Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Behavioral Sciences

Inference as a fundamental process in behavior

Ramon Bartolo and Bruno B Averbeck

In the real world, uncertainty is omnipresent due to incomplete or noisy information. This makes inferring the state-of-the-world difficult. Furthermore, the state-of-the-world often changes over time, though with some regularity. This makes learning and decision-making challenging. Organisms have evolved to take advantage of environmental regularities, that allow organisms to acquire a model of the world and perform model-based inference to robustly make decisions and adjust behavior efficiently under uncertainty. Recent research has shed light on many aspects of model-based inference and its neural underpinnings. Here we review recent progress on hidden-state inference, state transition inference, and hierarchical inference processes.

Address
Laboratory of Neuropsychology, NIMH/NIH, Building 49 Room 1B80, 49 Convent Drive MSC 4415, Bethesda, MD 20892-4415, United States

Corresponding author: Bartolo, Ramon (rbartolo@comunidad.unam.mx)

Current Opinion in Behavioral Sciences 2020, **38**:8–13

This review comes from a themed issue on **Computational cognitive neuroscience**

Edited by **Geoff Schoenbaum** and **Angela J Langdon**

https://doi.org/10.1016/j.cobeha.2020.06.005

2352-1546/Published by Elsevier Ltd.

Introduction

In a changing environment, learning, decision-making and cognitive control are critical functions for adaptive behavior under uncertainty. Making decisions involves integrating multiple pieces of information from multiple sources with varying degrees of certainty. When a given decision invariably produces the same outcome, inferring the consequences of choices is straightforward. In the real world, however, uncertainty is common because the information available to a decision-making agent is typically incomplete or hidden by noise.

When facing a novel environment, two scenarios are possible: a) the animal has to learn environmental contingencies from scratch, repeatedly sampling noisy information to form associations between choices and corresponding outcomes, using *model-free* reinforcement learning strategies to drive decision making [1]; or b) some prior information is available, allowing the animal to infer various environmental features using *model-based* learning strategies. Because environmental regularities occur, and these can be used to build priors, actively making inferences about the state-of-the-world is often the best solution.

Inference from incomplete information occurs at multiple levels of cognition. At the perceptual level, percepts are formed by combining noisy or incomplete sensory information with prior beliefs (i.e. models), acquired through experience, to infer features of a sensory object. Inference increases processing speed and reduces the energy necessary to make perceptual decisions [2,3]. Moreover, perceptual errors (e.g. hallucinations in mental disease) have been associated with faulty perceptual inference [4].

In higher cognition, animals form beliefs about the world from environmental regularities and use them to infer future outcomes and optimize decision-making. Bayesian-like computations, that combine prior probability distributions with currently available information are used to make inferences, though typically in a (mathematically) suboptimal way [5,6]. Here we review recent progress regarding the neural underpinnings of inference in decision making, focusing on state inference, state-transition inference, and hierarchically organized inference processes.

Inferring the current state-of-the-world

Without prior information about the potential outcomes of decisions, organisms first need to learn the values of actions (Box 1). Much attention has been devoted to this initial learning process. Reinforcement Learning (RL) is the brute force approach to estimating values of options or actions given the current environment, which results in the gradual development of choice preferences (Figure 1a). This strategy is called *model-free* learning, because it does not rely on prior beliefs. RL is driven by prediction errors signaled by midbrain dopamine neurons [7]. Though computationally inexpensive, RL requires multiple trials to form useful action-value estimates. The number of trials required increases with uncertainty. Gradually updating value through RL may be suboptimal in an ever-changing world. Using prior information can reduce the need for a large number of samples. A comparatively small number of trials provide information that, combined with prior knowledge, is enough to infer *hidden states* [8] at the cost of higher computational requirements [9].

Box 1 Basic concepts of Reinforcement Learning and Bayesian Inference of state transitions

Reinforcement Learning (RL) and Action Value: In RL, *value* represents the association strength between an action and its outcome and is updated after each trial in proportion to a Reward Prediction Error (RPE). RPE is the difference between the current *action value* and the last outcome. Thus, if action A has a high *value* for (i.e. is strongly associated with) positive outcome *O*, an omission of *O* after *A* results in a large negative RPE. Similarly, if action *B* has a low *value* for outcome *O*, obtaining *O* after *B* will give a large positive RPE. The RPE is weighted by a *Learning Rate* (δ), which can be fixed, or dependent upon the RPE sign, or variable across trials, and so on. After updating, *value* is transformed into a choice probability that drives the next choice. After a state transition in a reversal learning task, keeping δ constant, the number of reward omissions needed to bring down the action *value* to a level that makes the action unlikely is proportional to *value* before the state-transition (Figure 1b).

Rescorla–Wagner model: It is perhaps the simplest RL model. *Value* updates are given by:

$$v_i(k+1) = v_i(k) + \delta_f(R(k) - v_i(k))$$

Where v_i is the value estimate for option i, $R(k)$ is the outcome for the choice for trial k, and δ_f is an outcome-dependent learning rate parameter, where f indexes whether the current choice was rewarded ($R = 1$) or not ($R = 0$), that is, δ_{pos}, δ_{neg}.

Bayesian inference model for state transitions: Bayesian models incorporate task structure information, either implicitly or explicitly. For example, a model that infers that the outcome mapping will reverse across two options after some trial may be:

$$p(reversal|choices, outcomes) = \frac{f(choices, outcomes|reversal)p(reversal)}{p(choices, outcomes)}$$

This model includes explicit information in the form of a prior *p(reversal)*. Information is also implicit in the likelihood function *f(choices,outcomes|reversal)*, because it assumes that the outcome mapping can follow from two states. This results in a posterior probability distribution over trials that the outcome mapping has switched, that can be used to infer the trial at which the reversal occurred with a level of certainty determined by the level of noise (see Figure 1d) and the shape of the prior.

Mechanistic modeling and the neural substrates of *model-based* inference processes are currently a matter of intense research. Recent work has focused on overtrained subjects that have acquired a model of the world and can use it to efficiently infer future outcomes and make choices [10,11•,12]. For example, a recent study found fingerprints of state-inference during the execution of a foraging-like task [13••]. Mice had to seek rewards by nose-poking into one of two ports. On any given trial, only one port was active, delivering reward with probability $p < 1$. After each trial, the currently active port (i.e. current state) could be switched off, thus activating the other one, with a certain probability $q > 0$. With this design, a single reward means that the chosen port is active, while an omission does not necessarily indicate that the port is inactive. Thus, animals need to integrate information over several trials to solve the task. Analysis of behavior across training showed that, in early training, the number of consecutive reward omissions before the mice switched to the other port was proportional to the number of rewards at that port before it was switched off, consistent with *model-free*, gradual value updating (see Figure 1b). In contrast, during late training sessions the number of reward omissions needed to switch ports was independent of the number of rewards received at the port, consistent with *model-based* inference and accumulation of statistical evidence rather than gradual value updating. The *model-based* strategy was largely impaired when the Orbitofrontal Cortex (OFC) was inactivated, in agreement with other work suggesting that OFC is important for *hidden-state* representation [11•,14–17]. Interestingly, inactivation of the Anterior Cingulate Cortex (ACC) did not impair inference [13••] but multiplicatively increased the number of reward omissions before switching ports. Previous work has suggested that ACC is important for state inference [18,19••]. However, this dissociation fits the idea that ACC represents state-related information but not the state itself [20,21].

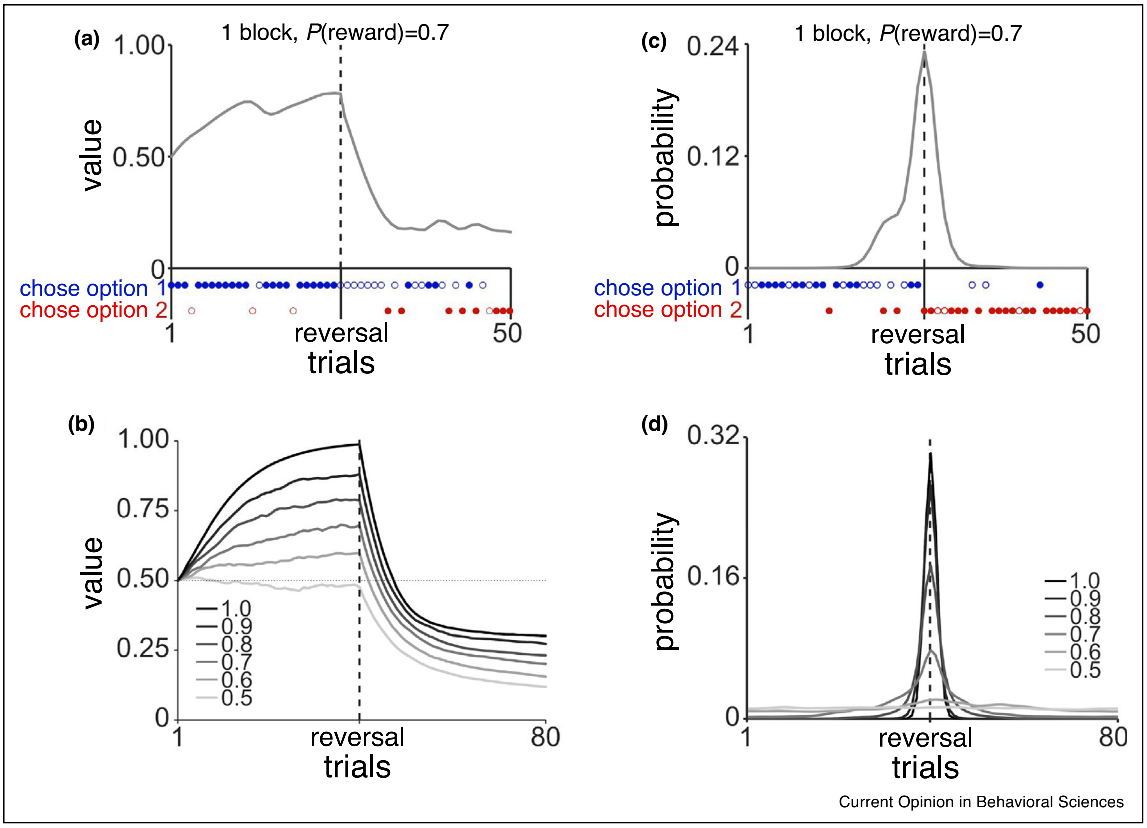
Other frontal regions have been implicated in state inference. For example, mouse infralimbic (IL) PFC is important for inferring states within a trial [22••]. Also, neural activity associated with 'explore' versus 'exploit' states has been observed in the Frontal Eye Fields [23••] and the dorsal-lateral PFC (dlPFC) represents current behavioral states (i.e. 'rules') [24]. Moreover, midbrain dopamine activity reflects beliefs about the current state [25,26], suggesting a role of dopamine in *model-based* inference, in addition to its well-known role in *model-free* learning.

Inferring state-transitions

In Vertechi *et al.* [13••] state transitions drove belief updates. If state transitions follow a pattern, a probability distribution over transitions can be estimated and used to infer future transitions, optimizing behavioral flexibility and robustness under high uncertainty (Figure 2).

Reversal learning tasks have been widely used to assess behavioral flexibility [11•,21,27]. Recent work has focused on overtrained animals, that have acquired a model of the reversal learning task [10,11•,12,28]. In a recent study, monkeys were overtrained on a reversal-learning task to investigate how knowledge of task structure was used to infer state transitions in a multifaceted

Figure 1

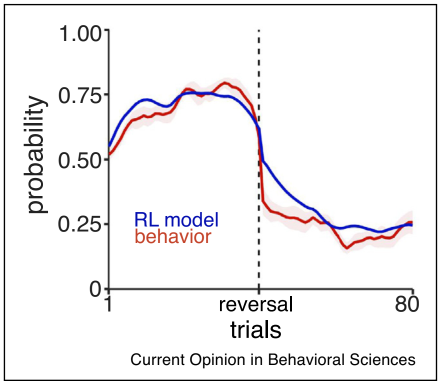


Reinforcement Learning versus Bayesian Inference at State Transitions. **(a)** Action Value for choice of option 1, in a two-armed bandit, across trials from a Reinforcement Learning simulation following the Rescorla–Wagner rule (see Box 1). Blue and Red circles indicate choices (option 1 or option 2) made on each trial. The choice delivered reward when circles are filled, and empty circles indicate the choice was not rewarded. Note that value decreases gradually across trials after the reward mapping was switched ('reversal'). **(b)** The reward rate (grayscale coded) determines the value of each choice. Note that, as in Ref. [13••], the number or trials needed to bring *Value* to a specific level (e.g. 0.5) is proportional to *Value* of the optimal choice before the reward mapping reversal, shifting the curves to the right. **(c)** Posterior probability distribution over trials from a Bayesian inference model (see Box 1). This model assumes that reversals can occur within a certain window during a block of trials (prior). The Posterior Probability that a reversal has occurred increases after reward omissions during the reversal window. Thus, rather than a gradual value update, evidence is accumulated under assumptions determined by prior knowledge. **(d)** Inference allows one to estimate robustly if a reversal has occurred, with a level of certainty (maximum posterior probability) that depends upon environmental noise (reward probability - grayscale). (a) and (c) show data for 1 simulated block. (b) and (d) show averages for 100 simulated blocks.

environment [29••]. The task was organized in blocks of trials. In each trial, monkeys chose between two images, with different reward probabilities, simultaneously presented to the left and right. To find the optimal choice, monkeys had to sample both options. Within each block, the reward mapping was reversed across options at a randomly selected trial within a consistent window near the middle of the block. A choice preference toward the optimal option developed gradually throughout the first few trials (Figure 2). This was well described by a classic RL model, as opposed to an inference process. Within the block the monkeys had to decide when the switch in the reward mapping occurred (i.e. the reversal), so they could switch their choices. After the reward mapping flipped, the behavioral reversal was abrupt, inconsistent with the cumulative *value* updating characteristic of *model-free* RL (Figure 2). A behavioral model that used Bayesian change-point inference (Box 1) captured the choice reversal more accurately (Figure 1c), implying that knowledge of the task structure was used to infer that the state-of-the-world (i.e. the best option) had switched. It is important to point out that inferring state transitions is different from inferring the current state, even though both processes may be intertwined.

As previously discussed, several regions of the PFC are important for state inference, and it seems that the dlPFC also codes state transitions in the context of a task where detecting switches in stimulus-outcome mappings is optimal. This inference process might be an approximation to

Figure 2



Observed behavior is not completely explained by model-free Reinforcement Learning. The adjustment of behavior after flipping the choice-outcome association in a reversal learning task is too abrupt (red curve) and cannot be tracked by an RL model (blue curve). The abrupt transition, nevertheless, is well captured by a Bayesian model (Box 1). Plot produced with data from Ref. [29••]. Solid red line is the mean choice behavior collected from 2 animals (48 blocks each), and the shaded region is the SEM across sessions ($n = 8$). Blue curve shows the predicted choice probabilities from a Rescorla–Wagner fit to the data.

Bayesian change-point inference. Neural ensembles in the dlPFC exhibited activity patterns that followed closely posterior probability estimates of choice preference reversal. In fact, the dlPFC signal predicted a reversal of choice preference from the time of the outcome of the previous trial (i.e. whether the previous choice was rewarded) until the decision in the trial in which behavior reversed. This suggests that the dlPFC contributes to evaluating evidence about state transitions during reversal learning.

A broad range of models exist between *model-free* RL, which has no task-structure information, and Bayesian models with full task-structure information. RL algorithms are considered mechanistic, since they estimate subjective *values* for each option based on Reward Prediction Errors (RPEs) that are signaled by the midbrain dopamine system [7,30]. In contrast, the Bayesian change-point inference model [29••] is a descriptive model that computes a posterior probability distribution (over the trials in the block) that the animal reversed its choice preference. The findings in dlPFC contribute to placing these formal models within a neural framework. However, the network mechanisms underlying these inference processes remain a topic for future research.

Learning to make optimal choices in multifaceted environments involves a network of cortical and subcortical areas, including PFC, amygdala, basal ganglia and thalamus [31,32]. Even simple learning tasks likely engage processes on multiple time-scales [33] from working and episodic memory [34,35], to dopamine-mediated plasticity or spike-timing mechanisms that operate on longer time-scales [36,37]. Reversal learning tasks [29••] may recruit both *model-free* mechanisms, presumably mediated by subcortical structures including the striatum and amygdala [28,38] and Bayesian-like *model-based* mechanisms, likely mediated by cortical structures. *Model-based* mechanisms in turn modulate midbrain dopamine RPEs [22••] and the integration of RPEs through striatal cholinergic activity [39]. It is becoming clear that dopamine RPE signaling is modulated by previously acquired models, suggesting a shared mechanism for both *model-free* and *model-based* learning [40]. To understand how multiple learning processes are orchestrated, more work will be necessary to dissect the specific contributions and interactions of multiple cortical and subcortical nodes.

Higher order rules: multi-level inference and hierarchical reasoning.

Most tasks used to study inference processes are designed in such a way that a single 'behavioral state' among all possible 'behavioral states' can be assumed during the whole task. In the real world, however, most of behavior can be represented as sequences of hierarchical decisions [41]. In the study by Bartolo and Averbeck [29••], state transitions were subordinated to a broader context. In any given block, the reward probability mapping was defined in one of two possible ways: in What blocks, reward probabilities were associated to the images, independent of their location (left/right). In Where blocks, reward probabilities were associated to the left or right target, independent of the specific image presented at that location. Because blocks of either type were randomly interleaved, monkeys had to perform inference at two levels: 1) infer the correct choice strategy (i.e. choose an image or choose a location), which is equivalent to inferring a task from a known set [8]; and 2) infer the reversal.

Another study addressed hierarchically structured decision processes explicitly using a task with two levels of inference [19••]. First, monkeys had to make a categorical perceptual judgement about the duration of an interstimulus interval, *long* or *short*. Second, the correct response was contingent upon a hidden task-state that animals had to infer: a) In state *PRO*, if the duration was judged as *long* or *short*, the corresponding correct response was *X* or *Y*, respectively; b) in state *ANTI*, the correct response mapping was reversed. Task-state alternated stochastically, and animals reported their belief about the current task-state before making the perceptual judgement. Reward was delivered only if both task-state and perceptual judgement were correct, thus reward omissions could be due to an incorrect perceptual judgement or an incorrect task-state inference. Therefore, the level of certainty about the perceptual judgment greatly influenced the inferred task-state. Since perceptual certainty could be experimentally manipulated, it was possible to dissociate it from the certainty about the task-state. The probability that the animals would report that the task-state had changed depended on the outcome history weighted by the perceptual certainty. Neural activity in both the Dorsal Medial Frontal Cortex (DMFC: supplementary eye field, pre-supplementary, and supplementary motor areas) and ACC showed associations to both outcome and perceptual certainty. Plus, ACC seemed to accumulate evidence for switching state, consistent with the idea that the ACC encodes the relative value of a choice under a specific state (see above) which can be transformed into evidence of a state transition.

In line with its role in inferring state-transitions, the PFC may be the substrate of meta-reinforcement learning processes [42], integrating information from many cortical and subcortical structures, and producing novel implementations of previously learned rules. Also, the amygdala and ventral-striatum contribute to inferring correct choices [10,28] in a way consistent with them mediating *model-free* learning processes [37,43–49]. Both can represent complex reward values, for example, the value of exploring novel options when mediating explore-exploit trade-offs [38]. Hierarchically organized inference processes provide a framework for the integration of dissimilar pieces of information to drive sequences of decisions. Current theories suggest that complex learning mechanisms, including hierarchical RL and *model-based* learning [50–52], may be mediated by the PFC.

Conclusion

Inference processes are ubiquitous in cognition, from the interpretation of sensory inputs to cognitive control. Inference is critical for adaptive behaviors in a changing and noisy environment, both for determining the current state and state transitions. Furthermore, learned behaviors can be considered sequences of states. Hierarchically organized inference processes are the fundamental component that shapes these sequences, thus having a fundamental role in behavior.

Conflict of interest statement

Nothing declared.

CRediT authorship contribution statement

Ramon Bartolo: Conceptualization, Writing - original draft, Writing - review & editing. **Bruno B Averbeck:** Funding acquisition, Writing - original draft, Writing - review & editing.

Acknowledgements

This work was supported by the Intramural Research Program, National Institute of Mental Health/N.I.H. (ZIA MH002928).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sutton RS, Barto AG: *Reinforcement Learning: An Introduction*. edn 1. MIT Press; 1998.

2. Parr T, Corcoran AW, Friston KJ, Hohwy J: **Perceptual awareness and active inference**. *Neurosci Conscious* 2019, **2019** niz012.

3. Press C, Kok P, D Yon: **The perceptual prediction paradox**. *Trends Cogn Sci* 2020, **24**:13-24.

4. Cassidy CM, Balsam PD, Weinstein JJ, Rosengard RJ, Slifstein M, Daw ND, Abi-Dargham A, Horga G: **A perceptual inference mechanism for hallucinations linked to striatal dopamine**. *Curr Biol* 2018, **28**:503-514 e504.

5. Tavoni G, Balasubramanian V, Gold JI: **What is optimal in optimal inference?** *Curr Opin Behav Sci* 2019, **29**:117-126.

6. Matsumori K, Koike Y, Matsumoto K: **A biased Bayesian inference for decision-making and cognitive control**. *Front Neurosci* 2018, **12**:734.

7. Schultz W: **Recent advances in understanding the role of phasic dopamine activity**. *F1000Res* 2019, **8**.

8. Collins AG, Frank MJ: **Cognitive control over learning: creating, clustering, and generalizing task-set structure**. *Psychol Rev* 2013, **120**:190-229.

9. Otto AR, Gershman SJ, Markman AB, Daw ND: **The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive**. *Psychol Sci* 2013, **24**:751-761.

10. Rothenhoefer KM, Costa VD, Bartolo R, Vicario-Feliciano R, Murray EA, Averbeck BB: **Effects of ventral striatum lesions on stimulus-based versus action-based reinforcement learning**. *J Neurosci* 2017, **37**:6902-6914.

11. • Groman SM, Keistler C, Keip AJ, Hammarlund E, DiLeone RJ, Pittenger C, Lee D, Taylor JR: **Orbitofrontal circuits control multiple reinforcement-learning processes**. *Neuron* 2019, **103**:734-746 e733.

Using a directionally specific viral ablation approach Groman *et al.* show that the orbitofrontal cortex (OFC) encodes separable reinforcement learning, playing a critical role in decision-making in dynamic environments. OFC neurons projecting to the Nucleus Accumbens incorporate negative outcomes into values. Amygdala neurons projecting to OFC incorporate positive outcomes into values. And OFC neurons projecting to the amygdala stabilize action values.

12. Farashahi S, Donahue CH, Khorsand P, Seo H, Lee D, Soltani A: **Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty**. *Neuron* 2017, **94**:401-414 e406.

13. •• Vertechi P, Lottem E, Sarra D, Godinho B, Treves I, Quendera T, Oude Lohuis MN, Mainen ZF: **Inference-based decisions in a hidden state foraging task: differential contributions of prefrontal cortical areas**. *Neuron* 2020, **106**:166-176 e166.

Using a foraging-like task with a hidden state and that allows to distinguish inference from simpler strategies, Vertechi *et al.* show that mice and humans adopt inference-based strategies that are tuned to the environment statistics. Optogenetic inactivation reveals that the orbito-frontal cortex is necessary for this inference process, while the anterior cingulate cortex seems to fine tune inference parameters to environmental statistics.

14. Schuck NW, Cai MB, Wilson RC, Niv Y: **Human orbitofrontal cortex represents a cognitive map of state space**. *Neuron* 2016, **91**:1402-1412.

15. Bradfield LA, Dezfouli A, van Holstein M, Chieng B, Balleine BW: **Medial orbitofrontal cortex mediates outcome retrieval in partially observable task situations**. *Neuron* 2015, **88**:1268-1280.

16. Costa VD, Averbeck BB: **Primate orbitofrontal cortex codes information relevant for managing explore-exploit tradeoffs**. *J Neurosci* 2020, **40**:2553-2561.

17. Sadacca BF, Wied HM, Lopatina N, Saini GK, Nemirovsky D, Schoenbaum G: **Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task**. *eLife* 2018, **7**.

18. Durstewitz D, Vittoz NM, Floresco SB, Seamans JK: **Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning**. *Neuron* 2010, **66**:438-448.

19. •• Sarafyazd M, Jazayeri M: **Hierarchical reasoning by neural circuits in the frontal cortex**. *Science* 2019, **364**:eaav8911.

Sarafyazd and Jazayeri used a task that involved hierarchically organized decision-making in two levels. First a perceptual decision, that determined the correct response for a second decision, which in turn was contingent upon a hidden state that had to be inferred. After a negative outcome, the cause is ambiguous. They show that, to resolve the ambiguity, the anterior cingulate cortex (ACC) accumulates evidence associated with inferring the hidden state.

20. Lin WH, Gardner JL, Wu SW: **Context effects on probability estimation**. *PLoS Biol* 2020, **18**:e3000634.

21. Kolling N, Wittmann MK, Behrens TE, Boorman ED, Mars RB, Rushworth MF: **Value, search, persistence and model updating in anterior cingulate cortex**. *Nat Neurosci* 2016, **19**:1280-1285.

22. •• Starkweather CK, Gershman SJ, Uchida N: **The medial prefrontal cortex shapes dopamine reward prediction errors under state uncertainty**. *Neuron* 2018, **98**:616-629 e616.

Starkweather *et al.* show that dopamine neurons, in addition to their role in signaling reward prediction errors, they incorporate hidden-state inference related signals under uncertainty. They also show that the medial prefrontal cortex mediates hidden state inference influence on dopamine signals, shedding light on a neural circuit underlying reinforcement learning under state uncertainty.

23. •• Ebitz RB, Albarran E, Moore T: **Exploration disrupts choice-predictive signals and alters dynamics in prefrontal cortex**. *Neuron* 2018, **97**:450-461 e459.

Ebitz *et al.* report that monkeys shift between exploratory and exploitative states, and that reduced prefrontal control may underlie behavioral flexibility. They show that spatially selective, choice predictive neurons in the prefrontal cortex do not predict choices before exploratory decisions. During exploration, prefrontal choice-predictive activity is disrupted, suggesting that exploration is associated to reward-dependent reorganization of neural dynamics.

24. Genovesio A, Brasted PJ, Mitz AR, Wise SP: **Prefrontal cortex activity related to abstract response strategies**. *Neuron* 2005, **47**:307-320.

25. Babayan BM, Uchida N, Gershman SJ: **Belief state representation in the dopamine system**. *Nat Commun* 2018, **9**:1891.

26. Starkweather CK, Babayan BM, Uchida N, Gershman SJ: **Dopamine reward prediction errors reflect hidden-state inference across time**. *Nat Neurosci* 2017, **20**:581-589.

27. Chau BK, Sallet J, Papageorgiou GK, Noonan MP, Bell AH, Walton ME, Rushworth MF: **Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques**. *Neuron* 2015, **87**:1106-1118.

28. Costa VD, Dal Monte O, Lucas DR, Murray EA, Averbeck BB: **Amygdala and ventral striatum make distinct contributions to reinforcement learning**. *Neuron* 2016, **92**:505-517.

29. •• Bartolo R, Averbeck BB: **prefrontal cortex predicts state switches during reversal learning**. *Neuron* 2020, **106**:1044-1054 S0896-6273(20)30231-2.

Bartolo and Avecbeck show that primates performing a two-armed bandit reversal learning task use knowledge of the task structure to infer state transitions, that is, reversals, reflected in an abrupt behavioral switch that is well described by a Bayesian change-point inference model. Furthermore, they show that there is a population signal in the dorsal-lateral prefrontal cortex that predicts the behavioral switches and that follows closely the posterior probability estimates from the Bayesian model. These findings provide insight for mechanistic modelling of state-transition inference processes.

30. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH: **A causal link between prediction errors, dopamine neurons and learning**. *Nat Neurosci* 2013, **16**:966-973.

31. Neftci EO, Averbeck BB: **Reinforcement learning in artificial and biological systems**. *Nat Mach Intell* 2019, **1**:133-143.

32. Lee D, Seo H, Jung MW: **Neural basis of reinforcement learning and decision making**. *Annu Rev Neurosci* 2012, **35**:287-308.

33. Averbeck BB: **Amygdala and ventral striatum population codes implement multiple learning rates for reinforcement learning**. *IEEE Symposium Series on Computational Intelligence* 2017.

34. Collins AG, Frank MJ: **How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis**. *Eur J Neurosci* 2012, **35**:1024-1035.

35. Gershman SJ, Daw ND: **Reinforcement learning and episodic memory in humans and animals: an integrative framework**. *Annu Rev Psychol* 2017, **68**:101-128.

36. Frank MJ: **Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism**. *J Cognit Neurosci* 2005, **17**:51-72.

37. Averbeck BB, Costa VD: **Motivational neural circuits underlying reinforcement learning**. *Nat Neurosci* 2017, **20**:505-512.

38. Costa VD, Mitz AR, Averbeck BB: **Subcortical substrates of explore-exploit decisions in primates**. *Neuron* 2019, **103**:533-545 e535.

39. Bell T, Lindner M, Langdon A, Mullins PG, Christakou A: **Regional striatal cholinergic involvement in human behavioral flexibility**. *J Neurosci* 2019, **39**:5740-5749.

40. Langdon AJ, Sharpe MJ, Schoenbaum G, Niv Y: **Model-based predictions for dopamine**. *Curr Opin Neurobiol* 2018, **49**:1-7.

41. Niv Y: **Learning task-state representations**. *Nat Neurosci* 2019, **22**:1544-1553.

42. Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M: **Prefrontal cortex as a meta-reinforcement learning system**. *Nat Neurosci* 2018, **21**:860-868.

43. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ: **Dissociable roles of ventral and dorsal striatum in instrumental conditioning**. *Science* 2004, **304**:452-454.

44. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ: **Cortical substrates for exploratory decisions in humans**. *Nature* 2006, **441**:876-879.

45. Hampton AN, Adolphs R, Tyszka MJ, O'Doherty JP: **Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex**. *Neuron* 2007, **55**:545-555.

46. Rudebeck PH, Ripple JA, Mitz AR, Averbeck BB, Murray EA: **Amygdala contributions to stimulus-reward encoding in the macaque medial and orbital frontal cortex during learning**. *J Neurosci* 2017, **37**:2186-2202.

47. Taswell CA, Costa VD, Murray EA, Averbeck BB: **Ventral striatum's role in learning from gains and losses**. *Proc Natl Acad Sci U S A* 2018, **115**:E12398-E12406.

48. Seo M, Lee E, Averbeck BB: **Action selection and action value in frontal-striatal circuits**. *Neuron* 2012, **74**:947-960.

49. Lee E, Seo M, Dal Monte O, Averbeck BB: **Injection of a dopamine type 2 receptor antagonist into the dorsal striatum disrupts choices driven by previous outcomes, but not perceptual inference**. *J Neurosci* 2015, **35**:6298-6306.

50. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ: **Model-based influences on humans' choices and striatal prediction errors**. *Neuron* 2011, **69**:1204-1215.

51. Doll BB, Simon DA, Daw ND: **The ubiquity of model-based reinforcement learning**. *Curr Opin Neurobiol* 2012, **22**:1075-1081.

52. Abe H, Seo H, Lee D: **The prefrontal cortex and hybrid learning during iterative competitive games**. *Ann N Y Acad Sci* 2011, **1239**:100-108.