REVIEW ARTICLE

Advances in modeling learning and decision-making in neuroscience

Anne G. E. Collins [1,3 ✉] and Amitai Shenhav [2,3 ✉]

An organism's survival depends on its ability to learn about its environment and to make adaptive decisions in the service of achieving the best possible outcomes in that environment. To study the neural circuits that support these functions, researchers have increasingly relied on models that formalize the computations required to carry them out. Here, we review the recent history of computational modeling of learning and decision-making, and how these models have been used to advance understanding of prefrontal cortex function. We discuss how such models have advanced from their origins in basic algorithms of updating and action selection to increasingly account for complexities in the cognitive processes required for learning and decision-making, and the representations over which they operate. We further discuss how a deeper understanding of the real-world complexities in these computations has shed light on the fundamental constraints on optimal behavior, and on the complex interactions between corticostriatal pathways to determine such behavior. The continuing and rapid development of these models holds great promise for understanding the mechanisms by which animals adapt to their environments, and what leads to maladaptive forms of learning and decision-making within clinical populations.

Neuropsychopharmacology; https://doi.org/10.1038/s41386-021-01126-y

INTRODUCTION

One of the brain's most fundamental functions is to guide the organism towards good outcomes and away from bad ones. Over the past few decades, research into the role of prefrontal cortex in these functions has flourished, driven not only by novel empirical observations (see Soltani & Koechlin [1], Monosov & Rushworth [2], and Friedman & Robbins [3] this issue (also see the chapter by Friedman &Robbins in this volume) but also by the increased availability of computational models to account for those findings, and to enable the development and testing of novel hypotheses. In this chapter, we review the significant progress that has been made towards developing better models for how animals learn about their environment (e.g., what kinds of actions they can take and what outcomes might result) and how they make decisions about the best course of action. We will describe how these models have advanced over the years to account for the types of complexity (e.g., action hierarchy) and constraints (e.g., limited cognitive resources) that characterize real-world learning and decision-making. Finally, we will discuss insights that have been gained through increased interaction between these two families of algorithms, as well as shared opportunities and pitfalls for applying these models to the understanding of psychological processes, neural circuitry, and clinical disorders.

COMPUTATIONAL MODELS OF LEARNING
Early models of learning

Computational models of learning have a long history in the cognitive sciences, dating back to behaviorism. Some of the first computational models of learning attempted to capture a family of phenomena observed in classical conditioning: a neutral (*conditioned*) stimulus (e.g., a bell, CS) becomes predictive of an appetitive (*unconditioned*) stimulus (e.g., food, US), after repeated exposure to the food following the bell; the bell then starts eliciting reflexive responses normally reserved for the US (e.g., salivation) [4]. The Rescorla-Wagner (RW) model [5] provided a mathematical formalism that could capture this *acquisition* phenomenon, as well as others such as *extinction* (unlearning when the CS stops being paired with the US), *blocking* (no learning for a new CS if it is paired with a CS that already predicts the US), and *overshadowing* (stronger learning for more salient stimuli). RW tracks the *prediction strength* W_i of each CS, and updates it at each occurrence in proportion to the prediction error, controlled by a shared learning rate ($0 < \alpha < 1$) and a specific saliency parameter ($0 < \beta_i < 1$). The prediction error takes into account the cumulative prediction made by all present CS, and the value of the outcome (positive in the presence of a US, 0 in its absence).

$$W_i(t+1) = W_i(t) + \alpha\beta_i[\text{outcome}(t) - (W_1(t) + \ldots + W_n(t))]$$

[1]Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Cognitive, Linguistic, & Psychological Sciences and Carney Institute for Brain Science, Brown University, Providence, RI, USA. [3]These authors contributed equally: Anne G. E. Collins, Amitai Shenhav.
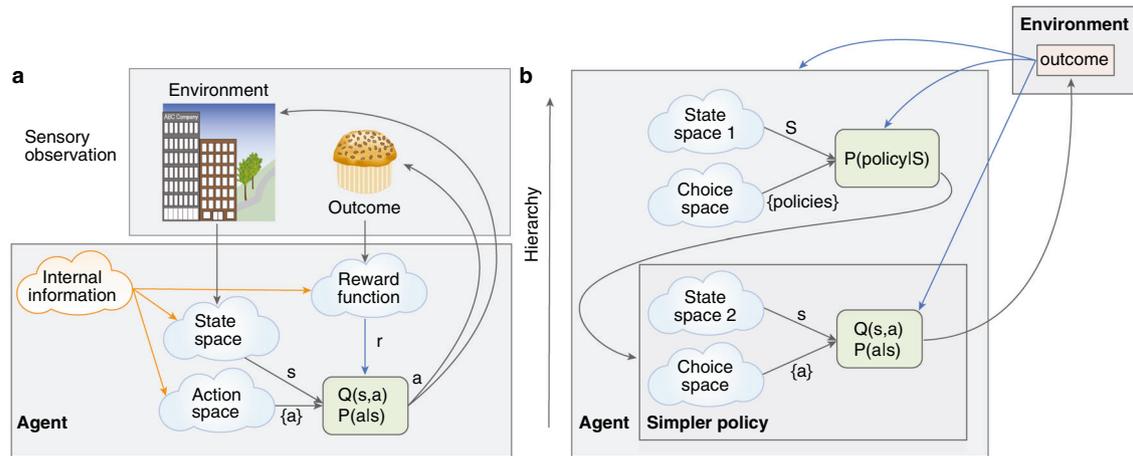✉email: annecollins@berkeley.edu; amitai_shenhav@brown.edu

Fig. 1 Levels of state and action representation during learning. A To obtain rewards (*r*), we need to learn what actions (*a*) to select in different states (*s*). However, observations from our environment (e.g., a busy street) are very complex. For efficient learning, they need to be converted into a lower dimensional, goal-relevant state space (e.g., the distance to a bakery, if we're hungry). Outcomes are also internally processed into goal-dependent rewards (e.g., a chocolate chip muffin might be rewarding when hungry, but not when on a diet). Orange arrows reflect that internal representations (e.g., memories, goals, learned representations, attention etc.) influence what states, choices and outcomes the agent considers. These variables are then passed to an adaptive decision making process (thick box), such as an RL model evaluating the cached value of different choices (e.g., buying a muffin vs. buying fruit salad; see also Fig. 2). The resulting reward drives learning (blue arrows). **B** An agent may represent states and choices at multiple levels of hierarchy, and learn to make choices over these different state-action pairs simultaneously. For example, an agent might make a more abstract choice (going to a bakery) in an abstract state (hungry in the street), which constrains lower level choices (state: in front of the bakery; choice: open the door of the bakery). The outcome will drive learning at multiple levels of the hierarchy, and can also influence the structure of the hierarchical representation (blue arrows).

Together, this simple, intuitive equation captured many aspects of classical conditioning, and represented a major starting point for the computational modeling of animal learning. However, this model also had limitations, of two major kinds. First, it was unable to account for some well-established findings in the domain of classical conditioning, such as second-order conditioning (where repeated exposure to a light followed by the bell lead to the light becoming conditioned too) or reinstatement (where learning was faster after undergoing extinction). Second, the model was focused on reflexive behaviors (e.g., salivating) rather than any type of choices, which behaviorists established could be trained in similar ways [6].

Since then, a broadening field of computational modeling has extended the family of computational models that RW is a part of, tackling some of its limitations (such as capturing second-order conditioning), and still grappling with others (such as reinstatement). This family of models falls under the umbrella term of *reinforcement learning* (RL), and defines the problem of learning as *learning to select actions in different states with the goal of maximizing the cumulative future (discounted) rewards*. A subfamily of RL algorithms (called "model-free"), solve this problem by estimating the value of states, actions, or state-action pairs in a way that is very similar to RW: the observation of a prediction error $\delta = r - V$ (the difference between observed and predicted outcome) drives learning in proportion to a learning rate α. The simplest such model, sometimes called "delta-rule", only considers the immediate reward as the outcome:

$$V(t + 1) = V(t) + \alpha\delta$$

where *V* could be the value of a state, or a state-action pair. To take into account sequential dependencies and future outcomes, as observed in second-order conditioning, the temporal-difference learning algorithm uses as a prediction not only the immediate outcome, but also the value of the next state, temporally discounted by γ:

$$\delta = r + \gamma V(s_{t+1}) - V(s_t)$$

Here, if $\gamma = 0$, the value of the next state is ignored, while if $\gamma = 1$, there is no time discounting—γ is usually set between those

extremes. The introduction of temporal difference (*TD*) was important in many ways. First, it enabled RL models to capture phenomena RW could not, such as second-order conditioning: because the bell has acquired value, observing the bell after the light will drive a prediction error (and thus learning) even in the absence of an immediate reward: $\delta = 0 + \gamma\ V(bell) - V(light)$. Second, seminal work in the 1990s showed that this temporal difference prediction error signal was a good quantitative model for the phasic firing of dopamine neurons [7, 8], making TD-RL a strong candidate for a computational model of learning at all three of Marr's levels of analysis [9, 10]: functional (the goal is to optimize future reward), algorithmic and representational (by estimating reward values with reward prediction error updates), and mechanistic (implemented by dopamine dependent signaling).

Indeed, since this foundational finding, a host of research has established the existence of a network of brain areas that appear to encode RL-like computations: the cortico-basal ganglia network, with dopamine-dependent plasticity. This loop circuit originates in cortex, and projects back to a similar area via projections to striatum and thalamus. There is now strong evidence that striatum encodes activity related to action value [11] in a way that causally influences choice [12]. Dopamine-dependent plasticity in striatum [13] is thought to be an important mechanism for the learning of those values [14, 15]. These parallel loops originate in multiple places in cortex, including multiple areas of prefrontal cortex, showing an important role for the cortex in representing states and actions over which RL computations operate [16]. For instance, research has shown that some choices can be particularly abstract and executive (prefrontal cortex-dependent), such as learning how to use working memory (i.e., which information to gate in or out [17]).

More recent research increasingly recognizes that the RL framework, on its own, is too limited to capture the flexibility of learning in human and non-human animals. In particular, PFC-dependent mechanisms appear to provide much nuance to the way information is used and organized to make decisions in a learning environment. In the next part, we present a survey of computational models of learning along two dimensions. First, some research keeps the overall structure of the RL learning

algorithms intact as a model of learning, but argues that more complex behavior can be explained by making this structure more subtle in various ways: how are state and action spaces defined? Might multiple RL computations co-exist? Second, other research argues that the existing RL architecture of learning needs to be supplemented with other computational architectures, expressing different optimization functions, and reflected by different underlying brain mechanisms. These might include working memory, planning and episodic memory.

Modeling complexities in learned representations

State spaces. Behaviorists focused on behavior, rather than on internal representations that had been previously studied through introspective methods; RL models of learning similarly tend to neglect such representations. Indeed, classic models of learning take the state space, the action space, and the reward function for granted, and focus on the algorithm that allows the estimation of the expected value and choice policy instead [18]. In particular, the state and action spaces are experimenter defined (e.g., state = {light on, bell ringing} in a Pavlovian experiment; action = {press, no press} in an instrumental conditioning task). However, from the point of view of the agent, neither is obvious (Fig. 1A). Sensory inputs in the real world are extremely complex and multi-dimensional—defining the state space based on those inputs would lead RL models to fail dramatically, due to the curse of dimensionality [19]. How, then, does the agent know that only the bell is relevant, as opposed to other sensory features? Recent work proposes new models that integrate attentional modules into RL algorithms to explain not only how we learn, but also what we learn about.

For instance, in an attentional learning task [20], participants were exposed to stimuli that varied along three dimensions with three features each (e.g., color, shape, and texture). In a given episode, only one dimension was relevant, and only one feature of this dimension led to reward with high probability—thus participants needed to learn what the relevant state space was, in addition to learning the value of those relevant states. Computational modeling showed that participants used attentional filtering to do so [20–22], and that this attentional filtering itself was learned over time [21]. These studies confirm that learning can be explained with a simple RL update over an internal representation of an attentionally derived state space, where prefrontal executive functions play a fundamental role in devising the state space. Other studies further confirm that the state space over which value is tracked changes dynamically over the course of learning [23]. These changes can account parsimoniously for surprising effects—for example, a simulation study offered a proof of principle that this "thinning" of the state space to only task-relevant features could account for observed ramping in dopamine signaling [24].

The process of defining a relevant state space is not necessarily limited to the reduction of attentional dimensionality. Instead, it is possible that some relevant states are not trivially linked to compressed sensory information. For example, even in the absence of current sensory information, one's memory of checking the weather forecast last night informs what they believe the current relevant state is for the task of dressing in the morning: Warm? Cold? Rainy? More generally, the latent unobservable states over which RL operates provide a useful modeling framework for capturing more flexible learning. These "belief states" [25–27] are the result of internal computations and cannot be directly inferred from current observations. The phenomenon of reinstatement following extinction offers a useful example of how such belief states extend RL models: classic RL models cannot capture this phenomenon, because they simply "unlearn" past associations during extinction. When belief states are incorporated into these models, the agent creates a new state during extinction, indicating a different context in which

the bell does not lead to reward anymore. At reinstatement, the agent can recognize a previous context where the bell led to rewards, and identify it as the current latent state, leading to faster learning [26].

There is currently a debate in the literature as to which brain regions represent this task-relevant state space [28, 29]. Recent attention has been given to the orbitofrontal cortex (OFC) as a potential locus [30, 31]. Indeed, a recent human fMRI study used representational similarity analysis [32] to show that OFC contained all the information needed to represent task-relevant states, in a way that related to performance. This has since also been observed in rodents [33], although OFC appears to orthogonally encode value representation as well [34]. However, there is also a host of recent literature implicating hippocampus in state representation, under the name of *cognitive maps*. An example of a cognitive map is the "successor representation", a compact representation of current and likely future states that supports fast and flexible learning with simple RL-like mechanisms, and is putatively represented in hippocampus [35, 36]. A recent proposal suggests that hippocampus and entorhinal cortex work together to represent a cognitive map that respects generalizable relational properties of the environment, but also to identify specific, non-generalizable properties [37]. It remains to be understood what distinct roles OFC vs. HC play in representing states, and how other regions contribute to this role.

Action spaces. As in the case of the state space, the action space is often taken for granted in RL models [18]: in a two-alternative forced choice, does an animal choose between repeating or switching, or between left and right? When typing, do I choose to move my index finger in a specific motor movement, or to press the key "T"? The process of considering *which options to choose from* is separate from the process of choosing itself [38]. One reason it is important to consider the action space relates again to the curse of dimensionality: with too many possible choices (e.g., if choice is encoded as continuous motor movements), learning becomes exponentially slower. A "divide and conquer" approach helps solve this: instead of considering all possible choices, we make choices between a few options at a high level of granularity (e.g., soup or salad), then conditioned on that choice, we make choices at a lower level of granularity (e.g., tomato or onion soup), and so on [39, 40]. (Fig. 1B).The hierarchical reinforcement learning framework formalizes this extension of RL by adding to the basic action space "temporally abstracted choices", or *options* [39, 41], which are local learnable policies that can be selected as a whole and then are followed to a specified termination. This can, for example, greatly simplify navigation problems, provided the right options exist: an agent can choose to go open the door, rather than making the sequence of choices to [stand up, turn right, [step] *3, open the door]. Indeed, by using established options, we can explore the environment more efficiently, transferring past knowledge encapsulated in the option [42]. There is evidence that the brain tracks options, as shown by option-specific learning signals in the striatum [43] and anterior cingulate cortex (ACC) [44].

Abstract choices to "divide and conquer" a problem need not be temporal abstractions, as stated in the options framework (Fig. 1B). Another line of research focuses on hierarchy in the degree of abstraction of a choice: for example, I can choose to answer an email (abstract choice), which translates into different motor choices in different situations (e.g., clicking an arrow, or a combination of keys). A gradient of areas in lateral prefrontal cortex represents decision-making problems at multiple levels of abstraction [45, 46]. Recent research has expanded RL models with such abstract choices, in the form of *rules* or *task-sets* [47, 48]. In these models, agents select rules, which then condition more concrete choices made in response to stimuli, and these policies are learned via RL. Computational models of hierarchical choice reveal an important role for medial prefrontal cortex in tracking

4

rules via an inference process [49], or by propagation of hierarchical errors [50, 51]. However, there is also evidence that rule selection, in some cases, could also proceed via its own RL process [48, 52]: indeed, there is evidence that the striatum tracks rule values [53], and a recent study showed behavioral evidence for parallel tracking of value at two levels of choice abstraction [54].

Hierarchical learning models need to jointly define state and action spaces at multiple levels of abstraction (Fig. 1B). For example, one set of models [49, 55] defines a "context space" of latent causes that condition the selection of task-sets, which then condition (via RL) the selection of actions over a simpler state space of visual stimuli. Other models include multiple (distinct) observable state spaces [48, 54]. The relevant state space might even be dependent on the selected rule, which includes an attentional filter on the stimulus features [52, 56]. In short, allowing flexibility on the state and action representations over which the RL system operates has provided computational models of learning with greatly increased flexibility, while keeping a bridge to a biologically plausible implementation.

Modeling constraints on learning

Classic model-free RL models, which store "cached" estimates of values, have been successfully extended by considering what state and action spaces they operate over, and how they can be dynamically modulated. The prefrontal cortex plays an important role in these functions, thus taking advantage of a phylogenetically old system with fairly simple computations in basal ganglia (RL), to support flexible, human-level efficient learning. However, other computational models of learning propose other contributions that do not only depend on the RL computations performed in cortico-basal ganglia loops.

Planning. One popular extension of classic RL models imported the notion of "model-based" RL from the AI literature [19]. Model-based RL describes algorithms that have access to a model of the environment—usually the transition function (*which state do I end up with when I select action a in state s?*) and the reward function. So equipped, such agents can perform planning by internally propagating forward the consequences of their choices, and are thus able to actively re-calculate the estimated value of a choice on the fly, even when the environment changes. However, the flexibility of these algorithms comes at the expense of greater demands on proactive forward planning, which relies on prefrontal executive functions. The contrast between these flexible but computationally expensive model-based algorithms and more rigid but computationally cheap model-free algorithms (which simply require storing and accessing the value of an action in a given state) inevitably drew parallels with similar dichotomies across research on slow/controlled ("System II") vs. fast/automatic ("System I") forms of reasoning and decision-making [57, 58]. These parallels were substantiated by demonstrating that people develop a bias towards model-free over model-based RL when their executive functions are taxed [59, 60].

While the term model-based RL takes the specific meaning of an RL model that uses a *transition model* to perform forward planning to estimate values, many other extensions of model-free RL have been proposed that involve the knowledge and use of a model of the environment outside of such planning [61]. This transition model is often used to infer a latent cause, such as the task-relevant state, and perform appropriate credit assignment [48, 61–63]. In most cases, the learning and encoding of the transition model itself has been related to prefrontal function. Knowledge of the structure of the environment can take multiple computational forms, but generally leads to more flexible adaptation in the face of changes in the environment. For example, the successor representation mentioned above [35, 36] summarizes knowledge about frequent transitions in the environment in a way that enables an agent to flexibly change their behavior when rewards change, without requiring planning.

Memory. The previous models still assume a theoretical "RL" framework, where the objective of the algorithm is to optimize long-term future reward, and thus the strategy is to estimate this expected reward, or to estimate a policy that ensures this objective [19]. However, humans also rely on other memory systems that are more general and not specialized in valuation. Arguably, evolutionary pressure likely ensures that such systems contribute to choices that maximize rewards, as shown in artificial neural networks that develop working memory-like activity when trained from rewards [64]. Recently, cognitive models of learning have leveraged this idea that memory systems not specialized in valuation might nevertheless also contribute to learning. As an example, there is now strong evidence that working memory sometimes contributes a large portion of the observed behavior in reinforcement learning tasks. Because information maintenance in working memory is limited in time, it cannot be considered a learning system per se-nevertheless, maintaining information over short periods of time can actively guide the choices made. Indeed, when feedback is reliable and participants only need to learn a few associations, working memory is the main contributor to learning [47, 65]. Reliance on working memory for learning decreases with cognitive load [47, 66] or with uncertainty, but remains an important contributor [67, 68]. It is important to note that in most cases such behavior can still be well captured with a "modified" RL model (e.g., where learning rate is dependent on load), but that this approach misattributes behavior to a single process rather than two, and hides the critical contributions of prefrontal working memory to learning.

There is now also strong evidence that episodic memory contributes to learning: indeed, if we can store individual memories of information from past trials and sample such past information from our memory, then we can use that information to guide our decision (e.g., when I made this choice in this state, I got rewarded - let's do that again). There is a wealth of research showing that such hippocampally-mediated processes contribute to learning in parallel with striatally-mediated processes [69], and interact with them [70, 71]. Indeed, recent research showed that such sampling processes make significant contributions to both "model-free" RL [72, 73] and "model-based" RL [74].

Meta-learning, strategies, heuristics. An important target for computational models of learning is also the domain of "meta-learning", or learning how to learn. One example of this is the dynamic modulation of one's learning rate such that they learn more from recent observations when the environment is changing rapidly, but then integrate information over longer time scales when the environment is changing slowly; this dynamic modulation is tracked in ACC [75]. How the brain integrates such information to dynamically modulate learning is still debated, as the complex inference processes required are not biologically plausible. Recent proposals have introduced heuristics that approximate such processes to a high degree of optimality, while staying closer to potential mechanisms [55, 76]. Other meta-learning models have proposed upregulation of learning rates by surprise in amygdala and PFC [77–79]. Additional meta-learning mechanisms include modulating the expected scale of rewards (reference point centering and range scaling) [80]; and identifying and making use of regularities in an environment—for example, using counterfactual learning when options are anticorrelated [81, 82]. Last, the early learning process (in the first few trials of a new situation) often relies on "meta-learned" heuristic strategies that are dependent on prefrontal function: indeed, it is an important period of exploration that reveals priors, biases and heuristics that we rely on to quickly acquire information [42, 83].
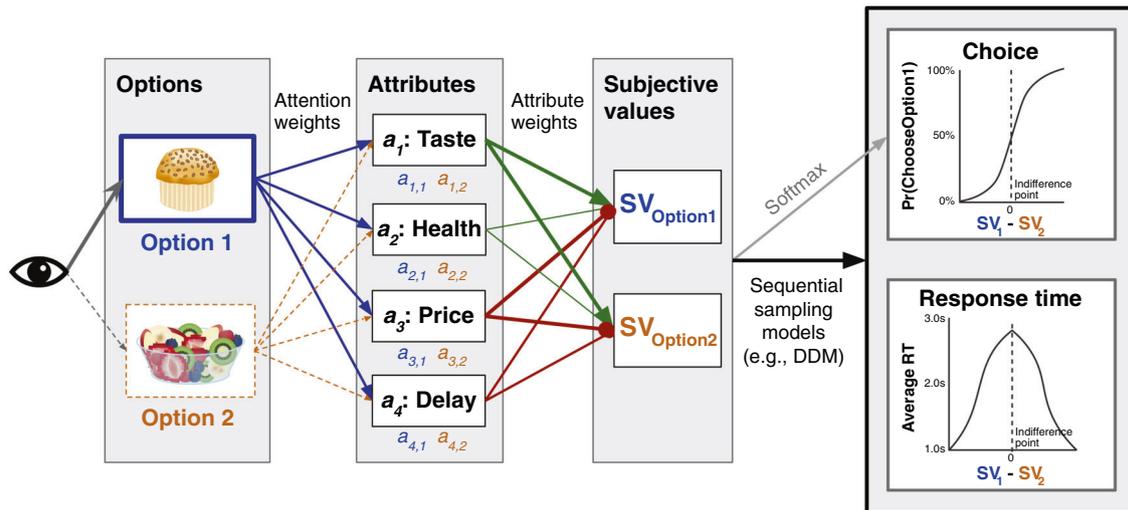
Fig. 2 Evaluating and selecting between options. When deciding between a pair of options (e.g., whether to go buy a muffin or fruit salad for breakfast), the agent must decompose those options into goal-relevant attributes (e.g., a1: tastiness, a2: healthiness, a3: price, a4: delay/distance), with each option carrying a value for that attribute (e.g., a1,1: tastiness of muffin, a1,2: tastiness of fruit salad, etc.). Those attribute values are weighted by top-down factors (e.g., which option/attribute is being attended to more closely) and bottom-up factors (e.g., how much that person cares about taste vs. health; how readily information about this attribute comes to mind). The weighted attribute values contribute positively (green arrows) or negatively (red arrows) to the overall subjective value (SV) of each option. The likelihood of choosing a given option (e.g., muffin) can be determined by performing a softmax based on the degree to which one SV is higher than the other. Alternatively, sequential sampling models (e.g., drift diffusion and leaky competing accumulator models—which dynamically accumulate information up to a threshold level (e.g., based on SVs and/or lower-level attribute values)—can be used to predict choice likelihood as well as expected distributions of response times.

COMPUTATIONAL MODELS OF DECISION-MAKING
Weighing up subjective values
Whereas models of learning address how we come to know something's value in a given environment, models of decision making address how we select between multiple options with different values. The intervening computations determine which inputs are relevant to the decision, how they will be differentially weighted, the process by which options are selected, and what outcome(s) result from this selection process (e.g., what kind of action is taken) (Fig. 2).

Early model-driven research in decision neuroscience borrowed heavily from the field of behavioral economics [84], where researchers had described the functions people applied when considering monetary outcomes that vary in their likelihood (risk) or in how long it will take to acquire them (delay). The goal here was to account for the fact that the same monetary amount (e.g., $100) mattered to someone to a different degree if it was a gamble versus a sure thing, and if they would be receiving it immediately or after some prolonged period of time.

Research examined the major elements that shaped these evaluations, and what specific shape the value transformation took (for more comprehensive reviews, see [85–89]). For instance, studies of *risk preferences* [90, 91] examined how people weighed the probability of a given outcome occurring (e.g., when offered a guaranteed $100 versus an equivalent gamble that offered 50% chance of $200 and 50% of $0). This work showed that people weighed risk nonlinearly, and in a way that opposed choosing a gamble with increasing risk (i.e., was *risk-averse*) when the possible outcomes were all positive, and demonstrated the opposite pattern (*risk-seeking*) when the possible outcomes were negative. This work additionally revealed that the overall subjective value of an option (SV) reflected a stronger weight on potential negative outcomes than positive outcomes (*loss aversion*):

$SV = V^{\alpha}$ for gains and $SV = \lambda(-V)^{\beta}$ for losses

where α vs. β represent sensitivity to magnitudes of potential gains vs. losses (V), and λ represents an additional weight on losses of any magnitude.

Studies of *delay discounting* examined how people overall down-weight the value of a reward the longer it takes to obtain it (e.g., potentially preferring $50 today over $100 in a year), with studies often showing that these delays are discounted hyperbolically, such that value demonstrates a steep drop-off after relatively short delays before it plateaus at a lower level [92]:

$$SV = \frac{V}{1 + kD}$$

where k is the degree to which a given reward (V) is discounted by a given delay (D).

Studies of *effort discounting* examined how people similarly discount reward value by the amount of physical [93] or cognitive [94] effort required to achieve that reward, again typically finding that these costs result in nonlinear down-weighting of rewards (i.e., increasing effort requirements have greater and greater costs) [87], for instance according to quadratic or sigmoidal functions of cost (C):

Quadratic costs: $SV = V - kC^2$

Sigmoidal costs: $SV = V\left(1 - \frac{1}{1 + e^{-k(C-p)}}\right)$

where p represents the inflection point of the sigmoid function.

Identifying these subjective weighting functions serves two purposes. First, it allows a researcher to estimate how an outcome's subjective value changes for a given person with variability in factors such as risk and delay. These weights effectively serve as an exchange rate across different types of options. This in turn allows the researcher to identify subjective equivalencies between pairs of options within a *common currency*, for instance that a person cares equally about receiving $10 today

and $20 in 6 months. Second, it provides a quantitative estimate of how people differ in their evaluation of these same variables (e.g., in their discount or cost parameters), providing a basis for research into individual differences in sensitivity to each. Early research in neuroeconomics used these models to identify correlates of each of these decision attributes—including risk [95–97], loss [98], delay [99, 100], and effort [101, 102]—and putative points of convergence, where a common currency-like signal could be found. The network of brain regions that most consistently tracks these re-weighted outcome values includes the ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex, and ventral striatum [103–105].

Transforming values into choice

Estimating the values of a set of options is a critical first step to modeling decision-making, requiring appropriate functions to translate the objective option properties/attributes into subjective values (Fig. 2). However, even once those values are determined, there needs to be an algorithm for determining which option to select. The most straightforward solution to this would be to select whichever choice has the highest subjective value (i.e., if the overall subjective values of Option A vs. Option B are $50 vs. $51, select Option B). However simple this solution might be, it is inconsistent with what is observed when people do select between options.

A basic property of findings on option selection is that people do not seem to always choose the better of two options, even when accounting for differences in subjective weighting of each of the choice attributes. If they did, we would expect to see that their choices followed a step function—as long as Option B has a higher value than Option A (as in the example above), they always choose Option B, and vice versa. Instead, people's choices follow a sigmoid-like pattern—more step-like when there is a large difference between the option values (e.g., Option B is much more valuable than Option A) but, as that difference narrows, people start to choose the higher-valued one with less consistency (i.e., are increasingly likely to choose the "objectively" lower-valued option). This pattern is similar to the well-characterized psychometric functions observed when discriminating between different percepts, and indeed comparable choice patterns have been shown within the same participants performing both perceptual and value-based decision-making tasks [106, 107].

To account for these inconsistencies, choice models incorporate a *noise* term, which determines the extent to which choices will be strictly determined by the higher-valued option versus by random chance:

$$Pr(ChooseOption1) = \frac{1}{1 + e^{-\beta(SV_1 - SV_2)}}$$

This function, referred to as a *softmax*, produces choices that are increasingly likely to favor the higher-reward option the greater the difference between the options (e.g., $10 vs. $1) and increasingly likely to choose either option equally (i.e., express indifference) when that difference is small (e.g., $10 vs. $9), with the steepness of that transition being determined by an inverse temperature parameter (β). The function can also generalize to choices between any number of options, and is regularly used to model choice in learning models described above.

A persistent question concerns the source of this decision noise. Some of the potential sources of such "noisy" behavior—which we will return to in a later section—relate to the selection process itself, in particular the extent to which participants are engaging in heuristic or strategic (e.g., "exploratory") behavior to choose an option other than the one that is seemingly best [108, 109]. However, another (and non-mutually exclusive) possibility is that this noisy behavior reflects noise in the information being processed (the value of one's options), consistent with random

utility theory from economics [110, 111] and more recent theories within psychology and neuroscience of evaluation as a constructive process [112, 113]. These theories have in turn facilitated the development of a newer class of decision-making models that attempt to simulate the dynamics by which individuals construct the value of their options.

Modeling complexities in the choice process

Evidence accumulation. By incorporating the appropriate functions for weighing values (e.g., exponential or hyperbolic) and for selecting between options based on those values (e.g., softmax), early decision-making models could successfully predict what the outcome of a given decision would be (i.e., what choice a participant would make when faced with a set of options). However, these functions for transforming inputs to outputs do not explain the dynamics by which this transformation occurred. As a result, these models are also unable to predict how long it would take to make a given choice. The next wave of decision-making models sought to fill these gaps by leveraging sequential sampling approaches that had been previously developed in the context of signal detection theory (as a method for determining how much information to collect before judging whether a signal is present or absent [114]) and subsequently applied to the study of cognition [115, 116].

According to these models, from the time a set of choice options appear, the decision-maker is accumulating *evidence* for or against each of those options; the more valuable a given option is, the more evidence is assumed to accumulate in its favor on average. This continues until the amount of evidence that has been accumulated exceeds some *threshold* for making a choice; the response time for that choice is indexed by the time at which that threshold is crossed (along with a fixed amount of time reflecting the time required to execute the response). Sequential sampling models are therefore able to project both how likely a person would be to make a given choice, and the distribution of possible times at which that decision will be made. Further, an inherent assumption about this accumulation process is that the evidence being sampled about each of the options is sampled with noise. This results in variability across choices for a given set of option values (e.g., sometimes resulting in choosing the less valuable option), thus accounting for characteristic inconsistencies in real-world choice.

While they all follow the same general principles just outlined, sequential sampling models vary along several dimensions, including whether sampling occurs continuously or in discrete time steps; the types of interactions they assume between the sampling streams; and their level of biological detail/plausibility [117, 118]. One of the most popular classes of models is the drift diffusion model (DDM [115, 119]), which assumes that decisions are made by continuously accumulating the relative evidence for one of two responses (i.e., how much better one seems than the other), referred to as the overall decision variable (DV):

$$DV(t) = DV(t) + w(SV_1 - SV_2) + noise(t)$$

where *w* represents the conversion between value difference and evidence accumulated, and *noise* at time *t* is typically drawn from a Gaussian distribution.

The decision threshold dictates how much more evidence one must accumulate for one response relative to the other. The DDM accounts for the observation that people are not only more likely to choose the better option when their options are more different in value (i.e., when there is much more evidence for one than the other) but they are also faster to do so. The DDM can also account for how the speed and accuracy (consistency) of a choice are affected by differences in value weighting functions (e.g., based on risk or delay preferences), choice/response biases (e.g., toward a predetermined default), and choice deadlines [120–124].

Choice competition. Whereas the DDM assumes that evidence is being accumulated in relative terms (i.e., the individual is only ever weighing the extent to which one option is better than another), other models allow for evidence to accumulate for each response in parallel, such that the decision is made once enough evidence is generated for any one of those responses. This property enables these models—such as the Leaky Competing Accumulator [125] and Feedforward Inhibition [126] models—to account for the finding that people generally respond faster when the overall strength of a given option (e.g., its value) is greater, independently of how much better that option is than the other ones available [127, 128]. By incorporating interactions across multiple input and output units (which can each be thought of as representing populations of neurons tuned for a given option feature or response type), models like these also offer greater biological plausibility than the DDM. More elaborate models, like those developed by Wang and colleagues [129] build further on these sequential sampling models to provide additional levels of biological detail, including interneuron-like units that capture additional inhibitory dynamics across units [130].

Over the past few decades, sequential sampling models have played an increasingly prominent role in modeling not only behavior but also patterns of neural activity during value-based decision-making. Signatures of this evidence accumulation process have been identified throughout the brain [131, 132], with some notable differentiation between neural circuits that appear to track the accumulation of information about which options are the most valuable (e.g., in orbital and/or ventromedial PFC [128, 133, 134]); which response is the best overall (e.g., in dorsomedial prefrontal or lateral intraparietal regions [131, 135, 136]); and what motor action to implement (e.g., in premotor and/or motor cortices [137–139]). However, as we discuss next, the interpretation of these signals as indexing evidence accumulation per se, rather than a covariate thereof, is a matter of significant debate [140].

Post-choice evaluation. Whereas models of decision-making have traditionally focused on the process leading up to a decision, with the outcome being the choice itself, an additional benefit of evidence accumulation models is that they can also generate predictions for how the decision process continues to unfold after the choice is made. In particular, these models can allow the accumulation process to continue indefinitely following the crossing of the decision threshold, and estimate the likelihood that this evidence would have continued to strengthen in favor of the action they chose, weakened, or even reversed such that they would have even had a change of mind (i.e., chosen another option instead) if given more time [141–145]. These extensions of the sequential sampling framework can be used to index *metacognitive* variables such as confidence/certainty in one's actions, and to examine how a person uses such information to subsequently correct their behavior, adjust their strategy, or otherwise improve their performance (e.g., attend more to options on future choices) [146–148]. Work has shown that a person can evaluate these same metacognitive signals of confidence/uncertainty *while* making their choice (e.g., based on the overall strength of evidence), and can make online adjustments to their choice strategy accordingly (e.g., increasing their decision threshold when less confident) [143, 145, 149, 150]. Metacognitive evaluations thus offer a prominent alternative explanation for putative neural correlates of evidence accumulation, including estimates of the decision variable itself [127, 140, 151–154].

Modeling complexities in the outputs of choice
The models above describe the transformation of input values to an ultimate choice. In most cases, this research has focused on choices that involve a simple discrete action as the ultimate choice output, such as a button press to indicate which of multiple options to select. However, these models can also be applied to more complex types of choice outputs, including those that are continuous rather than discrete, internally rather than externally directed, and/or linked hierarchically.

Control signals. Most actions in the real world require coordinating multiple different types of motor outputs (rather than, e.g., moving a single finger to press a button). They also require determining both the types of efferents involved (e.g., which muscles) and the intensity with which to engage those efferents (e.g., how strongly to flex or extend a muscle). Decisions involving such complex *control signals* can also be modeled as forms of value-based decision-making [155–157], where people are simultaneously evaluating possible outcomes of different combinations of control adjustments. These models incorporate the additional consideration that higher intensities of control (e.g., more vigorous actions) incur greater costs (e.g., because they demand more metabolic resources), which people experience as *effort*. The resulting cost-benefit analysis aims to produce control adjustments that both maximize potential future reward and minimize these costs (i.e., avoid using higher intensities of control than are necessary). Recent work has extended these models from decisions about physical actions (i.e., motor control) to decisions about how to invest one's mental resources (i.e., cognitive control) [158, 159], demonstrating that the same form of cost-benefit analysis can explain how people adjust their control allocation within and across tasks based on the current incentives and task demands [160]. Signatures of these control decisions have been found across regions of medial prefrontal cortex—ranging from premotor and supplementary motor areas to anterior midcingulate cortex—suggesting a posterior-to-anterior gradient of decision-making for motor-to-cognitive forms of control [161–164].

Response hierarchy. As discussed in an earlier section, real-world behavior is also structured hierarchically, in that specific actions often extend from higher-level goals (Fig. 1B). As a result, decisions we make about a particular course of action are constrained by decisions we have made about our goals, and vice versa. Goal selection can encompass a wide range of timescales, from short (e.g., whether to make coffee or tea) to long (e.g., whether to major in neuroscience or computer science). The sequence of actions that satisfy those goals can likewise vary from habituated sequences (e.g., making coffee) to a further set of hierarchically structured goals (e.g., choosing classes, performing homework assignments). Recent work has used sequential sampling models to describe evidence accumulation occurring across multiple levels of this hierarchy, with information about the value of individual sub-options/actions and their attributes accumulating to inform decisions about the higher-level choices; [165–167] neural signatures of these parallel accumulators were identified in similar neural circuits as mentioned earlier (e.g., dorsomedial PFC [166]). Additional work has used hierarchical choice models to simulate decisions to pursue an effortful (but rewarding) goal, and shown that simulated lesions to this model (which is proposed to describe interactions between the ACC and basal ganglia) reproduce characteristic motivational impairments observed with lesions to homologous circuits in animals [168].

Modeling constraints on decision-making
Attentional focus. We ordinarily face more options than we can process at once, and therefore must prioritize some over others. For instance, a restaurant menu or course catalog might introduce dozens or even hundreds of options to select from. To accumulate evidence for all of these at once would be challenging if not intractable. Recent work has sought to account for how we choose what information gets prioritized for evaluation (i.e., which options or features to attend to); how this attentional focus shifts over the course of a decision; and how it influences the way in which

options are evaluated to make our decision.

Models of decision-making have long acknowledged attention's role in shaping decisions [169–171]. When choosing between options that each have multiple attributes, as is typically the case, it was assumed that people assign different weights to these attributes based on their relative priority for the decision-maker (Fig. 2). For instance, a person choosing what to buy for lunch may differentially weigh the tastiness vs. healthiness of their food options; a person choosing a new car may differentially weigh speed vs. fuel efficiency; and a person choosing between job offers may weigh salary vs. location. The value of each option can be determined through a weighted combination of the attribute values:

$$SV_{OptionN} = \sum_a w_a \times AttributeValue_{a,N}$$

where w_a represents the weight a person places on attribute a, which scales the value of a given attribute (e.g., level of healthiness) for a given option N.

The process of selecting between these multi-attribute options has been modeled similarly to the selection process described earlier, with models differing in whether attributes first converge to determine each option's value or whether parallel streams of evidence drive responses based on a given attribute [118, 166, 169, 171–173].

These models capture the process by which people accumulate evidence about choice options (and their constituent attributes) under the assumption that all options are always given equal consideration while deciding. However, as the restaurant menu example makes clear, this assumption is implausible. Rather, people selectively attend to one or more options at a time while making their choice, potentially influencing the overall weight those options are given at that moment in the decision process before the person shifts their attention to other options. To account for the influence of these attentional dynamics on decision-making, an extension of the DDM was proposed [174] according to which the option(s) that are in the focus of attention are momentarily given more weight during evidence accumulation, whereas unattended options are momentarily down-weighted (scaled by a parameter θ):

$$DV(t+1) = DV(t) + w\left(SV_{attended} - \theta SV_{unattended}\right) + noise(t)$$

This process adjusts dynamically over the course of a decision such that, as attention shifts between the options over time, the value of the currently attended option (whether rewarding or aversive) is always magnified relative to the others. The predictions of this attentionally-weighted DDM have been validated using eye-tracking methods, showing that the likelihood of choosing a given option, and the amount of time it takes to choose it, both scale with the value of that option and the amount of time a participant spent (overtly) attending to it [174–176]. Debates persist about how often these patterns of behavior can instead be attributed to attention being a reflection rather than predictor of upcoming choice [177], but recent work suggests that both directions of causation are likely at play in many decisions [178].

Work in this area has further examined how people decide which options to attend to and for how long. For instance, normative (Bayesian) models have been put forward that build on the sequential sampling framework to propose that people dynamically adjust which options they are attending to based on which samples they think would be most informative for their choice [179–181]. In particular, the normative prediction is that people should attend most to options whose values are uncertain and potentially determinative of what choice they should make. Consistent with this prediction, work shows that people are more

likely to sample information about options that are of uncertain value [180, 182] and that people spend more time evaluating choice options when they are less certain about how they feel about them [183].

The weighting of choice attributes can be reflected in interactions between regions of lateral prefrontal cortex and vmPFC. Consistent with earlier work on the weighting of risk and other modifiers of potential outcomes, the vmPFC has been found to generally track the integrated value of a broad set of option attributes—such as the healthiness and tastiness of a given food item [184, 185]; the aesthetic and semantic value of a symbol; [186] and the utilitarian and emotional value of a given course of action [187, 188]—in each case reflecting the value of a given attribute according to the weight it was assigned by the decision-maker (but see, e.g., [189]). The weighting of those attributes is often associated with lateral PFC activity and its coactivation with vmPFC [184, 185].

Reference points. Our evaluation of an option is rarely done in isolation but rather with reference to other potentially relevant options. For instance, how good someone thinks a given sushi restaurant is might vary based on their recent experience with sushi from other restaurants, and based on what other restaurants are available in their local environment. Such reference points have been shown to result in apparent contradictions in people's choice behavior, whereby decisions are influenced by the presence of a seemingly irrelevant alternative within the choice set [190]. For instance, someone who chooses an apple over an orange when presented in isolation may choose the orange over the apple when these are presented in the context of some other option, like a lemon.

It turns out that many of these irrational-seeming *contrast effects* can be accounted for naturally by sequential sampling models that accumulate evidence over individual attributes of one's options, such as the leaky competing accumulator [125] and decision field theory [171]. These models allow competition to occur *within* a given attribute across options (e.g., price vs. fuel efficiency of three car options) so that, as attribute-specific evidence accumulates in support of a given option (e.g., Car A has high fuel efficiency), that accumulated evidence inhibits competing options. As a result, attributes of options that are seemingly irrelevant (i.e., dominated by all of the other options) exert an influence on how the attributes of these other options are evaluated, leading to reference-dependence in the ultimate choice [118, 169, 172, 191].

A separate class of recent models captures reference-dependent choice phenomena by varying the choice values themselves (the inputs to the choice) rather than the process by which those inputs compete with one another to generate a choice. Specifically, building on earlier models from research on visual attention [192, 193], it was proposed that the values of competing options may be normalized relative to a broader set to which the individual has been familiarized or which are currently being presented [194]. This *divisive normalization* account provides an alternate explanation for the effects of irrelevant alternatives on choice [195] (but see [196]) and makes additional predictions about how contexts shape choice, for example that moderately valued options will acquire greater value during periods when the decision-maker has been primarily viewing low-value options and will acquire lower value after primarily viewing high-value options [197]. The latter prediction, that valuation is sensitive to the history of past choice sets, is shared by a wider set of models, including models of Bayesian inference mentioned earlier, which assume that estimates of current choice values are influenced by prior expectations of potential values in one's environment [179, 183, 198].

Information accessibility. The models described above assume that evidence relevant to a given option is generated from some

source for as long as that option is being evaluated, and scaled by the extent to which that option is the focus of attention and by the range of other options to which it is being compared. However, it fails to address a more fundamental question: where does this evidence come from? Recent work has attempted to address this question both theoretically and experimentally, and in so doing has provided insight into what kind of information is privileged over the course of a decision. This work suggests that evidence arises from sampling episodic memories and that, as a result, the priority with which a given piece of evidence plays a role in decision making is determined by factors governing the accessibility of those memories [72, 73, 113, 199–201]. Recent studies have shown, for instance, that experiences of extreme outcomes (e.g., the largest gains and the largest losses while gambling) are most memorable, leading to choices that favor or oppose risk depending on the most extreme outcome in a given setting [202, 203]. Such effects can be accounted for by assuming that people simulate (sample) extreme outcomes more readily than others when accumulating evidence to make a choice [204].

CROSS-CURRENTS AND SHARED CHALLENGES ACROSS MODELS OF LEARNING AND DECISION-MAKING

Progress in modeling learning and decision-making has occurred partly through parallel streams of research, but the nature of the shared questions and methodologies has always required these streams to interact closely. For instance, to measure what has been learned about a set of options, researchers typically rely on decisions between those options (e.g., how reliably is one option preferred to another). From a practical perspective, this means that innovations in modeling decisions are ultimately *also* beneficial to models of learning. For instance, learning researchers have traditionally estimated parameters of the learning process (e.g., learning rate) based solely on the choice a participant makes when encountering a set of options. Building on innovations in modeling the dynamics of decision-making, recent work has shown that incorporating information about how long it took to choose between those options (e.g., as part of a generative decision model like the DDM or other sequential sampling models) can provide more robust estimates of that learning process when fitting data from a group or an individual [205–207]. This same approach can also be used in future work to better understand the role of selective attention during learning, leveraging existing sequential sampling models of attentional weighting during decision-making [174, 175]. Similarly, a vast array of research on learning shows that model fits improve by accounting for histories of past *choice* (e.g., whether the participant pressed the left or right button on the last trial) rather than only histories of past outcomes (e.g., [57, 208–210]), and that failing to account for such choice history effects can lead to misestimation and misinterpretation of standard RL learning rates [211].

As this last example makes clear, the benefits of studying how models of learning and decision-making intersect go far beyond the practicalities of model fitting. Indeed, learning is shaped by past decisions as much as decisions are shaped by past learning, and this occurs over multiple timescales and levels of action/outcome complexity. The impact of recent choices on behavior can thus reflect the combined influence of RL processes and the use of strategies or heuristics (like choice repetition or alternation) that optimize a different objective function entirely (e.g., learning about the task or simply getting out of the experiment sooner) [83]. Similarly, simultaneously accounting for choices and RTs not only improves model fits but can also lead to unifying new insights into the multiple processes being incorporated into a given model, such as working memory and RL [67].

How different systems of learning and decision-making interact with one another has been a major source of inquiry and debate at a much broader scale [62, 212]. For instance, separate sets of models have described the processes that underlie learning and action selection for habitual versus goal-directed behavior, with the former being determined through some form of averaging over the outcomes experienced [57] or actions taken [208] when previously in a given environment. By contrast, goal-directed decisions have been described by forms of model-based learning and decision-making outlined above. Recent work considers how these two sets of algorithms work together [213, 214] and/or in competition [208, 215, 216] to determine behavior [217]. Similar approaches have also been taken towards modeling interactions between goal-directed behavior and behavior that is driven by Pavlovian associations between a situation and potential outcomes (e.g., a dose of nicotine) and therefore biased towards reflexive ("impulsive") actions [218–221].

Another area of shared inquiry between models of learning and decision-making concerns sources of variability in choice behavior that are not otherwise explained by features of the environment. These are collectively denoted by the term decision "noise," and incorporated as such into algorithms for action selection (e.g., the softmax), but there are layers to this noise that can reflect elements of learning, inference, motor execution, and even strategy [108], making such noise part of a computational process approximating intractable computations [76]. For instance, work has shown that people preferentially engage in apparently noisy decision-making during periods where it is optimal for them to explore their environment, in those instances either deliberately choosing an option that currently seems worse (*directed exploration*) or choosing at random (*random exploration*) [109, 222, 223]. How these exploratory behaviors impact learning remains an underexplored area [224, 225].

Another area that has recently gained significant momentum examines how existing models can be augmented to account for the role that affective experiences play in shaping learning and decision making. Research on this topic varies from studies that treat mood and affect as an isolated set of processes that act on parameters of the models described earlier—such as the impact of stress/anxiety on model-based planning [60, 226]—to more recent work that examines how affect itself can be understood through the lens of these model parameters. As an example of the latter, Rutledge and colleagues [227] have recently proposed that affective states may reflect an integration over recent outcome prediction errors, such that people feel happier after experiencing a sequence of large reward prediction errors (see also [228]). Complementary research has examined how these changes in affective states feed back into one's estimates of future outcomes, resulting in biases in learning and choice [229–231].

CLINICAL IMPLICATIONS

Computational models of learning and decision making are an important tool to further our understanding of the mechanisms of cognition, and as such carry a great deal of promise for the study of clinical conditions. The nascent field of computational psychiatry aims to bridge the gap between neural substrates and behavior, cognition and emotions, by exposing the mechanisms that support them [232–235]. For this approach to be successful, it is crucial that the computational models reflect computational components that can indeed be related to brain function. In that sense, recent research often highlights both the risks and benefits of the computational psychiatry approach. For example, learning impairments in patients with schizophrenia are very task-dependent, in such a way that RL modeling has led to different conclusions across studies, despite their ability to capture behavior well [236]. Taking into account PFC function in the form of working memory, recent studies showed that these learning impairments could be attributed to impairments in working memory (and its downstream impact on learning performance) but not to impairments in RL mechanisms themselves [237, 238].

This reconciles previous findings by suggesting that different tasks recruited working memory to different degrees, leading to observed learning impairments in some tasks and not others. RL modeling without considerations of working memory would mistakenly attribute this to RL functioning. Similar observations have been made with respect to attention—it was shown that poorer learning performance in older adults during complex tasks was not a reflection of impaired RL mechanisms but impaired attention [239].

Variability in decision-making across clinical populations has been observed at the level of how a given population weighs the subjective value of decision options, for instance increased risk aversion in anxiety disorders [240, 241]; increased delay discounting in addiction [242, 243] and increased effort discounting in depression [244, 245]. However, it remains unclear whether the different choices these individuals are making (relative to healthy controls) reflect differences in how they inherently *value* relevant costs and benefits, or in how they generate information relevant to their decision (e.g., which attributes they attend to, which episodes are drawn from memory) and/or which heuristics/strategies they use to help make their choice. Careful modeling of the multiple contributors to performance on learning and decision making tasks—and PFC's role in each of these—is therefore essential to guiding research into which underlying neural substrates might be impaired.

FUTURE RESEARCH DIRECTIONS
Much research remains to be done to propose better quantitative models of learning and decision-making. We highlight here a few (non-exhaustive) directions for future research. First, as discussed above, recent research increasingly takes into consideration the fact that multiple systems contribute jointly to learning and decision-making, where each system could make a choice of its own, based on slightly different information. So far, this important crossing of boundaries across systems has mostly been limited to studying their competitive arbitration, and the "meta"-decisions around them: for example, when should I use planning rather than relying on cached values? However, an important question for future research is whether these systems are more tightly interlaced than a simple competitive interaction for choice. Preliminary evidence suggests that, even when a system is operational on its own, it can additionally be influenced by other systems: for example, the content of working memory appears to modify how RL reward prediction errors are computed [57, 246, 247], in such a way that when we can rely on WM to learn (e.g., under low load), we do not retain the information as well [65]. An important direction for future research will be to better understand how very disparate systems influence each other's computations. This will involve dissecting the mechanisms through which the same information is processed by different systems—for example, how outcome expectations are processed by the RL, WM, and episodic memory systems—and identifying the manner and extent to which information in one system feeds into the others.

Another area for future research is to better understand the representations that these computations operate over [18, 140, 248]. We focused earlier on state and action representation, where many questions remain open, but another (arguably even more fundamental) question pertains to how people represent *outcomes*, in particular what counts as a reward. It has become evident that the answer to this is less straightforward than often implied, as reinforcers are often context-dependent [112], and neural representations of choice value can be less reflective of how good one's options are (the *reward value* itself) than they are reflective of *how well-aligned those reward values are with one's immediate goal* (e.g., tracking reward value positively when the aim is to choose the best option, and tracking reward value negatively when the aim is to choose the worst option [127, 249, 250]). Indeed, it remains mysterious how humans are even able to so efficiently endow even the most arbitrary goals with value to support learning and decision-making [251]. Building on recent progress in understanding the role of context in shaping evaluation [252], future work should seek to better understand how one's evaluative goals (both extrinsic and intrinsic) shape their consideration of potential outcomes.

In addition to better understanding the evaluation of prospective outcomes, it is also important to better understand the discounting of those outcomes by prospective costs such as mental or physical effort. Extensive research has demonstrated when and how people discount outcome values by such costs, but *why* they do so (i.e., what gives rise to effort costs) remains mysterious. Previous theories that either form of effort cost primarily reflects bottom-up resource constraints (e.g., muscle fatigue, depleted glucose) have seen mixed evidence, and have given way to proposals that effort costs largely reflect top-down constraints on expected effort output [253–255]. The basis for these top-down constraints remains heavily debated, but includes value-based models proposing that effort costs and their temporal dynamics (e.g., fatigue) may reflect evaluations of the opportunity costs (i.e., the value of foregone alternatives) when engaging in a given form of action or task [253, 256, 257]. How these opportunity costs are estimated, and what costs this evaluation process incurs, are themselves open questions. A further challenge for these and alternate accounts of effort costs (eg, [258, 259]) remains accounting for the variety of situations in which effort appears to serve as a reward (i.e., something that people seek out) rather than or in addition to serving as a cost [260]. More fundamentally, the question of how to reliably estimate the costs (and/or reward) for effort—for instance, using combinations of behavior, self-report, physiological responses, and/or neural activity—also remains a pressing challenge for future work [159].

There has been tremendous progress in AI and its ability to excel in complex tasks previously thought to be human benchmarks. This invariably raises the question of whether these newer AIs could support better models of learning and decision making. Indeed, successful efforts to bridge deep neural networks with brain function in the domain of perception [261, 262] offer some promise. Early efforts in the domain of learning and decision making show that correlates of brain function, including PFC, can be found when deep RL agents perform complex tasks [263, 264]. However, at this point, it remains unclear if this approach will provide insights into underlying mechanisms. Instead, early attempts show that deep neural networks can be a useful analytical tool for predicting (rather than explaining) choices [265] or for fitting classic cognitive models [266, 267]. Biologically realistic neural networks, by contrast, remain an important area of future research for computational modeling of decision making and learning. Indeed, they offer the resolution necessary to test specific implementational predictions, such as the role of the subthalamic nucleus in response inhibition [149], or the role of excitatory-inhibitory balance in networks supporting decision making [263]. Such models bridge the gap between brain and behavior better than simpler cognitive models do, especially when considered in conjunction [14, 48], and may be a stepping stone to bringing this knowledge to AI. Extending this approach to modeling learning and decision-making functions that are more dependent on PFC function is an area of active interest [56, 268, 269], but remains limited and an important direction for future research.

CONCLUSIONS
Computational modeling is an important tool for developing a quantitative understanding of the mechanisms by which animals

adapt to their environments, and what leads to maladaptive forms of learning and decision-making within clinical populations. However, to fulfill this promise, it is essential that models successfully bridge across levels of analysis [9, 10], offering algorithms that both relate to brain mechanisms and explain the function of the computations. Recent research has successfully started moving away from more simplistic models to models that better encompass the complexity and constraints of learning and decision-making, as supported by multiple distinct, interacting neural systems. This direction remains important for future research, and holds great promise for a more nuanced understanding of adaptive decision-making.

REFERENCES

1. Averbeck B, O'Doherty JP. Reinforcement-learning in fronto-striatal circuits. Neuropsychopharmacology. 2021. https://doi.org/10.1038/s41386-021-01108-0.
2. Monosov IE, Rushworth MF. Interactions between ventrolateral prefrontal and anterior cingulate cortex during learning and behavioural change. Neuropsychopharmacology. 2021. https://doi.org/10.1038/s41386-021-01079-2.
3. Friedman, N.P., Robbins, T.W. The role of prefrontal cortex in cognitive control and executive function. Neuropsychopharmacology. (2021). https://doi.org/10.1038/s41386-021-01132-0.
4. Dickinson A, Mackintosh NJ. Classical conditioning in animals. Annu Rev Psychol. 1978;29:587–612.
5. Wagner AR, Rescorla RA. Inhibition in Pavlovian conditioning: application of a theory. Inhibition and learning. 1972:301–36.
6. Skinner BF. Conditioning and extinction and their relation to drive. J Gen Psychol. 1936;14:296–317.
7. Montague P, Dayan P, Sejnowski T. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. 1996;16:1936–47.
8. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997;275:1593–9.
9. Marr D. Vision: a computational approach. Freeman & Co.: San Francisco; 1982.
10. Niv Y, Langdon A. Reinforcement learning with Marr. Curr Opin Behav Sci. 2016;11:67–73.
11. Samejima K. Representation of action-specific reward values in the striatum. Science 2005;310:1337–40.
12. Tai L-H, Lee AM, Benavidez N, Bonci A, Wilbrecht L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. Nat Neurosci. 2012;15:1281–\s9.
13. Calabresi P, Picconi B, Tozzi A, Di Filippo M. Dopamine-mediated regulation of corticostriatal synaptic plasticity. Trends Neurosci. 2007;30:211–9.
14. Collins AGE, Frank MJ. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol Rev. 2014;121:337–66.
15. Frank MJ. By Carrot or by Stick: cognitive reinforcement learning in Parkinsonism. Science. 2004;306:1940–3.
16. Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci. 1986;9:357–81.
17. Hazy TE, Frank MJ, O'Reilly RC. Banishing the homunculus: making working memory work. Neuroscience 2006;139:105–18.
18. Rmus M, McDougle SD, Collins AG. The role of executive function in shaping reinforcement learning. Curr Opin Behav Sci. 2021;38:66–73.
19. Sutton RS, Barto AG. Reinforcement learning: an introduction. MIT Press: Cambridge, Mass; 1998.
20. Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, et al. Reinforcement learning in multidimensional environments relies on attention mechanisms. J Neurosci. 2015;35:8145–57.
21. Leong YC, Radulescu A, Daniel R, DeWoskin V, Niv Y. Dynamic interaction between reinforcement learning and attention in multidimensional environments. Neuron 2017;93:451–63.
22. Wilson RC, Niv Y. Inferring relevance in a changing world. Front Hum Neurosci. 2012;5.
23. Farashahi S, Xu J, Wu S-W, Soltani A. Learning arbitrary stimulus-reward associations for naturalistic stimuli involves transition from learning about features to learning about objects. Cognition. 2020;205:104425.
24. Song MR, Lee SW. Dynamic resource allocation during reinforcement learning accounts for ramping and phasic dopamine activity. Neural Netw. 2020;126:95–107.
25. Babayan BM, Uchida N, Gershman SJ. Belief state representation in the dopamine system. Nat Commun. 2018;9:1891.
26. Gershman SJ, Niv Y. Learning latent structure: carving nature at its joints. Curr Opin Neurobiol. 2010;20:251–6.
27. Gershman SJ, Uchida N. Believing in dopamine. Nat Rev Neurosci. 2019;20:703–14.
28. Niv Y. Learning task-state representations. Nat Neurosci. 2019;22:1544–53.
29. Sanders H, Wilson MA, Gershman SJ. Hippocampal remapping as hidden state inference. eLife. 2020;9:e51140.
30. Schuck NW, Wilson R, Niv Y. A state representation for reinforcement learning and decision-making in the orbitofrontal cortex. Goal-directed decision making. Elsevier; 2018. p. 259–78.
31. Wilson Robert C, Takahashi Yuji K, Schoenbaum G, Niv Y. Orbitofrontal cortex as a cognitive map of task space. Neuron. 2014;81:267–79.
32. Schuck Nicolas W, Cai Ming B, Wilson Robert C, Niv Y. Human orbitofrontal cortex represents a cognitive map of state space. Neuron. 2016;91:1402–12.
33. Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat Rev Neurosci. 2009;10:885–92.
34. Zhou J, Gardner MPH, Stalnaker TA, Ramus SJ, Wikenheiser AM, Niv Y, et al. Rat orbitofrontal ensemble activity contains multiplexed but dissociable representations of value and task structure in an odor sequence task. Curr Biol. 2019;29:897–907.e3.
35. Brunec, IK, & Momennejad, I Predictive representations in hippocampal and prefrontal hierarchies. BioRxiv. 2019;786434.
36. Momennejad I. Learning structures: predictive representations, replay, and generalization. Curr Opin Behav Sci. 2020;32:155–66.
37. Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, et al. The Tolman-Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. Cell. 2020;183:1249–63.e23.
38. Morris A, Phillips JS, Huang K, Cushman FA Generating options and choosing between them rely on distinct forms of value representation. Psychol Sci. in press.
39. Botvinick MM, Niv Y, Barto AG. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. Cognition. 2009;113:262–80.
40. Cooper RP, Shallice T. Hierarchical schemas and goals in the control of sequential behavior. Psychol Rev. 2006;113:887–916.
41. Solway A, Diuk C, Córdova N, Yee D, Barto AG, Niv Y, et al. Optimal behavioral hierarchy. PLoS Comput Biol. 2014;10:e1003779.
42. Xia L, Collins AGE. Temporal and state abstractions for efficient learning, transfer and composition in humans. Psychol Rev. 2021;128:643–66.
43. Diuk C, Tsai K, Wallis J, Botvinick M, Niv Y. Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. J Neurosci. 2013;33:5797–805.
44. Ribas-Fernandes José JF, Solway A, Diuk C, McGuire Joseph T, Barto Andrew G, Niv Y, et al. A neural signature of hierarchical reinforcement learning. Neuron 2011;71:370–9.
45. Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia. 2007;45:2883–901.
46. Koechlin E. The architecture of cognitive control in the human prefrontal cortex. Science. 2003;302:1181–5.
47. Collins AGE, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis: Working memory in reinforcement learning. Eur J Neurosci. 2012;35:1024–35.
48. Collins AGE, Frank MJ. Cognitive control over learning: creating, clustering, and generalizing task-set structure. Psychol Rev. 2013;120:190–229.
49. Donoso M, Collins AGE, Koechlin E. Foundations of human reasoning in the prefrontal cortex. Science. 2014;344:1481–6.
50. Alexander WH, Brown JW. Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex. Neural Comput. 2015;27:2354–410.
51. Zarr N, Brown JW. Hierarchical error representation in medial prefrontal cortex. NeuroImage. 2016;124:238–47.
52. Frank MJ, Badre D. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. Cereb Cortex. 2012;22:509–26.
53. Ballard I, Miller EM, Piantadosi ST, Goodman ND, McClure SM. Beyond reward prediction errors: human striatum updates rule values during learning. Cereb Cortex. 2018;28:3965–75.
54. Eckstein MK, Collins AGE. Computational evidence for hierarchically structured reinforcement learning in humans. Proc Natl Acad Sci. 2020;117:29381–9.
55. Collins A, Koechlin E. Reasoning, learning, and creativity: frontal lobe function and human decision-making. PLoS Biol. 2012;10:e1001293.
56. Badre D, Frank MJ. Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: evidence from fMRI. Cereb Cortex. 2012;22:527–36.
57. Daw Nathaniel D, Gershman Samuel J, Seymour B, Dayan P, Dolan Raymond J. Model-based influences on humans' choices and striatal prediction errors. Neuron. 2011;69:1204–15.

58. Stanovich KE, West RF. Individual differences in reasoning: implications for the rationality debate? Behav Brain Sci. 2000;23:645–65.

59. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND. Model-based choices involve prospective neural activity. Nat Neurosci. 2015;18:767–72.

60. Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND. Working-memory capacity protects model-based learning from stress. Proc Natl Acad Sci. 2013;110:20941–46.

61. Collins AGE, Cockburn J. Beyond dichotomies in reinforcement learning. Nat Rev Neurosci. 2020;21:576–86.

62. Miller KJ, Ludvig EA, Pezzulo G, Shenhav A. Re-aligning models of habitual and goal-directed decision-making. In: Bornstein AM, Morris RW, Shenhav A, editors. Goal-directed decision making: computations and neural circuits. Amsterdam: Elsevier; 2018. p. 407–28.

63. Moran R, Keramati M, Dayan P, Dolan RJ. Retrospective model-based inference guides model-free credit assignment. Nat Commun. 2019;10:750.

64. Yang GR, Joglekar MR, Song HF, Newsome WT, Wang XJ. Task representations in neural networks trained to perform many cognitive tasks. Nat Neurosci. 2019;22:297–306.

65. Collins AGE. The tortoise and the hare: interactions between reinforcement learning and working memory. J Cogn Neurosci. 2018;30:1422–32.

66. Viejo G, Khamassi M, Brovelli A, Girard B. Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. Front Behav Neurosci. 2015;9.

67. McDougle SD, Collins AGE. Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. Psychon Bull Rev. 2021;28:20–39.

68. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci. 2007;104:16311–16.

69. Poldrack RA, Clark J, Paré-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, et al. Interactive memory systems in the human brain. Nature. 2001;414:546–50.

70. Foerde K, Shohamy D. The role of the basal ganglia in learning and memory: Insight from Parkinson's disease. Neurobiol Learn Mem. 2011;96:624–36.

71. Wimmer GE, Braun EK, Daw ND, Shohamy D. Episodic memory encoding interferes with reward learning and decreases striatal prediction errors. J Neurosci. 2014;34:14901–12.

72. Bornstein AM, Khaw MW, Shohamy D, Daw ND. Reminders of past choices bias decisions for reward in humans. Nat Commun. 2017;8:15958.

73. Bornstein AM, Norman KA. Reinstated episodic context guides sampling-based decisions for reward. Nat Neurosci. 2017;20:997–1003.

74. Vikbladh OM, Meager MR, King J, Blackmon K, Devinsky O, Shohamy D, et al. Hippocampal contributions to model-based planning and spatial memory. Neuron. 2019;102:683–93.e4.

75. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. Nat Neurosci. 2007;10:1214–21.

76. Findling C, Chopin N, Koechlin E. Imprecise neural computations as a source of adaptive behaviour in volatile environments. Nat Hum Behav. 2021;5:99–112.

77. Brown VM, Zhu L, Wang JM, Frueh BC, King-Casas B, Chiu PH. Associability-modulated loss learning is increased in posttraumatic stress disorder. eLife. 2018;7:e30150.

78. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND. Differential roles of human striatum and amygdala in associative learning. Nat Neurosci. 2011;14:1250–2.

79. Nassar MR, Wilson RC, Heasly B, Gold JI. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J Neurosci. 2010;30:12366–78.

80. Bavard S, Lebreton M, Khamassi M, Coricelli G, Palminteri S. Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. Nat Commun. 2018;9:4503.

81. Boorman ED, Behrens TE, Rushworth MF. Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. PLoS Biol. 2011;9:e1001093.

82. Palminteri S, Lefebvre G, Kilford EJ, Blakemore S-J. Confirmation bias in human reinforcement learning: evidence from counterfactual feedback processing. PLOS Comput Biol. 2017;13:e1005684.

83. Mohr H, Zwosta K, Markovic D, Bitzer S, Wolfensteller U, Ruge H. Deterministic response strategies in a trial-and-error learning task. PLoS Comput Biol. 2018;14:e1006621.

84. Thaler RH. Behavioral economics: past, present, and future. Am Econ Rev. 2016;106:1577–600.

85. Stewart N, Reimers S, Harris AJ. On the origin of utility, weighting, and discounting functions: how they get their shapes and how to change their shapes. Manag Sci. 2015;61:687–705.

86. Mitchell SH. Discounting the value of commodities according to different types of cost. Choice, behavioural economics and addiction. 2003. p. 339–62.

87. Chong T, Bonnelle V, Husain M. Quantifying motivation with effort-based decision-making paradigms in health and disease. Prog Brain Res. 2016;229:71–100.

88. Tobler PN, Weber EU Valuation for risky and uncertain choices. In: Glimcher PW, Fehr E, editors. Neuroeconomics: Decision Making and the Brain: Second Edition. Academic Press; 2014. p. 149-72.

89. Kable JW. Valuation, intertemporal choice, and self-control. In: Glimcher PW, Fehr E, editors. Neuroeconomics: Decision Making and the Brain: Second Edition. Academic Press; 2014. p. 173-92.

90. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econom J Econom Soc. 1979;47:263–91.

91. Mata R, Frey R, Richter D, Schupp J, Hertwig R. Risk preference: a view from psychology. J Econ Perspect. 2018;32:155–72.

92. Berns GS, Laibson D, Loewenstein G. Intertemporal choice–toward an integrative framework. Trends Cogn Sci. 2007;11:482–8.

93. Klein-Flügge MC, Kennerley SW, Saraiva AC, Penny WD, Bestmann S. Behavioral modeling of human choices reveals dissociable effects of physical effort and temporal delay on reward devaluation. PLoS Comput Biol. 2015;11:e1004116.

94. Chong TTJ, Apps M, Giehl K, Sillence A, Grima LL, Husain M. Neurocomputational mechanisms underlying subjective valuation of effort costs. PLoS Biol. 2017;15:e1002598–28.

95. Levy I, Snell J, Nelson AJ, Rustichini A, Glimcher PW. Neural representation of subjective value under risk and ambiguity. J Neurophysiol. 2010;103:1036–47.

96. Preuschoff K, Bossaerts P, Quartz S. Neural differentiation of expected reward and risk in human subcortical structures. Neuron. 2006;51:381–90.

97. Tobler PN, O'Doherty JP, Dolan RJ, Schultz W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. J Neurophysiol. 2006;97:1621–32.

98. Tom SM, Fox CR, Trepel C, Poldrack RA. The neural basis of loss aversion in decision-making under risk. Science. 2007;315:515–8.

99. Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 2007;10:1625–33.

100. McClure SM, Laibson D, Loewenstein GF, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. Science. 2004;306:503–7.

101. Prévost C, Pessiglione M, Météreau E, Cléry-Melin M, Dreher J. Separate valuation subsystems for delay and effort decision costs. J Neurosci. 2010;30:14080–90.

102. Schmidt L, Lebreton M, Cléry-Melin M-L, Daunizeau J, Pessiglione M. Neural mechanisms underlying motivation of mental versus physical effort. PLoS Biol. 2012;10:e1001266.

103. Levy DJ, Glimcher PW. The root of all value: a neural common currency for choice. Curr Opin Neurobiol. 2012;22:1027–38.

104. Bartra O, McGuire JT, Kable JW. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage. 2013;76:412–27.

105. Seaman KL, Brooks N, Karrer TM, Dang L, Hsu M, Zald DH, et al. Neural subjective value representations across age and discount factors: time delay, physical effort, and probability discounting. Soc, Cogn, Affect Neurosci. 2018;13:449–59.

106. Oud B, Krajbich I, Miller K, Cheong JH, Botvinick M, Fehr E. Irrational time allocation in decision-making. Proc R Soc B. 2016;283:20151439–8.

107. Grueschow M, Polanía R, Hare TA, Ruff CC. Automatic versus choice-dependent value representations in the human brain. Neuron 2015;85:874–85.

108. Wyart V, Koechlin E. Choice variability and suboptimality in uncertain environments. Curr Opin Behav Sci. 2016;11:109–15.

109. Wilson RC, Bonawitz E, Costa VD, Ebitz RB. Balancing Exploration and Exploitation with Information and Randomization. Curr Opin Behav Sci. 2021;49-56:49–56.

110. Webb R. The (Neural) dynamics of stochastic choice. Manag Sci. 2019;64:230–55.

111. Becker GM, DeGroot MH, Marschak J. Stochastic models of choice behavior. Behav Sci. 1963;8:41–55.

112. Juechems K, Summerfield C. Where does value come from? Trends Cogn Sci. 2019;23:836–50.

113. Vlaev I, Chater N, Stewart N, Brown GDA. Does the brain calculate value? Trends Cogn Sci. 2011;15:546–54.

114. Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. Ann Math Stat. 1948;19:326–39.

115. Ratcliff R, Smith PL, Brown SD, McKoon G. Diffusion decision model: current issues and history. Trends Cogn Sci. 2016;20:260–81.

116. Shadlen MN, Kiani R. Decision making as a window on cognition. Neuron 2013;80:791–806.

117. Bogacz R. Optimal decision-making theories: linking neurobiology with behaviour. Trends Cogn Sci. 2007;11:118–25.

118. Teodorescu AR, Usher M. Disentangling Decision Models: From Independence to Competition. Psychol Rev. 2013;120:1–38.

119. Ratcliff R. A theory of memory retrieval. Psychol Rev. 1978;85:59–108.
120. Milosavljevic M, Malmaud J, Huth A, Koch C, Rangel A. The Drift Diffusion Model can account for the accuracy and reaction time of value-based choices under high and low time pressure. Judgm Decis Mak. 2010;5:437–49.
121. Hutcherson CA, Bushong B, Rangel A. A neurocomputational model of altruistic choice and its implications. Neuron. 2015;87:451–62.
122. Shenhav A, Straccia MA, Cohen JD, Botvinick MM. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. Nat Neurosci. 2014;16:1127–39.
123. Peters J, D'Esposito M. The drift diffusion model as the choice rule in intertemporal and risky choice: a case study in medial orbitofrontal cortex lesion patients and controls. PLoS Comput Biol. 2020;16:e1007615.
124. Clithero JA. Improving out-of-sample predictions using response times and a model of the decision process. J Econ Behav Organ. 2018;148:344–75.
125. Usher M, Mcclelland JL. The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev. 2001;108:550–92.
126. Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol. 2001;86:1916–36.
127. Frömer R, Dean Wolf CK, Shenhav A. Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. Nat Commun. 2019;10:4926.
128. Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MFS, Behrens TEJ. Mechanisms underlying cortical activity during value-guided choice. Nat Neurosci. 2012;15:470–6.
129. Wong KF, Wang XJ. A recurrent network mechanism of time integration in perceptual decisions. J Neurosci. 2006;26:1314–28.
130. Soltani A, Chaisangmongkon W, Wang XJ Neural circuit mechanisms of value-based decision-making and reinforcement learning. In: Dreher J, Tremblay L, editors. Decision Neuroscience. Academic Press; 2017. p. 163-76.
131. Hunt LT, Hayden BY. A distributed, hierarchical and recurrent framework for reward-based choice. Nat Rev Neurosci. 2017;18:172–82.
132. Enel P, Wallis JD, Rich EL. Stable and dynamic representations of value in the prefrontal cortex. Elife. 2020;9.
133. Padoa-Schioppa C. Neuronal origins of choice variability in economic decisions. Neuron. 2013;80:1322–36.
134. Rich EL, Wallis JD. Decoding subjective decisions from orbitofrontal cortex. Nat Neurosci. 2016;19:973–80.
135. Hare TA, Schultz W, Camerer CF, O'Doherty JP, Rangel A. Transformation of stimulus value signals into motor commands during simple choice. Proc Natl Acad Sci. 2011;108:18120–25.
136. Louie K, Glimcher PW. Separating value from choice: delay discounting activity in the lateral intraparietal area. J Neurosci. 2010;30:5498–507.
137. Gluth S, Rieskamp J, Buchel C. Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. J Neurosci. 2012;32:10686–98.
138. Gluth S, Rieskamp J, Buchel C. Classic EEG motor potentials track the emergence of value-based decisions. NeuroImage. 2013;79:394–403.
139. Pisauro MA, Fouragnan E, Retzler C, Philiastides MG. Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. Nat Commun. 2017;8:1–9.
140. Frömer R, Shenhav A. Filling the gaps: cognitive control as a critical lens for understanding mechanisms of value-based decision-making. PsyArXiv. 2021.
141. Pleskac TJ, Busemeyer JR. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychol Rev. 2010;117:864–901.
142. De Martino B, Fleming SM, Garrett N, Dolan RJ. Confidence in value-based choice. Nat Neurosci. 2013;16:105–10.
143. Desender K, Donner TH, Verguts T. Dynamic expressions of confidence within an evidence accumulation framework. Cognition. 2021;207:104522.
144. Resulaj A, Kiani R, Wolpert DM, Shadlen MN. Changes of mind in decision-making. Nature. 2009;461:263–6.
145. Yeung N, Summerfield C. Metacognition in human decision-making: confidence and error monitoring. Philos Trans R Soc Lond B Biol Sci. 2012;367:1310–21.
146. Desender K, Murphy P, Boldt A, Verguts T, Yeung N. A postdecisional neural marker of confidence predicts information-seeking in decision-making. J Neurosci. 2019;39:3309–19.
147. Rouault M, Dayan P, Fleming SM. Forming global estimates of self-performance from local confidence. Nat Commun. 2019;10:1141.
148. Desender K, Boldt A, Yeung N. Subjective confidence predicts information seeking in decision making. Psychol Sci. 2018;29:761–78.
149. Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF, et al. fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. J Neurosci. 2015;35:485–94.
150. Wiecki TV, Frank MJ. A computational model of executive control in frontal cortex and basal ganglia: multiple levels of analysis. Psychol Rev. 2013;120:329–55.
151. Frömer R, Shenhav A. Spatiotemporally distinct neural mechanisms underlie our reactions to and comparison between value-based options. BioRxiv. 2019; 609198.
152. Hanks TD, Kopec CD, Brunton BW, Duan CA, Erlich JC, Brody CD. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature. 2015;520:220–3.
153. Erlich JC, Brunton BW, Duan CA, Hanks TD, Brody CD. Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. eLife. 2015;4:e05457.
154. Shenhav A, Straccia MA, Botvinick MM, Cohen JD. Dorsal anterior cingulate and ventromedial prefrontal cortex have inverse roles in both foraging and economic choice. Cogn Affect Behav Neurosci. 2016;19:1286–91.
155. Wolpert DM, Landy MS. Motor control is decision-making. Curr Opin Neurobiol. 2012;22:996–1003.
156. Manohar SG, Chong TTJ, Apps MAJ, Batla A, Stamelou M, Jarman PR, et al. Reward Pays the Cost of Noise Reduction in Motor and Cognitive Control. Curr Biol. 2015;13:1707–16.
157. Todorov E. Efficient computation of optimal actions. Proc Natl Acad Sci USA. 2009;106:11478–83.
158. Shenhav A, Botvinick MM, Cohen JD. The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron 2013;79:217–40.
159. Shenhav A, Musslick S, Lieder F, Kool W, Griffiths TL, Cohen JD, et al. Toward a rational and mechanistic account of mental effort. Annu Rev Neurosci. 2017;40:99–124.
160. Frömer R, Lin H, Dean Wolf CK, Inzlicht M, Shenhav A. Expectations of reward and efficacy guide cognitive control allocation. Nat Commun. 2021;12.
161. Shenhav A, Cohen JD, Botvinick MM. Dorsal anterior cingulate cortex and the value of control. Nat Neurosci. 2016;19:1286–91.
162. Nachev P, Kennard C, Husain M. Functional role of the supplementary and pre-supplementary motor areas. Nat Rev Neurosci. 2008;9:856–69.
163. Venkatraman V, Huettel SA. Strategic control in decision-making under uncertainty. Eur J Neurosci. 2012;35:1075–82.
164. Cisek P, Pastor-Bernier A. On the challenges and mechanisms of embodied decisions. Philos Trans R Soc Lond B Biol Sci. 2014;369.
165. Solway A, Botvinick MM. Evidence integration in model-based tree search. Proc Natl Acad Sci. 2015;112:11708–13.
166. Hunt LT, Dolan RJ, Behrens TEJ. Hierarchical competitions subserving multi-attribute choice. Nat Neurosci. 2014;17:1613–22.
167. Solway A, Botvinick MM. Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. Psychol Rev. 2012;119:120–54.
168. Holroyd CB, McClure SM. Hierarchical control over effortful behavior by rodent medial frontal cortex: a computational model. Psychol Rev. 2015;122:54–83.
169. Busemeyer JR, Gluth S, Rieskamp J, Turner BM. Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. Trends Cogn Sci. 2019;23:251–63.
170. Tversky A. Elimination by aspects: a theory of choice. Psychol Rev. 1972;79:281–99.
171. Busemeyer JR, Townsend JT. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol Rev. 1993;100:432–59.
172. Bogacz R, Usher M, Zhang J, Mcclelland JL. Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multi-dimensional choice. Philos Trans R Soc Lond, Ser B Biol Sci. 2007;362:1655–70.
173. Harris A, Clithero JA, Hutcherson CA. Accounting for taste: a multi-attribute neurocomputational model explains the neural dynamics of choices for self and others. J Neurosci. 2018;38:7952–68.
174. Krajbich I, Armel C, Rangel A. Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci. 2010;13:1292–8.
175. Krajbich I, Rangel A. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. Proc Natl Acad Sci. 2011;108:13852–7.
176. Armel KC, Beaumel A, Rangel A. Biasing simple choices by manipulating relative visual attention. Judgm Decis Mak. 2008;3:396–403.
177. Cavanagh JF, Wiecki TV, Kochar A, Frank MJ. Eye Tracking and Pupillometry Are Indicators of Dissociable Latent Decision Processes. Journal of Experimental Psychology General. 2014;143:1476–88.
178. Westbrook A, van den Bosch R, Maatta JI, Hofmans L, Papadopetraki D, Cools R, et al. Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. Science. 2020;367:1362–6.
179. Callaway F, Rangel A, Griffiths TL. Fixation patterns in simple choice are consistent with optimal use of cognitive resources. PLoS Computational Biology. 2021;17:e1008863.
180. Callaway F, van Opheusden B, Gul S, Das P, Krueger P, Lieder F, et al. Human planning as optimal information seeking. PsyArXiv. 2021.

181. Jang AI, Sharma R, Drugowitsch J. Optimal policy for attention-modulated decisions explains human fixation behavior. eLife. 2021;10:e63436.

182. Polania R, Woodford M, Ruff CC. Efficient coding of subjective value. Nat Neurosci. 2019;22:134–42.

183. Frömer R, Callaway F, Griffiths TL, Shenhav A. Considering what we know and what we don't know: expectations and confidence guide value integration in value-based decision-making. in prep.

184. Hare TA, Camerer C, Rangel A. Self-control in decision-making involves modulation of the vmPFC valuation system. Science. 2009;324:646–48.

185. Hare TA, Malmaud J, Rangel A. Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. J Neurosci. 2011;31:11077–87.

186. Lim S-L, O'doherty JP, Rangel A. Stimulus value signals in ventromedial pfc reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. J Neurosci. 2013;33:8729–41.

187. Shenhav A, Greene JD. Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. J Neurosci. 2014;34:4741–9.

188. Hutcherson CA, Montaser-Kouhsari L, Woodward J, Rangel A. Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. J Neurosci. 2015;35:12593–605.

189. Tusche A, Hutcherson CA. Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. Elife. 2018;7:e31185.

190. Tversky A, Simonson I. Context-dependent preferences. Manag Sci. 1993;39:1179–89.

191. Tsetsos K, Usher M, Chater N. Preference reversal in multiattribute choice. Psychol Rev. 2010;117:1275–91.

192. Reynolds JH, Heeger DJ. The normalization model of attention. Neuron. 2009;61:168–85.

193. Summerfield C, Tsetsos K. Do humans make good decisions? Trends Cogn Sci. 2015;19:27–34.

194. Louie K, Glimcher PW. Efficient coding and the neural representation of value. Ann NY Acad Sci. 2012;1251:13–32.

195. Louie K, Khaw MW, Glimcher PW. Normalization is a general neural mechanism for context-dependent decision making. Proc Natl Acad Sci. 2013;110:6139–44.

196. Gluth S, Kern N, Kortmann M, Vitali CL. Value-based attention but not divisive normalization influences decisions with multiple alternatives. Nat Hum Behav. 2020;4:634–45.

197. Khaw MW, Glimcher PW, Louie K. Normalized value coding explains dynamic adaptation in the human valuation process. Proc Natl Acad Sci. 2017;114:12696–701.

198. Frömer R, Shenhav A Overriding first impressions: evidence for a reference-dependent and attentionally-weighted multi-stage process of value-based decision-making, in The 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making. 2019.

199. Bornstein AM, Pickard H. "Chasing the first high": memory sampling in drug choice. Neuropsychopharmacology. 2020;45:907–15.

200. Bakkour A, Zylberberg A, Shadlen MN, Shohamy D. Value-based decisions involve sequential sampling from memory. BioRxiv. 2018; 269290.

201. Vaidya AR, Badre D. Neural systems for memory-based value judgment and decision-making. J Cogn Neurosci. 2020;32:1896–923.

202. Ludvig EA, Madan CR, McMillan N, Xu Y, Spetch ML. Living near the edge: how extreme outcomes and their neighbors drive risky choice. J Exp Psychol Gen. 2018;147:1905–18.

203. Madan CR, Ludvig EA, Spetch ML. Remembering the best and worst of times: memories for extreme outcomes bias risky decisions. Psychon Bull Rev. 2014;21:629–36.

204. Lieder F, Griffiths TL, Hsu M. Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. Psychol Rev. 2018;125:1–32.

205. Pedersen ML, Frank MJ, Biele G. The drift diffusion model as the choice rule in reinforcement learning. Psychon Bull Rev. 2017;24:1234–51.

206. Ballard IC, McClure SM. Joint modeling of reaction times and choice improves parameter identifiability in reinforcement learning models. J Neurosci Methods. 2019;317:37–44.

207. Miletic S, Boag RJ, Trutti AC, Stevenson N, Forstmann BU, Heathcote A. A new model of decision processing in instrumental learning tasks. Elife. 2021;10:e63055.

208. Miller KJ, Shenhav A, Ludvig EA. Habits without values. Psychol Rev. 2019;126:292.

209. Lau B, Glimcher PW. Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav. 2005;84:555–79.

210. Urai AE, de Gee JW, Tsetsos K, Donner TH. Choice history biases subsequent evidence accumulation. Elife. 2019;8:e46331.

211. Katahira K, Toyama A. Revisiting the importance of model fitting for model-based fMRI: It does matter in computational psychiatry. PLoS Comput Biol. 2021;17:e1008738.

212. van der Meer M, Kurth-Nelson Z, Redish AD. Information processing in decision-making systems. Neuroscientist. 2012;18:342–59.

213. Sutton RS. First results with Dyna, an interesting architecture for learning, planning, and reacting. In: Miller T, RS Sutton, Werbos P, editors. Neural networks for control. Cambridge, MA: MIT Press; 1990. p. 179–9.

214. Gershman SJ, Markman AB, Otto AR. Retrospective revaluation in sequential decision making: a tale of two systems. J Exp Psychol Gen. 2012;143:182.

215. Keramati M, Dezfouli A, Piray P. Speed/accuracy trade-off between the habitual and the goal-directed processes. PLoS Comput Biol. 2011;7:e1002055.

216. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005;8:1704–11.

217. Kool W, Cushman FA, Gershman SJ. Competition and cooperation between multiple reinforcement learning systems. In: Bornstein AM, Morris RW, Shenhav A, editors. Goal-Directed Decision Making: Computations and Neural Circuits. Amsterdam: Elsevier; 2018. p. 153–78.

218. Bach DR, Dayan P. Algorithms for survival: a comparative perspective on emotions. Nat Rev Neurosci. 2017;18:311–9.

219. Dayan P, Niv Y, Seymour B, Daw ND. The misbehavior of value and the discipline of the will. Neural Netw. 2006;19:1153–60.

220. Cavanagh JF, Eisenberg I, Guitart-Masip M, Huys Q, Frank MJ. Frontal theta overrides pavlovian learning biases. J Neurosci. 2013;33:8541–8.

221. Rangel A. Regulation of dietary choice by the decision-making circuitry. Nat Neurosci. 2013;16:1717–24.

222. Wilson RC, Wang S, Sadeghiyeh H, Cohen JD. Deep exploration as a unifying account of explore-exploit behavior. PsyArXiv. 2020.

223. Wang S, Wilson RC. Any way the brain blows? The nature of decision noise in random exploration. PsyArXiv. 2018.

224. Schulz E, Gershman SJ. The algorithmic architecture of exploration in the human brain. Curr Opin Neurobiol. 2019;55:7–14.

225. Schulz E, Bhui R, Love BC, Brier B, Todd MT, Gershman SJ. Structured, uncertainty-driven exploration in real-world consumer choice. Proc Natl Acad Sci USA. 2019;116:13903–8.

226. Huys QJM, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS Comput Biol. 2012;8:e1002410.

227. Rutledge RB, Skandali N, Dayan P, Dolan RJ. A computational and neural model of momentary subjective well-being. Proc Natl Acad Sci. 2014;111:12252–57.

228. Bennett D, Davidson G, Niv Y. A model of mood as integrated advantage. PsyArXiv. 2020.

229. Eldar E, Rutledge RB, Dolan RJ, Niv Y. Mood as representation of momentum. Trends Cogn Sci. 2016;20:15–24.

230. Eldar E, Roth C, Dayan P, Dolan RJ. Decodability of reward learning signals predicts mood fluctuations. Curr Biol. 2018;28:1433–39 e7.

231. Vinckier F, Rigoux L, Oudiette D, Pessiglione M. Neuro-computational account of how mood fluctuations arise and affect decision making. Nat Commun. 2018;9:1708.

232. Moutoussis M, Eldar E, Dolan RJ. Building a new field of computational psychiatry. Biol Psychiatry. 2017;82:388–90.

233. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends Cogn Sci. 2012;16:72–80.

234. Huys Q, Maia TV, Frank MJ. Computational psychiatry as a bridge between neuroscience and clinical applications. Nat Neurosci. 2016;19:404–13.

235. Huys QJ, Guitart-Masip M, Dolan RJ, Dayan P. Decision-theoretic psychiatry. Clin Psychol Sci. 2015;3:400–21.

236. Deserno L, Sterzer P, Wustenberg T, Heinz A, Schlagenhauf F. Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia. J Neurosci. 2012;32:12–20.

237. Collins AGE, Brown JK, Gold JM, Waltz JA, Frank MJ. Working memory contributions to reinforcement learning impairments in schizophrenia. J Neurosci. 2014;34:13747–56.

238. Collins AGE, Albrecht MA, Waltz JA, Gold JM, Frank MJ. Interactions among working memory, reinforcement learning, and effort in value-based choice: a new paradigm and selective deficits in schizophrenia. Biol Psychiatry. 2017;82:431–9.

239. Daniel R, Radulescu A, Niv Y. Intact reinforcement learning but impaired attentional control during multidimensional probabilistic learning in older adults. J Neurosci. 2020;40:1084–96.

240. Charpentier CJ, Aylward J, Roiser JP, Robinson OJ. Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. Biol Psychiatry. 2017;81:1014–22.

241. Bishop SJ, Gagne C. Anxiety, depression, and decision making: a computational perspective. Annu Rev Neurosci. 2018;41:371–88.

242. Kirby K, Petry N, Bickel W. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. J Exp Psychol Gen. 1999;128:78–87.

243. Bickel WK, Miller ML, Yi R, Kowal BP, Lindquist DM, Pitcock JA. Behavioral and neuroeconomics of drug addiction: competing neural systems and temporal discounting processes. Drug Alcohol Depend. 2007;90:S85–S91.

244. Cooper JA, Arulpragasam AR, Treadway MT. Anhedonia in depression: biological mechanisms and computational models. Curr Opin Behav Sci. 2018;22:128–35.

245. Treadway MT, Buckholtz JW, Schwartzman AN, Lambert WE, Zald DH. Worth the 'EEfRT'? The effort expenditure for rewards task as an objective measure of motivation and anhedonia. PLoS One. 2009;4:e6598.

246. Collins AGE, Ciullo B, Frank MJ, Badre D. Working memory load strengthens reward prediction errors. J Neurosci. 2017;37:4332–42.

247. Collins AGE, Frank MJ. Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. Proc Natl Acad Sci USA. 2018;115:2502–7.

248. Eckstein MK, Wilbrecht L, Collins AGE. What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience. Curr Opin Behav Sci. 2021;41:128–37.

249. Frömer R, Sturmer B, Sommer W. The better, the bigger: The effect of graded positive performance feedback on the reward positivity. Biol Psychol. 2016;114:61–8.

250. Frömer R, Nassar MR, Bruckner R, Stuermer B, Sommer W, Yeung N. Response-based outcome predictions and confidence regulate feedback processing and learning. Elife. 2021;10:e62825.

251. McDougle SD, Ballard IC, Baribault B, Bishop SJ, Collins AG. Executive function supports single-shot endowment of value to arbitrary transient goals. Cereb Cortex. 2021.

252. Hunter LE, Daw ND. Context-sensitive valuation and learning. Curr Opin Behav Sci. 2021;41:122–7.

253. Kurzban R, Duckworth A, Kable JW, Myers J. An opportunity cost model of subjective effort and task performance. Behav Brain Sci. 2013;36:661–79.

254. Marcora SM, Staiano W. The limit to exercise tolerance in humans: mind over muscle? Eur J Appl Physiol. 2010;109:763–70.

255. Inzlicht M, Schmeichel BJ. What is ego depletion? Toward a mechanistic revision of the resource model of self-control. Perspect Psychol Sci. 2012;7:450–63.

256. Agrawal M, Mattar M, Cohen JD, Daw ND. The temporal dynamics of opportunity costs: a normative account of cognitive fatigue and boredom. Psychol Rev. in press.

257. Boureau Y-L, Sokol-Hessner P, Daw ND. Deciding how to decide: self-control and meta-decision making. Trends Cogn Sci. 2015;19:700–10.

258. Musslick S, Cohen JD. A mechanistic account of constraints on control-dependent processing: Shared representation, conflict and persistence. in *Proceedings of the 41st annual meeting of the Cognitive Science Society*. 849–55 (Montreal, CA, 2019).

259. Zenon A, Solopchuk O, Pezzulo G. An information-theoretic perspective on the costs of cognition. Neuropsychologia. 2019;123:5–18.

260. Inzlicht M, Shenhav A, Olivola CY. The effort paradox: effort is both costly and valued. Trends Cogn Sci. 2018;22:337–49.

261. Yamins DL, DiCarlo JJ. Using goal-driven deep learning models to understand sensory cortex. Nat Neurosci. 2016;19:356–65.

262. Botvinick M, Wang JX, Dabney W, Miller KJ, Kurth-Nelson Z. Deep reinforcement learning and its neuroscientific implications. Neuron. 2020;107:603–16.

263. Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, et al. Prefrontal cortex as a meta-reinforcement learning system. Nat Neurosci. 2018;21:860–+.

264. Cross L, Cockburn J, Yue Y, O'Doherty JP. Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. Neuron 2021;109:724–38.

265. Battleday RM, Peterson JC, Griffiths TL. Capturing human categorization of natural images by combining deep networks and cognitive models. Nat Commun. 2020;11:5418.

266. Fengler A, Govindarajan LN, Chen T, Frank MJ. Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. Elife. 2021;10(Apr):e65074.

267. Dezfouli A, Morris R, Ramos F, Dayan P, Balleine BW. Integrated accounts of behavioral and neuroimaging data using flexible recurrent neural network models. BioRxiv. 2018;328849.

268. Bouchacourt F, Buschman TJ. A flexible model of working memory. Neuron 2019;103:147–60 e8.

269. Alexander WH, Brown JW. Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex. Neural Comput. 2015;22:1–57.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to A.G.E.C. or A.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.