Cognitive, Affective, & Behavioral Neuroscience (2021) 21:35–57
https://doi.org/10.3758/s13415-020-00842-0

Neural Mechanisms of Human Decision-Making

Seth Herd[1,2] • Kai Krueger[1,2] • Ananta Nair[1,2] • Jessica Mollick[1,3,2] • Randall O'Reilly[1,4,2]

Accepted: 28 September 2020 / Published online: 6 January 2021

Abstract

We present a theory and neural network model of the neural mechanisms underlying human decision-making. We propose a detailed model of the interaction between brain regions, under a *proposer-predictor-actor-critic* framework. This theory is based on detailed animal data and theories of action-selection. Those theories are adapted to serial operation to bridge levels of analysis and explain human decision-making. Task-relevant areas of cortex *propose* a candidate plan using fast, model-free, parallel neural computations. Other areas of cortex and medial temporal lobe can then *predict* likely outcomes of that plan in this situation. This optional prediction- (or model-) based computation can produce better accuracy and generalization, at the expense of speed. Next, linked regions of basal ganglia *act* to accept or reject the proposed plan based on its reward history in similar contexts. If that plan is rejected, the process repeats to consider a new option. The reward-prediction system acts as a *critic* to determine the value of the outcome relative to expectations and produce dopamine as a training signal for cortex and basal ganglia. By operating sequentially and hierarchically, the same mechanisms previously proposed for animal action-selection could explain the most complex human plans and decisions. We discuss explanations of model-based decisions, habitization, and risky behavior based on the computational model.

Keywords Neural network · Decision-making · Basal ganglia · Model-based · Controlled processing

Introduction

Decision-making is of critical importance. In personal life, professional activities, and in government and military contexts, the quality of people's decisions is among the most important determinants of whether our outcomes are good, bad, or disastrous. As such, a great deal of scientific work has been directed at human decision-making, at multiple levels of analysis. In this paper, we advance a theory and neural network model of how specific brain networks give rise to both the power and pitfalls of human-level decision making. We build upon previous theories and models of specific brain systems and a wealth of existing functional and anatomical data from animal decision-making. We attempt to bridge from that level of detailed data to address human behavior and human neuroimaging data.

We explore the implications of our theory by manipulating and testing a neural network model of the relevant brain systems, implemented in the Leabra framework (O'Reilly & Munakata, 2000). We use this model to address several important issues in decision-making. We address the distinction between model-based and model-free decision-making, in computational and anatomical terms; the use of predictive models that match the structure of particular tasks; a mechanism for habitization; and some individual differences in biology and life experience that result in making risky decisions.

There is now a broad consensus about the critical role of the basal ganglia in helping to select actions. By learning over time from dopamine neuromodulation, it selects (or "gates") those actions which maximize reward and minimize negative outcomes (Barto, Sutton, & Anderson, 1983; Barto 1995; Mink, 1996; Graybiel, Aosaki, Flaherty, & Kimura, 1994; Joel, Niv, & Ruppin, 2002; Graybiel, 2005; Nelson & Kreitzer, 2014; Gurney, Prescott, & Redgrave, 2001; Frank, Loughry & O'Reilly, 2001; Brown, Bullock, & Grossberg, 2004; Frank, 2005; O'Reilly & Frank, 2006; see Collins and Frank, 2014 and Dunovan & Verstynen, 2016 for recent reviews). Computationally, this system is well-described by the

✉ Seth Herd
seth.herd@colorado.edu

[1] eCortex, Inc., Boulder, CO, USA

[2] University of Colorado, Boulder, CO, USA

[3] Yale University, New Haven, CT, USA

[4] University of California, Davis, Davis, CA, USA

Actor-Critic framework of Sutton and Barto (1981). In this framework, the basal ganglia action-selection system is the "actor," and the set of brain areas that produce phasic changes in dopamine are termed the "critic," which trains the actor according to its estimate of the value of actions.

More recently, there has been considerable interest in a higher-level, *model-based* form of action selection thought to depend on prefrontal cortical areas. This intuitively captures one special aspect of human decision-making: we seem to predict the outcomes of possible actions. This has been contrasted to the *model-free* nature of the learned associations in the basal ganglia system (Daw, Niv, & Dayan, 2005; Dayan and Berridge, 2014). Many people tend to think of this distinction in terms of separate, and perhaps competing, systems that enact *goal-directed* versus *habitual* behavior (Tolman, 1948; Balleine & Dickinson, 1998; Yin & Knowlton, 2006; Tricomi, Balleine & O'Doherty, 2009), where the basal ganglia is the habit system, and the prefrontal cortex is goal-directed. However, it has increasingly become clear that the basal ganglia plays a critical role in higher-level cognitive function (Pasupathy & Miller, 2005; Balleine, Delgado, & Hikosaka, 2007) and in goal-directed behavior (Yin, Ostlund, Knowlton, & Balleine, 2005).

We present an alternative model in which the basal ganglia and cortex are not separate, and do not compete, but interact to produce a spectrum of computations. These range between fully model-free (or *prediction-free*) to fully model-based (or *prediction-based*). We use the terms *prediction-based* and *prediction-free* to avoid a variety of accumulated terminological baggage, and an imperfect mapping to our proposed neural mechanism (see Discussion section and O'Reilly, Nair, Russin and Herd, 2020). This model is compatible with findings that model-free computational strategies show relatively more activity in basal ganglia, whereas model-based decisions produce more activity in PFC. This mapping is further addressed in the Discussion section.

In the context of these existing ideas and issues, we offer a specific theory of how brain systems computationally perform complex human decision-making. In this biologically based *Proposer-Predictor-Actor-Critic* framework, the prefrontal cortex and basal ganglia work together as an integrated system, and prediction-based computations are an optional additional step, rather than a separate system. We focus on the neural mechanisms of relatively complex decision-making (roughly, decisions that take a second, or more, and that are novel combinations of previously experienced elements) in distinction to much work that focuses on simpler perceptual decisions (Gold & Shadlen, 2007) and uses drift-diffusion mathematical models (O'Connell, Shadlen, Wong-Lin, & Kelly, 2018). We think that those decisions use similar neural systems but proceed using fully parallel neural computations. In contrast, we think that complex decisions demand serial consideration of each option, and we make that a central prediction of this model.

This framework provides an explicit attempt to account for the temporally extended, sequential nature of complex human-level decision-making. All of the existing computational models of basal ganglia in action selection (of which we are aware) process multiple options in parallel. Other verbal theories of basal ganglia function seldom explicitly address the serial/parallel distinction but seem to largely assume parallel competition. We think this characterization is likely correct for well-practiced tasks, including most animal laboratory tasks. However, we argue that making important decisions in complex, novel situations demands a slower, serial computational approach to maximize accuracy, flexibility, and transfer of prior learning. Computationally, parallel computations introduce *binding problems* in neural systems (Treisman, 1996) that can be resolved with additional learning or with serial attentional allocation. Intuitively, devoting all available brain systems to each option in turn is optimal for addressing difficult, new, and critical decisions.

Structure of Proposer-Predictor-Actor-Critic architecture

The *Proposer-Predictor-Actor-Critic* circuit (shown in Figure 1) is a theory of how basal ganglia works with prefrontal cortex. There are loops descending from different areas of frontal cortex through the basal ganglia and thalamus, which converge back to modulate the function of the same areas of frontal cortex (Alexander, DeLong, & Strick, 1986; Haber, 2010; Sallet et al., 2013; Haber, 2017). In this theory, this same functional circuit operates in each of the different levels of decision-making and action-selection associated with each different fronto-striatal area. We assume, with many others (e.g., Miller & Cohen, 2001) that complex decision-making consists of selecting (gating) working memory representations into an active state. Those representations then condition further steps in the decision-making process, including action-selection. Critically, we also argue that each such circuit also functions sequentially across multiple iterations within complex decision-making tasks. This is a *serial-parallel* model, in which parallel neural network computations are iterated serially so that each option can be evaluated with the full computational power available.

- The cortical *Proposer* (cortical areas depending on task domain; see Methods section) settles on a representation of one potentially rewarding action, plan, or task set (we use the term *Plan* for all of these, because the neural mechanisms are isomorphic). The parallel process of generating a candidate Plan involves neural activation across

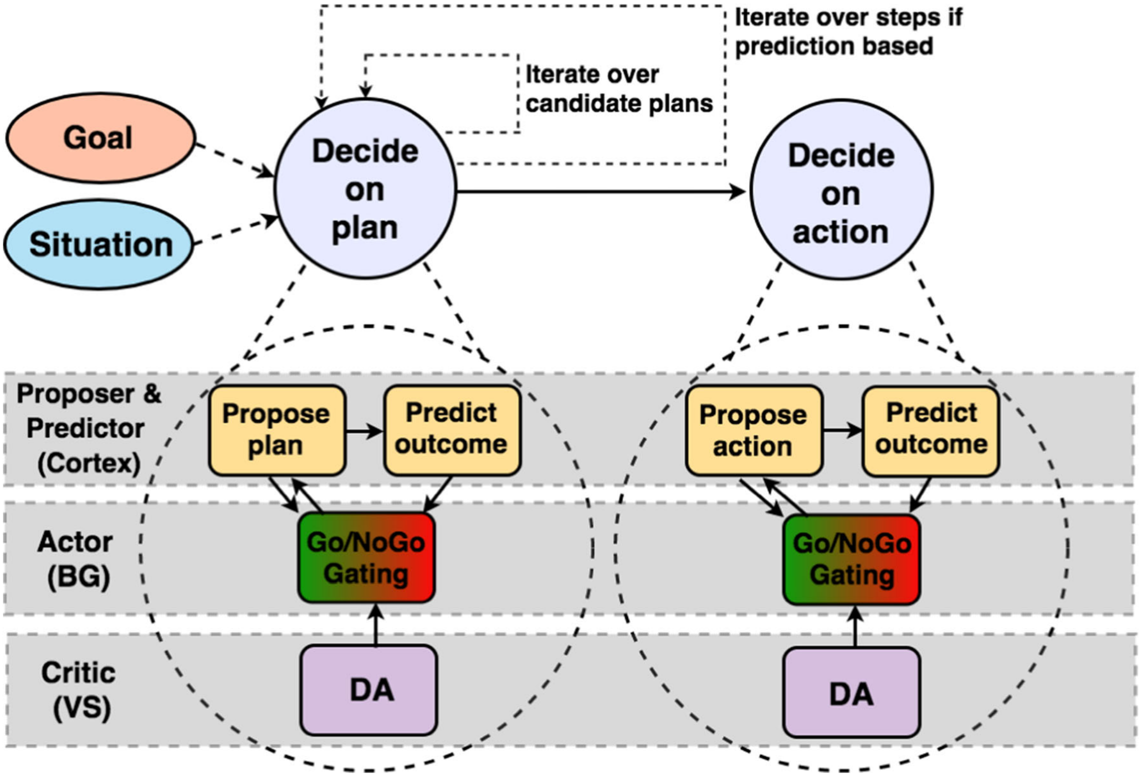

Figure 1. Structure of Proposer-Predictor-Actor-Critic architecture. The top section is a broad functional diagram, emphasizing the serially iterative and hierarchical nature of our proposed decision-making process. The bottom expands those functions and identifies the brain areas that perform each function. BG, Basal Ganglia; VS, Ventral Striatum (along with amygdala and related regions). This architecture and the associated brain regions are described in more detail in the Methods section. There are two parallel circuits with a hierarchical relationship: the outer loop circuit selects a plan, and the inner loop then selects an action appropriate to that plan.

multiple interconnected cortical areas, which integrates learned synaptic weights with the current external inputs (stimuli, context, affective and body states, etc.) to produce a plausible plan that represents at least a locally-maximal satisfaction of all these factors (Hopfield, 1984; Ackley, Hinton & Sejnowski, 1985; Rumelhart & McClelland, 1986; O'Reilly & Munakata, 2000; O'Reilly, Wyatte, Herd, Mingus, & Jilk, 2013). This process acts in a single parallel neural computation and so is relatively fast.

- A cortical *Predictor* (e.g., parietal and/or medial temporal lobe for motor actions) predicts *specific outcomes* of that plan in a given context or situation. This step is optional; it is engaged only when a prediction-based strategy is employed, which takes extra time, while (usually) providing additional accuracy and generalization to the decision. While this computational model does not capture it, we think this decision of whether to engage additional predictions is one of many sequential decisions in the overall decision making process.
- The *Actor*, consisting of a basal ganglia loop linked to the *Predictor* area, takes a predicted value as input and uses what it has learned from the reward history of similar predictions to accept or reject that plan. If the proposed plan is rejected, the Proposer proposes a different plan, and the process continues. The critical computational feature of this basal ganglia system, which is not well-supported in the cortical system, is the ability to boil everything down to two opposing evaluations: Go (direct) and NoGo (indirect), which critically allows the system to delay to consider other options. We propose that it often operates in a serial fashion in complex human decision-making. The basal ganglia uses cortical inputs, which enables it to function effectively in novel decision-making contexts where the history of learning is only applicable when it is associated through abstract representations developed by cortex.
- Once a plan is selected and an outcome is experienced, the *Critic* estimates the value of that outcome relative to its expectations for that situation as a *reward prediction error* (Schultz, 2016). This Critic is composed of a set of subcortical areas that function as a reward-prediction system, and it uses that reward prediction to discount the outcome's value, sending the result as a phasic dopamine signal. The Critic's dopamine signal trains both the Actor and the Proposer components, whereas the Predictor is trained by the specific outcome that occurred.

By binarizing (accept/reject), as well as sequentializing (considering one proposed plan at a time), this canonical decision circuit could scale out to arbitrarily complex decisions, in the same way that sequential computer programs can accommodate arbitrarily complex chains of logic. Parallel algorithms are faster but have more constraints based on interference and binding problems. Furthermore, interactions between isomorphic loops between cortex and BG, at different levels of the anterior-posterior PFC gradient of abstraction (Badre & D'Esposito 2007), enable this set of mechanisms to function semi-hierarchically, where higher-level decisions can be unpacked into subgoals and steps at lower levels.

Neural Evidence of Sequential Decision-Making

While most of the neuroscience data in animal models is consistent with the idea that multiple options are evaluated in parallel (Balleine, Delgado, & Hikosaka, 2007; Collins & Frank, 2014), there is some recent detailed neural recording data that suggest a more serial process in some more complex tasks. Hunt et al. (2018) concluded that monkey orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC) neurons represent the value of the currently attended stimulus. Parallel models, such as traditional drift-diffusion models, would seem to predict that both alternatives, or a difference of the two, should be reflected in the recording data. They used a two-alternative choice task, but with multiple attributes for each option, and recorded from monkey frontal neurons. The activation values of those neurons in OFC correlated with the identity and the value of the currently attended stimulus, and the ACC showed a more stepwise function, perhaps functioning as a belief updating and accept/reject signal in their paradigm.

OFC primarily represents the value of the currently fixated option; while previous cue values are represented above baseline level, the current cue representation is actually anticorrelated with previous cue values, indicating a roughly subtractive relationship. This indicates a comparison-with-current-best-option representation, as opposed to the value-summation representations assumed by parallel models. In other words, in the OFC, Hunt et al. found a signature of attention guided value comparisons as the monkey shifts its gaze from one location to the next. Hunt et al. interpret their data as subjects making sequential decisions of whether to accept or reject the option they currently think is best. The primary OFC representation data is shown in Figure 2a, and we address other aspects of their data in the Discussion section.

Rich and Wallis (2016) also examined firing patterns in OFC neurons and found that they largely represented the value of a single choice option at any given point in time. Activity was recorded in OFC while monkeys performed a two-alternative-forced-choice task, and that data was used to train a decoder to distinguish neural representations of those choices. The decoder's estimate of the presence of each option showed a strong reciprocal relationship; a strong representation of one option corresponded to a weak representation of the other option. This pattern of results indicates that the OFC primarily represented the value of one of the two options at any given time point. These data are also summarized in Figure 2b.

Despite the overall serial nature of the full loop of decision-making in our framework, there is still an important parallel selection of a single option among all others. On each iteration, the Proposer's learned connection weights select one plan using standard neural parallel processing. (this is described in detail in the subsection "Structure of Proposer-Predictor-Actor-Critic architecture"). Thus our model is a serial-parallel model, and we think the addition of strategic, discrete serial steps to parallel neural processing is one key component of human intelligence (Herd et al., 2014). This serial hypothesis is consistent with the finding that value representations in ventromedial prefrontal cortex and striatum in the early stages of decision making tasks correlate strongly with the outcome the animal will go on to select on that individual trial, rather than an average of all available outcomes,

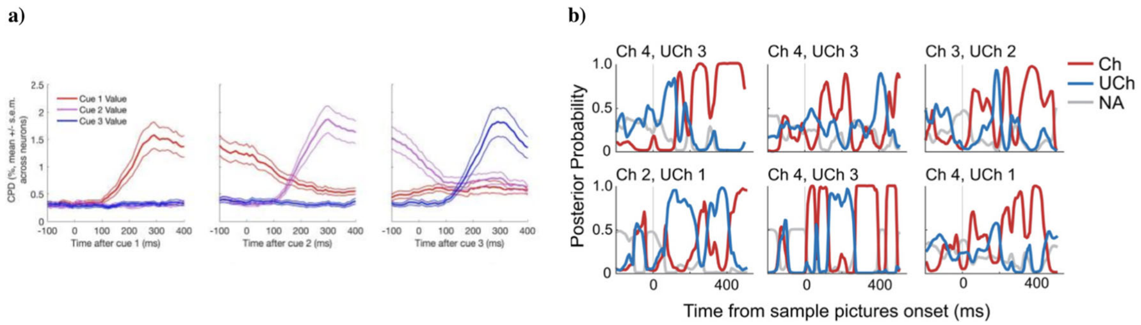

Figure 2. **a)** Data from Hunt et al. (2018) showing correlations between neural activity and cue values at each time point in monkey OFC during a two-option comparison task. **b)** Data from Rich and Wallis (2016). Decoder results on individual trials. Red lines are decoder estimates of the probability of a representation of the chosen option; blue lines are estimated probabilities of available of unchosen options; and gray lines are an average of the two remaining options that are not available on that particular trial. Results, and additional analysis, strongly indicate that OFC neurons alternate between representing each of the available options, and do not represent them in parallel.

as shown by the recording data above, and in Kable and Glimcher (2009).

We ran a full model, as described in detail below, and several variants that illustrate different possible decision-making strategies and brain computations accompanying them, and some possible individual differences in risk bias. Those simulations and results are as follows:

- Full model: Over the course of training, the model increasingly relies on the prediction-free Proposer component. Its performance speeds up as it requires the serial consideration of fewer plans by the Predictor, Actor, and Critic. This illustrates a smooth transition from more *controlled* to more *automatic* processing (Shiffrin & Schneider 1977) or habitual behavior (Tolman, 1948).
- No-Proposer: Plans are selected for consideration at random instead of through reward learning. This variant learns more slowly but generalizes to the held-out test set better, because it cannot transition to the more automatic, habitual mode of performance as the Proposer learns, as the full model does. Comparing this with the full model illustrates the advantages and disadvantages of the Proposer's contribution.
- No-Predictor, model-free: In this variant, the model's Proposer and Actor components (cortex and basal ganglia) perform model-free RL. This addresses the possibility that the system sometimes makes decisions without taking time to make any prediction about outcomes. This variant performs poorly on our full, complex task, but can perform well on simpler tasks and performs faster without the need to wait for an explicit prediction from the cortical Predictor layers. We think this is the fastest but least accurate mode of human decision-making, as proposed by Daw, Niv, and Dayan (2005).
- Value-only Predictor: In this variant, the two cortical prediction layers do not predict a specific Result or Outcome, but only the value of the result. This variant blurs the line between prediction-based and prediction-free strategies; it is technically performing a model-free computation, but it is using prediction. This fast, mixed mode of prediction can be performed in a single computational step. Thus, it performs faster but with poor generalization relative to the full model with its two-step Predictor component.
- Unstructured Predictor: In this variant, the Predictor component does not separate predictions into distinct steps that match the task structure. It performs nearly as well as the full model on the training set but much worse on novel combinations of situation and goal (the test set). This result illustrates the improved generalization produced by splitting predictions into steps that match the task domain.
- Basal ganglia parameter variations affecting risk. We shifted the balance between Go (D1) versus NoGo (D2) pathways in the Actor (basal ganglia) component. The model reproduces experimental results showing more risky decisions with more D1 influence, in accord with a variety of empirical results. We also modeled vicarious learning, in which people learn from others' experiences in risky domains and showed how different vicarious experiences (e.g., different public awareness campaigns), produce different behavioral risk profiles.

Table 1 (methods) summarizes all of the model variants, whereas Table 2 (discussion) summarizes their results and interpretations.

Next, we present the specific computational implementation of our overall framework, the above manipulations, and their results. In the *Discussion*, we consider the relationships between this framework and a variety of other approaches to understanding decision making across a range of different levels of analysis.

Methods

Modeling framework

Our model was created within the open source Leabra modeling framework (O'Reilly & Munakata, 2000; O'Reilly et al., 2012; O'Reilly, Hazy, Herd, 2016). The Leabra framework is a cumulative theory of cortical function, with variants covering subcortical function. It has been used to model a wide variety of cognitive phenomena and is an attempt to constrain a general theory of cortical function with as much evidence as possible. In this case, however, we have focused not on the specific contributions of the Leabra framework but the general computational properties. Our results hold true for a variety of parameter choices within the Leabra algorithm, and we believe that they should hold true for neural network models with similar architectures under a wide variety of learning rules, activation functions, and other parameter choices.

The Leabra framework uses point neurons with sigmoidal response functions. It is here (and most frequently in other work) run with rate-coded responses. This arrangement is similar to a number of other modeling frameworks designed to address similar levels of analysis—those reaching up to human cognition and behavior (Deco & Rolls 2002; Deco & Rolls 2003; Rumelhart & McClelland, 1986; Grossberg 2013, Brown, Bullock & Grossberg 2004; Collins and Frank, 2014). Leabra units are usually considered to be representative neurons among a much larger population. The full model here is of limited size. It contains a total of 975 units, and each of the three cortical processing layers (Proposer, State Predictor and Outcome Predictor) contain only 100 units each for the standard model. This small model is adequate to demonstrate our general points, but much larger models are needed to process complex real-world input (e.g., visual object recognition from images; O'Reilly, Wyatte, Herd, Mingus, & Jilk, 2013).

Task

We model an abstract decision-making task that could apply to various domains. A Situation leads to different Results, depending on the Plan that is selected. Each Result provides one motivationally relevant Outcome. If this Outcome matches the agent's current Goal, a reward is provided to the model. If it does not, no reward or punishment is given (except for use of stochastic punishment only for the risky decision-making model manipulations, described in that section). We used a deterministic task, but our model should apply to stochastic domains with little modification. There is an interesting question of whether the predictions sample stochastically from multiple likely outcomes or are a mix of possible outcomes as in "successor representations." Those questions are outside the scope of the current paper (Fig. 3).

For example, in spatial foraging, physical locations would be the Situations and Results, Plans would be for navigating between locations, and Outcomes and Goals would be physical requirements like food, water, and shelter. In the domain of social decision-making, Situations and Results would be social situations, such as being challenged on one's claims, while Plans would be general approaches to social interaction (agree, discuss, threaten, etc.), and Goals/Outcomes would be social goals such as gaining respect or getting agreement, etc. In complex tasks, Situations and Results would be task states, while Outcomes and Goals would be progress toward task subgoals. For instance, a Situation could be a certain board position in chess, whereas a Result would be a new board position, which could accomplish Goals of controlling the center of the board, freeing pieces, creating attacking pressure, or defending a vulnerable piece.

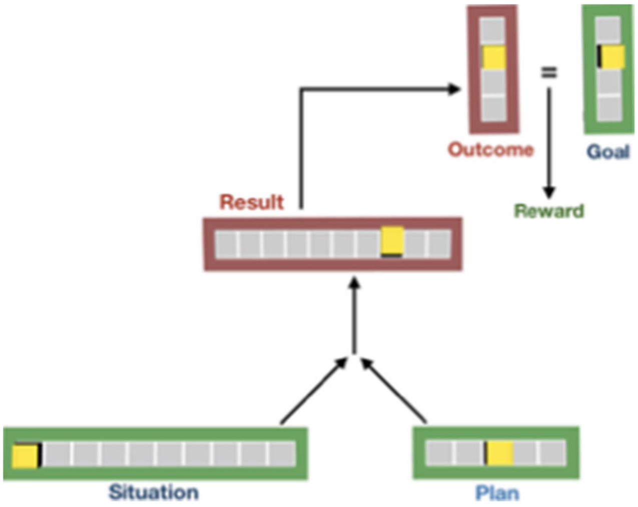

Figure 3. Task. The depicted are input and output layers of the model. The model's task is to choose a Plan that achieves an Outcome matching its current Goal, given a random starting Situation. Each Situation and Plan lead deterministically to a Result (conceptually, another Situation), which leads to one Outcome which is potentially rewarding (e.g., food, water, or shelter, or in a social context, agreement, submission, or appreciation). The model receives a reward signal if it chooses a Plan that matches its current, randomly chosen Goal. There are 10 Situations and Results, 5 Plans, and 4 Outcomes and Goals, for a total of 240 Plan-Situation-Goal combinations, each of which deterministically leads to success or failure.

The model interacts with the task as a set of localist input and output layers. There is one input unit each for the four possible Goals; ten input units for the Situations; ten units for the possible Results; and four units for the four possible Outcomes, each matching one Goal. Value is represented by 23 units capturing a continuous −1 to 1 scale.

Model computations and biological underpinnings

Different areas of cortex and basal ganglia have been shown to be involved in different types of decision-making. A core hypothesis of this theory is that, while different areas have representations of different domains, their circuits and computations are closely analogous. Thus, our task and model are both presented in relatively abstract terms. In the following section, we describe each component of the model in more detail, including its computations and the biological evidence upon which it is based (Fig. 4).

The Proposer model component learns to propose an appropriate plan based on the Situation and Goal inputs. Theoretically, this can be learned from a variety of signals. In the current model, we used the dopamine signal, such that connections from inputs to and from the Proposer layer are strengthened or weakened when the Critic produces a positive or negative dopamine signal based on its computed expectations, and the actual outcome. The Proposer layer of the model maps to different areas of cortex depending on the task domain, in accord with recording and imaging studies showing different specialized representations in different areas of prefrontal and motor cortex (Badre & D'esposito, 2007), and evidence accumulation for perceptual decision-making in areas specific to the task (reviewed in O'Connell, Shadlen, Wong-Lin, & Kelly, 2018).

If the proposed Plan is rejected, the Proposer produces a different plan. This is implemented with an accommodation function on the units in the Proposer layer; units that were active in the last trial become less active, allowing units representing a different Plan to dominate. In the reported simulations, a maximum of five Plans were considered before the simulation "timed out" and progressed to a new Situation and Goal combination.

The Predictor component consists of two areas and stages of prediction: state and outcome. The State Predictor learns to predict the Result (the state that would result from the current

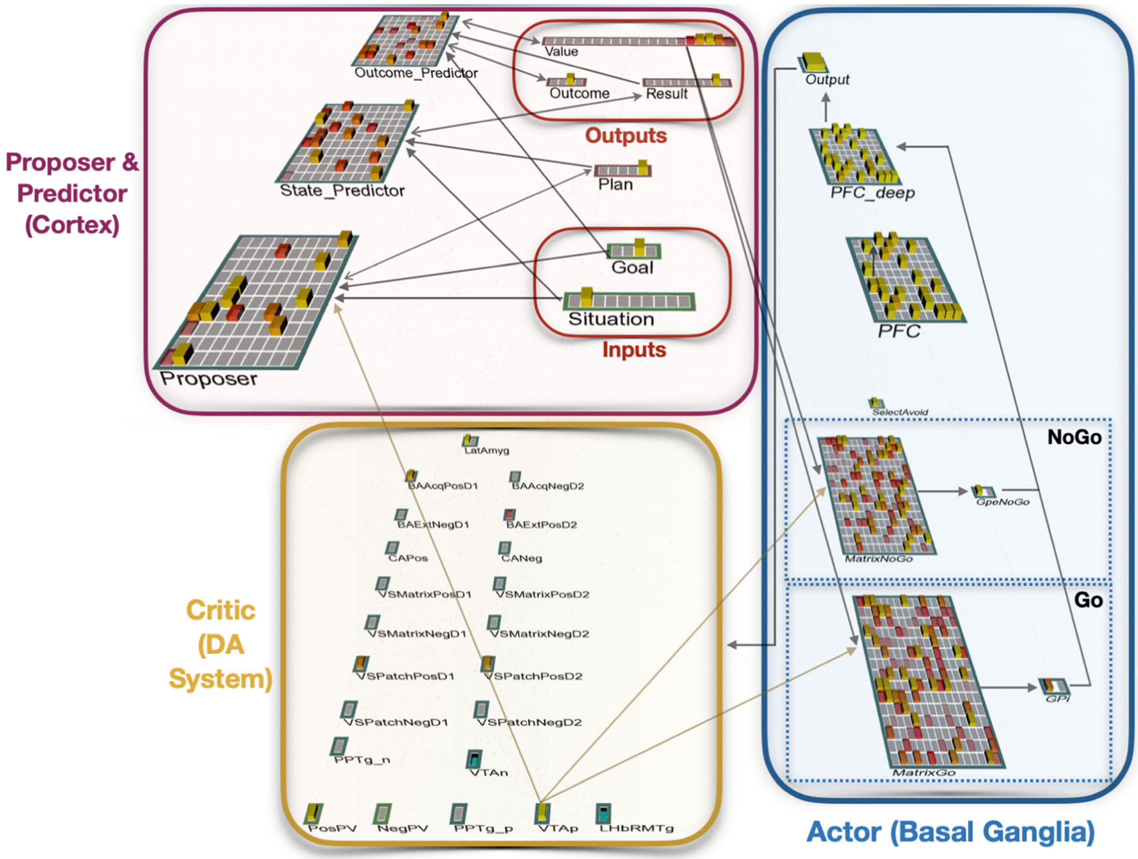

Figure 4. Model. The computational model has four broad functional divisions, identified with each element of our Proposer-Predictor-Actor-Critic terminology. Each small square is a simulated unit; colors are unit activations captured on one sample trial. The first two are different areas of the cortex (red outline) (O'Reilly, Wyatte, Herd, Mingus, & Jilk, 2013; O'Reilly, & Munakata, 2000; O'Reilly et al., 2016). The Proposer and Predictor each take the current Situation and Goal, and respectively propose a plan, and predict its outcome given those conditions. Those functions are both distributed across areas of cortex and medial temporal lobe; the areas that serve these roles will depend on the task domain. The Actor functional division (blue outline) is the basal ganglia loop attached to the relevant areas of cortex (Hazy, Frank, and O'Reilly 2006; Herd et al. 2013). It uses reinforcement learning (RL) to "gate" or enhance signals from specific areas of cortex. It thus adds a RL-driven selection mechanism that acts to enhance ("accept") representations in frontal cortex or reject them to allow consideration of other options. The final subsystem is the reward prediction system (yellow outline) (Hazy, Frank, O'Reilly, 2010; Mollick et al., in press; Hazy, Frank, & O'Reilly, 2007). This system is the Critic; it acts as a teacher of the Actor and Proposer systems. Yellow arrows depict dopamine connections that train those areas. The Critic learns to predict rewards and discounts the dopamine signal of those predicted rewards. Gray arrows indicate all-to-all learning connections.

Situation and the Plan currently under consideration). The Outcome Predictor layer (identified with OFC) predicts the Outcome (potential reward, e.g., food, social dominance) of that Result (state) and the reward value of that Outcome (1 if it matches the current Goal, 0 if it does not). Thus, the two Predictor layers are each making a prediction in their own domain, based on the current Situation and Goal, and the Plan currently under consideration. Each of those environmental variables are presented to the network as an input from the simulated task—each in a localist coding in which a single unit corresponds to a single variable identity. The actual Result, Outcome, and Value of that combination is presented as an output and training signal if the Actor component accepts the current plan. The Predictor areas of our model pass their predictions to the basal ganglia through their output Value layer. The Predictor areas, like the proposer, will map to different cortical or medial temporal areas depending on the specific domain of decision-making.

Note that this model assumes that OFC represents both specific outcomes (e.g., food vs. water, Rudebeck & Murray, 2014) and outcome value (Cai & Padoa-Schioppa, 2014). Our model would show OFC units relating to predicted task state (Wilson, Takahashi, Schoenbaum, & Niv, 2014), because that layer takes the predicted state as an input. Our model predicts that state prediction is primarily performed by other cortical areas.

In our main model the basal ganglia layer receives input only from the Value prediction layer of the cortex, which represents the summarized output of cortical prediction. This single input provides an ideal signal for the basal ganglia to learn from, because after learning the task structure, the cortex is able to make a very accurate prediction of the reward value of the current candidate Plan. Cortical inputs to basal ganglia are known to be diverse and integrative (Haber, 2010), so we think that in reality the basal ganglia often integrates information from multiple sources to make a final decision.

The Actor component in the basal ganglia determines whether to accept or reject the current plan. In our current model, the basal ganglia is composed of the Matrix and Globus Pallidus (GP) layers, with the GP layer which also includes the computational roles of the substantia nigra and thalamus in the basal ganglia loops (Hazy, Frank, and O'Reilly, 2006). The Go and NoGo pathways (Schroll & Hamker; 2013; Hazy et al., 2007; Collins & Frank, 2014) compete to make a decision, based on weights learned through dopamine reward signals from the Critic component of the model, described below. Thus, the weights for the Go pathway support accepting the current Plan, and those in the NoGo pathway support rejecting the current Plan. The Actor in our full model receives connections only from the Value layer—the output layer of the Predictor component. This was a practical choice to allow better performance (four inputs proved difficult for our implementation of BG to learn). In the actual brain, we would expect that basal ganglia to receive inputs from a variety of cortical areas, allowing it to make decisions without the support of a prediction of value from cortex (Haber, 2010). Our no-predictor model variant indeed includes those connections from input layers.

The PFC and Output layers also are elements of the Actor and represent the downstream effects of selecting the proposed plan. When the GPe layer activates (which in turn happens when the Matrix comparison favors the Go layer over the NoGo layer), the PFC layer is gated into the PFC_deep layer, which in turn activates the Output layer, which we interpret as the model activating the Plan currently under consideration. This architecture is modeled in accord with our existing *PBWM* theory of working memory (Hazy, Frank, and O'Reilly, 2006; Herd, Hazy, Chatham, Brant, & Friedman, 2014), reflecting the idea that a plan would usually be maintained in an active state so that it can bias further processing accordingly (Miller & Cohen, 2001; Herd, Banich & O'Reilly, 2006). As output layers of this model, they do not feed back and play a functional role, except as an input to the Critic component, giving it the information that a plan was accepted and will be pursued, and so a reward may be forthcoming.

The Critic component of the model is adapted from the PVLV model of reward prediction, detailed in Hazy, Frank, & O'Reilly (2010) and Mollick et al. (in press). In the current model, we use the Primary Value (amygdala and ventral striatum) components, which predict rewards via projections originating in the Output layer, and use this prediction to discount predicted rewards when they occur. Thus, expected rewards generate lower levels of phasic dopamine bursts and unexpected failures produce phasic dips. These effects are well-documented empirically, as reviewed in Schultz (2013). We did not use the Learned Value portion of the model, which drives Conditional Stimulus (CS) dopamine. In this one-step task, it is redundant with a trace learning rule we use in the Actor. This learning rule affects those units that triggered the previous action and allows them to learn from the dopamine signal that resulted at the time of US. Using this hypothesized trace learning rule means that CS dopamine is only necessary for learning to achieve subgoals. The Critic and Actor portions of this model would presumably function almost identically if we had used CS dopamine instead of trace learning, so that distinction is outside the scope of the current work.

The reward-prediction function of the Critic system is crucial, because it is the only source of negative learning signals in our primary task. As in many human decision-making domains, there is no explicit punishment. Pursuing a plan that does not accomplish the current goal does not cause any direct, physical harm. The critic discounts the actual reward, in effect subtracting the predicted reward from that actually received. It is only by receiving less reward than predicted, and experiencing a dopamine dip based on expectation, that the agent knows that it has committed an error. Without the reward prediction Critic system in place, the model learns only from success and, as a result, learns to accept every proposed Plan. There is direct evidence that reward prediction signals in the striatum can be influenced by prediction-based computations in this way (e.g., Simon & Daw, 2011; reviewed in Doll, Simon & Dayan, 2012).

In addition to our training, we used a holdout test set to test generalization. During training, a subset of Situation/Goal pairs were withheld and never shown. This allowed us to test generalization or transfer to novel combinations of situations and goals. In testing mode, those withheld pairings were presented, and no weight updates were performed. Testing for 5 epochs is interleaved every 25 epochs during all simulations to obtain a learning curve for generalization performance for the held out examples. Of the possible 40 Situation/Goal pairs, for any given simulation, 4 pairs were left out. For each simulation, different Plan/Situation to Result to Outcome pairings were chosen at random, and a different training/testing split was chosen.

Results

Serial prediction of outcomes

We tested the model's match to empirical data by performing an analysis similar to one Rich & Wallis (2016) performed on their neural recording data from monkey OFC. This study is described in the Introduction section above. In our analysis, we looked at activations in the Value layer on a cycle by cycle basis, correlating its activations with the idealized activation

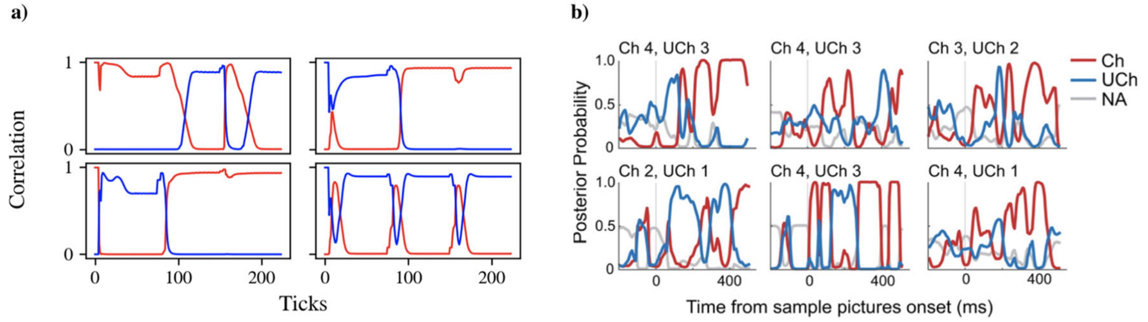

Figure 5. Value representations over time, illustrating the serial consideration of one action at a time, in model and empirical data. **a)** Model. Cosine similarity to canonical Value representations, plotted over time in one instance. In this trial, the model considered two Plans whose outcomes did not match the current Goal, and the Predictor component correctly predicted a low Value, before correctly predicting the rewarding Outcome of the third Plan it considered. **b)** Similar analysis of monkey OFC representations during decision-making. Cosine similarity of current activations to average of neural activity at time of decision (Rich & Wallis, 2016). Each analysis shows a relatively clean representation of the predicted outcome of one option at each moment in time, indicating a serial prediction computation.

for a good and a bad option. The Value layer represents the predicted value of the predicted outcome based on the plan being considered and captures the value-relevant representations of our model's equivalent of OFC. This analysis shows how the value prediction evolves over time, including how it shifts as the plan the model is considering changes after the Actor rejects a plan.

Rich and Wallis (2016) calculated cosine similarities between neural recordings from OFC at each point in time for every trial, and the average activity when a choice was accepted. In Fig. 5 below, both analyses shows a relatively clean representation of the predicted outcome of one option at each moment in time, which is strongly indicative of serial performance.

While this is not surprising, as we designed the model to perform a serial analysis of options, the match between this data and of Rich & Wallis lends support to their hypothesis that their monkeys were serially considering response options and that their OFC (and likely, other linked, contributing brain systems) was predicting outcomes of one possible response at a time.

Basic Model Results

We report several results from our model, demonstrating its basic functionality and accounting for central aspects of the empirical data. We first discuss the performance of the primary ("full") model (discussed in detail above) and then compare that model to several variants. Each comparison illustrates a different computational aspect of the full model's performance.

Each of our results were taken from 50 runs using different random starting connection weights, train/test splits, and ordering of trials and random selection of Plan for consideration for the no-Proposer models.

The State Predictor learns to correctly predict a Result (94.5% ± 0.4 SEM) given an input of a Situation and a Plan, whereas the Outcome Predictor learns to predict both an Outcome (91.3% ± 0.8) and a Reward (98.4% ± 0.08) given a predicted Result (from the State Predictor layer) and a Goal. All performance statistics are taken from 50 model runs with different random seeds, averaging performance between epochs 500 to 600. We report standard errors of the mean for all uncertainties.

Because their roles are each split out separately, this becomes a relatively easy learning task. Note that the learning task is made harder by the fact that these layers only learn when an option is selected, so that sampling is uneven; as the model learns to select correct options, it ceases selecting and learning about nonrewarding options. This bias toward exploitation versus exploration can negatively affect learning. It can prevent learning about new combinations of Situation and Plan, and it causes the cortical Predictor layers to partially "overwrite" their correct predictions about nonrewarding (and so decreasingly sampled) Plan-Situation combinations until they are incorrectly predicted. As a result, the basal ganglia Actor starts to select these nonrewarding options, which then triggers self-correcting new learning. Thus, the model has an intrinsic tendency to titrate between exploration and exploitation, and by approximately maximizing reward in the short term, it never achieves maximal performance. For example, the full model averaged between epoch 500 to 600 over 50 batches chooses an optimal plan in 95.9% ± 0.4% of trials on the training set.

Computational shift and speedup with experience

Our model shows some transition from slower, more prediction-based to faster, more prediction-free computations. This transition has long been a topic of interest in psychology under the terms *controlled versus automatic behavior* (Shiffrin & Schneider 1977; Cohen, Dunbar, & McClelland, 1990). It also has been addressed in terms of a shift between *goal-directed* to *habitual* behavior (Tolman, 1948; Tricomi, Balleine & O'Doherty, 2009); however, that distinction does

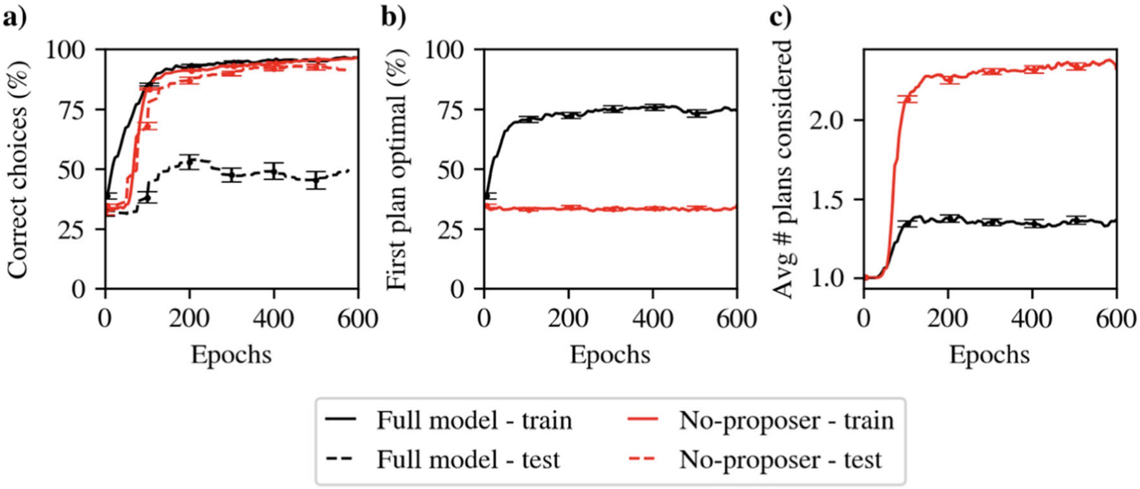

Figure 6. Full model speedup and no-Proposer model. **a)** Test behavior for full (black lines) versus no-Proposer (red lines) models. The no-Proposer model considers a randomly selected candidate plan on each time step. It performs about as well as the full model on the training set and performs almost as well on the generalization set as the training set. **b)** Fraction of first candidate plan correct for the Proposer layer (black line) versus no-Proposer (red black line). The Proposer learns across the course of 600 epochs to select rewarding candidate plans. **c)** Number of candidate plans considered. The full model considers fewer plans per decision, on average since its Proposer component learns to select useful plans to consider first. Thus, the Proposer portion of the model allows it to perform faster as it learns, at the cost of worse generalization performance.

not map directly to prediction-based versus prediction-free computations, as illustrated by the fact that our prediction-free Proposer component takes the current Goal into account and learns to produce candidate Plans that accomplish that goal. Because it selects a candidate Plan based on its weights, learned from the relationship between Situation, Goal, and previously performed Plans, the Proposer model component is performing a prediction-free computation. This mismatch between the definitions of prediction- or model-based and goal-directed behavior is discussed further in O'Reilly et al (2020).

As the Proposer component learns, we found that it reduces the number of Plans that the full model considers, and so saves substantial time, as shown in Figure 6b and c. Thus, to the extent that the Proposer can propose a good plan early in the consideration of options, it can significantly reduce the overall decision-making time. We think this speedup is made more dramatic in some cases by not waiting for the Predictor; we model that possibility in the no-Predictor variant model architecture (explicitly modeling the mechanisms that control whether or not to use a Predictor component are planned for future work).

Our full model generalized above chance but rather poorly on the held-out test set (47.7% ± 3.4). This generalization performance seems surprisingly poor given the model's computations. Because it separately predicts Outcomes given Plan and Situation, and Reward given Outcome and Goal, it should be able to produce near-perfect generalization to the test set of withheld Situation-Goal combinations. In contrast, the no-Proposer model, in which plans are considered by random draw instead of learned selection, performed very well on the generalization test (92.0% ± 1.1%) (Figure 6a). We found that the full model sometimes generalizes incorrectly despite thorough training because the basal ganglia has learned a bias to accept whatever plan the Proposer proposes. This happens since the Proposer often proposes the correct plan on the first try during training (74.2% ± 1.4% after 500 epochs of training). Because the Proposer is never trained on the held-out test set combinations of Situation and Goal, it proposes a correct plan at below chance levels in the generalization test (22.4% ± 3.1%), and yet the basal ganglia will have an increased tendency to approve the first candidate plan. This produces worse performance on generalization vs. training (47.7% ± 3.4% on the held-out test set vs. 95.9% ± 0.4% on the training set; Figure 6a). We take this as a serious proposal for how habitization works: basal ganglia becomes less important as cortex more reliably proposes a good-enough plan on the first try.

Model Variants

We ran several model variants to explore and illustrate the computational functions of each component, and proposed explanations for several phenomena. Table 1 provides an overview of each model variant for reference; full explanations are provided below.

No-Predictor, Model-Free Model

Our primary, full model addresses how the brain might perform prediction-based decision-making. To address how the human brain might perform faster, but less accurate decision-

Table 1. Summary of model variants

Features	Full model	No proposer	Value only predictor	No predictor model	Unstructured predictor	Risk: Go/No Go	Risk: vicarious
Proposer	Yes	No	Yes	No	Yes	No	No
Predictor	Yes	Yes	Yes – model free prediction	No	Yes - does not break predictions into steps	Yes	Yes
Inputs to basal ganglia	Value	Value	Value	Situation, goal, and plan	Value	Value	Value
Punishment	No	No	No	No	No	25% .2 reward	25% .5 reward
Task	Single rewarding goals	Single rewarding goals	Single rewarding goals	Single rewarding goals	Single rewarding goals	Multiple rewarding goals	Multiple rewarding goals
Pre-training	No	No	50 epochs	No	No	No	No
Additional differences			Larger cortical layers	Changed gain factor (without performs at chance)	Larger cortical layers	Run with various gain factor values	

making, we ran two comparison models, one with no explicit prediction, and another which predicts only the Value of a Plan-Situation-Goal combination, without predicting the Result or the Outcome.

In the first comparison model, we simulated the original hypothesis of Daw, Niv, & Dayan (2005) that the basal ganglia performs prediction-free decision-making. This subcortical model-free version of the model uses only the basal ganglia to accept or reject each plan, with no prediction from the Predictor component. Without any contribution from the Predictor component, the Value layer has no activation. Thus, it was necessary to change the connectivity of the model by introducing a direct connection from the *Plan, Situation and Goal* to the *MatrixGo* and *MatrixNoGo* layers, and removed the connection from the *Value* layer. This model performed poorly with default parameters (33.1% ± 0.7% on the training set, around the empirically determined chance performance level of 32.3% ± 0.5%). This appears to be the result of a bias toward accepting plans when the basal ganglia matrix layer has more input layers. We adjusted the threshold level in weighing the matrix Go versus NoGo activities, from 0.1 to 0.5, and saw better performance of 51.7% ± 0.8% for training. This variant performed poorly, at 23.6% ± 1.8% on the testing set, as expected, because it does not segment the task into predicting Results and Outcomes, and so should not be able to perform above chance on the holdout test set (it actually performs worse than chance, because it has learned to respond correctly to combinations not in the test set). When we added a punishment value of 0.5 to all Outcomes that did not match the current goal, the model's performance on the training set improved to 69.9% ± 1.2%. This appears to be the result of further reducing the Go bias early in learning. This variant still performed at or below chance levels on the testing set, 21.1% ± 2.3%.

Our model thus predicts an upper bound on the types of decisions that can be learned by the basal ganglia without predictive input from the cortex, in line with other predictions from other theories of model-free decision-making.

We agree with the hypothesis that some decisions are made in a model-free mode that relies primarily on basal ganglia (Daw, Niv, & Dayan, 2005). However, we would predict that even in those situations, the basal ganglia are making their decision using highly processed cortical representations of sensory/state information as input, so cortex participates heavily even in "prediction-free" decisions. It may be difficult to fully eliminate any form of cortical model-learning from even the most basic forms of human decision-making, consistent with evidence reviewed by Doll et al. (2012).

One advantage of prediction-free decision-making is that it should allow faster performance, because the basal ganglia need not wait for a prediction (or whole series of predictions) from cortex or medial temporal lobe. Consistent with this model, Oh-Descher et al. (2017) have observed a shift from cortical to subcortical activity when time pressure was increased, accompanied by a shift in decision style to use a simpler set of criteria.

Value-only Predictor model

People also may switch to strategies with less outcome prediction when the task structure becomes more complex, increasing the difficulty and demands of explicitly predicting outcomes (Kool, Cushman & Gershman, 2018). We illustrate another such possible strategy with our next model variant. In this hypothesized decision-making model, the cortex participates in model-free but prediction-based, decision-making. In this hypothesized strategy or model, the computational power of the cortex is used to learn and predict which actions will be successful, but without making specific predictions about outcomes. This model variant instead makes a simpler, one-step prediction of the value of a Plan in a given Situation. This type of prediction is model-free according to the commonly used definition, because it does not include a prediction of a specific outcome. The prediction should be faster (because it demands fewer cognitive steps) but less useful in novel situations.

This version of the model therefore learns about the value of a plan in accomplishing a given Goal in a given Situation, but without using a serial process to predict specific outcomes, in a specific order, as our primary model does. Instead, the two cortical layers work in series, to produce a more powerful "deep" network with two hidden layers, which takes in all of the relevant information, and produces only a predicted Value as an output. This model is similar to other deep networks in that error-driven learning feeds back from the output, to the final layer, and from there to the first layer, allowing both layers to work in series to produce the output using two stages of neural representation. We think that humans may sometimes use the cortex and basal ganglia in this technically model-free, but prediction-based way. This computational approach brings the computational power of cortex to bear, without requiring individual, time-consuming steps for each specific predictive step necessary to arrive at a likely outcome in a complex task.

This model performs well on the initial training set (92.9% ± 0.6% optimal Plans chosen after training), but it does not generalize above chance level on the testing set (25.9% ± 3.3% vs. an empirical chance level of 32.3% ± 0.5% (Figure 7). (Measured as average performance during an all select pre-training with random plan choice.) This illustrates one key advantage of prediction-based decision-making, when predictions are organized into separate steps: it can produce the correct decision on the very first encounter with a new combination of known elements (e.g., starting a known maze with the reward in a new but known location, etc.), because each element's outcome is predicted separately.

This version of the model is at a second disadvantage: without making specific predictions, it cannot be organized to learn separately about the two steps of prediction for this task. During this pretraining, the model learned about the values of uniformly sampled random Plans for each Goal and Situation combination, before the Proposer, Actor, and Critic components were allowed to choose and learn. This biases the experiences of the Predictor component.

Without this pretraining advantage, the model performed somewhat worse on the primary task (86.9% ± 0.8%); because of this worse training performance in the absence of pretraining, we do not draw conclusions from its even worse (roughly at-chance) generalization performance.

Unstructured Predictor comparison model

Humans appear to be capable of decomposing their predictions of outcomes into multiple discrete steps. For instance, in planning our day we may predict that taking the northbound freeway will get us downtown, then, in a separate cognitive step, predict that being downtown will allow us to meet a potentially valuable contact for lunch near where they work. This decomposition of problem space into sensible subcomponents offers substantial computational advantages (at the likely cost of slower performance).

In reality, we think that humans can use a flexible and unlimited number of predictive steps, but for simplicity our computational model always uses two steps. The State Predictor layer predicts a Result, and the Outcome Predictor layer uses that prediction to predict the Outcome linked to that Result (and the reward value that results from that Outcome in combination with the current Goal). To illustrate this advantage, we ran another comparison model that does not explicitly separate those two predictive steps. In this version, there are still two cortical layers, but they are organized in a strictly serial manner: the first layer receives all inputs (Plan, Situation, and Goal) and projects to the second layer, which projects to each prediction layer: Result, Outcome, and Value. Thus, this comparison model is similar to a standard deep network with two hidden layers. The cortical layers are larger (400 units vs. 100 in the main model) to give this model a better chance of performing well. When the model is so arranged, it performs somewhat worse than the full model on the training set (89.0% ± 0.7%) and dramatically worse on the generalization tests (20.6% ± 2.5%) (Figure 7). (When the two hidden layers are held to the same size, the model performs slightly worse: training 81.9% ± 1.2%; testing 27.7% ± 2.9%.)

This result is interesting in relation to recent progress in artificial neural networks. It is certainly possible for a neural network to decompose a complex problem into its components and so achieve good generalization through error driven learning. This computational principle has been demonstrated

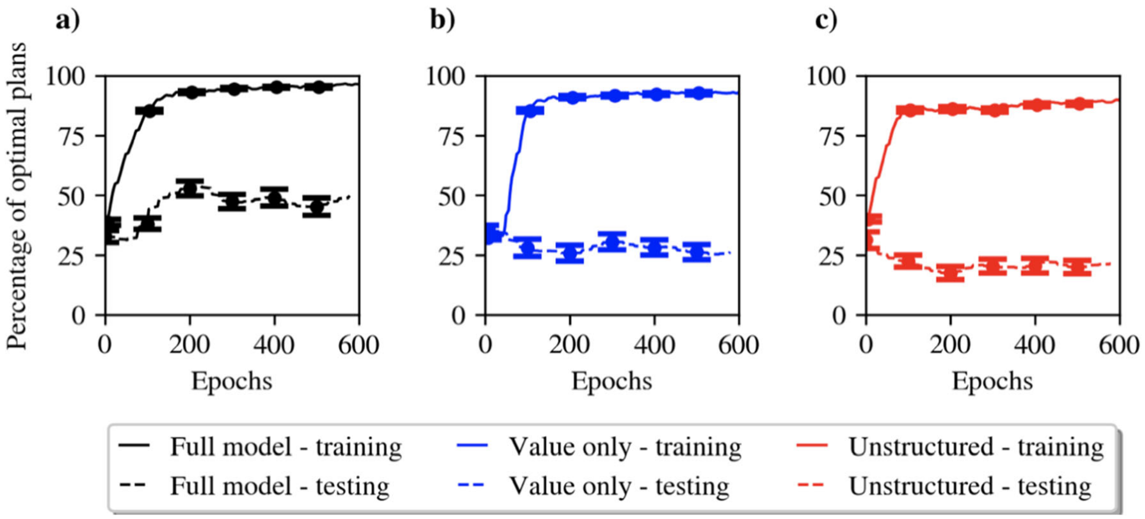

Figure 7. Reduced Prediction Model Results. The full model performs at around 95% accuracy on the full task (i.e., rejecting poor Plans until choosing a Plan that achieves the current Goal for each Situation/Goal combination). It performs reasonably well on the testing set of reserved, never-before-seen combinations of Goal and Situation, tested every 25 epochs). Two versions of an unstructured prediction model performed much worse on the testing set. Each of these models used the two predictor layers in series, without the intermediate State Prediction step trained independently. The Value-only model was trained to produce only the value, whereas the unstructured model produced all of the same predictions as the full model but without separating those by layer. One layer acts as a hidden layer, and the output layer predicts State, Outcome, and Value. This very different performance on the hold-out test set illustrates one key advantage of prediction-based processing with structured predictions in discrete steps: it can produce correct decisions on the very first experience with new combinations of known causal factors.

by the recent success of deep networks on visual object recognition (Krizhevsky, Sutskever, & Hinton, 2012; Wang et al., 2017). Such learning may even produce internal predictions of outcomes in the hidden layers, despite being trained only to predict their value, as addressed in the Discussion section. However, those networks receive millions of training trials, while biological brains experience only a handful to thousands of decisions in each domain (Lake et al., 2017). Creating a priori divisions in predictive steps can reduce the amount of training needed for generalization to new situations dramatically if those divisions match the structure of the real world. Exactly how those predictive steps are created to match the world's structure is an outstanding question, but it is likely to include directing attention to different types of outcomes.

Applications to Risky Decision-Making

One notable application of a mechanistically detailed theory of human decision-making is in understanding how humans make bad decisions. Taking large risks, such as driving while intoxicated, risky sex, and dangerous drug use are all areas in which bad decisions produce enormous societal and personal costs. Risky decisions in political and military domains can produce even worse impacts. While the model as it stands cannot fully describe the complex processes by which humans make such decisions, it does still offer some potential explanations of factors leading to more risk-averse or risk-tolerant styles of decision-making.

Our comparison between the full model and the no-Predictor variant shows how the human brain can support two distinct approaches to making decisions: a fast approach, without explicit predictions of outcomes, and a slower but more accurate prediction-based approach. This speed/accuracy tradeoff between model/prediction-free and model/prediction-based computations appears to be a common conclusion in related work. If this is correct, one major factor in mitigating risky decision-making is to use interventions that encourage a careful, prediction-based approach when decisions may have serious consequences. While this suggestion may seem obvious, it is not clear that existing interventions have fully explored this strategy.

We performed additional modeling of another potential factor in risky decision-making: individual differences in propensity to approach and avoid. These individual differences have been characterized as the two opponent systems, the Behavioral Inhibition System and Behavior Approach System (BIS/BAS). There are many individual genetic and environmental influences that could affect approach and avoidance behaviors; we manipulated a fairly basic parameter controlling our model's Go versus NoGo behavior.

To address risky decision-making, we modified our task to reflect more closely real-world situations in which risky behavior has been identified and studied. In most such real-world domains, there are rare, highly negative outcomes, balanced against more frequent, smaller rewards. To approximate that profile, we modified our task and rewards to produce more small rewards and a few large punishments. We used the same basic task structure but made three goals (instead of one) rewarding for each trial and added stochastic large punishment when the model selected the remaining single bad outcome for its current goal on any trial. We rewarded the model (with a 0.2 reward value) for achieving an Outcome that matched any of those three randomly selected Goals. We stochastically punished the model (with a punishment of 1) on 25% of the failing trials (in which it arrived at an Outcome that matches the one currently invalid Goal). This produced a total base rate of 6.25% (1/16) punishment trials and a base rate of punishment to reward amounts of 12.5% punishment.

We then manipulated the model (our full model, with all components as depicted in Figure 2). Differences in reward learning and decision-making have been observed in many studies when drugs and optogenetics have been used to differentially strengthen the Go (D1-receptor-dependent) and NoGo (D2-receptor-dependent) pathways in dorsal striatum (corresponding to the Actor basal ganglia model component of our model) (Frank, Seeberger, & O'Reilly 2004; Moustafa, Cohen, Sherman, & Frank 2008; These results show the expected difference. In line with published experimental results suggesting that strengthening D1 pathways leads to increased responding to rewarding options, while strengthening D2 pathways leads to increased avoidance of bad options (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007). This also is consistent with results showing that driving D1 neurons in the Go pathway directly causes motor actions (Sippy, Lapray, Crochet, & Petersen, 2015).

We simulated this by manipulating a gain factor moderating the competition between those pathways in our simulated Globus Pallidus internal segment (GPi layer), the final step in the basal ganglia loop that decides whether to use the current candidate plan (Go) or reject this plan and consider a different candidate (NoGo). To show this, we measure the avoidance/acceptance of bad options. Furthermore, to isolate the effect of the manipulation on the GPi, we ran the model without the proposer to ensure a balanced sampling of good and bad plans. Strengthening the D2-driven NoGo pathway in our model relative to the D1-driven Go pathway leads to a more cautious, less risky behavioral profile in which fewer options are selected overall (Figure 8, below).

Vicarious Learning for Risky Decisions

In many domains, including risky decisions, people seem to base their decisions not on their own experience but on what they have learned about the experiences of others. They might simulate the experiences of others in enough detail to produce a relatively complete physiological response to good or bad outcomes, including dopamine release; in this case, our model of learning for decision-making would function identically whether experiences were personal or vicarious. Alternatively, in some cases people may learn information in the abstract, without experiencing the physiological responses.

We simulated such a learning by separately training the Actor (basal ganglia) and cortical portions of the model. The cortical model was trained using a training set of oversampled good or bad plans chosen with an 80-20% probability, whereas the Actor was trained using a 50-50% training set. These probabilities were achieved with an algorithmic proposer. This phase of training modeled early learning of good and bad experiences in a variety of domains and training the Predictor component (cortex) on our risky decision task, described above, but in this case with a reward value of 0.5 for each correct choice (the choice of 0.2 reward for the Go vs. NoGo manipulation was chosen to show behavioral differences more clearly, but using 0.5 reward also showed smaller differences). This training simulated vicarious experience with the abstract semantics of this task, for instance, hearing about others' successes and failures in gambling, inebriated driving, risky sex, etc. We then tested the model on making decisions on that task but with all learning turned off.

This test provided a measure of model behavior on risky decision-making in a domain that has only been experienced vicariously (e.g., the first time a teenager chooses whether to ride with an inebriated driver). We demonstrated that a model trained on mostly negative outcomes made many fewer risky decisions than a model trained mostly on positive outcomes (7.9 ± 1.1% chance of choosing a bad plan for the negatively over sampled model versus 46.6% ± 2.4% for the positively over sampled model). This result follows straightforwardly from the model's distinction between Actor and Predictor

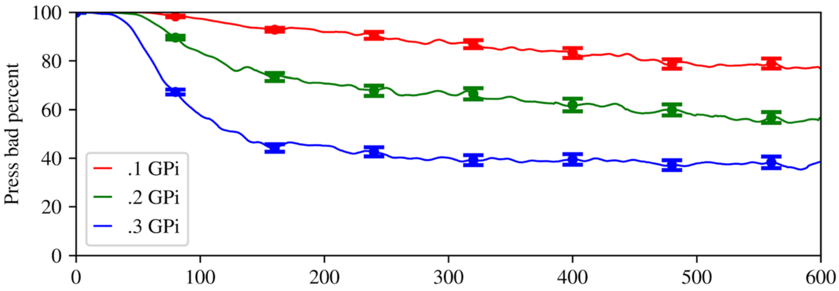

Figure 8. Go versus NoGo pathway Manipulations. Risky decision making model with increasing strength of NoGo pathway (modeling D2 receptors in dorsal striatum). Models with higher relative D2/NoGo pathway strength accept fewer bad options, creating an overall less risky behavioral profile.

components. This demonstration illustrates how those components map to abstract knowledge and full experience (whether direct or simulated) of rewarding and punishing outcomes.

Discussion

We have presented a computational implementation of our *Proposer-Predictor-Actor-Critic* theory of how the core loops between the frontal cortex and basal ganglia support a variety of complex human-level decision-making unfolding over multiple sequential steps. We have shown that our model can learn to accurately propose and predict outcomes in a multidimensional decision-making task that captures important aspects of real-world tasks and that using predictions of specific outcomes produces better performance. Furthermore, the model was able to generalize its knowledge to novel task configurations that it was never trained on. This generalization ability also depends critically on its predictive-learning components. As training proceeded, the proposer aspect of the model became much better at generating appropriate plans. This affected the balance of exploration vs. exploitation, while also significantly speeding up decision-making performance with expertise. Both of these phenomena emerged out of the more basic properties of the model and provide important first-principles predictions about the dynamics of learning in people and other animals.

Finally, our model accords with the data of Hunt et al. (2018), Rich & Wallis (2016), and others who have found evidence that the OFC and related brain systems encode reward predictions about whatever option is currently attended, relative to other options. Hunt et al. (2018) also interpret their recordings from ACC to indicate that the ACC is also considering a single option, the best one encountered thus far, and accepting or rejecting that option. Because the monkeys gave a left-vs.-right manual movement response, we find it somewhat surprising that the ACC did not act as a parallel accumulator of the estimated relative value of each action. The ACC has a good deal of real estate devoted to manual left-vs.-right movement, and their task was quite well-practiced.

Their data are not conclusive, because they did not perform strong tests for neural representations of the other options. Their finding of a large accept-reject signal in even a relatively simple task may indicate that, even for relatively simple and fast decisions, people do not accumulate relative values for multiple options in parallel, but make a fully serial consideration of one option at a time, despite the nonoptimality of doing so in purely computational terms. This would be consistent with some theories of decision-making (some reviewed by Hayden, 2018); inconsistent with others (some reviewed by Turner, Schley, Muller, & Tsetsos, 2018); and entirely orthogonal to many aimed at the psychological level, with less specification of mechanisms and specific computations.

In the remaining sections, we consider the relationship between our framework and other existing frameworks, and associated empirical data, and then enumerate a set of testable empirical predictions from our model. Our theory is intended to be integrative across a great deal of animal and human work and associated theories, so it is compatible with many of those we discuss. Our model is intended to provide a specific hypothesis of the neural mechanisms, so it is compatible with many theories of human decision-making specified at more abstract levels of analysis. No theory of which we are aware covers all of the same ground or bridges levels to the same extent. We note the differences between this and other related theories. Table 2 summarizes the key results we compare to other theories.

Relation to theories of model-based vs. model-free decision-making

Overall, our framework has a significant amount of overlap with the model-free versus model-based framework (MFMB) originally elaborated by Daw et al. (2005) and widely discussed in the subsequent literature. Both frameworks include an essential role for predicting future sensory and other states based on internal "models" of the environment, along with a critical role for dopamine-mediated learning to select "good" versus "bad" actions. However, whereas the MFMB framework is based on a dichotomy between these two types of processes, our framework emphasizes their synergistic interactions within the context of the characteristic fronto-striatal circuit replicated across many frontal areas. The cortical and basal ganglia components of this circuit each contribute unique and important computational functions to the overall decision-making process, supported by their unique neural properties.

Thus, we think that the dichotomy envisioned in the MFMB framework is typically much more of a continuum, with model-like elements likely to be involved in many cases to varying extents. Our model, and this proposed continuum, is consistent with key findings from fMRI studies by Daw, Gershman, Seymour, Dayan, & Dolan (2011) and Simon & Daw (2011), where the striatum seemed to be involved in prediction-based decision-making, in that its activations correlate with predictable outcomes. In addition, Kool, Cushman, & Gershman (2016) found that people typically exhibit a complex mixture of both model-free and model-based profiles.

Predictions of the model

One important distinction between the current theory and the majority of work on human decision-making is our reliance on detailed depth-recording and tractography data. The

Table 2 Summary of results

Model type	Train	Test	Purpose	Key results
Full model (proposer & predictor)	95.9%	47.7%	To demonstrate the full working of the serial proposer-predictor-actor-critic neural decision making model	Transitions automatically and smoothly from serial prediction based to proposer based and thus enables faster decisions
No proposer (only predictor)	95.8%	92.0%	To demonstrate the generalization capabilities of a pure predictor based model	Better generalisation due to only using prediction based strategy. Results in slower decisions due to unordered considerations of plans
Value only predictor (proposer & model free predictor)	92.9%	25.9%	To demonstrate differences between value only predictions and state outcome predictions (model free but prediction based)	Cortex can perform complex value only predictions; good performance on overtrained cases, but no generalization
No predictor model (no proposer & no predictor)	69.9%	21.1%	To demonstrate the role of subcortical brain areas in prediction free decision making	There is an upper bound on the types of decisions that can be learned by the BG without predictions
Unstructured predictor (proposer & one step predictor)	89.0%	20.6%	To demonstrate an anatomical separation of state and value prediction.	Division into predictive steps matching the task speeds generalization dramatically
Risk: Go/No Go (no proposer)	0.1-73.4% 0.2-79.1% 0.3-84.5%	N/A	To demonstrate potential individual difference in the Go and NoGo pathway for the actor	Reducing the threshold for Go decisions increases risky behavior.
Risk: Vicarious (no proposer)	Neg - 95.5% Pos - 82.9%	N/A	To demonstrate the effect vicarious decision making of others has on individual decision making	Training on mostly negative outcomes resulted in fewer risky decisions than training mostly on mostly positive outcomes

component models were created based largely on animal work. While those animals are not trained to perform tasks as complex as those we address and model, that work provides vastly more detailed information on neural computations than can be obtained from humans. Our central argument here is that the same circuits that perform action-selection in animals are sufficient to explain complex decision-making in humans, when they are employed serially, and have access to the more abstract representations available in the human brain (see also Haber & Knutson, 2010 on this mapping). Our model makes strong predictions about the interaction between cortex and basal ganglia, and neuroimaging and neuropsychological approaches with human subjects should be able to test those predictions and so falsify or support this theory.

Our primary predictions are the computational divisions-of-labor that we have emphasized throughout: cortex produces a candidate plan; other cortical regions may produce a model of the predicted outcomes; and connected loops of basal ganglia make a final decision to use that candidate or reject it (at least for the moment) and consider another candidate in a serial, iterative process. Perhaps our most central prediction is that human decision-making, and the predictions upon which it relies, work serially in relatively novel domains.

It is somewhat challenging to test this prediction. Our model is not easily falsifiable solely through behavioral data. The history of research in visual search serves as a useful analogy, as much behavioral work attempted to distinguish serial from parallel processes, apparently with no success (Wolfe 2003). However, neuroimaging data or neural recording data could falsify our theory and model, just as more recent recording data has provided strong evidence on the serial/parallel question in visual search (Eimer 2014). Tracking precise timing of representational content in humans is challenging with current neuroimaging techniques, and allowing subjects to select their own ordering of sequenced steps makes this problem more difficult. Interpreting detailed animal recording data can be difficult; for instance, the neural recording data of Lorteije et al. (2015) were reinterpreted to fit models of one-step parallel process (Hyafil & Moreno-Bote, 2017). However, animal data (Hunt et al. 2018; Rich & Wallis, 2016) and clever human behavioral designs (e.g., using priming at different times during decision-making in concert with self-reported ordering of predictions) can bring evidence to bear on this prediction.

While the current model has two prediction steps hardwired, we think that another similar decision may be required to perform each predictive, model-creation step. These subdecisions may be performed by the same Actor component of basal ganglia, which might use the same dopamine reinforcement signal to learn to delay a final decision long enough for a cortical prediction to play a role, or to accept the Proposer's first Plan to perform quickly under time pressure. It also is possible that the decision to perform more predictive steps may use an anatomically distinct but functionally analogous loop of the same canonical circuit, a loop of basal ganglia associated with an area of cortex that uses domain-appropriate learning to predict specific outcomes. This anatomical and computational question awaits further computational and empirical work. These possibilities are certainly differentiable empirically, but testing them with existing methods would be challenging.

While existing theories and evidence suggest that OFC often does predict reward value and stimuli that are closely linked to reward (reviewed in Rudebeck & Murray, 2014), we assume that predictions of Results will occur in different brain regions for tasks in different domains, even when their task structure is identical. While existing evidence is consistent with this prediction, it is currently insufficient to eliminate the alternative hypothesis that predictions of outcomes are always made in the same brain regions, regardless of domain. For instance, state prediction error signals have been observed in the intraparietal sulcus and in several areas of lateral PFC when people either observed or performed a spatially arranged state selection task (Gläscher, Daw, Dayan, & O'Doherty, 2010). This prediction can be further tested with relatively straightforward neuroimaging methods.

Relation to other theories of decision-making

There are a number of other existing theories of how cortex and basal ganglia contribute to decision-making. Our model and theory draw from those previous theories but has distinctions from each.

Dayan (2007) has explored some consequences of a theory much like ours for sequencing complex behavior. He implemented a simple model of reinforcement learning, which he identified with BG, PFC, and hippocampus, and trained it to produce multi-step behavior. He identified the usefulness of such a system in producing complex behavior based on verbal instructions. That work was based on the neural and biological model of O’Reilly & Frank (2006), upon which our current theory is also based. Although it does not directly address decision-making, that theory is focused on how the basal ganglia’s “gating” of information into working memory can produce arbitrarily complex behavior. Thus, this theory is very closely related to the current one.

Solway & Botvinick (2012) present a computational theory of prediction-based and prediction-free decision-making that is closely related to ours. Their model similarly includes PFC, BG, OFC, and amygdala, performing predictions for both outcomes and reward value, and includes an instantiation as a learning neural network model. The biggest difference between our models is that theirs performs action-selection as a parallel process, whereas ours predicts outcomes for only one option at a time. This is a central feature of our model; we believe that parallel plan selection can produce fast and useful actions when there is sufficient experience with a specific decision but that a serial prediction process is a key component of human generalization of knowledge in complex domains. Most theories that take detailed empirical data (e.g., animal single-cell recordings) into account similarly propose a parallel consideration of multiple options. Our theory holds that adapting this system to consider options serially allows humans to make decisions in more complex and novel domains.

Another interesting difference is that their theory focuses on the Bayesian inversion of a model of outcomes and task-space. This roughly equates to backward-chaining from desired outcomes to actions, while our theory currently only addresses forward chaining. It seems likely that humans can perform both types of chaining; the circumstances favoring and computational constraints surrounding each strategy remain a question for future work. Solway & Botvinick’s model also differs from ours in identifying the hippocampus and medial temporal lobe (MTL) with the outcome-prediction component. We think the MTL is probably involved in both proposing plans and predicting outcomes when there is little experience with the task domain, so that one-shot learning is necessary. The current task and model include cortex in both Proposer and Predictor roles, for simplicity.

Collins and Frank (2014) present a model with substantial overlap with our model of the basal ganglia (and indeed they share a common ancestry; Frank, Loughry, & O’Reilly, 2001). Their OpAL model accounts for incentive effects of dopamine that our model does not, and proposes a theory, in accord with empirical evidence, of how background dopamine levels can focus basal ganglia decision-making on opportunities versus rewards in a flexible and useful way. Like Solway & Botvinick (2012), and many other theories based in part on neural networks and animal data, their theory addresses a parallel action selection process, which we think is employed for well-practiced decisions (like many laboratory tasks), whereas ours proposes a serial process for more novel and complex task spaces.

Daw & Dayan (2014) offer a theory so compatible that our model could be considered a proposal for the mechanistic implementation of their computational level theory. They do not specifically address the serial versus parallel issue but appear to assume a serial prediction process. They propose that model-based (prediction-based in our terminology) decision-making relies on sparse sampling of predicted paths through problem-spaces, which is critical in problem spaces that are too complex to allow for a full start-to-finish prediction of every possible action. This is a critical computational requirement, and our model partially but incompletely addresses this point. Our model does not directly confront this issue, because it currently works only in a limited problem-space in which a full prediction can be made for each candidate Plan. However, our model does include an element that can help in addressing this problem. Our Proposer component selects plans for more detailed outcome prediction, in a fast, parallel neural constraint satisfaction process. This efficiency also contributes to effective sparse sampling of a plan space to focus on areas that are likely to be productive.

Daw & Dayan (2014) propose a different mechanism contributing to this same computational efficiency: prediction-free estimates of state value are employed in complex tasks to avoid the time commitment of following every model through task-space to a conclusion. We propose that the basal ganglia makes prediction-free decisions at every step of model building. Our current model does not perform this function, but we intend to capture this in future extensions of the model. Similarly, we propose that the decision to use cortical, prediction-based reasoning is itself made by the same canonical circuit of linked cortical and basal ganglia loops. Consistent with this prediction is evidence that deciding on task strategy activates the frontopolar cortex and inferior lateral PFC (Lee, Shimojo, & O'Doherty, 2014).

Daw, Niv, & Dayan (2005), followed by many others, propose a categorical distinction in which PFC performs prediction-based computations and basal ganglia performs prediction-free decisions. In our model and theory, basal ganglia contributes to both types of decisions, and cortex also may (see the cortical prediction-free model section). Subsequent empirical work has called this strict separation of systems into question. Cortex now appears to be heavily involved in habitual behaviors (Ashby, Turner, & Horvitz, 2010), and there is strong evidence that basal ganglia plays a critical role in higher-level cognitive function (Pasupathy & Miller, 2005; Balleine, Delgado, & Hikosaka, 2007) and in goal-directed behavior (Yin, Ostlund, Knowlton, & Balleine, 2005). Our model builds upon that evidence. Our cortical Predictor component performs an optional extra step, adding information to the prediction-free system, in distinction from that proposal of two parallel and competing systems. This theory and follow-up work (Daw & Dayan 2014) (along with most verbal theories of decision-making) do not directly address the parallel versus serial distinction upon which we focus, although their theoretical treatment of human decision-making seems consistent with assuming a largely serial process, in which each prediction adds a nontrivial time cost.

Buschman & Miller (2014) present a theory with substantial overlap but substantial differences from ours and that of Daw & Dayan (2014). They focus on a related but separate distinction: BG learns concrete associations quickly, whereas PFC slowly learns more abstract concepts for decision-making. Our model does not currently address this distinction; doing so is a promising avenue for future work. Buschman & Miller propose several possible computational advantages of the interaction between PFC and BG, none of which map closely to our Proposer-Predictor-Actor distinction. In particular, they propose that such loops may allow for stereotyped sequences of actions or thoughts to be strung together, in analogy to the well-studied role of BG in contributing to sequences of motor actions. Our model is consistent with this role but does not currently address it. They also make the suggestion that PFC's ability to capture abstract concepts allows the BG's action-selection to work in more abstract domains. While we propose a collaborative model of action selection between PFC and BG, we agree that this expansion of animal action-selection to complex human decision-making relies on the learning and use of abstract representations in PFC, another computational advantage that we do not directly explore in the current model.

Our proposed mapping from prediction-based versus prediction-free decision-making to anatomy is compatible with that proposed by Khamassi & Humphries (2012); however, our mapping of the distinction from goal-directed to model-based computation is not. They propose that dorsomedial striatum participates in model-based action selection, while dorsolateral striatum participates in model-free action selection. Our model would occupy more anterior areas of frontal cortex, and our basal ganglia maps to different areas of dorsal medial striatum, because it receives from OFC (Balleine, Delgado, & Hikosaka, 2007) and performs goal-directed behavior. It is critical to note that our model performs goal-based behavior even in the no-predictor variant whose computations are prediction-free, and model-free (by what appears to be the most common definition of the term). This confusion is one reason we prefer the term prediction-based to model-based. We address this issue in more detail in O'Reilly et al. (2020).

Koechlin & Hyafil (2007) also present a theory broadly consistent with ours but specific to cortical contributions; they do not address the role of basal ganglia or dopamine. They focus on human anterior prefrontal cortex (APFC) and review evidence showing its importance in complex, branching decision-making tasks. Donoso, Collins, & Koechlin (2014) present a related theory of how APFC is involved in strategy testing in complex tasks. These are both consistent with our theory, although our current model does not directly address the distinction; we propose that the same circuits we outline here are also at work in the APFC, and the same serial process is used to select strategies (or plans in our terminology) based on their previous success.

Our model and theory are also broadly compatible with Botvinick & Weinstein's (2014) review of work in hierarchical reinforcement learning. We make similar arguments for the computational advantages of making decisions hierarchically: selecting a broad plan, then subgoals, and only finally selecting actions. While our theory is compatible with that work, our focus here is on outlining the neural mechanisms that instantiate that computational process, and our current model does not progress through such a hierarchical decision process. Expanding the model to perform such a process is another topic for future work.

A great deal of work has been devoted to mathematical theories of decision-making. Because those theories make no contact with detailed neuroscience data, there is (so far) little contact between the current theory and that mathematical level of analysis. Because our theory allows for a variety of decision-making strategies, composed of different cognitive steps, it could be used to match results from many of the wide variety of models that are still being debated (Hastie & Dawes, 2010). We view those theories as working on a separate but complementary level to this one.

One notable exception is the drift-diffusion or sequential sampling framework, which has been closely related to anatomy and neural data. The mathematics of drift diffusion have been argued to closely characterize neural firing in relevant areas of cortex during perceptual decisions (e.g., lateral intraparietal lobe during a task based on movement; Hanks, Ditterich, & Shadlen, 2006; and other perceptual cortical areas for different perceptual tasks; O'Connell et al., 2018). Neural activity in basal ganglia have also been shown to closely map to a drift diffusion model (Ding & Gold, 2013). It has been argued that the basal ganglia's structure matches an extended version of drift diffusion that takes into account the evidence accumulated for other options to achieve optimality under certain conditions (Bogacz & Gurney, 2007). Although we have not simulated a task with progressive information accumulation, we believe our model is consistent with those findings and theories. We would map the cortical accumulation of evidence primarily to the Proposer model component, with the basal ganglia Actor component merely following, because reward history has an uncomplicated relation to perceptual category in those tasks.

Where the current theory disagrees with sequential sampling models is the serial/parallel distinction when they are applied to complex tasks (Busemeyer & Townsend, 1993). Even when they incorporate serial attention, sequential sampling models are fundamentally parallel in their accumulation of evidence (Diederich, & Oswald 2014). While such models may capture some aspects of complex decisions quite well, we argue that they do not match either the computations or underlying mechanisms involved in human decision-making over time scales longer than about a second.

The perceptual decision tasks usually modeled by drift diffusion models are relatively simple and well-practiced. We would actually predict that such a task would be approached in parallel, with separate neural populations in cortex and basal ganglia accumulating evidence for each option simultaneously. The present model does not capture such parallel computations, but many others do (e.g., Collins & Frank, 2014). We think serial consideration of each option is necessary to perform complex and novel decisions involving multiple factors. It seem possible for neural learning to bind together the information relevant to each option when it is relatively simple and well-practiced (e.g., leftward motion to pressing a button with the left hand), but much more difficult with multiple factors (varying goals and situations, as in our task). Furthermore, such parallel consideration becomes undesirable when a decision is important enough to merit devoting all available cognitive resources to fully consider each option, and time is available to do so.

Risky and biased decision-making

Our manipulations addressing risky decision-making are a modest first effort to apply our model and theory to addressing the individual differences, or risk factors, for risky decision-making. Our manipulation of D1/Go versus D2/NoGo pathway strength in dorsal striatum matched previous empirical results (Stopper, Khayambashi, & Floresco, 2013). We think that these results suggest the utility of such a rich and detailed model in addressing the biological and environmental causes of risky behavior, but they certainly are not comprehensive. It remains for future work to expand on those predictions. Simulating a different task that better captures real-world risky decision-making would be useful in more fully capturing that phenomena. Such a task would be closer to a gambling task: payoffs are highly stochastic, with relatively little to learn about good and bad options.

The second key factor in human risky decision-making is capturing how humans learn about highly risky activities without directly experiencing the worst consequences. It seems clear that humans learn from vicarious experience; for instance, hearing about an auto accident caused by drunk driving seems to change behavior, despite a lack of firsthand experience with the outcome. Mental simulation has strong theoretical and empirical support (Reviewed in Barsalou, 2008). The precise mechanisms of vicarious learning are an important outstanding question, one we hope to address in future work. In particular, it is not known whether the dopamine system participates in vicarious learning, or whether that direct signal of reward and predicted reward is reserved for real rewards. Understanding how vicarious learning works in relation to risky decisions should have important implications for interventions, because most interventions involve communicating information about outcomes, rather than actual, experienced outcomes.

Risky decision-making has a good deal of overlap with bad, or biased, decision-making. Many paradigms do not distinguish risk-seeking behavior (in which some individuals prefer risk-reward tradeoffs in which high risks produce equally great average rewards) from bad or biased decision-making, in which some individuals simply make bad decision in certain domains, such as domains in which bad outcomes simply outweigh good ones, such as gambling against the house. Those decisions also are classified as examples of cognitive biases.

The current theory makes two important behavioral predictions regarding sources of biases. The first prediction is that some risky decision-making results from a failure to make a prediction-based decision. This strategy results from a tradeoff: the construction of useful predictive models is relatively time-consuming, and so not worth performing for less important decisions. Such a time tradeoff has been proposed as one major component in mental effort effects (Shenhav, Musslick, Lieder, Kool, Griffiths, Cohen, & Botvinick, 2017; Kurzban, Duckworth, Kable, & Myers, 2013). Impulsive individuals, who are more vulnerable to making risky decisions (Białaszek, Gaik, McGoun, & Zielonka, 2015), may be biased in their preference for this time-saving tradeoff.

The second prediction is that, even when some amount of predictive model-building is performed, the model, and therefore the decision, will be biased by several factors. Most predictive models will by necessity be incomplete, since most domains do not allow for a construction of all paths to all outcomes in finite time (Daw & Dayan, 2014). The subset of outcomes that are predicted may be nonrandom and biased. One important bias should arise from motivated reasoning effects. For instance, if it is more pleasant to think about positive outcomes, people will include more pleasant than unpleasant outcomes in their model than an unbiased estimate. This bias should occur because our model posits that outcome prediction models are created based on further iterations of the same canonical circuit, other areas of PFC and BG "decide" whether to create each predictive step in the model.

The behavior of the basal ganglia Actor component of each such predictive loop is shaped by and so ultimately under the control of dopamine reward signals. Because those signals sum total predicted reward across time and dimensions (Schultz, 2013), this system does not produce locally optimal decisions. Of particular importance, decision-making in challenging domains appears to be highly biased toward perceived social reward, which is one type of *motivated reasoning* (Kahan, Jenkins-Smith, Braman, 2011). For instance, one may anticipate a proximal monetary reward for getting the right answer to a simple math problem but also anticipate a social reward from peers for getting the answer that accords with their political beliefs (Kahan et al 2011; Kahan, Peters, Dawson, & Slovic, 2017). Motivated reasoning has been proposed as an underlying cause for flawed decisions of enormous consequence, such as the intelligence community deciding that Iraq likely possessed weapons of mass destruction (Jacobson, 2010). Understanding the neural basis of decision-making should help us understand and compensate for motivated reasoning, and perhaps other important biases. Our current model does model basal-ganglia based and dopamine-trained control components of the loops involved in creating predictions, so capturing such an effect remains for future work.

Conclusions

We have presented a relatively computationally and mechanistically detailed theory of how human beings make decisions. We presented a computational model that captures the core of that theory, as a canonical brain circuit. The mechanisms we proposed for this microcircuit are based on extensive empirical work on animal action selection, so the most central proposal is that human complex decision-making uses similar mechanisms and computations, enhanced by using more abstract representations, and more iterative and hierarchical steps. To allow such iteration to accumulate useful sub-decisions into a complex decision, each step must work serially by considering a single proposed action, plan, or conclusion at a time.

We term that canonical decision circuit a Proposer-Predictor-Actor-Critic model, in which the cortex proposes a potential plan or action, the basal ganglia acts to accept or reject that plan, and the amygdala and associated subcortical systems acts as a critic to gauge the success of that plan relative to expectations. This basic process can be enhanced to incorporate specific predictions of outcomes by the use of additional iterations of such a circuit to serve as a Predictor component, which can provide additional information to the Actor at the cost of extra time. When it uses this process of creating predictions, the circuit is performing model (or prediction)-based computations; when the Predictor component is not involved, the computations are largely model (or prediction)-free (although the learning in the other components may induce some neurons to represent likely outcomes, and so constitute a limited form of model-based processing).

While the empirical support for such a canonical circuit, and its use in human decision-making is indirect, we think it is quite strong. However, it remains for future work to investigate how such a circuit might work in detail, and whether and how a series of relatively simple choices can aggregate to create the most complex human planning, decision-making, and thinking.

Acknowledgments This work was funded by NSF grant 5R01GM109996-03, "Integrated Cognitive Modelling for Risky Decision-Making"; Office of Naval Research contract D00014-12-C-0638, "Neural Mechanisms of Human Adaptive Executive Control," and ONR contract N00014-18-C-2067, "Capturing the Power and Pitfall of Human Decision-Making."

The authors thank Stephen Read and Thomas Hazy for extensive discussions of the theory and Prescott Mackie for extensive programming and design of an early version of the model.

The model version reported here is available at: https://github.com/ecortex/HerdKruegerNairEtAl20.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, *9*(1), 147–169.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*, 357–381.

Ashby, F. G., Turner, B. O., & Horvitz, J. C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. *Trends in Cognitive Sciences, 14*(5), 208-215.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience, 19*(12), 2082–2099.

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*, *27*(31), 8161-8165.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4-5), 407-419.

Barsalou, L. W. (2008) Grounded cognition. *Annual Review of Psychology,* 59:617–45

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. L. Davis & D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia.* Cambridge, MA: MIT Press.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics* (5), 834-846.

Białaszek, W., Gaik, M., McGoun, E., & Zielonka, P. (2015). Impulsive people have a compulsion for immediate gratification—certain or uncertain. *Frontiers in Psychology*, 6, 515.

Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, *19* (2), 442-477.

Botvinick, M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Phil. Trans. R. Soc. B*, *369* (1655), 20130480.

Brown, J., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, *17*, 471–510.

Buschman, T. J., & Miller, E. K. (2014). Goal-direction and top-down control. Phil. Trans. R. Soc. B, 369(1655), 20130471.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, *100*(3), 432.

Cai, X., & Padoa-Schioppa, C. (2014). Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron*, *81*(5), 1140-1151.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychological Review, 97(3), 332-361.

Collins, A. G., & Frank, M. J. (2014). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychological Review, 121(3), 337.

Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. Phil. Trans. R. Soc. B, 369(1655), 20130478.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron, 69(6), 1204-1215.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience, 8(12), 1704.

Dayan, P. (2007). Bilinearity, rules, and prefrontal cortex. *Frontiers in Computational Neuroscience*, *1*, 1.

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. Cognitive, Affective, & Behavioral Neuroscience, 14(2), 473-492.

Deco, G., & Rolls, E. T. (2002). Object-based visual neglect: a computational hypothesis. *European Journal of Neuroscience*, *16* (10), 1994–2000.

Deco, G., & Rolls, E. T. (2003). Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *The European Journal of Neuroscience*, *18*, 2374–2390.

Diederich, A., & Oswald, P. (2014). Sequential sampling model for multiattribute choice alternatives with random attention time and processing order. *Frontiers in Human Neuroscience, 8*, 697.

Ding, L., & Gold, J. I. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron*, *79* (4), 640-649.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology* , *22* (6), 1075-1081.

Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science*, *344* (6191), 1481–1486.

Dunovan, K., & Verstynen, T. (2016). Believer-Skeptic meets Actor-Critic: Rethinking the role of basal ganglia pathways during decision-making and reinforcement learning. *Frontiers in Neuroscience*, *10*, 106.

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in cognitive sciences*, *18*(10), 526-535.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*, 51–72.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1* (2), 137-160.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104* (41), 16311-16316.

Frank M. J., Seeberger L. C., O'Reilly R. C. By carrot or by stick: Cognitive reinforcement learning in parkinsonism. Science. 2004;306:1940–43.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66* (4), 585-595.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30*.

Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Current Opinion in Neurobiology*, *15*. http://www.ncbi.nlm.nih.gov/pubmed/16271465

Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science* , *265* (5180), 1826–1831.

Grossberg, S. (2013). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, *37*, 1–47. https://doi.org/10.1016/j.neunet.2012.09.017

Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, *84* (6), 401-410.

Haber, S. N. (2010). Integrative Networks Across Basal Ganglia Circuits. In *Handbook of Behavioral Neuroscience* (Vol. 24, pp. 535-552). Elsevier.

Haber, S. N. (2017). Anatomy and connectivity of the reward circuit. In J-C. Dreher and L. Tremblay (Eds.) *Decision Neuroscience* (pp. 3-19). Academic Press.

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology, 35(1), 4-26

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience*, *9*(5), 682.

Hastie, R., & Dawes, R. M. (2010). *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. Sage.

Hayden, B. Y. (2018). Economic choice: the foraging perspective. Current Opinion in Behavioral Sciences, 24, 1-6.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, *139*, 105–118.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 362(1485), 1601-1613.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience & Biobehavioral Reviews*, 34(5), 701-720.

Herd, S. A., Banich, M. T., & O'reilly, R. C. (2006). Neural mechanisms of cognitive control: An integrative model of Stroop task performance and fMRI data. *Journal of Cognitive Neuroscience*, *18* (1), 22-32.

Herd, S. A., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014). A neural network model of individual differences in task switching abilities. *Neuropsychologia*, *62*, 375-389.

Herd, S. A., Krueger, K. A., Kriete, T. E., Huang, T. R., Hazy, T. E., & O'Reilly, R. C. (2013). Strategic cognitive sequencing: a computational cognitive neuroscience approach. *Computational intelligence and neuroscience*, *2013*.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences USA*, *81*, 3088–3092.

Hunt, L. T., Malalasekera, W. N., de Berker, A. O., Miranda, B., Farmer, S. F., Behrens, T. E., & Kennerley, S. W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature Neuroscience*, *21* (10), 1471.

Hyafil, A., & Moreno-Bote, R. (2017). Breaking down hierarchies of decision-making in primates. *eLife*, *6* , e16650.

Jacobson, G. C. (2010). Perception, memory, and partisan polarization on the Iraq War. *Political Science Quarterly*, *125* (1), 31-56.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, *15* (4-6), 535-547.

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: Consensus and controversy. *Neuron* , *63* (6), 733–745. https://doi.org/10.1016/j.neuron.2009.09.003

Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, *14* (2), 147-174.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1* (1), 54-86.

Khamassi, M., & Humphries, M. D. (2012). Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. *Frontiers in Behavioral Neuroscience*, *6*, 79.

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science* , *318* (5850), 594-598.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off? *PLOS Comput Biol*, *12*(8), e1005090. https://doi.org/10.1371/journal.pcbi.1005090

Kool, W., Cushman, F. A., & Gershman, S. J. (2018). Competition and cooperation between multiple reinforcement learning systems. In *Goal-Directed Decision Making* (pp. 153-178). Academic Press.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36* (6), 661-679.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron* , *81* (3), 687-699.

Lorteije, J. A., Zylberberg, A., Ouellette, B. G., De Zeeuw, C. I., Sigman, M., & Roelfsema, P. R. (2015). The formation of hierarchical decisions in the visual cortex. *Neuron*, *87* (6), 1344-1356.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, *50* (4), 381-425.

Mollick, J., Hazy, T., Herd, S., Mackie, P., Nair, A., Krueger, K., & O'Reilly, R. (in press) A Systems-Neuroscience Model of Phasic Dopamine. *Psychological Review*

Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in parkinsonism. *Journal of Neuroscience*, *28* (47), 12294-12304.

Nelson, A. B., & Kreitzer, A. C. (2014). Reassessing models of basal ganglia function and dysfunction. *Annual Review of Neuroscience*, *37*, 117-135.

O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging neural and computational viewpoints on perceptual decision-making. *Trends in neurosciences*, *41*(11), 838-852.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18* (2), 283-328.

O'Reilly, R. C., Hazy, T. E., & Herd, S. A. (2016). The {Leabra} Cognitive Architecture: How to Play 20 Principles with Nature and Win! In S. Chipman (Ed.), *Oxford Handbook of Cognitive Science*. Oxford University Press. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199842193.001.0001/oxfordhb-9780199842193-e-8

O'Reilly, R. C., & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2012). *Computational Cognitive Neuroscience*. Wiki Book, 1st Edition, URL: http://ccnbook.colorado.edu. http://ccnbook.colorado.edu

O'Reilly, R. C., Nair, A., Russin, J., & Herd, S. A. (2020). How Sequential Interactive Processing Within Frontostriatal Loops Supports a Continuum of Habitual to Controlled Processing. Frontiers in Psychology, 11, 380.

O'Reilly, R.C., Wyatte, D., Herd, S., Mingus, B. & Jilk, D.J. (2013). Recurrent Processing during Object Recognition. *Frontiers in Psychology, 4,* 124.

Oh-Descher, H., Beck, J. M., Ferrari, S., Sommer, M. A., & Egner, T. (2017). Probabilistic inference under time pressure leads to a cortical-to-subcortical shift in decision evidence integration. *NeuroImage*, *162* , 138-150.

Pasupathy, A., & Miller, E. K. (2005). Different time courses for learning-related activity in the prefrontal cortex and striatum. *Nature*, *433*, 873–876.

Rich, E. L., & Wallis, J. D. (2016). Decoding subjective decisions from orbitofrontal cortex. *Nature Neuroscience*, *19*(7), 973.

Rudebeck, P. H., & Murray, E. A. (2014). The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron*, *84* (6), 1143-1156.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: explorations in the microstructure of cognition. Volume 1, foundations.

Sallet, J., Mars, R. B., Noonan, M. P., Neubert, F. X., Jbabdi, S., O'Reilly, J. X., … & Rushworth, M. F. (2013). The organization of dorsal frontal cortex in humans and macaques. *Journal of Neuroscience*, *33* (30), 12255-12274.

Schroll, H., & Hamker, F. H. (2013). Computational models of basal-ganglia pathway functions: focus on functional neuroanatomy. *Frontiers in Systems Neuroscience*, *7*, 122.

Schultz, W. (2013). Updating dopamine reward signals. Current Opinion in Neurobiology, 23 (2), 229-238.

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience*, *17* (3), 183.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. Annual review of neuroscience, 40, 99-124.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review*, *84* (2), 127.

Simon, D. A., & Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *Journal of Neuroscience*, *31* (14), 5526-5539.

Sippy, T., Lapray, D., Crochet, S., & Petersen, C. C. (2015). Cell-type-specific sensorimotor processing in striatal projection neurons during goal-directed behavior. *Neuron*, *88* (2), 298-305.

Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychological Review*, *119* (1), 120.

Stopper, C. M., Khayambashi, S., & Floresco, S. B. (2013). Receptor-specific modulation of risk-based decision making by nucleus accumbens dopamine. Neuropsychopharmacology, 38(5), 715.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychological Review*, *88*(2), 135.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55* (4), 189.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6* (2), 171-178.

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. European Journal of Neuroscience, 29(11), 2225-2232.

Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, *125* (3), 329.

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., & Tang, X. (2017). Residual attention network for image classification. arXiv preprint arXiv:1704.06904.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267-279.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, *7*(2), 70–76.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat Rev Neurosci*, *7*(6), 464–476.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, *22* (2), 513-523.