PERSPECTIVE | FOCUS
https://doi.org/10.1038/s41593-019-0470-8

nature neuroscience

Learning task-state representations

Yael Niv

Arguably, the most difficult part of learning is deciding what to learn about. Should I associate the positive outcome of safely completing a street-crossing with the situation 'the car approaching the crosswalk was red' or with 'the approaching car was slowing down'? In this Perspective, we summarize our recent research into the computational and neural underpinnings of 'representation learning'—how humans (and other animals) construct task representations that allow efficient learning and decision-making. We first discuss the problem of learning what to ignore when confronted with too much information, so that experience can properly generalize across situations. We then turn to the problem of augmenting perceptual information with inferred latent causes that embody unobservable task-relevant information, such as contextual knowledge. Finally, we discuss recent findings regarding the neural substrates of task representations that suggest the orbitofrontal cortex represents 'task states', deploying them for decision-making and learning elsewhere in the brain.

The ubiquitous problem of representation learning

Imagine standing on a street corner and preparing to cross the street on your way home (Fig. 1a). Even in the calmest of neighborhoods, your sensory systems will confront a staggering amount of information that may or may not be relevant for the decision of whether to go or to wait. Computationally, avoiding getting run over is daunting. Nevertheless, you can probably complete the street-crossing task successfully even while talking to a friend or mentally planning your afternoon. What allows our brains to make decisions in complex, multidimensional environments with such ease and efficiency?

We argue that the brain solves seemingly complex tasks by learning efficient, low-dimensional representations that simplify these tasks. A useful task representation will focus on aspects of the environment that are critical to correct performance of the task. That is, it will include all factors that are causally related to the outcome of our actions, for instance, the speed and distance of the closest oncoming car. At the same time, the task representation will gloss over all other information: the colors of the cars, the shops across the street, etc. Ignoring input dimensions (color, shape) that are irrelevant for task performance and concentrating on the few dimensions that are critical (speed, distance) allows us not only to make rapid decisions, but also to generalize learning as widely as possible. That is, correctly ignoring irrelevant aspects of the environment will allow learning from one experience to inform decision-making in other scenarios that share relevant features with the current experience and differ only in the irrelevant ones.

Efficient representations are task-specific. When crossing the street, you should represent and act upon the speed and distance of cars, whereas when hailing a taxi, you should represent the color of the car and whether or not the medallion light is on. How does the brain construct a representation for each of our numerous tasks? Summarizing a decade of our own research, here we discuss findings that suggest that two processes are critical to learning task representations: selective attention to only the relevant observable aspects of a task[1,2] and augmentation of these with hidden (unobservable) aspects from memory[3,4]. The learned task state, we will then argue, resides in the orbitofrontal cortex[5,6], which relays to other brain areas a pared-down task representation tailored for each decision.

Reinforcement learning as a framework for decision-making

In recent years, ideas from the computational field of reinforcement learning (RL)[7,8] have revolutionized the study of learning and decision-making in the brain. In RL, tasks are defined as a set of states with action-dependent transitions between them and with rewards (that can also be zero or negative) for each state. The agent starts at some state of the environment and traverses states based on (potential or probabilistic) transitions, collecting rewards (or paying costs) throughout. Tasks do not have unique state representations; for instance, Fig. 1b,c details two alternative representations of the same street-crossing task. In Fig. 1b, the representation includes four states (ovals), each defined by a location. Figure 1c represents the location 'south sidewalk' as two distinct states, depending on whether a car is approaching or not, thus expanding the representation to five states. As can be seen from the action-dependent transitions in the figure (arrows, with actions in rectangles and probabilities of each transition) and as we detail below, some representations are more useful than others. In particular, by using the representation in Fig. 1c to select actions, one can ensure reaching home safely, by choosing A2 (wait) while in state S1b (car approaching) until such as time as a transition to S1a (no car approaching) occurs, at which point one can choose A1 (go). In contrast, given the state representation in Fig. 1b, there is no action policy that will get you home while ensuring you don't get hit by a car (and therefore hospitalized).

RL algorithms provide a host of asymptotically converging methods for learning optimal behavioral policies that will maximize reward and minimize punishment within a given state representation[7,8]. Several reviews have detailed the correspondence between these algorithms and putative neural substrates, as well as challenges to this framework as an explanation of learning and decision-making by humans and other animals (for example, refs. [9–11]). Briefly, in model-free RL, most algorithms are variations on the idea of using prediction errors to learn from experience a value for each state (or for each action in each state; called 'state-action Q-values'[12]) that summarizes the sum of future rewards that can be expected if one is in this state (and takes a particular action). Actions are then chosen that have a higher value and thus are expected to ultimately lead to more reward. In model-based RL, instead of learning values from experience, one uses experience to learn a model of the task: the transition probabilities between states and the probability of reward in each state (for example, the diagram in Fig. 1b). The internal model is then used in a mental simulation to calculate the expected long-term return from different action options and to choose the most rewarding one[13]. Evidence suggests that the brain uses both model-free and model-based RL[14,15], with dopamine-dependent learning in the dorsal parts of the basal ganglia implementing the former[16,17] and prefrontal–hippocampal–medial basal ganglia circuitry implementing

Psychology Department and Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, USA. e-mail: yael@princeton.edu

the latter[18–22], perhaps aided by dopamine released from the noradrenergic locus coeruleus[23,24]. However, the strict division between the two algorithms in the brain, and their neural substrates, has recently been challenged[25].

Reinforcement learning relies on representation of tasks as sequences of states

Designing the correct state space for each task is critical in RL[26–28]. First, different state representations will lead to different optimal action policies, some of which may never lead to the desired goal. For example, if you represent the street crossing task as in Fig. 1b, you will stay on the curb forever, whereas the representation in Fig. 1c will get you home safely. Second, RL methods suffer from the 'curse of dimensionality', whereby the time (or iterations of learning) needed to solve the task scales exponentially with the number of states of the task[29]. If the state representation in Fig. 1 included all pixels in the visual field, there would be numerous alternatives for S1, and it would take very many street crossings to estimate the values of different actions for each of the states accurately. Thus, to make learning in RL simulations feasible, state representations are typically crafted by hand, by an expert, so as to accurately describe the task with as few states as possible[27,28] (for example, ref. [30]). The overwhelming majority of RL implementations, whether used to play backgammon[31] or to model animal learning (for example, ref. [32]), assume a state space and are not concerned with learning it (but see refs. [28,33]). A question then remains regarding how living agents know what to represent[28] in order to use neural RL to solve these (and other) tasks.

Work in the booming field of deep learning has been specifically focused on learning useful, reduced representations, often not in the service of RL, but rather for classification tasks (for example, labeling of images or face recognition). The marriage of deep learning and RL (that is, learning representations that are useful for goal-directed-policy learning) has led to exciting breakthroughs in artificial intelligence, such as mastering the game Go[34] and playing Atari games better than humans[35]. However, these implementations require millions of iterations to learn representations and therefore do not provide insight regarding how humans and animals learn task representations[36]. Moreover, even deep learning networks cannot flexibly solve new tasks that they have not been specifically programmed to learn[37], to the extent that these require a different representation of the same inputs (for example, the task of hailing a taxi in Fig. 1d).

A useful state representation for RL can be derived from the true causal structure of the task: it should include all the environmental features that determine (causally) whether actions will lead to (long-term) task-relevant outcomes (rewards and punishments), while any dimension or feature of the environment that is not causal to these outcomes can be ignored and generalized over[27]. Learning such a minimal representation from trial and error is not trivial due to the large number of features that are potentially relevant to each task[28,38]. These include not only all the currently observed features, but also past events and actions that may be causally relevant to future outcomes, for example, the fact that you hurt your foot yesterday and thus can only walk slowly across the street in Fig. 1a. Moreover, while the environment typically provides feedback for actions, there is no direct feedback for the representations underlying these actions, making it harder to determine whether our current state representation is suitable for the task at hand or can be further improved. Nevertheless, the brain seems to learn appropriate representations for new tasks almost effortlessly and in few trials[39]. How is representation learning achieved in the brain of humans and animals? Below, we summarize findings from several studies that were aimed at elucidating basic principles of neural representation learning, making note of open questions where progress can be made in the near term.

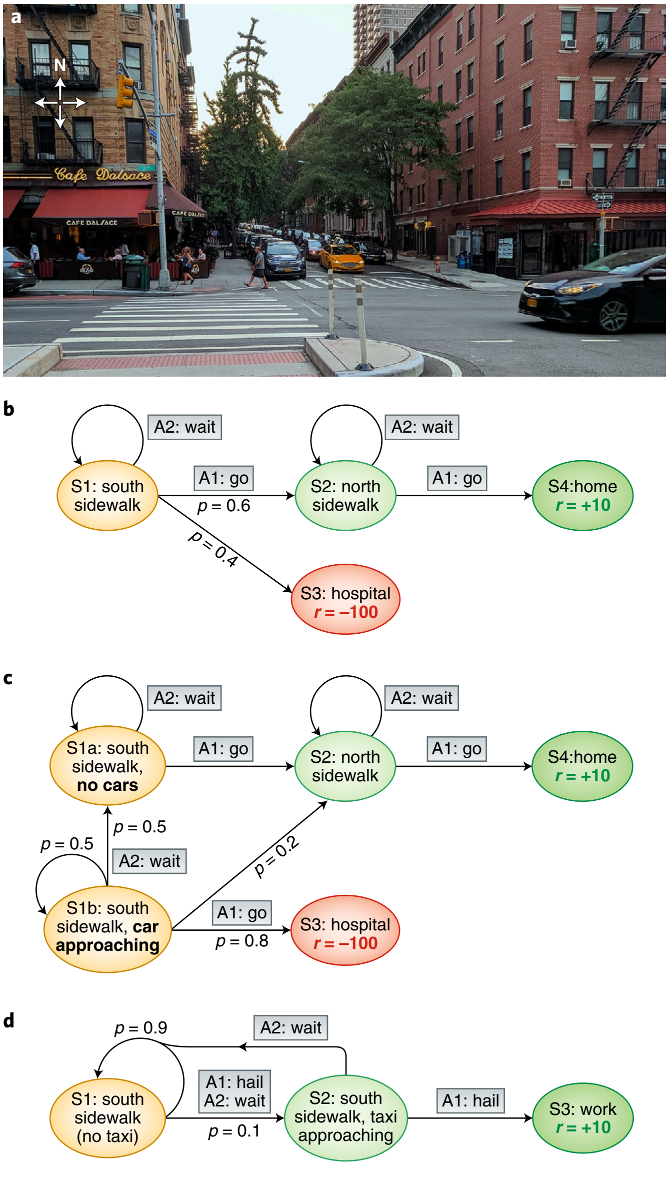
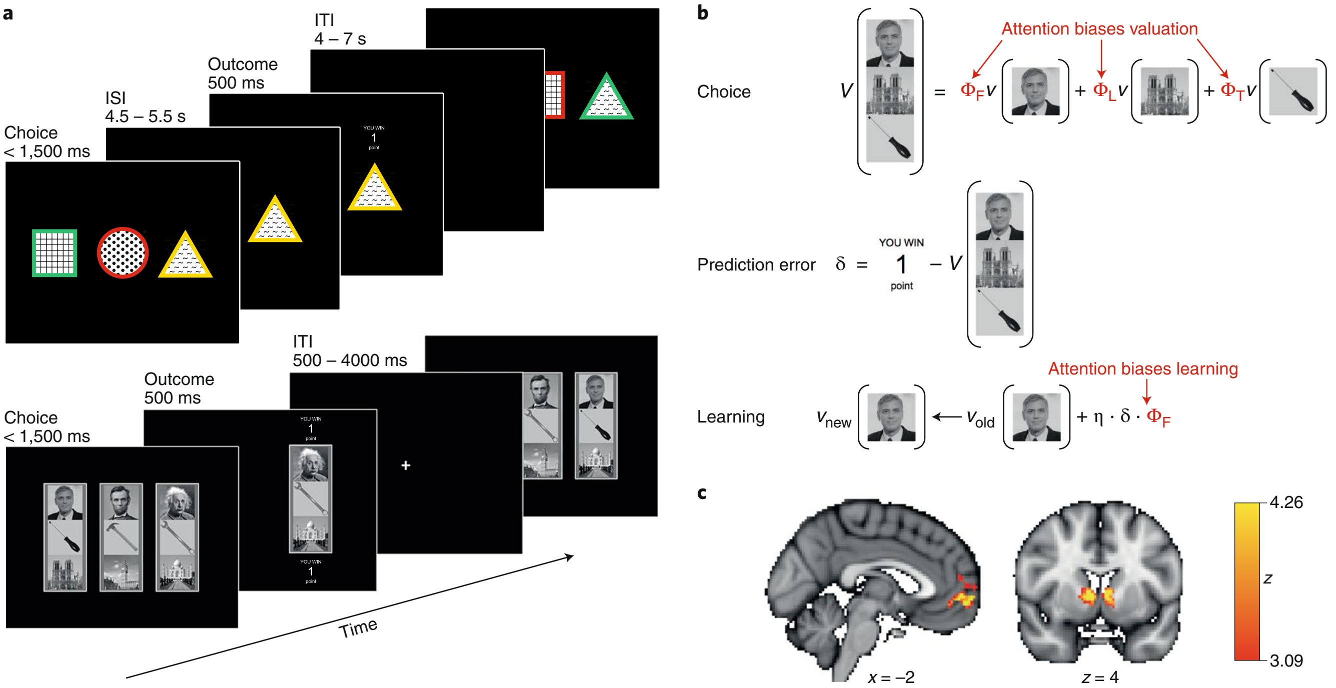


Fig. 1 | Task representations. a, The setting: a street corner. You are on the south side, just outside the picture. **b**, State diagram of the 'going home' task from your egocentric point of view. States are in circles, actions in gray rectangles, arrows denote state transitions, and reward outcomes are in color within the respective state ($r = 0$ if not stated). At the first state (S1, yellow: south sidewalk), choosing A1 (go) will result in a transition to S2 (light green; north sidewalk) with probability $p = 0.6$, and a transition to S3 (red; hospital, due to being run over) with $p = 0.4$. Since hospitalization is accompanied by a very aversive outcome ($r = -100$) that overwhelms the appetitive outcome of making it home safely (S4, green, $r = 10$), the optimal policy is to wait at S1 indefinitely. **c**, An alternative state representation for the 'going home' task that divides S1, the state of standing on the south sidewalk, into two states: S1a, in which no car is in sight and crossing the street is perfectly safe; and S1b, in which a car is approaching and crossing the street is dangerous (only $p = 0.2$ for the transition to the north sidewalk when choosing A1, and $p = 0.8$ for the transition to the hospital). Waiting is the optimal policy in S1b, and going is optimal at S1a. Eventually, you will get home safely. **d**, Representation of an alternative task in the same setting: hailing a taxi to go to work. Transitions in S1 occur regardless of your chosen action (to hail or to wait), however, in S2 the action of hailing a taxi will get you to work, whereas waiting will bring you back to S1.



$$V = \Phi_F V + \Phi_L V + \Phi_T V$$

$$\delta = 1 - V$$

$$V_{new} \leftarrow V_{old} + \eta \cdot \delta \cdot \Phi_F$$

Fig. 2 | The Dimensions Task. a, Two variants of the Dimensions Task, one with overlaid color, shape and texture dimensions (top), and another with rectangle stimuli comprising face, tool and scene dimensions (bottom) that allow direct measurement of attention via eye-tracking and multivariate pattern analysis from visual areas responsive to faces, tools and scenes. **b**, Attention-weighted reinforcement learning. A value is learned for each of the nine features. The values are combined, weighted by attention to each of the face, tool and scene dimensions, to determine the total value (expected reward) of a rectangle stimulus. Once the outcome is observed, a prediction error is calculated as the difference between the outcome and the expected reward (value). This prediction error is used to update the values of each of the chosen features, weighed by attention to each dimension. ϕ, attention weight; δ, prediction error; η, learning rate or step size parameter. **c**, Neural activations corresponding to areas that correlate significantly better with values (left) and prediction errors (right) from the attention-weighted algorithm than with those from from a similar algorithm with uniform (1/3) weights for each of the dimensions. Figure adapted from ref. [1], Society for Neuroscience, and ref. [2], Cell Press.

Dimensionality reduction of task representations using attention

As a highly simplified laboratory analog of a task such as street-crossing, where only a few dimensions of the environment determine the outcome for actions, we developed the ‘Dimensions Task’[1,2] (Fig. 2a). On each trial, the participant was asked to choose one of three visual stimuli defined along three dimensions (for example, shape, color and texture) to obtain reward. Importantly, participants were instructed that only one dimension (for example, shape) was relevant to determining rewards and that within this dimension one ‘target’ feature (for example, circle) would lead to a reward 75% of the time while other features had only a 25% chance of leading to reward. Every few trials (for example, every 30 trials), the relevant dimension and target feature changed, and the participant was notified (‘a new game is now starting’). This task, like the extradimensional–intradimensional set-shifting task in animals[40], is a multidimensional extension of the multiple-option choice task (*n*-armed bandit task) used in many neuroeconomics studies[41,42] and is a probabilistic version of the Wisconsin card sorting task[43], albeit with rapid, signaled, changes of the relevant dimension. Our goal was to use this task, together with a host of computational models, to investigate in detail how humans learn by trial and error what the relevant dimension is for representing task stimuli.

Findings from the Dimensions Task showed that participants employed attention mechanisms to create a lower-dimensional representation that constrained learning and decision-making to dimensions that were deemed relevant for obtaining reward[1,2]. In particular, formal model comparison showed that participants did not learn as Bayesian ideal observers on the one extreme, nor did they apply RL to each stimulus as a whole on the other extreme (for example, learning a value for a green triangle with stripes and a separate value for a green triangle with polka dots)[1]. Instead, participants learned the task using attention-weighted RL[2,28]: stimulus dimensions were selectively weighted by attention both when determining the value of each option and when learning from prediction errors (Fig. 2b). These weights can be seen as constructing a lower-dimensional task representation. For instance, if the weight for one dimension (for example, color) is 1 and the other two weights are 0, the task representation will include only the color of the stimuli, ignoring shape and texture. Other weights will prioritize dimensions less exclusively, for instance strongly representing shape while mostly but not completely ignoring color. In this way, attention dynamically generates a representation that focuses on the dimensions deemed most relevant to task performance at any given time point[44].

This work has highlighted the involvement of the frontoparietal attention network (including the intraparietal sulcus, precuneus and dorsolateral prefrontal cortex)[45,46], which is classically associated with attentional control, in changing attention from one dimension of the task to another during learning[2] and therefore in switching between alternative task representations. Blood-oxygen level-dependent (BOLD) signals in both value-sensitive areas in the dorsolateral prefrontal cortex and prediction-error-sensitive areas in the ventral striatum showed significantly higher correlation with values and prediction errors (respectively) derived from the attention-weighted model as compared to a model that attends equally to all three dimensions[2] (Fig. 2c), further supporting the idea that task representations are dynamically modified using attention[44,47,48].

Box 1 | Open questions: representation learning in real time, in the brain

A major outstanding question regards how trial-by-trial feedback shapes attention (which, in turn, shapes task representations). Our findings show that the interaction between attention and learning is bidirectional, with high-value features attracting attention and attention determining how value is accrued to different features[2]. Still, these findings do not explain how attention is determined trial-by-trial, that is, how feedback (for example, rewards and prediction errors) is used to determine whether to switch the focus of attention on the next instance[27,28,48].

Computationally, it is not known what statistics of the task (and our performance on it) convey that we are focusing on the wrong task dimensions[97]. What evidence is indicative of too narrow a focus of attention that must be widened, and vice versa? It is likely that this type of evidence transcends a single trial. For instance, one way to define the quality of a representation is by the entropy of motivationally important outcomes predicated on that representation, with lower entropy (i.e., more deterministic predictions) being preferable. For example, the representation in Fig. 1b is less effective than that in Fig. 1c in predicting our chances of reaching the other side of the street safely. However, this, and similar hypotheses, require testing.

Neurally, it is not yet known whether selective attention operates directly to shape the funneling-in of cortical afferents to the striatum (so that irrelevant, unattended, input dimensions do not contribute to striatal RL as the striatum does not 'know' that these dimensions existed in stimuli), computations within the striatum (for example, through weighting the contribution of different input dimensions to the valuation of the current state) or cortical representations (for example, in the OFC) that, in turn, affect what the striatum represents and learns about. In perception, expectations and attention affect neural responses to stimuli both before (with elevated activity for predicted stimuli) and after stimulus onset (with predicted stimuli eliciting less activation than surprising ones due to 'expectation suppression')[98]. These neural effects have been formalized within the computational framework of predictive coding as effects on initial starting points and drift rates in Bayesian evidence accumulation models[98]. While we have shown that attention to input dimensions influences both valuation and updating in RL[2], an understanding of how this manifests neurally, and indeed a detailed comparison of computational models of this influence, awaits future work.

Other open questions include what brain areas are involved in learning new representations (rather than representing known task states) and how arbitrary 'pointers' to task states that can differ radically from task to task are implemented in the OFC. Better insight into the nature of state representations in the OFC will hopefully help explain how the OFC can switch rapidly and dynamically between representations of different tasks and, indeed, how completely new task representations can be generated in the OFC in moments when people are explicitly instructed about the structure of a new task (for example, as was done in ref. [6]). Ideas about how the hippocampus represents spatial maps that differ from environment to environment[99,100] may provide important clues regarding the representation of abstract cognitive maps of tasks. Moreover, understanding how unexpected feedback shapes learning of task representations may shed light on, for example, the role of dopamine in prefrontal cortices and in the hippocampus.

An important question that the surveyed research has not addressed is how the general structure of inputs (some of which are not task-relevant), which is learned through unsupervised learning, interacts with learning task-relevant representations. For instance, in the Dimensions Task, all features are present on every trial, but this is not the case when crossing the street. An important hint that the barking dog is irrelevant for the latter task is probably the fact that street crossing is seldom accompanied by a barking dog.

Finally, we do not yet know what constraints our brain has adapted to and been able to take advantage of: what are the statistics of natural tasks? Can the brain safely assume that, even in highly multidimensional environments, only a few dimensions are relevant to any given task? Presumably, decision-making systems evolved to be tailored to the set of tasks that we are routinely faced with. Similarly to the breakthroughs in understanding vision that followed the quantification of statistics of natural scenes, a clear description of the statistics of natural tasks might revolutionize our understanding of the neural basis of high-level learning and decision-making.

However, it is not yet clear how humans determine what to attend to or how attention changes from trial to trial based on actions and outcome feedback (Box 1).

The role of inference and context in shaping task representations

Dimensionality reduction through selective attention addresses one aspect of the representation learning problem: the existence of extraneous information in sensory input. An opposite problem is that of missing data: not all task-relevant information is readily available. For instance, in some countries you should look right rather than left in search of oncoming traffic; still in others, to cross the street you should not wait for a gap in traffic but rather walk at a steady, slow pace and cars will slow down and create a gap around you. Crosswalks may nevertheless look deceptively similar in these different cases. More generally, context—location, time, phase of the task, as well as internal goals and past knowledge—is often critical to learning and decision-making, but is not necessarily perceptually observable.

Put differently, a ubiquitous problem in learning a task representation is that externally similar situations can require different action policies in different contexts and thus should be represented as separate states depending on context. Conversely, some seemingly different situations may be equivalent in terms of the causal structure of the task and thus should be aliased into a single state to allow generalization[27,28,49,50]. In real-world tasks, incoming information is rarely delineated into trials with clear, punctate stimuli that flash on and off to signal the current state, and thus a fundamental problem in learning is determining which experiences should be considered together (i.e., represented as the same state, allowing learning from one experience to affect behavior in the other) or learned about separately[4,51]. For instance, in the example in Fig. 1, the question is whether instances of standing on the south sidewalk should be delineated into two states based on the presence or absence of oncoming traffic, as in Fig. 1c (when attempting to go home), or whether they can all be considered as a single state (as is the case in Fig. 1d, when going to work, although in that case instances in which a taxi is approaching need to be represented as a separate state). Regardless of whether the context is observable or not (in the example in Fig. 1, the task goal can be considered a context), it is clear that situations must be grouped into states differently for different contexts, and this context- or task-specific grouping must be learned.

Since no two experiences are exactly alike, one idea is that incoming information is clustered according to similarity (i.e., similar experiences are assigned to the same cluster), with each cluster effectively

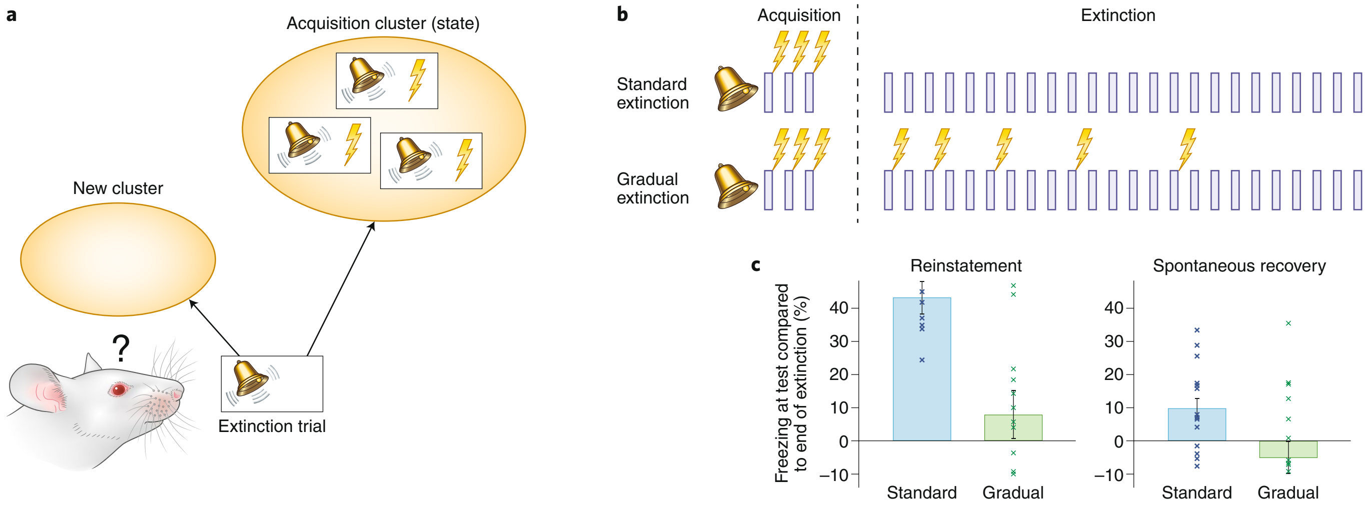


Fig. 3 | Latent-cause inference in representation learning suggests a mechanism for altering fear memories. a, When confronted with the first extinction trial, the animal determines, based on the (dis)similarity of this trial to previous experiences (acquisition trials), whether to attribute it to the acquisition cluster or to a new cluster. Clusters correspond to states and are determined based on inferred latent causes. In the inference process, cluster assignment is probabilistic, such that an event can be partially assigned to more than one cluster. **b**, A gradual extinction protocol aimed at making extinction more similar to acquisition, thus coaxing the animal to assign all trials to a single cluster. Top: standard extinction, with no shocks in the extinction phase. Bottom: gradual extinction in which shocks taper off gradually in the extinction phase. **c**, In two experiments, rats showed significant return of fear after reinstatement with a reminder shock (left) or due to spontaneous recovery (right) following standard extinction (freezing at test compared to last four trials of extinction), whereas after gradual extinction rats showed no increase of fear at test. Figure adapted from ref. [56], Frontiers Media.

forming a state on which RL operates (Fig. 3a). This intuitive but powerful idea has been formalized within a statistically optimal theory (based on Bayesian inference) of how an animal or human might generate task states in their internal representation of the world, based on experience. The idea is that we start from a single state (cluster) and expand our representation as needed when new events differ substantially from what we have experienced so far[3,52]. In this way, the high-level contextual structure of a task can be learned.

Such clustering-based representation learning, together with traditional associative learning that treats each cluster as a distinct state (for example, learning a single association or value based on all stimuli assigned to that cluster), can explain why it is difficult to change old associations. Consider the notorious ineffectiveness of extinction procedures in fear conditioning: after a rat is exposed to pairing of a tone and a shock such that the tone comes to elicit freezing behavior due to prediction of an upcoming shock, attempts to weaken the association between the tone and the shock by repeatedly presenting the tone without shocks reduce fear only temporarily. Studies demonstrate that the fear response returns over time ('spontaneous recovery')[53], with changes in context ('renewal')[54] or after a reminder shock ('reinstatement')[55]. Our framework explains the ineffectiveness of extinction as resulting from the animal's inference that extinction trials and acquisition trials are likely generated by different 'latent causes' and thus should be clustered into separate states (Fig. 3a)[3]. This idea provides a normative explanation for the long-held view that in extinction the animal learns a new 'tone→no shock' association, rather than modifying the old 'tone→shock' association[54].

Indeed, increasing the similarity between the acquisition phase and the extinction phase through gradual extinction can increase the effectiveness of extinction. Gershman and colleagues[56] used the principle of similarity-based representation learning to modify fear memories in rats by making the change from shock to no shock gradual (in contrast to the commonly used abrupt-extinction paradigm; Fig. 3b). Their goal was to prevent the rats from generating a new task state in the extinction phase and thus to cause the 'tone→no shock' experiences to influence the previously acquired 'tone→shock' association. Gradual extinction resulted in more persistent loss of fear, as measured by lack of reinstatement or spontaneous recovery of fear at test (Fig. 3c). In human fear conditioning, a meta-analysis showed that individuals who extinguished a fear association faster showed more spontaneous recovery of fear[57]. This may be because those who showed faster extinction had more readily inferred a new latent cause in extinction. As a result, they attributed fewer extinction trials to the acquisition cause, leaving the original fear memory intact, as was evident in the later recovery of fear.

Similarity-based clustering as a general principle of representation learning

Clustering of experience according to similarity for the purposes of delineating task states seems to be a general phenomenon in learning, going beyond situations that involve strong aversive reinforcers. For instance, in a perceptual decision-making task[58], humans were asked to quickly estimate the number of circles on a screen (Fig. 4a). The true number of circles was then revealed, such that participants could estimate the number of circles on subsequent trials by learning the mean number of circles in the task (the number of circles was drawn from a Gaussian distribution, making the mean of previous trials the best estimate for the next trial). Unbeknownst to participants, there were two types of trials signaled by different colors of the circles and drawn from different Gaussian distributions (for example, blue circles in one trial type and green circles in the other; Fig. 4a,b). The resulting number estimates (Fig. 4c) suggested that when the distributions associated with the two trial types were similar, participants grouped the different colored trials together and averaged their statistics during learning, as if they all belonged to one colorblind state. In contrast, when the two means were dissimilar, participants seemed to separate trials into single-color states (as in Fig. 4b), separately learning about circles of each color[58].

Beyond their effect on delineating learning, clusters of experiences that are blended together in one state can be viewed as comprising a single multi-event 'memory trace'. The boundaries of

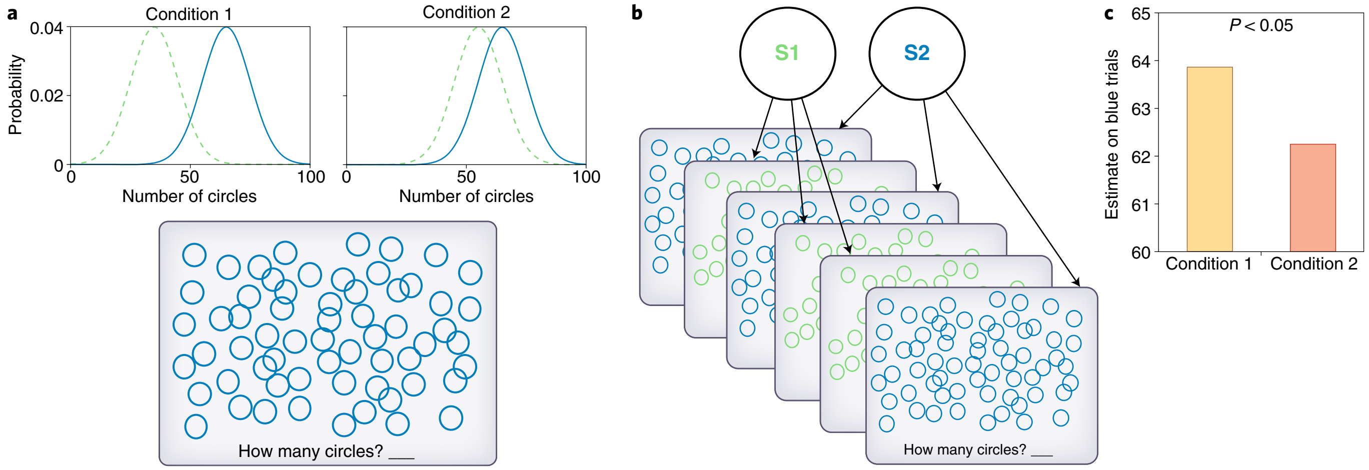


Fig. 4 | Humans spontaneously use similarity to infer the latent structure of task. a, The circles task[58]. Human participants were asked to key in a double-digit guess for the number of circles on the screen (bottom). In condition 1, the number of circles was drawn either from a Gaussian of mean 65 or from a Gaussian of mean 35 (top left). In condition 2, means were 65 and 55 (top right). Conditions occurred in blocks, with each Gaussian associated with a different color of the circles. Blocks of the two conditions were randomly intermixed within participant (8 blocks of each, 20 stimuli per block). Each block involved two different colors; we use only blue and green to illustrate that the mean-65 trials were identical in both conditions. **b**, Participants' guesses on the mean-65 trials in the two conditions suggested that they spontaneously inferred whether the task involved two latent causes (states), each generating stimuli of one color (depicted) or whether there was only one latent cause generating all trials in a block. **c**, In condition 1, stimuli were sufficiently different to warrant two latent causes, such that learning about blue circles was segregated from learning about green circles, and the estimate on blue trials was close to their true mean of 65. In contrast, in condition 2, the greater overlap between the distributions resulted in stimuli that were more similar across colors, which encouraged participants to infer a single latent cause and thus ignore the color of stimuli and guess a number of circles that was closer to the global mean of 60. Figure adapted from ref. [58], Frontiers Media.

memory traces may therefore also reflect the process of representation learning: if the current experience is similar enough to previous experiences, it should be added to a previous memory trace (thereby modifying that memory, i.e., causing learning[59]), whereas a novel experience should spur the creation of a new memory trace. To test this, participants were presented with line segments that changed gradually throughout a block of the experiment. Critically, in half of the blocks, among the small trial-to-trial changes of the line segment, one larger change ('jump') was embedded. The authors hypothesized that this jump would cause a splitting of memory traces, thereby protecting the memories of the line segments in the beginning of the block from interference by the segments later in the block. Indeed, when tested for memory of one of the initial line segments in the block, participants were more accurate in blocks that involved a jump as compared to blocks that only involved gradual changes. Computational modeling suggested that in the jump condition, two memory traces were created, whereas blocks with no jump resulted in a single memory trace[60]. The idea that latent-cause inference determines not only RL, but also the organization and modification of memories, can explain many extant phenomena in the literature on 'reconsolidation' of memory[61].

The neural substrates of representation learning

Lesion and developmental data in rodents suggest that the hippocampus plays an important role in determining when to create a new state (cluster) or update an old one[3,61–65]. The hippocampus is important for detecting novelty[66,67] and therefore may participate in computing and signaling the (dis)similarity of different experiences. Animals with hippocampal lesions do not show context-sensitivity of extinction[62] or other related conditioning phenomena such as latent inhibition[63], instead behaving as if extinction and acquisition trials are all clustered in one state. Animals with an underdeveloped hippocampus (for example, young humans and rodents) also generalize learning widely[64,65,68,69], consistent with an oversimplified representation with a small number of states[3].

Other work implicates a different brain area—the orbitofrontal cortex (OFC)—in the representation of learned task states[70,71]. Many decision-making-related functions have been attributed to the OFC, including response inhibition, somatic markers and, most dominantly, outcome and/or economic value expectancies[5,71]. However, lesions[72–75], inactivation[76], electrophysiological recording data[77–81] and functional MRI (fMRI) findings[6,82–84] suggest that the OFC is critical for representing the current state of the task when this state is not immediately evident from sensory information and likely conveys this representation to the striatum for RL[5]. For instance, in an fMRI study in which human participants were asked to infer what quadrant of a safari they were in based on evidence of animals recently seen, the similarity structure of optimal location inferences over time correlated best with the similarity structure of activations in the OFC[83]. In electrophysiological recordings from OFC in an odor-guided go/no-go task in rodents where different odors appeared in sequences that made the task equivalent to traversing a (virtual) T-maze, state similarity showed that OFC represented the detailed state structure of the task, above and beyond other quantities such as expected rewards that could also be extracted from the neural dynamics[81].

Both electrophysiological[77] and lesion data[5] suggest that even without a functioning OFC the striatum has access to a state representation that is bound to external stimuli ('observable states'; unobservable information about timing is also likely computed in the striatum[85]). In contrast, inferred states that incorporate internal information—for example, from working memory, intended actions or future goals—and are based on latent-cause inference as discussed above ('partially observable states'), seem to require the OFC[5,86,87]. This can explain why decision making becomes more OFC-dependent as tasks rely more on inference processes necessary to determine the underlying hidden state of a task.

To test this hypothesis, Schuck and colleagues designed a task in which correct performance required representing partially observable states[6]. Human participants had to judge the age (old or young)

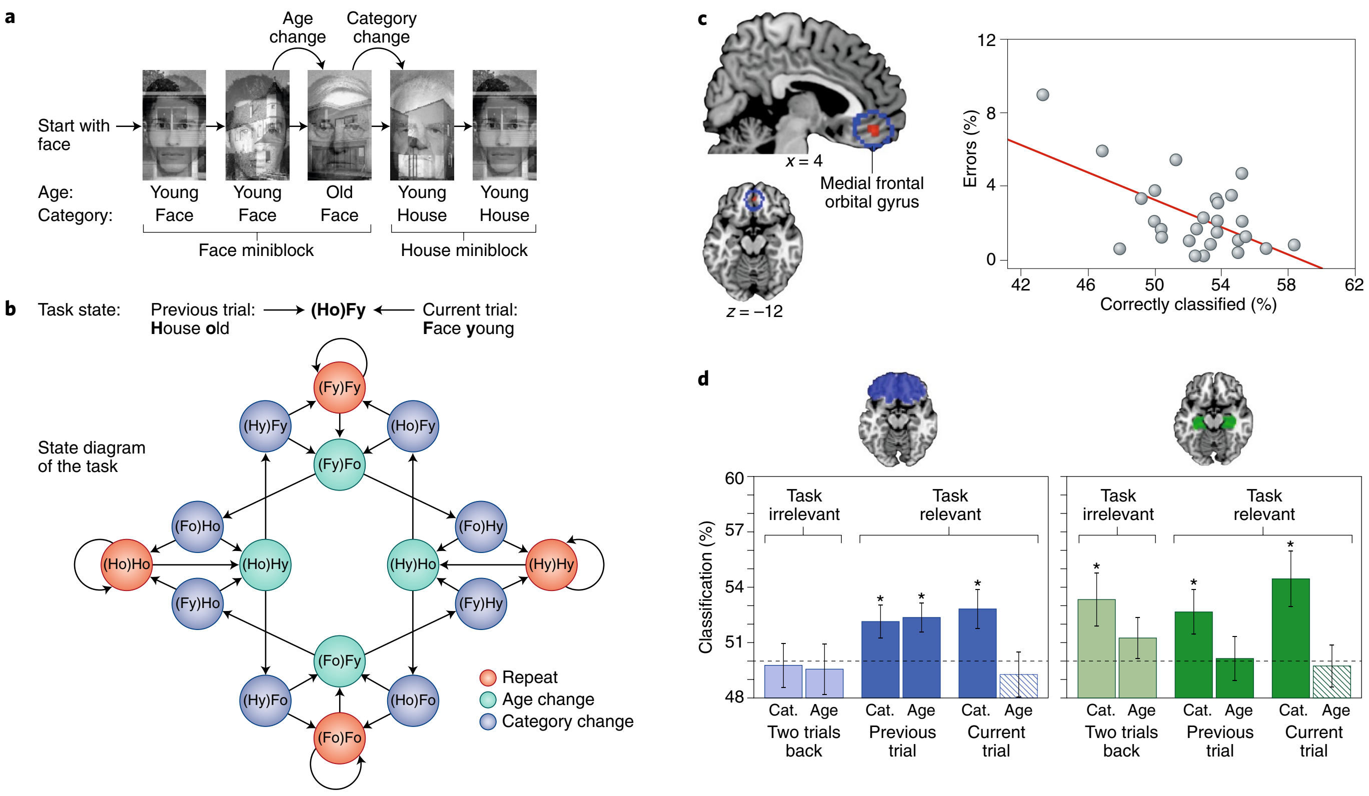


Fig. 5 | The orbitofrontal cortex represents the current state of the task. a, An age-judgement task with hidden states. Participants must judge the age (young or old) of one category (for example, faces) until the age in that category changes. The subsequent trial then starts a miniblock of judging the alternate category (for example, houses) until the age in that category changes, and so forth. **b**, These instructions create a 16-state task in which each state includes the current category to be judged, the age of the current stimulus in that category, the age in the previous trial (for comparison with current age) and the category in the previous trial (as age comparison is not needed in the beginning of a miniblock). All state components are unobservable except for current age. Each state transitions to one of two other states, with equal probability. **c**, The OFC (left) was the only brain area from which all unobservable components could be classified, and classification accuracy there (in an anatomically defined region of interest including the whole OFC) correlated with lower behavioral error rates (right). **d**, Only unobservable, task relevant, features were decodable in the OFC (left), in contrast to the hippocampus, where only category was decodable, for several trials back (right). Error bars indicate s.e.m.; dashed horizontal line shows chance baseline; $^*P < 0.05$ compared to chance, one-tailed. Figure adapted from ref. [6], Cell Press, and ref. [88], AAAS.

of either a face or a house presented overlaid on each other, with the category to be judged determined by the previous trial (Fig. 5a). The initial category to be judged was instructed. Thereafter, participants were asked to continue judging that same category as long as the age of the stimuli in the judged category remained the same. Once the age of the currently-judged category differed from the age in the previous trial, they were instructed to switch to judging the alternative category on the next trial. Participants performed the task well (<5% errors), suggesting that they correctly represented, at each time point, the state of the task (Fig. 5b). A multivariate classifier for the 16 task states could reliably decode the current state of the task from the OFC, and from that area only (Fig. 5c), despite the fact that there were no rewards or reward expectations in the task. The fidelity of task representations in the OFC correlated with behavioral performance (Fig. 5c), and on error trials the correct state was decoded significantly below chance. Moreover, irrelevant aspects of the task (for example, information from two trials back) were not decodable in the OFC (Fig. 5d), supporting the selectivity of the orbitofrontal representation in capturing only task-relevant information, as is required from a minimal state representation. This was in contrast to other brain areas, such as the hippocampus and the dorsolateral prefrontal cortex, that represented some (but not all) task-relevant aspects and some task-irrelevant information[6] (Fig. 5d). However, recent evidence suggests that replay of sequences of task states in the hippocampus at rest improves orbitofrontal representations[88].

As befits such a critical brain function, the above findings implicate a variety of brain areas in representation learning: attention networks determine perceptual similarity, which is augmented by previously remembered information (for example, to determine novelty) in the hippocampus and segmented into latent causes that are represented as separate states in the OFC. The dorsolateral prefrontal cortex, in turn, is implicated in switching between state spaces when the global task representation changes[89]. Even with these clues in place, however, it is not yet clear which brain areas coordinate the online learning of a task-state representation, the elusive process for which we also do not yet know the computational algorithm (Box 1).

Carving the world at its task-relevant joints

Reinforcement learning—how feedback from the world, and particularly unexpected feedback, is incorporated into future predictions—is fairly well understood in the brain. What is less clear is the fundamental process of representation learning: how we learn to carve streams of ongoing experience into task states that correctly encompass all that is relevant to the task at hand in a minimal representation that generalizes learning as widely as possible[27,28]. RL cannot occur without a state representation, and different representations can render a task extremely simple or exquisitely complex. Moreover, since only actions are given feedback in the form of rewards and punishments, the all-important task representation must be learned without direct feedback, extracted from the overall statistics of the task, the environment and the agent's performance.

In this Perspective, we focused on three lines of work attempting to elucidate the algorithms and neural substrates of representation learning. We first suggested that selective attention, rather than arising from neural constraints, can in fact be viewed as a mechanism that allows rapid learning: selective attention solves the curse of dimensionality in RL by reducing the dimensionality of task states and focusing only on those dimensions that are causally important for the task at hand. By blurring out irrelevant dimensions, selective attention allows us to generalize over them and employ our learning and decision-making processes more efficiently. Thus, selective attention helps overcome what is perhaps the most fundamental constraint: we can afford only a limited amount of experience because our lives are finite and the passage of time is irreversible.

Indeed, within the non-relenting stream of experiences, no two are exactly alike. Thus, learning from past experience entails generalization—using experience from one situation to inform us about a (slightly) different situation. In learning the boundaries of generalization, we implicated a clustering process that assumes that similar experiences belong to the same state (cluster), while also allowing for a growing representation when faced with novel experiences[3,52]. We presented empirical evidence suggesting that this similarity-based clustering process is intertwined with learning and that the learned representations affect both memory and decision-making. Learned attention interacts with the clustering process by affecting similarity: attention to a dimension, such as the speed of cars, effectively stretches that perceptual axis so that situations with slightly different speeds may be separated into distinct clusters, whereas ignoring a dimension, such as the color of cars, will mean that different colors will not be considered dissimilar. This similarity then affects both the organization of memory and how new experience is combined with old knowledge, through learning.

One conclusion from these studies is that, in some cases, slower learning is better. This is particularly important when the goal of learning is to modify previous knowledge when the environment has changed (for example, from threatening to safe). Fast learning can easily be achieved by postulating a new state of the task, but this prevents the modification of previously held beliefs associated with the old task states. In general, learning occurs when we are confronted with information that is not similar to previous experience, and thus creates a prediction error. A small prediction error will lead to little learning. A very large prediction error will lead to generation of a new state. Thus, to impact an old state with new information, the information must be unexpected, but not too much so. This principle of activating the old state and providing non-confirming information to update the predictions contingent on that state, which we took advantage of in our gradual extinction experiment[56], is also the basis of memory-modification methods in cognitive behavioral therapy for post-traumatic stress disorder[90].

The learned state representations are multimodal, potentially incorporating information from any sensory modality, as well as from memory. We reviewed evidence suggesting that the OFC, a prefrontal brain area that receives widespread sensory, limbic and higher-order afferents[91], is well-poised to represent the abstract identification of the current state[6]. The idea is that the representation in the OFC is akin to a 'pointer' in computer science: an abstract link to information represented in other brain areas, which identifies the current state. Information that is irrelevant to the current state (for example, the color of oncoming cars) would not affect the orbitofrontal representation. In contrast, information that, if different, would change the current state (for example, the speed of the closest car), would presumably be decodable in the OFC[92]. Importantly, according to our theory, this is not necessarily because the OFC represents this information per se, but because different settings of this variable lead to different states. This hypothesis, unlike the dominant view that the OFC represents expected (reward) outcomes[93–96], suggests that expected reward will be decodable in the OFC only to the extent that it is part of the state representation (which, in fact, it often is).

In sum, we suggest a dynamic interplay between RL in the basal ganglia, adaptive attention processes in the frontoparietal attention control network and memory processes that reflect the learned structure of the environment and shape orbitofrontal state representations. In this framework, selective attention is not a limitation of the neural learning system, but an adaptive mechanism that allows rapid learning. Memory is similarly seen as an active process that does not simply mirror the external environment, but rather reflects inference regarding causal relationships in the environment.

Research on representation learning is still in its infancy both in neuroscience and in machine learning, where a shift is underway from solving specific problems (like playing Go or designing a self-driving car) to designing general artificial intelligence that can adaptively learn to represent and solve new tasks. As such, many critical questions are yet unanswered (Box 1). The findings we have discussed here suggest that representation learning involves the dynamic interplay of cognitive functions that have traditionally been studied separately from each other. Understanding how information flows between these systems will help explain the amazing adaptive capabilities of humans that go orders of magnitude above and beyond simplified laboratory tasks. In any case, understanding representation learning—this computationally daunting task that our brain so marvelously excels at—will be fundamental to any complete theory of learning in the brain.

Received: 14 April 2019; Accepted: 19 July 2019;
Published online: 24 September 2019

References

1. Niv, Y. et al. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* **35**, 8145–8157 (2015).
2. Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V. & Niv, Y. Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron* **93**, 451–463 (2017).
3. Gershman, S. J., Blei, D. M. & Niv, Y. Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010).
4. Gershman, S. J., Norman, K. A. & Niv, Y. Discovering latent causes in reinforcement learning. *Curr. Opin. Behav. Sci.* **5**, 43–50 (2015).
5. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267–279 (2014).
6. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human orbitofrontal cortex represents a cognitive map of state space. *Neuron* **91**, 1402–1412 (2016).
7. Sutton, R.S. & Barto, A.G. *Reinforcement Learning: An Introduction*. (MIT Press, 2018).
8. Kaelbling, L. P., Littman, M. L. & Moore, A. W. Reinforcement learning: a survey. *J. Artif. Intell. Res.* **4**, 237–285 (1996).
9. Daw, N.D. & Tobler, P.N. Value learning through reinforcement: the basics of dopamine and reinforcement learning. in *Neuroeconomics*. (eds. Glimcher, P. W. & Fehr, E.) 283–298 (Academic Press, 2014).
10. Daw, N.D. & O'Doherty, J.P. Multiple systems for value learning. in *Neuroeconomics*. (eds. Glimcher, P. W. & Fehr, E.) 393–410 (Academic Press, 2014).
11. Niv, Y. & Langdon, A. Reinforcement learning with Marr. *Curr. Opin. Behav. Sci.* **11**, 67–73 (2016).
12. Watkins, C. J. C. H. & Dayan, P. Q-learning. *Mach. Learn.* **8**, 279–292 (1992).
13. Friedrich, J. & Lengyel, M. Goal-directed decision making with spiking neurons. *J. Neurosci.* **36**, 1529–1546 (2016).
14. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
15. Keramati, M., Smittenaar, P., Dolan, R. J. & Dayan, P. Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proc. Natl Acad. Sci. USA* **113**, 12868–12873 (2016).
16. Barto, A.G. Adaptive critics and the basal ganglia. in *Models of Information Processing in the Basal Ganglia* (eds. Houk, J. C., Davis, J. L. & Beiser, D. G.) 215–232 (MIT Press, 1995).
17. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).

18. Yin, H. H., Knowlton, B. J. & Balleine, B. W. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* **166**, 189–196 (2006).
19. Miller, K. J., Botvinick, M. M. & Brody, C. D. Dorsal hippocampus contributes to model-based planning. *Nat. Neurosci.* **20**, 1269–1276 (2017).
20. Vikbladh, O. M. et al. Hippocampal contributions to model-based planning and spatial memory. *Neuron* **102**, 683–693.e4 (2019).
21. McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y. & Schoenbaum, G. Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J. Neurosci.* **31**, 2700–2705 (2011).
22. Boorman, E. D., Rajendran, V. G., O'Reilly, J. X. & Behrens, T. E. Two anatomically and computationally distinct learning signals predict changes to stimulus-outcome associations in hippocampus. *Neuron* **89**, 1343–1354 (2016).
23. Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D. & Kandel, E. R. Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proc. Natl Acad. Sci. USA* **113**, 14835–14840 (2016).
24. Rouhani, N., Norman, K. A. & Niv, Y. Dissociable effects of surprising rewards on learning and memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **44**, 1430–1443 (2018).
25. Langdon, A. J., Sharpe, M. J., Schoenbaum, G. & Niv, Y. Model-based predictions for dopamine. *Curr. Opin. Neurobiol.* **49**, 1–7 (2018).
26. Ponsen, M., Taylor, M.E. & Tuyls, K. Abstraction and generalization in reinforcement learning: a summary and framework. in *International Workshop on Adaptive and Learning Agents (ALA 2009): Adaptive and Learning Agents*. (eds. Taylor M.E. & Tuyls K.) 1–32 (Springer, 2010).
27. Canas, F. & Jones, M. Attention and reinforcement learning: constructing representations from indirect feedback. *Proc. Annu. Meet. Cogn. Sci. Soc.* **32**, 1264–1269 (2010).
28. Jones, M. & Canas, F. Integrating reinforcement learning with models of representation learning. *Proc. Annu. Meet. Cogn. Sci. Soc.* **32**, 1258–1263 (2010).
29. Bellman, R. *Dynamic Programming* (Princeton University Press, 1957)
30. Sutton, R.S. Generalization in reinforcement learning: Successful examples using sparse coarse coding. in *Advances in Neural Information Processing Systems* (eds. Touretzky, D. S., Mozer, M. C. & Hasselmo, M. E.) 1038–1044 (1996).
31. Tesauro, G. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput.* **6**, 215–219 (1994).
32. Ludvig, E. A., Sutton, R. S. & Kehoe, E. J. Evaluating the TD model of classical conditioning. *Learn. Behav.* **40**, 305–319 (2012).
33. McCallum, R. A. Hidden state and reinforcement learning with instance-based state identification. *IEEE Trans. Syst. Man Cybern. B Cybern.* **26**, 464–473 (1996).
34. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
35. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
36. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
37. Wang, J.X. *et al.* Learning to reinforcement learn. Preprint at *arXiv* https://arxiv.org/abs/1611.05763 (2016).
38. Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychol. Rev.* **124**, 301–338 (2017).
39. Griffiths, T. L., Chater, N., Kemp, C., Perfors, A. & Tenenbaum, J. B. Probabilistic models of cognition: exploring representations and inductive biases. *Trends Cogn. Sci.* **14**, 357–364 (2010).
40. Dias, R., Robbins, T. W. & Roberts, A. C. Primate analogue of the Wisconsin Card Sorting Test: effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behav. Neurosci.* **110**, 872–886 (1996).
41. Frank, M. J., Seeberger, L. C. & O'reilly, R. C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**, 1940–1943 (2004).
42. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
43. Milner, B. Effects of different brain lesions on card sorting. *Arch. Neurol.* **9**, 90–100 (1963).
44. Kruschke, J. K. ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* **99**, 22–44 (1992).
45. Petersen, S. E. & Posner, M. I. The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* **35**, 73–89 (2012).
46. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215 (2002).
47. Kruschke, J.K. Learning involves attention. in: *Connectionist Models in Cognitive Psychology* (ed. Houghton, G.) 113–140 (Psychology Press, 2005).
48. Kruschke, J. K. Toward a unified model of attention in associative learning. *J. Math. Psychol.* **45**, 812–863 (2001).
49. McCallum, R.A. Instance-based utile distinctions for reinforcement learning with hidden state. in *Machine Learning Proceedings 1995* (eds. Prieditis, A. &Russell, S.) 387–395 (Morgan Kaufmann, 1995).
50. Collins, A. G. & Frank, M. J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).
51. Langdon, A.J., Song, M. & Niv, Y. Uncovering the 'state': tracing the hidden state representations that structure learning and decision-making. *Behav. Processes* https://doi.org/10.1016/j.beproc.2019.103891 (2019).
52. Love, B. C., Medin, D. L. & Gureckis, T. M. SUSTAIN: a network model of category learning. *Psychol. Rev.* **111**, 309–332 (2004).
53. Rescorla, R. A. Spontaneous recovery. *Learn. Mem.* **11**, 501–509 (2004).
54. Bouton, M. E. Context and behavioral processes in extinction. *Learn. Mem.* **11**, 485–494 (2004).
55. Rescorla, R. A. & Heth, C. D. Reinstatement of fear to an extinguished conditioned stimulus. *J. Exp. Psychol. Anim. Behav. Process.* **1**, 88–96 (1975).
56. Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M. H. & Niv, Y. Gradual extinction prevents the return of fear: implications for the discovery of state. *Front. Behav. Neurosci.* **7**, 164 (2013).
57. Gershman, S. J. & Hartley, C. A. Individual differences in learning predict the return of fear. *Learn. Behav.* **43**, 243–250 (2015).
58. Gershman, S. J. & Niv, Y. Perceptual estimation obeys Occam's razor. *Front. Psychol.* **4**, 623 (2013).
59. Preminger, S., Blumenfeld, B., Sagi, D. & Tsodyks, M. Mapping dynamic memories of gradually changing objects. *Proc. Natl Acad. Sci. USA* **106**, 5371–5376 (2009).
60. Gershman, S. J., Radulescu, A., Norman, K. A. & Niv, Y. Statistical computations underlying the dynamics of memory updating. *PLOS Comput. Biol.* **10**, e1003939 (2014).
61. Gershman, S. J., Monfils, M. H., Norman, K. A. & Niv, Y. The computational nature of memory modification. *eLife* **6**, e23763 (2017).
62. Ji, J. & Maren, S. Hippocampal involvement in contextual modulation of fear extinction. *Hippocampus* **17**, 749–758 (2007).
63. Honey, R. C. & Good, M. Selective hippocampal lesions abolish the contextual specificity of latent inhibition and conditioning. *Behav. Neurosci.* **107**, 23–33 (1993).
64. Yap, C. S. & Richardson, R. Extinction in the developing rat: an examination of renewal effects. *Dev. Psychobiol.* **49**, 565–575 (2007).
65. Yap, C. S. & Richardson, R. Latent inhibition in the developing rat: an examination of context-specific effects. *Dev. Psychobiol.* **47**, 55–65 (2005).
66. Knight, R. Contribution of human hippocampal region to novelty detection. *Nature* **383**, 256–259 (1996).
67. Kumaran, D. & Maguire, E. A. Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus* **17**, 735–748 (2007).
68. Mednick, S. A. & Lehtinen, L. E. Stimulus generalization as a function of age in children. *J. Exp. Psychol.* **53**, 180–183 (1957).
69. Droit-Volet, S., Clément, A. & Wearden, J. Temporal generalization in 3- to 8-year-old children. *J. Exp. Child Psychol.* **80**, 271–288 (2001).
70. Wikenheiser, A. M. & Schoenbaum, G. Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nat. Rev. Neurosci.* **17**, 513–523 (2016).
71. Stalnaker, T. A., Cooch, N. K. & Schoenbaum, G. What the orbitofrontal cortex does not do. *Nat. Neurosci.* **18**, 620–627 (2015).
72. Izquierdo, A., Suda, R. K. & Murray, E. A. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* **24**, 7540–7548 (2004).
73. Chudasama, Y. & Robbins, T. W. Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. *J. Neurosci.* **23**, 8771–8780 (2003).
74. Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H. & Rushworth, M. F. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* **65**, 927–939 (2010).
75. Tsuchida, A., Doll, B. B. & Fellows, L. K. Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *J. Neurosci.* **30**, 16868–16875 (2010).
76. Lak, A. et al. Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* **84**, 190–201 (2014).
77. Takahashi, Y. K. et al. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat. Neurosci.* **14**, 1590–1597 (2011).
78. Blanchard, T. C., Hayden, B. Y. & Bromberg-Martin, E. S. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron* **85**, 602–614 (2015).
79. Stalnaker, T. A. et al. Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nat. Commun.* **5**, 3926 (2014).

80. Farovik, A. et al. Orbitofrontal cortex encodes memories within value-based schemas and represents contexts that guide memory retrieval. *J. Neurosci.* **35**, 8333–8344 (2015).
81. Zhou, J. et al. Rat orbitofrontal ensemble activity contains multiplexed but dissociable representations of value and task structure in an odor sequence task. *Curr. Biol.* **29**, 897–907.e3 (2019).
82. Howard, J. D., Gottfried, J. A., Tobler, P. N. & Kahnt, T. Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proc. Natl Acad. Sci. USA* **112**, 5195–5200 (2015).
83. Chan, S. C., Niv, Y. & Norman, K. A. A probability distribution over latent causes, in the orbitofrontal cortex. *J. Neurosci.* **36**, 7817–7828 (2016).
84. Hampton, A. N., Bossaerts, P. & O'Doherty, J. P. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* **26**, 8360–8367 (2006).
85. Takahashi, Y. K., Langdon, A. J., Niv, Y. & Schoenbaum, G. Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. *Neuron* **91**, 182–193 (2016).
86. Bradfield, L. A., Dezfouli, A., van Holstein, M., Chieng, B. & Balleine, B. W. Medial orbitofrontal cortex mediates outcome retrieval in partially observable task situations. *Neuron* **88**, 1268–1280 (2015).
87. Takahashi, Y. K., Stalnaker, T. A., Roesch, M. R. & Schoenbaum, G. Effects of inference on dopaminergic prediction errors depend on orbitofrontal processing. *Behav. Neurosci.* **131**, 127–134 (2017).
88. Schuck, N. W. & Niv, Y. Sequential replay of nonspatial task states in the human hippocampus. *Science* **364**, eaaw5181 (2019).
89. Sharpe, M. J. et al. An integrated model of action selection: distinct modes of cortical control of striatal decision making. *Annu. Rev. Psychol.* **70**, 53–76 (2019).
90. Foa, E. B. & Kozak, M. J. Emotional processing of fear: exposure to corrective information. *Psychol. Bull.* **99**, 20–35 (1986).
91. Wallis, J. D. Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* **30**, 31–56 (2007).
92. Schoenbaum, G., Setlow, B. & Ramus, S. J. A systems approach to orbitofrontal cortex function: recordings in rat orbitofrontal cortex reveal interactions with different learning systems. *Behav. Brain Res.* **146**, 19–29 (2003).
93. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
94. Padoa-Schioppa, C. Neurobiology of economic choice: a good-based model. *Annu. Rev. Neurosci.* **34**, 333–359 (2011).
95. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27**, 9984–9988 (2007).
96. McNamee, D., Rangel, A. & O'Doherty, J. P. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat. Neurosci.* **16**, 479–485 (2013).
97. Nosofsky, R. M. Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* **115**, 39–61 (1986).
98. Summerfield, C. & de Lange, F. P. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745–756 (2014).
99. Colgin, L. L., Moser, E. I. & Moser, M. B. Understanding memory through hippocampal remapping. *Trends Neurosci.* **31**, 469–477 (2008).
100. Leutgeb, J. K. et al. Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron* **48**, 345–358 (2005).

Acknowledgements

I am grateful to my lab members, past and present, for their creative, methodical and incredibly revealing work on representation learning in the brain. I thank A. Langdon, N. Rouhani, and J. Zarate for helpful comments on a previous draft. This work was funded by grant W911NF-14-1-0101 from the Army Research Office and grant R01DA042065 from the National Institute on Drug Abuse.

Competing interests

The author declares no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to Y.N.

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