HHS Public Access

Author manuscript

Trends Cogn Sci. Author manuscript; available in PMC 2020 April 01.

Published in final edited form as:

Trends Cogn Sci. 2019 April ; 23(4): 278–292. doi:10.1016/j.tics.2019.01.010.

Holistic Reinforcement Learning: The Role of Structure and Attention

Angela Radulescu[1,2], **Yael Niv**[1,2], and **Ian Ballard**[3,*]

[1]Psychology Department, Princeton University

[2]Princeton Neuroscience Institute, Princeton University

[3]Helen Wills Neuroscience Institute, University of California, Berkeley

Abstract

Compact representations of the environment allow humans to behave efficiently in a complex world. Reinforcement learning models capture many behavioral and neural effects, but do not explain recent findings showing that structure in the environment influences learning. In parallel, Bayesian cognitive models predict how humans learn structured knowledge, but do not have a clear neurobiological implementation. We propose an integration of these two model classes in which structured knowledge learned via approximate Bayesian inference acts as a source of selective attention. In turn, selective attention biases reinforcement learning towards relevant dimensions of the environment. An understanding of structure learning will help resolve the fundamental challenge in decision science: explaining why people make the decisions they do.

Keywords

representation learning; corticostriatal circuits; category learning; striatum; rule learning; bayesian inference; dopamine; approximate inference

How do We Learn what to Learn About?

The complex, multidimensional nature of the external world presents humans with a basic challenge: learning to represent the environment in a way that is useful for making decisions. For example, a wine neophyte on her first trip to Napa could learn that white wines are more refreshing in hot weather. This experience would give rise to a useful distinction between white wines and all other wines, regardless of other dimensions such as the grape varietal or winery. Such **representation learning** (see Glossary) often involves dimensionality reduction — the pruning out of distinctions that are unlikely to be important [1]. For example, after multiple wine tastings, our wine enthusiast might start paying attention to the type of grape, but still ignore the color of the label on the bottle. For any given task, a learner

*Corresponding Author. 132 Barker Hall # 3190, Berkeley, CA 94720. iancballard@gmail.com.

should ideally form a compact representation of the environment that includes all relevant input from either the immediate sensory snapshot or memory. Making too few distinctions may lead to ignoring differences between situations that require different behavior. Making too many distinctions may lead to poor generalization to new, related situations. Understanding how learners arrive at a representation of their environment is a fundamental challenge in cognitive science.

Reinforcement learning algorithms are powerful descriptors of how humans learn from trial and error and substantial progress has been made in mapping these algorithms onto neural circuits [2–4]. Yet reinforcement learning approaches still lack an account of how humans learn task representations in multidimensional environments. In particular, recent work has highlighted two phenomena that current models do not fully capture: selective attention to a subset of features in the environment and the use of structured knowledge. These findings come from two literatures that have seen relatively little crosstalk, reinforcement learning models of decision-making and Bayesian models of category learning. We summarize recent advances at the intersection of the two literatures, with the goal of emphasizing parallels between them and motivating new research directions that can establish how neural circuits give rise to useful structured representations of the external world. We propose that representations of task structure learned through approximate Bayesian inference are the source of selective attention during learning. At the neural level, such structured representations in prefrontal cortex determine which aspects of the environment are learned about via dopaminergic modulation of corticostriatal synapses.

Probing How Humans Learn Task Representations

A burgeoning literature has begun to address the question of how humans learn task representations in two different, but related domains: reinforcement learning and category learning. In a typical multidimensional reinforcement learning task, participants are given a choice between stimuli that vary along several dimensions, and observe a reward outcome after every choice (Table 1, yellow). For example, the participant might be presented with a red square and a blue triangle, choose the red square, and receive a binary reward [5]. Often, reward probability does not uniformly depend on all features. For instance, one feature may be more predictive of reward than others [6,7]. Given only partial information about which features matter, participants are instructed to maximize reward in a series of sequential decisions. By contrast, in category learning tasks participants are required to sort multidimensional stimuli one at a time into one of several categories (Table 1, green). Category membership usually depends on the presence or absence of one or more features, as well as on the relationship between features. For example a red square would be classified as a “dax” if “all red objects are daxes” or as a “bim” if “all red circles are daxes” [8]. Finally, a related class of “weather prediction” tasks (Table 1, blue) ask participants to predict one of two outcomes (e.g. rain or shine) given multiple cues and a non-uniform mapping between cues and outcome probabilities [9,10].

Such tasks differ in framing (e.g. decision-making vs. categorization), the size of the observation space (i.e. how many dimensions stimuli can vary on), the nature of the feedback (scalar reward vs. a category label; stochastic vs. deterministic) and the

instructions the participant receives about the structure (Table 1). But they also are alike in that each trial consists of a **perceptual observation,** an **action** and a **reward outcome**. And they share the key property that the participant needs to disambiguate observations by learning and representing an appropriate mapping between perceptual observations and **environmental states** [4,11–13]. In other words, performance depends on learning to carve the perceptual observation space into a compact **state representation** appropriate for the task. In this review, we refer to the process of learning a mapping from observations to states as representation learning (note that there also exist other kinds of representation learning problems, such as that of learning the transition structure between states (i.e. model-based learning) [14]).

Several studies have tackled the question of what kinds of state representations humans use. These studies built reinforcement learning agents ("models") that simulated actions trial-by-trial, and compared these predictions to human data [5,7–10,15–17]. Importantly, models that have been proposed vary in the state representation they learn over (Fig. I, Box 1). For example, in object-based reinforcement learning [5,18], the agent maintains a value for each combination of features (Fig. I left). In other words, there is a one-to-one correspondence between unique percepts and states. In feature-based reinforcement learning (a form of function approximation [19]), the agent maintains values for all possible features in the environment and generalizes across observations by combining the predictions of constituent features (e.g., by summing them; Fig. I right). Recent work on multidimensional reinforcement learning has suggested that humans do not use one strategy exclusively. Instead, participants tended to adopt feature-based learning when information about features is predictive and object-based learning when it is not [5,10], and did not rely on object-based representations in a task in which a single feature is more predictive of reward [7] (cf. [20]). These findings suggest that state representations can flexibly adapt to task structure.

Attention Shapes State Representations

How can it be that people use different representations in different learning tasks? One possible explanation is that selective attention dynamically shapes state representations. Selective attention is defined as the preferential processing of a subset of environmental features at any stage between perception and action. Several recent studies have reported that attended features influence actions more strongly than unattended features, and are learned about more readily [7,17,21,22]. Moreover, this interaction between attention and learning is bidirectional: attentional biases are dynamically adjusted as a function of reward history [22].

Selective attention could sculpt state representations for reinforcement learning via known neurobiological mechanisms. Reinforcement learning is thought to occur via dopamine-dependent plasticity of synapses from cortical neurons onto striatal neurons [23]. In response to an unexpected reward, dopamine neurons fire in proportion to the reward prediction error [2] and release dopamine onto their striatal targets [24], facilitating long-term potentiation of corticostriatal synapses [25]. If cortical neurons are firing in response to a sensory cue, such as a tone, at the same time as a surprising reward, then their synapses onto striatal neurons will strengthen (and vice versa for a negative prediction error, which causes long-term

depression of synapses). If attention is directed to a subset of sensory features, then the cortical response to those features will be stronger and more precise [26]. Because these neurons are firing more, their synapses onto striatal neurons will be strengthened more in response to unexpected reward than those of unattended features. Attention could therefore bias reward-driven value learning towards a subset of features in the sensory environment.

If selective attention shapes the representations used for learning, then an important empirical question is, which aspects of the observation space are subject to attentional selection? Two prominent theories suggest different targets for attention in reinforcement learning. The Mackintosh model proposes that attention tracks stimulus features that are most predictive of reward [27]. The Pearce and Hall model suggests instead that attention is directed to features that learners are most uncertain about [28]. Hybrid approaches suggest that both processes occur independently [29,30], while integration models posit that a stimulus' learned attentional salience is directed by both predictiveness and uncertainty [31,32], with the balance of the two possibly different at the choice versus learning stage of a decision [33,34]. Additionally, features of the stimulus, such as its visual salience, can capture attention in a bottom-up manner [35] and such features ought to be learned about more readily. While a topic of active debate, how the brain dynamically directs attention during either choice or learning remains an open question [36].

Insights from Structure Learning

Humans use knowledge about the world to scaffold learning. This knowledge can take the form of abstract concepts [37], domain knowledge [38], relational maps of the environment [39,40], or hierarchically structured knowledge of task rules [41]. Knowledge about structure endows humans with a remarkable ability to generalize behavior to new environments. It also enables rapid "one-shot" learning of new concepts and "learning to learn" [42,43], the ability to learn more quickly after experience with a class of learning problems. Cognitive psychology research has successfully modeled such structure learning phenomena using Bayesian cognitive models.

Probabilistic programming models (Fig. 1, top row) construct rules (or concepts) from compositions of features with simple primitives (e.g., "for all", "if", "and", "or") [37,42,44]. In category learning tasks, rules are constructed from the stimulus features and simple logical operations (e.g. "(blue and square) or red") [8,37,45]. Given a hypothesis space of all possible rules, the posterior probability is calculated via Bayesian inference, considering both a rule's complexity and the proportion of examples it correctly categorizes. Rules are, a priori, exponentially less likely as the expressive length of the rule increases. Probabilistic programming models are an instance of a class of models that describe concept-learning data by assuming a compositional rule-based structure (e.g., RULEX [46]). The unique contribution of probabilistic programming models is that they use statistical inference to build concepts that are appropriate for the learning problem. This flexibility allows them to explain phenomena outside of categorization, such as how people infer causal relationships between events [47,48] or how people infer others' intentions from their actions [49] or choice of words [50].

Yet category learning tasks only subtly differ from decision-making experiments that are modeled with reinforcement learning (Table 1). A recent study of human category learning highlighted this relationship by directly comparing a Bayesian probabilistic programming model with several reinforcement learning models that learn over different state representations [8,37]. While reinforcement learning models can learn categorization rules if given the correct state representation, the Bayesian model better predicted human choices. In contrast to reinforcement learning, which learns over a predefined state representation and updates uninformative states indefinitely, probabilistic programming models settle on the rules that parsimoniously describe observations.

Bayesian non-parametric models (Fig. 1, middle row) also highlight how Bayesian inference can explain learning phenomena that have eluded reinforcement learning (cf. [51]). Bayesian non-parametric models group perceptual observations into unobserved “latent causes” (or clusters) [52–55]. For example, consider a serial reversal learning task in which the identity of the high-reward option sporadically alternates. In such tasks, animals initially learn slowly and but eventually learn to rapidly respond to contingency changes [56]. Bayesian non-parametric models learn this task by grouping reward outcomes into two latent causes: one in which the first option is better, and one in which the second option is better. Once this structure is learned, the model displays one-shot reversals after contingency changes because it infers that the latent cause has changed. This inference about latent causes in the environment has also shed light on several puzzling conditioning effects. When presented with a neutral stimulus such as a tone followed by a shock, animals eventually display a fear response to the tone. The learned fear response gradually diminishes when the tone is later presented by itself (i.e. in extinction), but often returns after some time has passed. This phenomenon is known as spontaneous recovery. Bayesian non-parametric models attribute spontaneous recovery to the inference that extinction signals a new environmental state. This prevents old associations from being updated [57]. Bayesian nonparametric models also predict that gradual extinction will prevent spontaneous recovery, a finding borne out by empirical data [57]. In gradual extinction, the model infers a single latent state and gradually weakens the association between that state and aversive outcome, thereby abolishing the fear memory.

In both probabilistic programming and Bayesian nonparametric models, learning is biased by a prior that favors simpler representations over complex ones. For example, a probabilistic programming model is biased towards rules in which a single feature is relevant for classification. This simplicity prior is appropriate across tasks and domains. This stands in contrast to reinforcement learning models that require definition of an appropriate state space for each task. A simplicity bias is also consistent with findings that suggest that trial-and-error learning follows a pattern whereby simpler feature-based state spaces precede more complex object-based spaces [5,58], and explains why classification becomes harder as the number of relevant dimensions grows [59]. The findings outlined in this section illustrate both the importance of structured knowledge in learning and the utility of Bayesian cognitive models for explaining how this knowledge is acquired. However, they raise an important question: how does this knowledge about task structure interface with the neural systems that support reinforcement learning?

Bridging Structure Representations and Neural Models of Reinforcement Learning

We propose a conceptual model that links reinforcement learning with a structure learning system in a neurobiologically plausible architecture (Fig. 2). We base our framework on connectionist models of the basal ganglia-prefrontal cortex circuit [41,60,61]. These models describe how rules [41,62,63] or working memory content [61,64] are selected via known corticostriatal circuitry (see [61] for neuroanatomical detail). Antero-lateral prefrontal cortical pools can represent different rules, working memory content, or hypotheses about task structure. These different pools compete via mutual lateral inhibition. The outcome of this competition is biased by the relative strength of each pools' connectivity with the striatum. Pools with stronger cortical-striatal connectivity will generate a stronger striatal response, which in turn increases the strength of thalamic feedback onto these pools. This recurrent circuit allows a pool representing a task rule to inhibit competing pools and control behavior [65]. If an unexpected reward occurs, dopamine release in the striatum strengthens the synapses from the most active cortical pool. In this way, rule representations that lead to reward are more likely to win over alternative representations in the future [41,66]. This model describes how a reinforcement learning system could gate the representation of a hypothesis about task structure into cortex. Our central proposal is that this hypothesis is the source of top-down selective attention during learning (Box 2).

Hypotheses about task structure can constrain feature-based reinforcement learning by directing attention to specific component features and not others. For example, a hypothesis that "red stimuli are daxes" would increase the strength and fidelity of the representation of color in sensory cortex. If an unexpected outcome follows a red square, the heightened representation of "red" will cause a larger update to the corticostriatal projections from "red" neurons than from "square" neurons. As a result, reinforcement learning will operate over a feature-based representation with biased attention to the color "red". If later in the task the hypothesis is updated to "red squares are daxes", then both red and square features will be attended to more strongly than other features. In this way, rules can sculpt the state representation underlying reinforcement learning.

Reinforcement learning could, in turn, contribute to the selection of hypotheses via two mechanisms. First, learning can adjust the corticostriatal weights of projections from cortical pools representing alternative rules, as has been proposed by a recent neural network model [41]. Second, reinforcement learning over features represented in sensory cortex can contribute to rule selection. For instance, the rule "red squares are bims" will be influenced by simple reinforcement learning linking "red" with "bim" and "square" with "bim". Even if the learner is using the rule "red objects are bims", the incidental reinforcement of "square" every time a red square is correctly classified as "bim" will cause higher activation of "red squares" relative to "red circles" neural pools in the cortex. This differential activation will increase the likelihood that the subject switches to the correct rule. This mechanism is consistent with the finding that even when subjects are given the correct categorization rule, features of the stimulus that are not part of the rule exert an influence on decisions [20]. When a rule is discarded, reinforcement learning would also support the selection of an

alternative rule that has explained past observations. This mechanism eliminates the need to remember all previous trials and evaluate alternative rules against these memories, thereby endowing the hypothesis testing system with implicit memory.

By designating hypotheses as the source of top-down attention, this model provides a mechanistic account of how reinforcement learning is influenced by both structured knowledge and attention. This idea is closely related to recent work suggesting that working memory contents in lateral prefrontal circuits act as the source of top-down attention to the constituent sensory circuits [67]. Indeed, working memory plays an important role in constraining reinforcement learning [68], and our model predicts that learning is influenced by the number of hypotheses that can be simultaneously considered [69].

Computational Feasibility of Bayesian Inference

Our model claims that hypotheses about task structure are considered in a manner consistent with Bayesian rational models. However, the computations underlying these models are generally intractable [70]. For instance, in probabilistic programming models, the agent must repeatedly compute the likelihood of all previous observations over all possible rules. Recent work has attempted to address this problem using sampling algorithms that approximate solutions to Bayesian inference [71,72]. One such algorithm is the **particle filter**.

A particle filter can approximate any given Bayesian model by using a finite number of particles, each of which expresses a particular hypothesis about the state of the world [73,74]. For example, in a category learning task, each particle represents a single categorization rule. After an observation, the particle either samples a new hypothesis or stays with the current one. This decision depends on how likely the observation is under the current hypothesis. A particle encoding the belief that “red stimuli are daxes” would be more likely to stay with its current hypothesis after observing a red square that is a “dax”, and more likely to switch to a new hypothesis after observing a red square that is a “bim”. This update algorithm is computationally simple because it only incorporates each particle’s belief about the world (e.g., one classification rule).

In addition to their computational simplicity, particle filters are an appealing model for representation learning because they can preferentially sample simpler rules. Moreover, they capture the phenomenological report that people consider alternative hypotheses [75]. Particle filters also closely resemble a serial hypothesis testing model that has previously been shown to describe human behavior in a multidimensional decision-making task [76]. In addition, they provide a single framework for implementing representation learning over different types of models, including both probabilistic programming models and Bayesian nonparametric models [69,71].

Given an infinite number of particles, particle filters converge to the true posterior probability. Remarkably, recent work has demonstrated that the use of a single or very few particles can describe human behavior well. This is because humans face a practical problem: rather than learning the true probability distribution over all possible rules, people

need only find a rule that explains enough observations to make good decisions [77]. This could explain why behavior across an entire group may be Bayes-optimal, but individual choices are often not [78]. If each individual tracks just one or a few hypotheses, only the group behavior will aggregate over enough “particles” to appear Bayes-optimal [78].

Our neural model proposes that hypotheses are gated by corticostriatal circuitry that is, in turn, influenced by reinforcement learning. This architecture could form the basis of a particle filter algorithm. Specifically, particle filters sample hypotheses based on how well each hypothesis accounts for previous observations. Feature weights learned via reinforcement learning could enable the sampling of hypotheses that have already explained some observations. Unlike particle filter accounts of sensory integration, which propose that individual spikes of feature-selective neurons represent particles [79–82], in our model particles correspond to distributed prefrontal representations of rules. The particle filter algorithm is a flexible mechanism for inference that could apply to different timescales (from milliseconds to trials) and different types of problems (e.g., perception and categorization).

Although corticostriatal connectionist models can exhibit properties similar to a Bayesian structure-learning model [41], the corticostriatal gating mechanism need not perfectly implement a particle filter, and the differences may be informative [77]. For example, in a task where the motor response mapping varies (e.g., “bim” is sometimes the left-hand and sometimes the right-hand response), a corticostriatal gating model would correctly predict that if recent right-hand selections were rewarded, the subject is more likely to respond “right” regardless of category of the current stimulus [83]. A particle filter implementing a probabilistic programming model of representation learning would not predict this effect. A fruitful area for future research will be to examine other ways in which constraints imposed by the corticostriatal architecture can predict deviations from Bayesian inference.

Related Models

Our proposal that prefrontal cortex and striatum interact to support structure learning is related to the longstanding idea that the brain contains multiple, competing learning systems [84,85] (see COVIS for computational implementation [86]). Previous work has shown that in tasks where categorization rules are difficult to verbalize, (e.g., respond left to squares that are more red than they are circular), performance is supported by incremental learning in corticostriatal circuits [87,88]. In contrast, performance of explicit rule-learning tasks depends on prefrontal cortex [89]. Our proposal also relies on the distinction between the kinds of representations and learning supported by the prefrontal cortex and striatum. However, we propose that prefrontal rule representations act as a source of top-down attention that sculpts the state representation over which reinforcement learning operates.

Rule-based categorization models stand in contrast to clustering approaches to learning. In clustering models, stimuli are clustered according to the similarity of their features. This clustering can be biased by attention, such that stimuli that differ in unattended features nonetheless cluster together (e.g., in the attention-learning covering map model; “ALCOVE” [90]). An association between a cluster (e.g., all previous red stimuli) and a category label

(e.g., “dax”) implicitly encodes a categorization rule (“red stimuli are daxes”). This mechanism for building representations has been proposed as a solution to the representation learning problem in reinforcement learning [91]. A related model, the supervised and unsupervised stratified adaptive incremental network (“SUSTAIN”), suggests that a single cluster is activated by each stimulus (akin to a single rule representation in our model) [92]. SUSTAIN shares many similarities to the Bayesian nonparametric models described above and more empirical work is necessary to adjudicate between their mechanisms for forming clustered representations. Importantly, both SUSTAIN and ALCOVE describe a bidirectional relationship between attention and representation learning that is influenced by prediction errors, in line with the framework we suggest here.

The critical difference between clustering models and the model that we describe is the nature of the representation (clusters versus concepts). Because probabilistic programming models perform inference over compositions of concepts, they can handle a broader range of tasks. Consider the task of learning to tie one’s shoelaces. Probabilistic programming models would treat each of the operations that can be applied to a shoelace as a concept, and could learn how to tie knots from compositions of these concepts. These compositions can be rapidly applied to solve new problems, such as tying a bow on a gift or triple-knotting one’s shoelaces before a hike. Categorization tasks may be a special case in which the predictions of cluster-based and rule-based models converge. Moreover, these models may map onto partially distinct neural systems, with the hippocampus and medial temporal lobe cortex supporting learning based on similarity to past exemplars [92,93], and the prefrontal cortex supporting learning based on concepts. Recent work showed that multivariate hippocampal representations of stimuli are similar to predictions of the SUSTAIN model [21]. In category learning, the top-down attention mechanism we propose may also influences hippocampal clustering. Indeed, there is elevated prefrontal-hippocampal functional connectivity during category learning [21].

The role of the hippocampus and surrounding cortex in representation learning is likely to extend beyond the clustering of past experiences. Conjunctive representations of multiple features in the hippocampus can support reinforcement learning over configurations of features [10,94] and the selection of hypotheses relating to conjunctions of features. Further, episodic retrieval of individual past choices and outcomes, as well as previous task rules, has a strong influence on decisions [95,96]. Retrieval is both influenced by and influences top-down attention. As a result, retrieval is likely to interact with the model we propose. For example, retrieval drives reinstatement of cortical representations of features [97], which could lead to reinforcement learning over features that are not present in the environment. Finally, the hippocampus and entorhinal cortex form spatial maps of the environment in the service of spatial learning [98]. Recent modeling and empirical work has emphasized a role for the hippocampus and entorhinal cortex in forming cognitive maps of tasks and a specific function for the hippocampus in signaling that the environment has changed enough to necessitate forming a new state representation [39,93,99]. Thus, a pressing question for the field is what distinguishes the representations of task structure between the hippocampus and prefrontal cortex.

Concluding Remarks

We have outlined recent findings showing the extent to which reinforcement learning is constrained by attention and by the underlying representation of the structure of the environment. We propose that attention is a key mechanism that sculpts the sensory representations supporting learning. Bayesian cognitive models explain unique aspects of behavior in categorization tasks that are very similar to the tasks used to study reinforcement learning. An exciting possible unification of these research threads is that abstract conceptual knowledge that forms the basis of these cognitive models drives top-down attention during learning. Our conceptual model makes several testable predictions. First, particle filter or related approximations to Bayesian cognitive models should describe the diversity of individual subject behaviors in representation learning tasks [69]. Second, application of a rule should be associated with increased sensory cortical responses to the constituent features of that rule, and increased reinforcement learning about them. Third, values learned via reinforcement learning should influence which rules are selected in rule learning tasks.

Manipulations that are known to influence attention and working memory should also influence representation learning. For example, interference from a dual task may reduce the number of hypotheses about task structure (e.g. rules in a categorization task) that can be considered simultaneously. A quantitative prediction of our model is that this would reduce the accuracy of rule learning. From a neuroscience perspective, dual tasks degrade the quality of representations held in working memory [100], which could cause learners to forget rules or implement them more noisily. Another possible consequence of a dual task is that attention is biased against complex rules. This could actually improve learning for tasks with a simple structure (e.g. tasks where a single feature is relevant for categorization or reward), because the dual task prevents the learner from considering complex hypotheses that may have otherwise interfered with learning. In addition, features that drive bottom-up attentional capture, such as exogenously salient or mnemonically important features, should be more likely be incorporated into hypotheses and, as a result, drive top-down attention. The impact of these features should depend on their rule-relevance: if salient features are relevant to the task rules, they should accelerate representation learning.

Answering these and related questions (see Outstanding Questions Box) will help to illuminate how rule representations in prefrontal cortex influence ongoing neural processing in the service of adaptive behavior. More broadly, it will help us to understand how humans build rich representations that are suited to both their environment and their goals, a question central to our understanding of cognition in both health and disease [101].

Acknowledgements

This work was supported by NIMH R01 MH063901 and Army Research Office W911NF-14-1-0101. We are grateful to Maria Eckstein, Adam Eichenbaum, Kevin Miller and Mingyu Song for comments on earlier drafts of this manuscript, and to Nathaniel Daw and Steven Piantadosi for insightful discussions.

Glossary

Action
A response the participant makes, e.g. choosing an option, labeling a stimulus or predicting an outcome.

Bayesian non-parametric models
A class of Bayesian cognitive models that group observations into sets of unobservable latent causes, or clusters.

Environmental state
A subset of environmental features relevant to the agent's goal, e.g. the feature red being more predictive of reward.

Particle filters
A class of sampling methods for approximating arbitrary probability distributions in a sequential manner, by maintaining and updating a finite number of particles (hypotheses).

Perceptual observation
A stimulus, potentially with multiple features.

Probabilistic programming models
A class of Bayesian cognitive models that reason over structured concepts such as rules.

Reinforcement learning
A class of algorithms that learn an optimal behavioral policy, often through learning the values of different actions in different states.

Representation learning
The process by which learners arrive at a representation of environmental states.

Reward outcome
Consequence of an action, e.g. a reward or category label.

Reward prediction error
The difference between the reward outcome and what was expected, used as a learning signal for updating values of states and actions.

State representation
The agent's internal representation of the environmental state.

References

1. McCallum AK and Ballard D (1996), Reinforcement learning with selective perception and hidden state., University of Rochester. Dept. of Computer Science
2. Schultz W et al. (1997) A neural substrate of prediction and reward. Science 275, 1593–1599 [PubMed: 9054347]
3. Niv Y (2009) Reinforcement learning in the brain. J. Math. Psychol
4. Langdon AJ et al. (2018) Model-based predictions for dopamine. Curr. Opin. Neurobiol 49, 1–7 [PubMed: 29096115]

5. Farashahi S et al. (2017) Feature-based learning improves adaptability without compromising precision. Nat. Commun DOI: 10.1038/s41467-017-01874-w
6. Roiser JP et al. (2009) Do patients with schizophrenia exhibit aberrant salience? Psychol. Med 39, 199–209 [PubMed: 18588739]
7. Niv Y et al. (2015) Reinforcement learning in multidimensional environments relies on attention mechanisms. Journal of Neuroscience 35, 8145–8157 [PubMed: 26019331]
8. Ballard I et al. (2017) Beyond Reward Prediction Errors : Human Striatum Updates Rule Values During Learning.
9. Akaishi R et al. (2016) Neural Mechanisms of Credit Assignment in a Multicue Environment. Journal of Neuroscience 36, 1096–1112 [PubMed: 26818500]
10. Duncan K et al. (2018) More Than the Sum of Its Parts: A Role for the Hippocampus in Configural Reinforcement Learning. Neuron 98, 645–657.e6 [PubMed: 29681530]
11. Wilson RC et al. (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81, 267–278 [PubMed: 24462094]
12. Schuck NW et al. (2016) Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. Neuron 91, 1402–1412 [PubMed: 27657452]
13. Kaelbling LP et al. (1998) Planning and acting in partially observable stochastic domains. Artif. Intell 101, 99–134
14. Daw ND et al. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci 8, 1704–1711 [PubMed: 16286932]
15. Daw ND (2009) Trial-by-trial data analysis using computational models. In Decision making, affect, and learning: Attention and performance XXIII pp. 3–38
16. Wunderlich K et al. (2011) The human prefrontal cortex mediates integration of potential causes behind observed outcomes. J. Neurophysiol
17. Marković D et al. (2015) Modeling the Evolution of Beliefs Using an Attentional Focus Mechanism. PLoS Comput. Biol 11, 1–34
18. Gluck M. a. et al. (2002) How do people solve the “weather prediction” task?: individual variability in strategies for probabilistic category learning. Learn. Mem 9, 408–418 [PubMed: 12464701]
19. Sutton RS and Barto AG (1998) Reinforcement Learning: An Introduction. IEEE Trans. Neural Netw 9, 1054–1054
20. Hahn U et al. (2010) Exemplar similarity and rule application. Cognition 114, 1–18 [PubMed: 19815187]
21. Mack ML et al. (2016) Dynamic updating of hippocampal object representations reflects new conceptual knowledge. Proceedings of the National Academy of Sciences 113, 13203–13208
22. Leong YC et al. (2017) Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. Neuron 93, 451–463 [PubMed: 28103483]
23. Calabresi P et al. (2007) Dopamine-mediated regulation of corticostriatal synaptic plasticity. Trends Neurosci 30, 211–219 [PubMed: 17367873]
24. Haber SN and Knutson B (2009) The Reward Circuit: Linking Primate Anatomy and Human Imaging. Neuropsychopharmacology 35, 4–26
25. Shen W et al. (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. Science 321, 848–851 [PubMed: 18687967]
26. Maunsell JHR and Treue S (2006) Feature-based attention in visual cortex. Trends Neurosci 29, 317–322 [PubMed: 16697058]
27. Mackintosh NJ (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. Psychol. Rev 82, 276–298
28. Pearce JM and Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol. Rev 87, 532–552 [PubMed: 7443916]
29. LePelley ME and McLaren IPL (2004) Associative history affects the associative change undergone by both presented and absent cues in human causal learning. J. Exp. Psychol. Anim. Behav. Process 30, 67–73 [PubMed: 14709116]

30. Pearce JM and Mackintosh NJ (2010) Two theories of attention: A review and a possible integration. In Attention and associative learning: From brain to behaviour pp. 11–39
31. Esber GR and Haselgrove M (2011) Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. Proc. Biol. Sci 278, 2553–2561 [PubMed: 21653585]
32. Nasser HM et al. (2017) The Dopamine Prediction Error: Contributions to Associative Models of Reward Learning. Frontiers in psychology 8, 244 [PubMed: 28275359]
33. Dayan P et al. (2000) Learning and selective attention. Nature neuroscience 3, 1218–1223 [PubMed: 11127841]
34. Gottlieb J (2012) Perspective Attention, Learning, and the Value of Information. Neuron 76, 281–295 [PubMed: 23083732]
35. Itti L and Koch C (2001) Computational modelling of visual attention. Nat. Rev. Neurosci 2, 194–203 [PubMed: 11256080]
36. Le Pelley ME et al. (2016) Attention and associative learning in humans: An integrative review. Psychol. Bull 142, 1111–1140 [PubMed: 27504933]
37. Goodman ND et al. (2008) A rational analysis of rule-based concept learning. Cogn. Sci 32, 108–154 [PubMed: 21635333]
38. Tenenbaum JB et al. (2006) Theory-based Bayesian models of inductive learning and reasoning. Trends Cogn. Sci 10, 309–318 [PubMed: 16797219]
39. Stachenfeld KL et al. (2017) The hippocampus as a predictive map. Nat. Neurosci 20, 1643 [PubMed: 28967910]
40. Hodges J (1995) Memory, Amnesia and the Hippocampal System. J. Neurol. Neurosurg. Psychiatry 58, 128
41. Collins AGE and Frank MJ (2013) Cognitive control over learning: creating, clustering, and generalizing task-set structure. Psychol. Rev 120, 190–229 [PubMed: 23356780]
42. Lake BM et al. (2015) Human-level concept learning through probabilistic program induction. Science 350, 1332–1338 [PubMed: 26659050]
43. Kemp C et al. (2010) Learning to Learn Causal Models. Cogn. Sci 34, 1185–1243 [PubMed: 21564248]
44. Tenenbaum JB et al. (2011) How to Grow a Mind: Statistics, Structure, and Abstraction. Science 331, 1279–1285 [PubMed: 21393536]
45. Piantadosi ST (2011) Learning and the language of thought, Massachusetts Institute of Technology.
46. Nosofsky RM and Palmeri TJ (1998) A rule-plus-exception model for classifying objects in continuous-dimension spaces. Psychon. Bull. Rev 5, 345–369
47. Goodman ND et al. (2011) Learning a theory of causality. Psychol. Rev 118, 110–119 [PubMed: 21244189]
48. Griffiths TL et al. (2011) Bayes and blickets: effects of knowledge on causal induction in children and adults. Cogn. Sci 35, 1407–1455 [PubMed: 21972897]
49. Goodman ND. Cause and intent: Social reasoning in causal learning; Proceedings of the 31st annual conference of the cognitive science society; 2009.
50. Frank MC and Goodman ND (2012) Predicting pragmatic reasoning in language games. Science 336, 998 [PubMed: 22628647]
51. Schmajuk NA and DiCarlo JJ (1992) Stimulus configuration, classical conditioning, and hippocampal function. Psychol. Rev 99, 268–305 [PubMed: 1594726]
52. Gershman SJ and Blei DM (2012) A tutorial on Bayesian nonparametric models. J. Math. Psychol 56, 1–12
53. Gershman SJ et al. (2015) Discovering latent causes in reinforcement learning. Current Opinion in Behavioral Sciences 5, 43–50
54. Soto FA et al. (2014) Explaining compound generalization in associative and causal learning through rational principles of dimensional generalization. Psychol. Rev 121, 526–558 [PubMed: 25090430]
55. Gershman SJ et al. (2014) Statistical Computations Underlying the Dynamics of Memory Updating. PLoS Comput. Biol 10,

56. Costa VD et al. (2015) Reversal learning and dopamine: a bayesian perspective. J. Neurosci 35, 2407–2416 [PubMed: 25673835]
57. Gershman SJ et al. (2013) Gradual extinction prevents the return of fear: implications for the discovery of state. Front. Behav. Neurosci 7, 164 [PubMed: 24302899]
58. Choung O-H et al. (2017) Exploring feature Dimensions to Learn a New Policy in an Uninformed Reinforcement Learning Task. Sci. Rep
59. Shepard RN et al. (1961) Learning and memorization of classifications. In Psychological monographs: General and applied 75
60. Cohen JD et al. (2002) Computational perspectives on dopamine function in prefrontal cortex. Curr. Opin. Neurobiol 12, 223–229 [PubMed: 12015241]
61. O'Reilly RC and Frank MJ (2006) Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput 18, 283–328 [PubMed: 16378516]
62. Frank MJ and Badre D (2012) Mechanisms of Hierarchical Reinforcement Learning in Corticostriatal Circuits 1: Computational Analysis. Cereb. Cortex 22, 509–526 [PubMed: 21693490]
63. Villagrasa F et al. (2018) On the role of cortex-basal ganglia interactions for category learning: A neuro-computational approach. J. Neurosci DOI: 10.1523/JNEUROSCI.0874-18.2018
64. Todd MT et al. (2009) Learning to Use Working Memory in Partially Observable Environments through Dopaminergic Reinforcement. Adv. Neural Inf. Process. Syst 21,
65. Alexander GE and DeLong MR (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual review of .. at <http://www.annualreviews.org/doi/pdf/10.1146/annurev.ne.09.030186.002041>
66. Graybiel AM et al. (1994) The basal ganglia and adaptive motor control. Science 265, 1826–1831 [PubMed: 8091209]
67. Kiyonaga A and Egner T (2013) Working memory as internal attention: toward an integrative account of internal and external selection processes. Psychon. Bull. Rev 20, 228–242 [PubMed: 23233157]
68. Collins AGE and Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur. J. Neurosci 35, 1024–1035 [PubMed: 22487033]
69. Lloyd K. Why Does Higher Working Memory Capacity Help You Learn?; Proceedings of the 39st annual conference of the cognitive science society; 2017.
70. Kwisthout J et al. (2011) Bayesian intractability is not an ailment that approximation can cure. Cogn. Sci 35, 779–784 [PubMed: 21609357]
71. Sanborn AN et al. (2010) Rational approximations to rational models: alternative algorithms for category learning. Psychol. Rev 117, 1144–1167 [PubMed: 21038975]
72. Sanborn AN and Chater N (2016) Bayesian Brains without Probabilities. Trends in cognitive sciences 20, 883–893 [PubMed: 28327290]
73. Liu JS and Chen R (1998) Sequential Monte Carlo Methods for Dynamic Systems. J. Am. Stat. Assoc 93, 1032–1044
74. Doucet A et al. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. Stat. Comput 10, 197–208
75. Armstrong SL et al. (1983) What some concepts might not be. Cognition 13, 263–308 [PubMed: 6683139]
76. Wilson RC and Niv Y (2012) Inferring relevance in a changing world. Front. Hum. Neurosci 5, 1–14
77. Lieder F et al. (2018) Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. Psychol. Rev 125, 1–32 [PubMed: 29035078]
78. Courville AC and Daw ND (2008) The rat as particle filter In Advances in Neural Information Processing Systems 20 (Platt JC et al., eds), pp. 369–376, Curran Associates, Inc.
79. Huang Y and Rao RPN (2016) Bayesian Inference and Online Learning in Poisson Neuronal Networks. Neural Comput DOI: 10.1162/NECO

80. Kutschireiter A et al. (2017) Nonlinear Bayesian filtering and learning: a neuronal dynamics for perception. Sci. Rep 7, 8722 [PubMed: 28821729]
81. Legenstein R and Maass W (2014) Ensembles of spiking neurons with noise support optimal probabilistic inference in a dynamically changing environment. PLoS Comput. Biol 10, e1003859 [PubMed: 25340749]
82. Lee TS and Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J. Opt. Soc. Am. A Opt. Image Sci. Vis 20, 1434–1448 [PubMed: 12868647]
83. Lau B and Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. J. Exp. Anal. Behav 84, 555–579 [PubMed: 16596980]
84. Poldrack RA et al. (2001) Interactive memory systems in the human brain. Nature 414, 546 [PubMed: 11734855]
85. Squire LR (2004) Memory systems of the brain: A brief history and current perspective. Neurobiol. Learn. Mem 82, 171–177 [PubMed: 15464402]
86. Ashby FG et al. (1998) A neuropsychological theory of multiple systems in category learning. Psychol. Rev 105, 442–481 [PubMed: 9697427]
87. Ashby FG and Maddox WT (2011) Human category learning 2.0. Ann. N. Y. Acad. Sci 1224, 147–161 [PubMed: 21182535]
88. Ashby FG and Valentin VV (2017) Multiple systems of perceptual category learning: Theory and cognitive tests In Handbook of Categorization in Cognitive Science (Second Edition) pp. 157–188
89. Waldron EM and Ashby FG (2001) The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. Psychonomic bulletin & review 8, 168–176 [PubMed: 11340863]
90. Kruschke JK (1992) ALCOVE: an exemplar-based connectionist model of category learning. Psychol. Rev 99, 22–44 [PubMed: 1546117]
91. Jones M and Canas F (2010), Integrating reinforcement learning with models of representation learning., in Proceedings of the Annual Meeting of the Cognitive Science Society, 32
92. Love BC et al. (2004) SUSTAIN: a network model of category learning. Psychol. Rev 111, 309–332 [PubMed: 15065912]
93. Gershman SJ et al. (2017) The computational nature of memory modification. Elife 6,
94. Ballard IC et al. Hippocampal Pattern Separation Supports Reinforcement Learning. Nature Communications, *in press*
95. Bornstein AM and Norman KA (2017) Reinstated episodic context guides sampling-based decisions for reward. Nature Publishing Group 20, 997
96. Bornstein AM et al. (2017) Reminders of past choices bias decisions for reward in humans. Nat. Commun 8, 15958 [PubMed: 28653668]
97. Nyberg L et al. (2000) Reactivation of encoding-related brain activity during memory retrieval. Proc. Natl. Acad. Sci. U. S. A 97, 11120–11124 [PubMed: 11005878]
98. Shapiro ML and Eichenbaum H (1999) Hippocampus as a memory map: synaptic plasticity and memory encoding by hippocampal neurons. Hippocampus 9, 365–384 [PubMed: 10495019]
99. Behrens TEJ et al. (2018) What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. Neuron 100, 490–509 [PubMed: 30359611]
100. Kiyonaga A et al. (2017) Neural Representation of Working Memory Content Is Modulated by Visual Attentional Demand. J. Cogn. Neurosci 29, 2011–2024 [PubMed: 28777056]
101. Diehl MM et al. (2018) Toward an integrative perspective on the neural mechanisms underlying persistent maladaptive behaviors. Eur. J. Neurosci 48, 1870–1883 [PubMed: 30044022]

Highlights

- Recent advances have refined our understanding of reinforcement learning by emphasizing roles for attention and for structured knowledge in shaping ongoing learning.
- Bayesian cognitive models have made great strides towards describing how structured knowledge can be learned, but their computational complexity challenges neuroscientific implementation.
- Behavioral and neural evidence suggests that each class of algorithms describe unique aspects of human learning.
- We propose an integration of these computational approaches in which structured knowledge learned through approximate Bayesian inference acts as a source of top-down attention, which shapes the environmental representation over which reinforcement learning occurs.

Outstanding Questions Box

- How do reinforcement learning and cortical representations of structure interact to control behavior? Can systematic deviations from Bayesian rationality be explained by reinforcement learning? Can we model why different tasks bias the arbitration between learning systems?
- Are the predictions of corticostriatal connectionist models of rule learning borne out by neural and neuroanatomical data? For example, do anterior cortical projections influence the stimulus-response coding profiles of striatal neurons targeted by motor cortex? Do thalamocortical projections mediate the activation of cortical neurons representing task rules?
- What are the regional differences in task structure learning and representation across prefrontal cortex? In particular, what are the different roles for orbitofrontal cortex and lateral prefrontal cortex in representation learning?
- How does working memory affect representation learning? Specifically, does interference from working memory contents influence hypothesis selection and top-down attention? Do multiple hypotheses or rules interfere with one another due to working memory constraints? Do individual differences in working memory capacity relate to representation learning?
- How do relational, spatial, and episodic knowledge in the hippocampus support or compete with reinforcement learning? Under what conditions does retrieval interfere with or support representation learning?
- Can a union of Bayesian cognitive models with reinforcement learning provide new ideas about education and classroom learning? How can attention be directed to facilitate the discovery of a novel concept?
- Can structure learning account for maladaptive behaviors in psychiatric and substance abuse disorders?

Box 1. State Representation Effects on Learning

In addition to learning how to select appropriate actions, humans and animals learn from trial and error which features of our environment are relevant for predicting reward. Formally, this converts the space of perceptual observations into a **state representation** suitable for the problem at hand.

Imagine for example that on the first trial of a categorization task, you correctly name a red square with a vertical stripe a “dax” (Fig. I, adapted from [8]). The expected value associated with the action “dax” for the current stimulus, V(“dax”|stimulus), can be updated based on the difference between the reward and the initial expected value of that action, i.e., the reward prediction error RPE = R − V_0(“dax”|stimulus), scaled by a learning rate η.

But the update depends on the state representation. For instance, you could either update the entire object (Fig. I left)

$$V_{new}(\text{``dax"} \mid \text{red circle with vertical stripe}) = V_{old}(\text{``dax"} \mid \text{red circle with vertical stripe}) + \eta \cdot \text{RPE}$$

or update individual features (Fig. I right):

$$V_{new}(\text{``dax"} \mid \text{red}) = V_{old}(\text{``dax"} \mid \text{red}) + \eta \cdot \text{RPE} \qquad \text{(and equivalently for ``square" and ``vertical")}$$

These different assumptions about the state representation will lead to diverging reward expectations. If you next encounter a red circle with a horizontal stripe and you have an object-based state representation, you will not expect reward for saying “dax” because you have never encountered this stimulus before. If you have a feature-based representation, you will expect a reward for saying “dax”, because the same action in response to a different red stimulus previously led to reward.

Box 2. How Selective Attention May Emerge from Structure Learning

When stimuli are multidimensional, the state representation underlying reinforcement learning is shaped by selective attention [22]. Total reward expectation can be computed as a weighted sum of expectations from each component feature:

$$V(\text{``dax"} \mid \text{red circle}) = \Phi_{\text{color}} V(\text{``dax"} \mid \text{red}) + \Phi_{\text{shape}} V(\text{``dax"} \mid \text{circle})$$

where V are reward expectations and Φ are attention weights. But how are the weights determined? These weights can be thought of as indexing the allocation of attention to each feature. We illustrate how these attention weights can emerge from inferring latent structure:

Probabilistic programming models (Fig. 1 top row) construct rules from compositions of perceptual features and logical primitives (e.g., “and”, “or”, “not”). Returning to the categorization problem from Fig. I reduced to 2 dimensions (color: red or blue, and shape: circle or square), say you observe evidence in favor of the rule that “red objects are daxes”. Given this rule, you can collapse across the shape dimension and only attend to the color of each object ($\Phi_{\text{color}} = 1$, $\Phi_{\text{shape}} = 0$, Fig. 1 left column, bottom panel). If on the other hand you believe that “squares are daxes”, you can ignore color and only attend to shape ($\Phi_{\text{color}} = 0$, $\Phi_{\text{shape}} = 1$, Fig. 1 middle column, bottom panel). Finally, if your rule is that “red squares are daxes”, you must be able to distinguish across both dimensions ($\Phi_{\text{color}} = 0.5$, $\Phi_{\text{shape}} = 0.5$, Fig. 1 right column, bottom panel).

Bayesian non-parametric models (Fig. 1 middle row) propose an alternative mechanism for structure learning, which is to group observations into clusters, or “latent causes” that generate linked observations. When a latent cause is “on”, it tends to emit linked observations. For example, if you infer that latent variable *y* causes both red and “dax”, you can ignore shape ($\Phi_{\text{color}} = 1$, $\Phi_{\text{shape}} = 0$, Fig. 1 left column, bottom panel), since only the presence of red is relevant for determining whether *y* is active and will also cause “dax”. If on the other hand you infer that latent variable *y* causes both square and “dax”, you can ignore color ($\Phi_{\text{color}} = 0$, $\Phi_{\text{shape}} = 1$, Fig. 1 middle column, bottom panel). Finally if both red and square are related to “dax” via *y*, attention should be allocated to both color and shape, as both dimensions provide information about the likelihood of *y* being active and also causing “dax” ($\Phi_{\text{color}} = 0.5$, $\Phi_{\text{shape}} = 0.5$, Fig. 1 right column, bottom panel).

Both probabilistic programming models and Bayesian non-parametric models offer normative solutions to the problem of learning to represent structure in the environment. Understanding whether and how the different representations they require (rules vs. clusters) map onto neural circuits may help adjudicate between the two model classes.

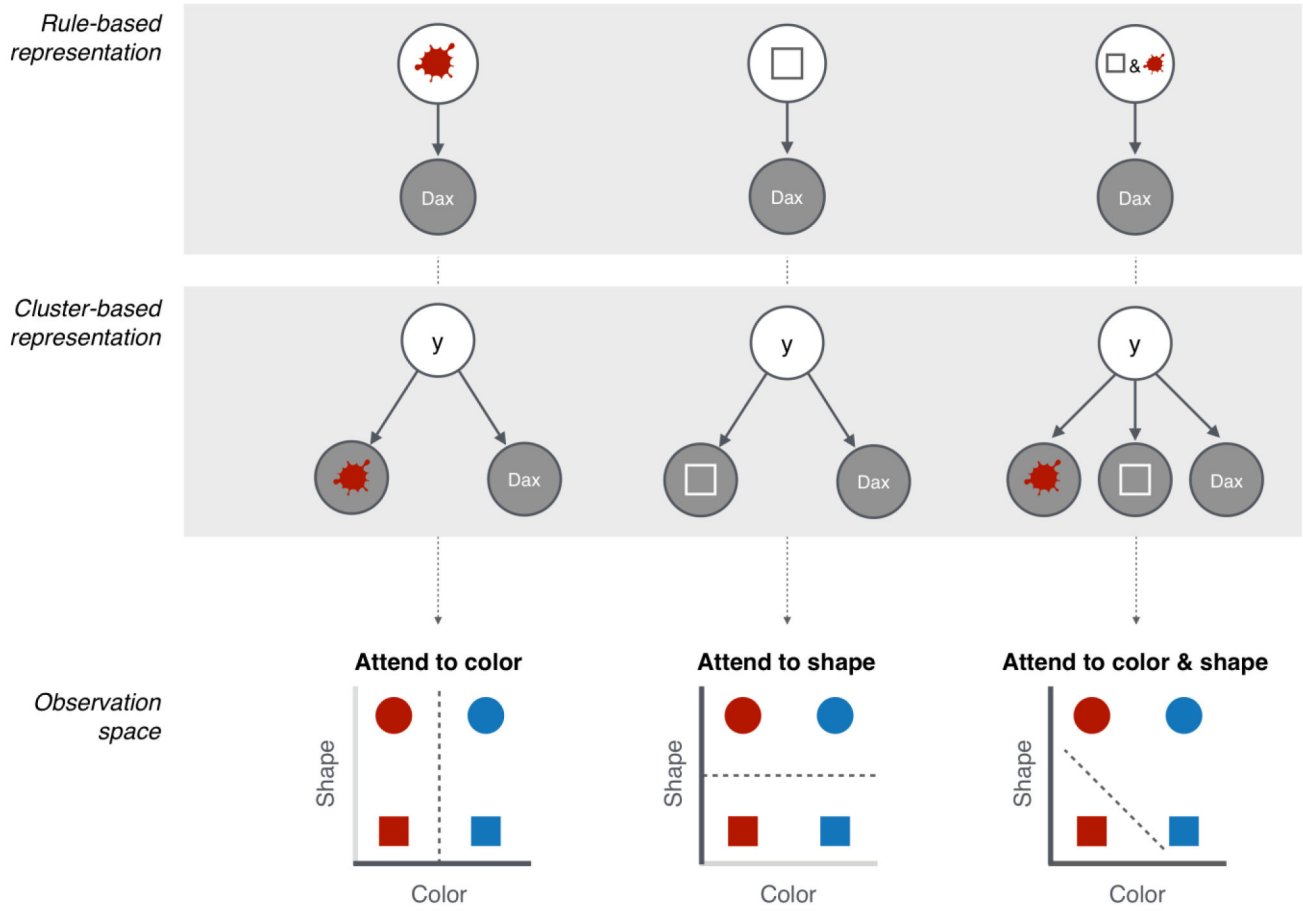

Figure 1: Structure learning guides selective attention.

Rules inferred by probabilistic programming models (top row) or clusters inferred by Bayesian non-parametric models (middle row) lead to different perceptual distinctions in observation space (bottom row). **Left column**: if the agent infers that red objects are "daxes", or that red and "dax" cluster together, then she can ignore shape and only attend to color when categorizing a stimulus as a "dax" or a "bim". **Middle column**: similarly if the agent infers that squares are "daxes" (middle left), or that square and "dax" cluster together, then she can ignore color and only attend to shape. **Right column**: finally if the agent infers that red squares are "daxes", or that red, square and "dax" cluster together, then she should attend to both color and shape.

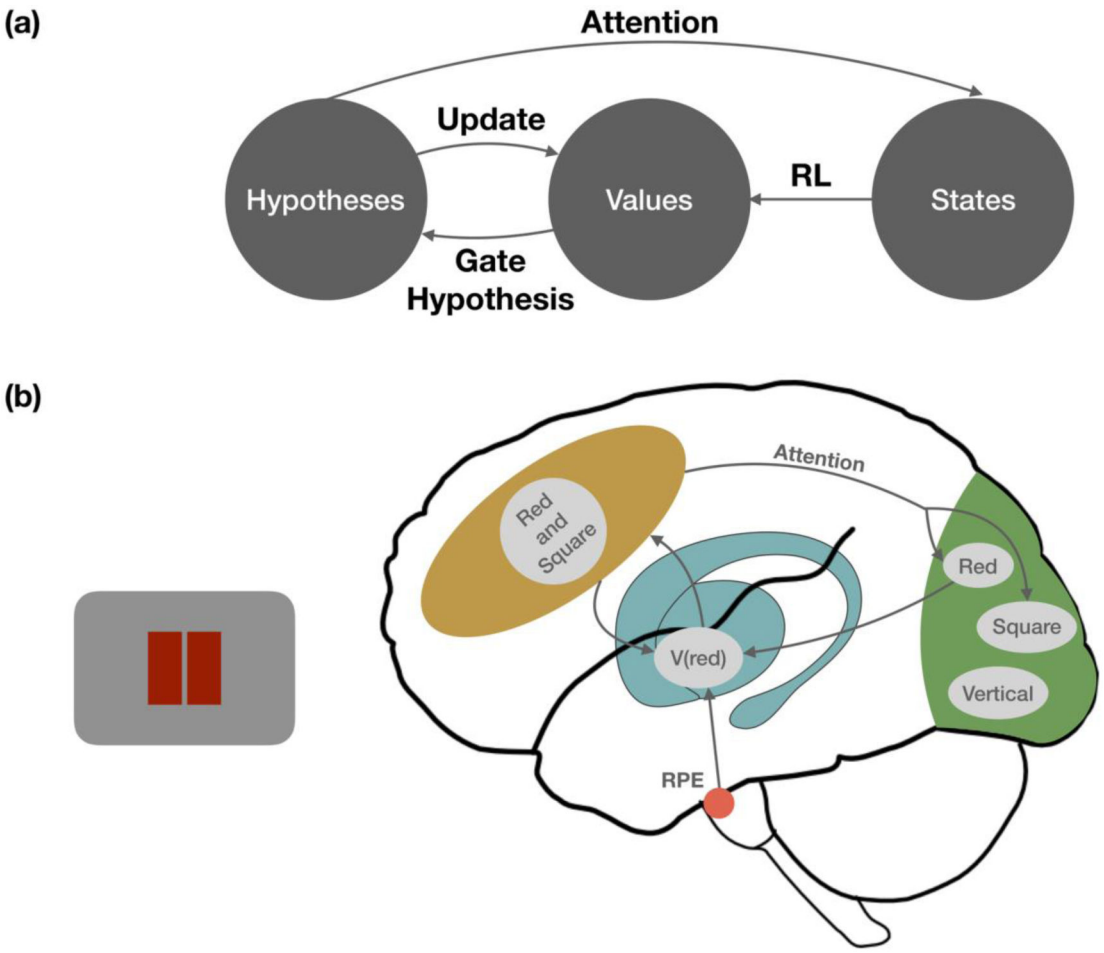

Figure 2: Conceptual model of structure learning.

(a). We propose that 1) hypotheses about task structure are the source of top-down attention and 2) attention sculpts the state representation by prioritizing some features of the environment for reinforcement learning. Existing models suggest that state values (or (state,action) pair values) are learned via reinforcement learning. In turn, learned values about states participate in the gating of which hypotheses are considered. Specifically, hypotheses that are consistent with the high-value (state, action) pairings are more likely to be considered. Finally, prediction errors in response to violations of these hypotheses help to update state values. **(b)** A simple model showing how the interacting systems architecture in **(a)** could be realized in different neural circuits. Yellow area corresponds to the lateral prefrontal cortex, blue to the basal ganglia, green to sensory cortex, and red to the dopaminergic midbrain. A prefrontal hypothesis that “red and square” is the correct categorization rule biases top-down attention to the “red” and “square” features in sensory cortex, which in turn increases learning about these features in response to reward prediction errors (RPE). In turn, values stored in the striatum influence prefrontal rule selection.

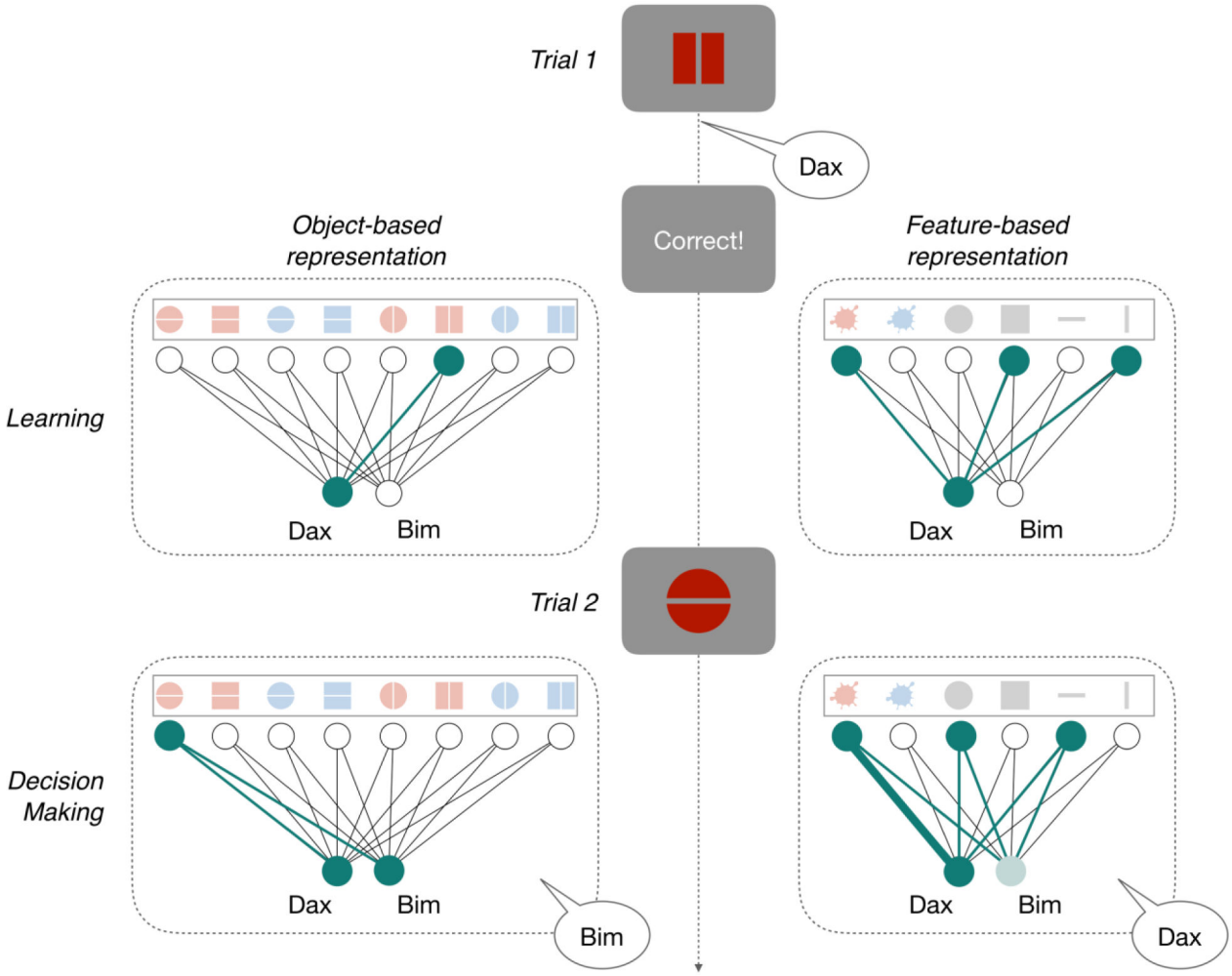

Figure I: Possible state representations for a prototypical representation learning task.

Perceptual observations varying along 3 dimensions (color: red or blue, shape: circle or square, and orientation: horizontal or vertical) can be categorized as “daxes” or “bims”. On trial 1, the participant learns from correctly categorizing the red square with a vertical stripe as a “dax”. She can represent the stimulus either as an object (left) or a composition of features (right). An object-based representation leads to strengthening the association between the object and “dax”, while a feature-based representation leads to strengthening the association between the component features and “dax”. On trial 2, the participant must categorize the red circle with a horizontal stripe. If she has an object-based representation, she is indifferent between “dax” and “bim” (and hence might randomly choose “bim”), while if she has a feature-based representation, she would be more likely to say “dax” due to the previously learned association between red and “dax”. Thicker lines show a stronger association between an input-level representation and the output label; active input units, and their connection to output units, are shown in teal.

Table 1.
Studies addressing representation learning

A representative set of studies that have addressed how humans learn mappings from percepts to states from trial and error. Shown here are distinctions between studies in the size of the observation space, the number of actions available to the subject, and the relationship between stimuli and rewards.

Paper	Observation space	Action space	Reward function
Akaishi et al. 2016 [9]	4 cues	2-alternative weather prediction task	Probabilistic binary outcome contingent on different cue combinations; 2 cues more predictive than the others
Ballard et al. 2017 [8]	3 dimensions × 2 features	2-alternative categorization task	Deterministic binary reward contingent on correctly categorizing stimulus based on 6 possible rules
Choung et al. 2017 [58]	3 dimensions × 2 features	2-alternative go-nogo task	Different cue combinations lead to deterministic positive (+10) or negative (−10) rewards, or to probabilistic rewards (+10 or −10)
Duncan et al. 2018 [10]	4 cues	2-alternative weather prediction task	Probabilistic binary outcome; 2 different environments: separable (both individual cues and combinations are predictive) vs. inseparable (only cue combinations are predictive)
Farashahi et al. 2017 [5]	2 dimensions × 2 features	2-alternative forced-choice task	Probabilistic binary reward; 2 different environments: generalizable (one dimension is on average more predictive) vs. non-generalizable (both dimensions are equally predictive)
Mack et al. 2016 [21]	3 dimensions × 2 features	2-alternative categorization task	Deterministic binary reward for correctly categorizing stimulus based on a diagnostic feature rule or a disjunctive rule
Niv et al. 2015 [7]	3 dimensions × 3 features	3-alternative forced-choice task	Probabilistic binary reward with high probability if subject selects “target feature”, and low probability otherwise