learning; **D**: anterior cingulate and vmPFC for model-based learning. Displayed are significant results from the meta-analysis of model-based learning, model-free learning and their conjunction/contrast.

Table 2
Significant regions of activation for model-free and model-based decisions.

Cluster	Cluster Size (mm3)	**Brain Regions (Peak)**	BA	x	y	z	ALE	Z value
Model-free								
1	3032	L Globus Pallidus		−12	6	−10	0.048	8.08
		L Amygdala		−20	−4	−14	0.018	4.22
		L Amygdala		−24	−4	−16	0.016	3.94
2	2680	R Caudate Head		12	10	−8	0.058	Infinity
Model-based								
1	4640	R Caudate Head		12	10	−8	0.052	Infinity
		L Caudate Head		−12	10	−8	0.037	7.12
2	1144	Dorsal Anterior Cingulate		−6	38	−8	0.018	4.54
		Ventral Anterior Cingulate	24	2	30	−14	0.015	3.95
3	1088	Anterior Prefrontal Cortex	9	−2	54	16	0.017	4.31
		Medial Frontal Gyrus	9	−8	60	10	0.015	4.02
		Dorsal Anterior Cingulate	32	−2	44	8	0.012	3.52
4	760	Medial Frontal Cortex	32	2	52	0	0.018	4.49
Model-free ∩ Model-based								
1	2096	Caudate Head		12	10	−8	0.052	N.A.
2	1392	Lateral Globus Pallidus		−12	8	−8	0.036	N.A.
Model-free > Model-based								
1	752	L Lateral Globus Pallidus		−18.1	−2.4	−9.6	N.A.	2.58
		Medial Globus Pallidus		−9	0	−12	N.A.	2.33
		R Superior Temporal Gyrus	34	−22	0	−14	N.A.	2.05
2	16	Caudate Head		4	10	−7	N.A.	1.64
Model-based > Model-free								
1	232	Medial Frontal Gyrus	9	−6	53	13	N.A.	1.88
			9	−6	54	18	N.A.	1.88
			9	−0.3	50.3	13.7	N.A.	1.64
2	24	Anterior Cingulate	25	4	20	−8	N.A.	1.75
3	16	Anterior Cingulate		−4	16	−7	N.A.	1.64
4	16	Anterior Cingulate		8	22	−6	N.A.	1.88
5	16	Dorsal Anterior Cingulate		9	20	−8	N.A.	1.75

Note: ALE: activation likelihood estimation; L: left; R: right; BA: Brodmann area.

of model-based signatures in lateral PFC, accompanied by a bias toward more model-based choices (Deserno et al., 2015a; Lee et al., 2014). Individuals with high ventral striatum presynaptic dopamine are also characterized by a diminished coding of ventral striatal model-free prediction errors (Daw et al., 2005). Ventral striatum may also be involved in the arbitration that chooses between the two strategies (Deserno et al., 2015a). The abovementioned findings may shed light on the specific role of the striatum during model-based processing. In addition, it has also been hypothesized that flexible, model-based choices may be accomplished using computations that are homologous to those used in model-free learning (Russek et al., 2017). The common activity in the striatum provides neural evidence to support the view that model-based choices can arise from the same dopaminergic-striatal circuitry that carries out model-free learning. Our results highlight the role of the ventral striatum in both types of learning and challenge the view that ventral striatum encodes only pure model-free signals.

4.2. Model-free learning specificity

We found that model-free learning is also associated with caudate and globus pallidus. The globus pallidus is a major component of the basal ganglia, with main inputs from the dorsal striatum, and direct outputs to the substantia nigra (Difiglia et al., 1982; Gillies et al., 2017). A previous meta-analysis showed that prediction errors in reversal learning tasks activate the caudate body and lateral globus pallidus across studies (Yaple and Yu, 2019). The basal ganglia are dominantly involved in prediction error in temporal-difference (TD) learning accounts of the dopaminergic system (Schultz et al., 1997). For example, in the actor-critic framework, the prediction error signals in basal ganglia drive individuals to take actions that are followed by a reward. Our findings suggest that model-free processing is not only limited to the ventral striatum but is extended to other reward-related brain regions, e.g., the caudate and globus pallidus.

4.3. Model-based learning specificity

Using model-based representations, humans and animals form mental representations of their environment. These representations contain information about what actions to take and consequentially which outcomes are expected. When an individual is confronted with one event, one can use these representations to look forward and predict specific features of the upcoming event, such as its timing, probability, and expected utility. Model-based processing is cognitively demanding and has been proposed to engage subregions of the prefrontal cortex. Specifically, our analysis revealed that model-based learning is associated with activity in the vmPFC. Animal research has shown that goal-directed control is associated with vmPFC, using reinforced devaluation and contingency manipulation (Balleine and O'Doherty, 2010; de Wit et al., 2009; Valentin et al., 2007). Human neuroimaging studies using similar paradigms also consistently identify the vmPFC in model-based inference (de Wit et al., 2009; O'Doherty, 2011; Tanaka et al., 2008; Valentin et al., 2007). It is worth noting that our analysis did not include all studies investigating goal-directed behaviors. Rather, we limited our analysis to include studies that specify model-based vs. model-free computations. Further studies may further compare model-based decision making with other types of goal-directed decision making.

4.4. Multiple systems of model-based and model-free learning

Although the activation of the hippocampus in the model-based learning process has been reported (Bornstein and Daw, 2012, 2013; Sebold et al., 2017), our meta-analysis revealed no such concordance. One possibility is that hippocampus is only involved in processing the long-lasting effects of transitions on learning and hence, model-based learning may not require the hippocampus-based memory system. Another possibility is that the standard two-step task is rather abstract with no explicit spatial association between stages (Daw et al., 2011). A newer task design with an explicit, spatial relation between stages may activate the hippocampus. Vikbladh et al. (2019) showed that both model-based planning and boundary-driven place memory share a common mechanism, which is affected in epilepsy patients treated using unilateral anterior temporal lobectomy with hippocampal resection (Vikbladh et al., 2019). The spatial association in this task may relate to stronger hippocampal contribution to model-based planning. Future research may use other tasks to pinpoint the multiple neural systems that contribute to model-based learning in different contexts.

4.5. Limitations

Although our study narrowly focuses on distinguishing model-based vs. model-free learning, this distinction parallels other synonymous terms, such as explicit vs. implicit (Seger and Miller, 2010), habit learning vs. episodic memory (Gershman and Daw, 2017), and unconscious vs. conscious learning (Duss et al., 2014). These taxonomies resemble model-based vs. model-free differentiation. Nevertheless, each term describes some unique features of learning, which is beyond the scope of the current research. The two learning systems may also interact with each other, such that model-free signals are informed by task structure, reflecting 'model-based' value estimates, whereas the model-based system may receive inputs from the model-free system as a means to optimize future choices. The meta-reinforcement learning framework posits that the dopamine system within the striatum trains the PFC to operate as its own free-standing learning system (Wang et al., 2018). It is also worth noting that model-based vs. model-free taxonomy of learning is not the only characterization to describe complex choices. Several control algorithms with distinct features have also been proposed to understand decision making such as defensive behaviour (Bach and Dayan, 2017; LeDoux and Daw, 2018). How the model-based and model-free systems interact with other control systems is an intriguing question for future research.

Moreover, model-based and model-free processes are naturally imbedded in many other learning tasks, including the Wisconsin Card Sorting Test (Glascher et al., 2019), reversal learning tasks (Yaple and Yu, 2019), and categorical learning tasks (Seger and Miller, 2010). However, previous neuroimaging studies using those paradigms did not provide contrasts that precisely isolated model-based vs. model-free components. Future research may expand our study by comparing model-based vs. model-free learning with other synonymous terms and comparing different paradigms in this framework.

4.6. Clinical implications

These meta-analyses performed in the current study may inform neuroimaging research in learning-related psychiatry disorders. Deficits in model-based learning have been proposed to be characteristics in addition or psychiatric disorders, including obsessive-compulsive disorder (OCD) and depression. Individuals with OCD have difficulty inhibiting an established response when faced with new contingencies, in terms of switching attention from one dimension of a stimulus to another or by suppressing or reversing to a previously rewarded response (Chamberlain et al., 2008; Remijnse et al., 2006). Individuals with OCD also show neurocognitive deficits in attentional/extra-dimensional set--shifting, affective set-shifting/reversal learning, and task shifting (Britton et al., 2010; Gruner and Pittenger, 2017; Gu et al., 2008). OCD patients rely exclusively on habit-like decision-making during reinforcement-driven learning (Gillan and Robbins, 2014). Inflexible reliance on habitual decision making in OCD may reflect a functional deficit in the mechanism associated with context-appropriate dynamic arbitration between model-free and model-based decision making. Cognitive inflexibility in OCD might result from a deficiency in mechanisms underlying goal-oriented processes or an over-reliance on mechanisms subserving habitual processes (Gillan and Robbins, 2014).

Another clinical group that may be relevant here are humans with substance dependence; those who are chronically exposed to substances and whom often have difficulty making adaptive, flexible choices (Jentsch et al., 2002; Stalnaker et al., 2009). For addicted individuals, overtraining may make performance insensitive to changes in reward value in reinforcer devaluation tasks (Balleine and Dickinson, 1998; Colwill and Rescorla, 1988; Holman, 1975; Killcross and Coutureau, 2003; Tricomi et al., 2009). A recent study with rats provides direct evidence that model-free and model-based decision-making processes are involved in distinct aspects of addiction vulnerability and pathology (Groman et al., 2019).

Using similar logic, researchers have also examined model-free vs. model-based RL in healthy subjects with depressive symptoms (Blanco et al., 2013; Otto et al., 2013; Maddox et al., 2014; Radenbach et al., 2015). For example, Blanco et al. (2013) found that depressed patients used model-based RL less often than non-depressed individuals, and data from those with higher depression scores had a better fit to the

model-free RL model. Another study showed that depressive individuals tend to have faster, more accurate and more frequent use of reflexive strategies (Maddox et al., 2014). The hypothesis that the model-based RL may be attenuated has yet to be tested by computational research in depression, although it has been studied in healthy subjects under stressful situations (Otto et al., 2013; Radenbach et al., 2015). Taken together, the examination of model-free and model-based RL strategies is not only necessary for healthy humans, but also for particular clinical populations that show difficulty in learning. Perhaps as more literature becomes available, further examination can be performed.

5. Conclusion

Our results confirm the role of the globus pallidus and striatum in model-free learning. Our findings support theories positing an important role for the vmPFC in model-based learning, while also providing a new perspective on the roles of ventral striatum by showing that ventral striatum is involved in both model-free and model-based learning. Our research applies meta-analysis on model-informed neuroimaging research, highlighting the promise of integrating computational fMRI studies to understand the neural basis of some general psychological processes.

Funding

Ministry of Health (MOH) Singapore National Medical Research Council (NMRC) (OFYIRG17may052 to [R.Y.]).

CRediT authorship contribution statement

Yi Huang: Conceptualization, Writing - original draft. **Zachary A. Yaple:** Writing - review & editing. **Rongjun Yu:** Conceptualization, Writing - review & editing, Supervision.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116834.

References

Anggraini, D., Glasauer, S., Wunderlich, K., 2018. *Neural signatures of reinforcement learning correlate with strategy adoption during spatial navigation. Sci. Rep. 8, 10110.
Bach, D.R., Dayan, P., 2017. Algorithms for survival: a comparative perspective on emotions. Nat. Rev. Neurosci. 18, 311–319.
Balleine, B.W., Dickinson, A., 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37, 407–419.
Balleine, B.W., O'Doherty, J.P., 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35, 48–69.
Beierholm, U.R., Anen, C., Quartz, S., Bossaerts, P., 2011. *Separate encoding of model-based and model-free valuations in the human brain. Neuroimage 58, 955–962.
Blanco, N.J., Otto, A.R., Maddox, W.T., Beevers, C.G., Love, B.C., 2013. The influence of depression symptoms on exploratory decision-making. Cognition 129, 563–568.
Bornstein, A.M., Daw, N.D., 2012. *Dissociating hippocampal and striatal contributions to sequential prediction learning. Eur. J. Neurosci. 35, 1011–1023.
Bornstein, A.M., Daw, N.D., 2013. *Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. PLoS Comput. Biol. 9, e1003387.
Britton, J.C., Rauch, S.L., Rosso, I.M., Killgore, W.D., Price, L.M., Ragan, J., Chosak, A., Hezel, D.M., Pine, D.S., Leibenluft, E., Pauls, D.L., Jenike, M.A., Stewart, S.E., 2010. Cognitive inflexibility and frontal-cortical activation in pediatric obsessive-compulsive disorder. J. Am. Acad. Child Adolesc. Psychiatr. 49, 944–953.
Chamberlain, S.R., Menzies, L., Hampshire, A., Suckling, J., Fineberg, N.A., del Campo, N., Aitken, M., Craig, K., Owen, A.M., Bullmore, E.T., 2008. Orbitofrontal dysfunction in patients with obsessive-compulsive disorder and their unaffected relatives. Science 321, 421–422.
Colwill, R.M., Rescorla, R.A., 1988. Associations between the discriminative stimulus and the reinforcer in instrumental learning. J. Exp. Psychol. Anim. Behav. Process. 14, 155.
Corbit, L.H., Balleine, B.W., 2011. The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. J. Neurosci. 31, 11786–11794.
Corbit, L.H., Muir, J.L., Balleine, B.W., 2001. The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. J. Neurosci. 21, 3251–3260.
Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. *Model-based influences on humans' choices and striatal prediction errors. Neuron 69, 1204–1215.
Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8, 1704.
Daw, N.D., O'Doherty, J.P., 2013. Multiple Systems for Value Learning. Neuroeconomics: Decision Making and the Brain, second ed. Elsevier Inc., pp. 393–410
de Wit, S., Corlett, P.R., Aitken, M.R., Dickinson, A., Fletcher, P.C., 2009. Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J. Neurosci. 29, 11330–11338.
Decker, J.H., Otto, A.R., Daw, N.D., Hartley, C.A., 2016. From creatures of habit to goal-directed learners: tracking the developmental emergence of model-based reinforcement learning. Psychol. Sci. 27, 848–858.
Deserno, L., Huys, Q.J., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A.A., Dolan, R.J., Heinz, A., Schlagenhauf, F., 2015a. *Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proc. Natl. Acad. Sci. Unit. States Am. 112, 1595–1600.
Deserno, L., Wilbertz, T., Reiter, A., Horstmann, A., Neumann, J., Villringer, A., Heinze, H.J., Schlagenhauf, F., 2015b. *Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. Transl. Psychiatry 5, e659.
Dickinson, A., 1985. Actions and habits: the development of behavioural autonomy. Philos. Trans. R. Soc. Lond. B Biol. Sci. 308, 67–78.
Difiglia, M., Pasik, P., Pasik, T., 1982. A Golgi and ultrastructural study of the monkey globus pallidus. J. Comp. Neurol. 212, 53–75.
Doll, B.B., Duncan, K.D., Simon, D.A., Shohamy, D., Daw, N.D., 2015. *Model-based choices involve prospective neural activity. Nat. Neurosci. 18, 767–772.
Doll, B.B., Simon, D.A., Daw, N.D., 2012. The ubiquity of model-based reinforcement learning. Curr. Opin. Neurobiol. 22, 1075–1081.
Dunne, S., D'Souza, A., O'Doherty, J.P., 2016. *The involvement of model-based but not model-free learning signals during observational reward learning in the absence of choice. J. Neurophysiol. 115, 3195–3203.
Duss, S.B., Reber, T.P., Hanggi, J., Schwab, S., Wiest, R., Muri, R.M., Brugger, P., Gutbrod, K., Henke, K., 2014. Unconscious relational encoding depends on hippocampus. Brain 137, 3355–3370.
Eickhoff, S.B., Laird, A.R., Fox, P.M., Lancaster, J.L., Fox, P.T., 2017. Implementation errors in the GingerALE Software: description and recommendations. Hum. Brain Mapp. 38, 7–11.
Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum. Brain Mapp. 30, 2907–2926.
Fermin, A.S., Yoshida, T., Yoshimoto, J., Ito, M., Tanaka, S.C., Doya, K., 2016. *Model-based action planning involves cortico-cerebellar and basal ganglia networks. Sci. Rep. 6, 31378.
Garrison, J., Erdeniz, B., Done, J., 2013. Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev. 37, 1297–1310.
Gershman, S.J., Daw, N.D., 2017. Reinforcement learning and episodic memory in humans and animals: an integrative framework. Annu. Rev. Psychol. 68, 101–128.
Gillan, C.M., Robbins, T.W., 2014. Goal-directed learning and obsessive–compulsive disorder. Phil. Trans. Biol. Sci. 369, 20130475.
Gillies, M.J., Hyam, J.A., Weiss, A.R., Antoniades, C.A., Bogacz, R., Fitzgerald, J.J., Aziz, T.Z., Whittington, M.A., Green, A.L., 2017. The cognitive role of the globus pallidus interna; insights from disease states. Exp. Brain Res. 235, 1455–1465.
Glascher, J., Adolphs, R., Tranel, D., 2019. Model-based lesion mapping of cognitive control using the Wisconsin Card Sorting Test. Nat. Commun. 10, 20.
Glascher, J., Daw, N., Dayan, P., O'Doherty, J.P., 2010. *States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66, 585–595.
Groman, S.M., Massi, B., Mathias, S.R., Lee, D., Taylor, J.R., 2019. Model-free and model-based influences in addiction-related behaviors. Biol. Psychiatr. 85, 936–945.
Gruner, P., Pittenger, C., 2017. Cognitive inflexibility in obsessive-compulsive disorder. Neuroscience 345, 243–255.
Gu, B.M., Park, J.Y., Kang, D.H., Lee, S.J., Yoo, S.Y., Jo, H.J., Choi, C.H., Lee, J.M., Kwon, J.S., 2008. Neural correlates of cognitive inflexibility during task-switching in obsessive-compulsive disorder. Brain 131, 155–164.
Han, W., Tellez, L.A., Niu, J., Medina, S., Ferreira, T.L., Zhang, X., Su, J., Tong, J., Schwartz, G.J., Van Den Pol, A., 2016. Striatal dopamine links gastrointestinal rerouting to altered sweet appetite. Cell Metabol. 23, 103–112.
Holman, E.W., 1975. Some conditions for the dissociation of consummatory and instrumental behavior in rats. Learn. Motiv. 6, 358–366.
Jentsch, J.D., Olausson, P., De La Garza 2nd, R., Taylor, J.R., 2002. Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. Neuropsychopharmacology 26, 183–190.
Killcross, S., Coutureau, E., 2003. Coordination of actions and habits in the medial prefrontal cortex of rats. Cerebr. Cortex 13, 400–408.
Kroemer, N.B., Lee, Y., Pooseh, S., Eppinger, B., Goschke, T., Smolka, M.N., 2019. *L-DOPA reduces model-free control of behavior by attenuating the transfer of value to action. Neuroimage 186, 113–125.
LeDoux, J., Daw, N.D., 2018. Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. Nat. Rev. Neurosci. 19, 269–282.

Lee, S.W., Shimojo, S., O'Doherty, J.P., 2014. *Neural computations underlying arbitration between model-based and model-free learning. Neuron 81, 687–699.
Maddox, W.T., Chandrasekaran, B., Smayda, K., Yi, H.-G., Koslov, S., Beevers, C.G., 2014. Elevated depressive symptoms enhance reflexive but not reflective auditory category learning. Cortex 58, 186–198.
McNamee, D., Liljeholm, M., Zika, O., O'Doherty, J.P., 2015. *Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate fMRI study. J. Neurosci. 35, 3764–3771.
Miller, K.J., Botvinick, M.M., Brody, C.D., 2017. Dorsal hippocampus contributes to model-based planning. Nat. Neurosci. 20, 1269–1276.
Nebe, S., Kroemer, N.B., Schad, D.J., Bernhardt, N., Sebold, M., Müller, D.K., Scholl, L., Kuitunen-Paul, S., Heinz, A., Rapp, M.A., 2018. *No association of goal-directed and habitual control with alcohol consumption in young adults. Addiction Biol. 23, 379–393.
O'Doherty, J.P., 2011. Contributions of the ventromedial prefrontal cortex to goal-directed action selection. Ann. N. Y. Acad. Sci. 1239, 118–129.
Otto, A.R., Raio, C.M., Chiang, A., Phelps, E.A., Daw, N.D., 2013. Working-memory capacity protects model-based learning from stress. Proc. Natl. Acad. Sci. Unit. States Am. 110, 20941–20946.
Padoa-Schioppa, C., Assad, J.A., 2006. Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223–226.
Radenbach, C., Reiter, A.M., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.-J., Deserno, L., Schlagenhauf, F., 2015. The interaction of acute and chronic stress impairs model-based behavioral control. Psychoneuroendocrinology 53, 268–280.
Reiter, A.M., Heinze, H.J., Schlagenhauf, F., Deserno, L., 2017. *Impaired flexible reward-based decision-making in binge eating disorder: evidence from computational modeling and functional neuroimaging. Neuropsychopharmacology 42, 628–637.
Remijnse, P.L., Nielen, M.M., van Balkom, A.J., Cath, D.C., van Oppen, P., Uylings, H.B., Veltman, D.J., 2006. Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. Arch. Gen. Psychiatr. 63, 1225–1236.
Russek, E.M., Momennejad, I., Botvinick, M.M., Gershman, S.J., Daw, N.D., 2017. Predictive representations can link model-based reinforcement learning to model-free mechanisms. PLoS Comput. Biol. 13, e1005768.
Schacter, D.L., Addis, D.R., Buckner, R.L., 2007. Remembering the past to imagine the future: the prospective brain. Nat. Rev. Neurosci. 8, 657.
Schultz, W., 2001. Book review: reward signaling by dopamine neurons. Neuroscientist 7, 293–302.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275 (5306), 1593–1599.
Sebold, M., Nebe, S., Garbusow, M., Guggenmos, M., Schad, D.J., Beck, A., Kuitunen-Paul, S., Sommer, C., Frank, R., Neu, P., Zimmermann, U.S., Rapp, M.A., Smolka, M.N., Huys, Q.J.M., Schlagenhauf, F., Heinz, A., 2017. *When habits are dangerous: alcohol expectancies and habitual decision making predict relapse in alcohol dependence. Biol. Psychiatr. 82, 847–856.
Seger, C.A., Miller, E.K., 2010. Category learning in the brain. Annu. Rev. Neurosci. 33, 203–219.
Simon, D.A., Daw, N.D., 2011. *Neural correlates of forward planning in a spatial decision task in humans. J. Neurosci. 31, 5526–5539.
Stalnaker, T.A., Takahashi, Y., Roesch, M.R., Schoenbaum, G., 2009. Neural substrates of cognitive inflexibility after chronic cocaine exposure. Neuropharmacology 56 (Suppl. 1), 63–72.
Stoianov, I.P., Pennartz, C.M.A., Lansink, C.S., Pezzulo, G., 2018. Model-based spatial navigation in the hippocampus-ventral striatum circuit: a computational analysis. PLoS Comput. Biol. 14, e1006316.
Sutton, R.S., Barto, A.G., 1998. Introduction to Reinforcement Learning. MIT press, Cambridge.
Tanaka, S.C., Balleine, B.W., O'Doherty, J.P., 2008. Calculating consequences: brain systems that encode the causal effects of actions. J. Neurosci. 28, 6750–6755.
Tricomi, E., Balleine, B.W., O'Doherty, J.P., 2009. A specific role for posterior dorsolateral striatum in human habit learning. Eur. J. Neurosci. 29, 2225–2232.
Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage 16, 765–780.
Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M., Fox, P., 2012. Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. Hum. Brain Mapp. 33, 1–13.
Valentin, V.V., Dickinson, A., O'Doherty, J.P., 2007. Determining the neural substrates of goal-directed learning in the human brain. J. Neurosci. 27, 4019–4026.
van der Meer, M.A., Johnson, A., Schmitzer-Torbert, N.C., Redish, A.D., 2010. Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. Neuron 67, 25–32.
Vikbladh, O.M., Meager, M.R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., Daw, N.D., 2019. Hippocampal contributions to model-based planning and spatial memory. Neuron 102, 683–693 e684.
Wang, J.X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J.Z., Hassabis, D., Botvinick, M., 2018. Prefrontal cortex as a meta-reinforcement learning system. Nat. Neurosci. 21, 860–868.
Wimmer, G.E., Daw, N.D., Shohamy, D., 2012. *Generalization of value in reinforcement learning by humans. Eur. J. Neurosci. 35, 1092–1104.
Worthy, D.A., Davis, T., Gorlick, M.A., Cooper, J.A., Bakkour, A., Mumford, J.A., Poldrack, R.A., Todd Maddox, W., 2016. *Neural correlates of state-based decision-making in younger and older adults. Neuroimage 130, 13–23.
Wunderlich, K., Dayan, P., Dolan, R.J., 2012. *Mapping value based planning and extensively trained choice in the human brain. Nat. Neurosci. 15, 786–791.
Yaple, Z.A., Yu, R., 2019. Fractionating adaptive learning: a meta-analysis of the reversal learning paradigm. Neurosci. Biobehav. Rev. 102, 85–94.