Human Representation Learning

Angela Radulescu[1,2], **Yeon Soon Shin**[2], **Yael Niv**[1,2]

[1]Department of Psychology, Princeton University, Princeton, New Jersey 08544, USA;

[2]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA

Abstract

The central theme of this review is the dynamic interaction between information selection and learning. We pose a fundamental question about this interaction: How do we learn what features of our experiences are worth learning about? In humans, this process depends on attention and memory, two cognitive functions that together constrain representations of the world to features that are relevant for goal attainment. Recent evidence suggests that the representations shaped by attention and memory are themselves inferred from experience with each task. We review this evidence and place it in the context of work that has explicitly characterized representation learning as statistical inference. We discuss how inference can be scaled to real-world decisions by approximating beliefs based on a small number of experiences. Finally, we highlight some implications of this inference process for human decision-making in social environments.

Keywords

representation learning; learning selective attention; memory

1. INTRODUCTION

Millennial decision-making can be complicated. We have a myriad of gamified apps that provide us with ever more goals to strive toward. Our attention is in high demand by different facets of social media. Our working memory is taxed by constant multitasking. To go about our daily lives, we must somehow discount the distal sense of catastrophe surrounding climate change (and the not-so-distal one brought on by a global pandemic). Yet, even in these conditions, we strive to make sense of the world and make good decisions. At the core of this flexibility is the ability to select relevant sources of information and use them for appropriate action selection.

Humans learn more from their experiences than just how to behave in different situations; they also learn to organize experiences into internal representations that facilitate future behavior. This organization depends on two core cognitive functions: selective attention and memory. In this review, we discuss evidence that attention and memory can be thought of as

interacting components of representation learning. We show how humans rely on different forms of statistical inference to organize past experiences and how this inference process gives rise to compact representations of tasks that guide action in complex environments. We illustrate the generality of this framework by showing that people spontaneously rely on such inference in social environments.

Representation learning: the process of learning a useful and compact mapping between observations and states in a specific task. Usefulness can be measured by how efficiently one can solve a task given the current representation.

1.1. Partially-Observable Markov Decision Processes

We first lay out a framework for defining information selection from the point of view of a goal-directed agent that is trying to make optimal decisions in a multidimensional world.

A Partially-Observable Markov decision process (POMDP) formalizes an agent's internal model of the way the world unfolds throughout a task (Kaelbling, Littman, & Cassandra, 1998; Russell & Norvig, 2002; Sutton & Barto, 2018; note that while POMDPs are often considered a formalization of the actual generative structure of the environment, here we focus on POMDPs as formalizing the agent's internal representation of a task). POMDPs consist of tuples $\{S, A, T, R, O\}$. We can define the state space S as a set of features (known as dimensions or attributes in the psychology literature) relevant to the agent's goal. For instance, when making tea, the state may include the location of the cup, the existence or absence of different ingredients already in the cup, and the temperature of the water in the kettle. Given a set of possible actions A (e.g., pour water into the cup), the transition function $T = p(s_{t+1}|s_t, a_t)$ describes how the agent believes the state will change after an action, that is, the probability that each state will unfold given the current state and action. The reward function $R = p(r_t|s_t, a_t)$ takes in the current state and action and returns r, a scalar representing the immediate utility of performing that action in that state (when making tea, reward is obtained only when drinking the prepared tea). Transition and reward functions can be either deterministic or probabilistic.

Importantly, in a POMDP, the state of the world is not directly observable. Instead, agents have access to noisy observations emitted by the state as per the observation function $O = p(s_t|o_{1:t}, a_{1:t})$, which they can use to infer the underlying state. However, observations do not uniquely identify states. For instance, the cup of tea will look identical before and after sugar was added to the hot water. As a result of this aliasing, observations are not a sufficient statistic for determining the probability that different events will unfold in the future (e.g., will the tea be sweet, and therefore highly rewarding). POMDPs, in this sense, are a formal representation of many real-world situations in which we must infer, using current and past observations and actions, the current state of the world. Observation, transition and reward functions together comprise a "world model" – a model of how the world would unfold henceforward, as per the current knowledge of the agent (Hamrick 2019). Based on this model, the agent can simulate different decisions in order to plan future action.

One might think of this process of inferring states from observations in two different ways. In the machine-learning POMDP literature, the observation function encodes a probability

distribution over true states of the environment. Alternatively, we do not have to assume that a true Markov state that encompasses all that is necessary to determine T and R actually exists (R. Sutton, personal communication) and can instead interpret S_t as a state representation internal to an agent that balances accurately capturing the causal relationships between events in the world and the computational and representational constraints of the agent. Under this view, representation learning is the process of developing a mapping from raw observations to state representations that are appropriate for the current task. The usefulness of this mapping, and the inferred underlying states, can be determined by experience (McCallum 1997).

1.2 The curse of dimensionality

To make correct actions, many reinforcement-learning algorithms postulate that agents estimate, either through planning or through trial and error, the future value (i.e., the expected sum of future rewards) contingent on taking different actions at each state of the world. For instance, the value of adding sugar to an otherwise sugar-less cup of tea may be high, but the value of adding sugar once sugar had already been added is very low. There is much evidence that humans and animals learn such values from trial and error, and the neural substrates of this learning process are well mapped (Barto 1995, Collins & Frank 2014, Daw et al. 2005, Joel et al. 2002, Montague et al. 1996, Niv 2009, Schultz et al. 1997, Sutton 1988). But the same algorithms that explain behavior and neural activity on simple learning tasks learn much more slowly as the dimensionality of the environment grows (Bellman 1957, Sutton 1988)—too slow to accurately account for real-world learning (Lake et al. 2017). This is because representing all features of the environment quickly leads to a combinatorial explosion, yielding too many different states for which values and policies need to be learned (Figure 1).

Nevertheless, given multidimensional observations, an agent could compress the state space by representing only some features and not others. This would improve learning through generalization – what is learned in one situation (adding sugar to a red mug of tea) can be used to estimate values and choose actions in other situations (making tea in a glass, or in a blue mug, or at the office rather than at home), as long as those differences (cup material, color and location) are not represented as part of the state. In the rest of this review, we discuss work on human representation learning: how humans learn a useful and compact mapping between observations and states that affords efficient learning.

2. ATTENTION AND MEMORY CONSTRAINTS IN REPRESENTATION LEARNING

At the core of representation learning is a need for efficient compression of experiences into a small number of states that summarize the task-relevant information in a way that supports efficient learning of action policies. The optimal level of compression depends on the task, and is also influenced by the capacity of the agent's attention and memory systems (Luck & Vogel 1997, Ungerleider & Kastner 2000). Recent work suggests that both selective attention (Leong, Radulescu et al. 2017, Mack et al. 2016, Markovi et al. 2015, Niv et al. 2015) and memory (Bornstein et al. 2017, Collins & Frank 2012) independently

influence learning and action selection. However, a fundamental feature of both systems is their limited capacity: We cannot attend equally to all the features of an observation, and we cannot remember every observation we experience. This gives rise to a credit-assignment problem for meta-decisions (Griffiths, Lieder, & Goodman, 2015; Sutton, 1984): Which features of the observation space are important for inferring latent state and predicting future reward, and should thus survive attentional selection? And which past events (observations and actions) are necessary for correctly predicting state and reward dynamics and should therefore be remembered for later recall?

Credit-assignment problem: The problem of correctly allocating credit (or blame) for an outcome to antecedent actions or states.

2.1. Learning to Attend

Within psychology and neuroscience, attention has classically been defined as the selective processing of a subset of features at any stage between sensation and action (Gottlieb 2012, Lindsay 2020). We suggest that a useful way to think about the need for selective attention is as a solution to the curse of dimensionality (Section 1.2)—attending to irrelevant features of the environment will cause a proliferation of states that will slow down learning (Niv 2019). We next review literature that has formalized attention learning as a policy solution for the curse of dimensionality, allowing agents' decision about what is relevant to inferring latent states change with experience (Figure 1). By selectively attending to only some aspects of the environment, an agent can easily generalize behavior across multiple different observations o, as long as they are similar along attended dimensions.

Recent theories of dimensional attention learning are rooted in the study of human category learning (Jones & Cañas 2010, Kruschke 1992): the process of grouping experiences into meaningful mental representations. In a classic categorization study, Shepard et al. (1961) showed that the ease with which we learn categories depends on the structure of the categorization problem. Human participants had to learn six types of category structures, each constructed from three binary stimulus dimensions (similar to Figure 1). Type I was based on a unidimensional rule, so attending to a single dimension was sufficient for perfect classification; type II instantiated an exclusive– or (XOR) logical rule in which membership depends upon whether one (not both) of two specific dimension values is present. In this case, correct classification depended on attending to two dimensions. Types III, IV, and V had some regularity, but an exception to the rule was always present. To perform well, attention to all three dimensions was necessary. Finally, type VI could only be learned by memorizing the mapping between individual items and category labels; that is, no generalization across the eight stimuli was possible. Human categorization performance reflected the underlying structure: People were fastest at learning type I problems and slowest at learning type VI problems.

To explain this pattern of results, the influential attention learning–covering map (ALCOVE) model formalized two key cognitive principles (Kruschke 1992): (*a*) The basis of classification is similarity to all exemplars in a category, and (*b*) classification requires dynamic selective attention to different stimulus dimensions. To accomplish the second principle, ALCOVE uses error-driven learning to learn attention weights that dictate

how much each dimension should factor into the similarity computation. Notably, for the benchmark data set (Shepard et al. 1961), dynamic selective attention is critical for reproducing the patterns of human behavior, outperforming approaches based on rule learning (Nosofsky et al. 1994). The ALCOVE model is one example in a class of models that formalize dimensional attention in a connectionist framework, and show how attention could be learned via an error-correcting mechanism (Cohen et al. 1990, Jones & Cañas 2010, Roelfsema & Ooyen 2005). The model has also led to the insight that learners restrict their attention to dimensions that are needed to succeed at the task at hand (McCallum 1997, Rehder & Hoffman 2005).

In a striking example of such representation learning, Schuck et al. (2015) instructed participants to press a button to indicate the location of a patch of colored squares within a reference frame: regardless of the color of the squares, if the patch was closer to the bottom left or top right, one key was to be pressed, whereas if the patch was closer to the bottom right or top left, a different key was to be pressed. Initially, the task was perceptually difficult albeit conceptually straightforward. At some point during the task, unbeknownst to the participants, a deterministic mapping between the color of the patch and the correct response was introduced—all patches that were near the bottom left or top right were red, whereas the other patches were blue. Despite extended practice with the location-based policy, some participants spontaneously adapted their state representation to the new structure of the task, using color to respond faster. The strategy shift was preceded by an increase in color information content in the medial prefrontal cortex (Schuck et al. 2015). These results directly show that humans can quickly and flexibly change strategies from one feature to another in the absence of explicit instructions, as had been hinted at by the category learning work.

Notably, categorization tasks, and other commonly used dimensional attention learning tasks such as the Wisconsin Card Sorting Task in humans (Berg, 1948; Milner, 1963; Steinke, Lange, Seer, & Kopp, 2018), and the intradimensional/extradimensional shift task in animals (Birrell & Brown, 2000), commonly test attention learning with deterministic feedback (Radulescu et al. 2019a). To address attention learning in POMDPs, we focus on studies that use probabilistic reinforcement, as such an impoverished feedback signal places stringent requirements on the representation learning algorithm.

Several recent studies directly address the interaction between attention and reinforcement learning in probabilistic settings. Tasks in this literature bear a strong resemblance to category learning tasks, in that learning a response policy can be construed as categorizing multidimensional stimuli according to the response that maximizes reward for each. An important methodological development for studying representation learning in these tasks has been the adoption of trial-by-trial model fitting (Bishara et al., 2010; Daw, 2011; Markovi, Gläscher, Bossaerts, O'Doherty, & Kiebel, 2015; Wilson & Collins, 2019). This method allows for inferring the dynamics of internal states (such as attention), predicting individual participants' future behavior given those states (e.g., what they will choose, or how quickly) and searching for brain correlates of computational components underlying ongoing behavior.

Using this approach, Niv et al. (2015) have demonstrated that representation learning recruits the same brain areas known to engage in goal-directed attentional selection (Corbetta & Shulman 2002). A subsequent study that directly measured dimensional attention using a combination of functional MRI and eye tracking showed that humans choose and learn based on an attentionally filtered state representation that dynamically changes as a function of experience (Leong, Radulescu et al. 2017). Indeed, the same attentional-control network that is engaged during instructed attentional set shifting is sensitive to learned attention, broadening the role of cortical networks classically thought of as supporting visual attention to include representation learning (Scolari et al. 2015).

We note here that the question of attentional selection has also been longstanding in associative learning, albeit with a focus on attention as enhancing learning about features of a stimulus, or making decisions based on some features and not others, rather than subselecting features for processing and representing. Two prominent theories emphasize opposing factors that might determine how attention is allocated between competing stimuli: the consistency with which a feature predicts reward (Mackintosh, 1975), and its inverse, uncertainty about the prediction (Pearce & Hall, 1980). The balance between these two factors might depend on the purpose of selective attention, with attention to predictive features affecting action selection, and attention to uncertain features enhancing allocation of learning from feedback (Dayan, Kakade, & Montague, 2000; Gottlieb, 2012; Grossberg, 1987). These theories emphasize how quickly one should update reward expectations, whereas attention-learning theories in the categorization literature focus more on how we choose what to attend to.

Finally, current models of human attention learning are limited to selecting among a finite set of possible features, rather than modifying representational spaces to accommodate new ones. But a more general form of representation learning would require flexibly changing representational spaces, a problem we turn to in section 2.2.1.

2.2. Learning to Remember

Reducing the size of the representation space using attention is one half of representation learning. The other half requires the opposite operation: augmenting the current observation, for example using information stored in memory (Barron et al. 2013, Biderman et al. 2020, Bornstein & Norman 2017, Shadlen & Shohamy 2016). For instance, remembering our recent actions when making tea will helpfully resolve whether the state of the world already includes sugar in the cup, and whether water has been boiled or not. This gives rise to two additional representation learning problems: 1) learning what to store in working memory to augment the current observation (e.g., recent sugar-related actions), and 2) organizing past experience in long-term memory in a way that facilitates retrieval of their summary statistics in the relevant circumstances (Figure 2). In particular, for the latter, rather than encoding each observation into memory, say, by order of appearance, it is useful to infer which future situations current information might be relevant to—that is, to categorize observations by the states to which they pertain—and update the summary statistics of that state in long-term memory with the current observation. When the state (and thus the prior observation) becomes relevant again, it can be retrieved and used, and potentially updated with new,

pertinent information (Gershman et al. 2017). This ability to actively select different aspects of our memories in the service of decision-making depends on the interaction between multiple memory systems that may operate on different timescales (Brunec et al., 2018; Hoskin, Bornstein, Norman, & Cohen, 2019).

2.2.1. Organizing memory around latent causes—As is the case with attention, adaptive accounts of memory cast memory formation as a problem of generalization, but in time, rather than in perceptual space (Anderson & Milson 1989). From a reinforcement learning perspective, remembering can be viewed as a meta-decision that affects an agent's construal of state. The broader question of representation learning can therefore be phrased as follows: Which past and present sensory data should the agent use to define the state?

Latent-cause inference provides a statistical framework for answering this question by grouping related experiences into latent (i.e., hidden) causes (Courville et al. 2006, Gershman et al. 2015). Latent-cause inference treats memory formation as a clustering problem, grouping experiences into memory traces based on similarity. Experiences that belong to the same latent cause can then be summarized, stored and recalled as part of the same memory trace (Gershman et al. 2017). One important insight of latent-cause inference theory is that creating new latent causes in memory depends on the extent to which new experiences are different enough from old ones to trigger the formation of a new latent cause. This organizational principle has helped unify a broad range of phenomena in psychology and neuroscience, most notably fear conditioning, categorization, and event segmentation (Anderson 1991; Gershman et al. 2010; 2017; Sanborn et al. 2010; Shin & DuBrow 2020; Franklin et al., 2020).

To illustrate how latent-cause inference can organize the way we use past memories to direct prediction and action in a new situation, consider compound generalization, the process by which animals and humans combine past learning about individual stimuli to form new predictions when encountering a combination of these stimuli. Experimentally, in compound generalization the agent first learns the predictive value of two simple stimuli. The agent is then tested on a novel compound that shares features with both of the simple stimuli. The degree to which the agent generalizes experience from simple to compound stimuli is measured through the resulting behavior.

The study of compound generalization has long been driven by a debate between elemental and configural theories. In elemental processing, a compound stimulus is treated as the sum of its parts, each with its own predictive value; in configural processing, the parts are integrated into a unique perceptual whole for which predictions are learned separately, with behavior depending both on the compound and on each of its elements/subcompounds, weighted by their similarity to the current configuration. Elemental and configural theories fundamentally differ in the degree of generalization they assume (Soto et al. 2014): Because individual cues are expressed independently in elemental theories, elemental generalization is stronger than configural generalization.

Soto et al. (2014) integrated elemental and configural theories into an account of compound generalization based on latent-cause inference. Their model grouped simple stimuli into

latent causes based on their statistical association with reward (Courville et al. 2006, Navarro 2006, Shepard 1987, Soto et al. 2014). When the agent encountered a novel compound stimulus, it used the features of the compound stimulus to infer whether one or several latent causes are present. Critically, recognizing that even repeated presentations of the same stimulus (in training) may lead to somewhat different perceptions trial by trial, latent causes in this model were associated with whole regions in perceptual space. These are called consequential regions, as they describe a region in perceptual space that leads to the same reward consequences. According to the model, experiences in the same consequential region are stored together at learning and recalled together at decision time. Indeed, the model does not remember every experience – it summarizes them in sufficient statistics (the extent of the consequential region, and the probability of reward) that are associated with the latent cause. From this organization follows generalization: the extent to which a particular experience will generalize to related ones depends on the structure and size of the consequential regions. This principle explains, for instance, the effect of sensory modality on elemental versus configural processing (Melchers et al. 2008). When two stimuli are from different modalities, they are more likely to be clustered into two separate consequential regions (leading to summation of predictions when the compound is presented); whereas when two stimuli are from the same modality, they are most often clustered into a single consequential region (leading to averaging of predictions for the compound) (Soto et al. 2014).

Viewing memory organization through the lens of latent-cause inference suggests that signals that disambiguate latent causes will be useful in organizing memory encoding and retrieval. One such signal is the context (i.e., the set of features that remain relatively constant over time). Indeed, observations encoded in a similar context are putatively bound in memory by the context in which they were encoded and tend to jointly influence decisions (Bornstein & Norman 2017). In fact, one can argue, based on latent-cause inference theory, that context can be defined as the currently active latent cause. And, as predicted by latent-cause theory, recent work suggests that events that deviate from our expectations—and therefore signal a new latent cause—help organize memories into distinct memory traces (Rouhani et al. 2018, 2020).

2.2.2. Using memory to augment state representations.—Memory also influences state inference and representation by augmenting immediately available sensory observations with previous ones that are germane to inferring the current state or latent cause. Todd et al. (2009) articulated this problem in the form of a POMDP that defines its internal state as the current observation augmented with a working memory component. Their model learned what recent observations are important to store in working memory through trial and error: The action space included not only external actions (i.e., pressing a button) but also memory actions (i.e., replacing current working memory content with a new observation). Learning the optimal memory policy for this POMDP was similar to learning external actions through trial and error and updating state values and action policies based on prediction errors. The algorithm learned to behave optimally in a benchmark human sequence-learning task and exhibited computational constraints similar to those seen in human working memory.

Collins et al. (2014) used a different implementation of working memory to model a deterministic instrumental learning task with a varying number of states (stimuli) and three possible actions. Their agent had two independent components: a standard action value-learning module and a working-memory module that retained the most recent observations and decayed over time. On each trial, the agent's policy was a mixture of the two components, where the mixing proportion was determined by the stimulus set-size (how many states humans had to learn about) and a fixed working-memory capacity. Collins et al. (2014) found that incorporating a limited-capacity working memory into the model captured human behavior better than did simple reinforcement learning. However, this model did not learn what to store in working memory.

Neither of the above models incorporated the idea of latent causes into the structure of memories used to augment reinforcement learning. Future work on representation learning stands to benefit from modifying these theories of working memory to account for the latent structure of memories, which can itself change with experience.

2.2.3. Memory for relational structure.—Models of memory that we have discussed so far address the question of which memories should be included in which state. But another aspect of memory that may be critical for generalization is its relational structure. Within the same state, humans group memories and concepts in relation to one other, and use this structure to make predictions about the future (Behrens et al., 2018; Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013). One recent proposal by Whittington et al. (2020) combines one-shot Hebbian memory with a mechanism for learning the relational structure between individual memories. This mechanism enables the model to address a memory based on its position in a graph. Since the graph structure is abstracted away from sensory input, the model can efficiently generalize to new, unseen environments. Evidence suggests that humans can indeed learn to represent the temporal structure of memories as graphs, and use this knowledge to infer relationships that they do not directly experience (Garvert, Dolan, & Behrens, 2017; Mark, Moran, Parr, Kennerley, & Behrens, 2020).

3. APPROXIMATE INFERENCE IN REPRESENTATION LEARNING

In the above discussion, we eschewed implementational constraints on latent-cause inference. However, these are nontrivial. In particular, in this framework, the agent's interpretation of the world relies on inferring a probability distribution over possible states given current and past observations. Exact Bayesian inference of this distribution is typically intractable in most cases (Niv et al. 2015, Roy et al. 2005). For this reason, sampling algorithms have recently gained popularity as neurally-plausible mechanistic accounts of how Bayesian inference may be realized in the brain online as new information comes in (Fiser & Lengyel, 2019; Haefner & Berkes, 2016; Sanborn & Chater, 2016; Sanborn, Griffiths, & Navarro, 2010; Schneegans, Taylor, & Bays, 2020; Speekenbrink, 2016).

Sequential sampling algorithms: algorithms that approximate a probability distribution using a finite set of samples and update the samples after each experience

Taking a sequential sampling view of learning explains the puzzling finding that even though learning behavior is often consistent with Bayesian inference at the group level, individuals are much more variable in their actions during a single learning episode, and learning can be an all-or-nothing process (Bower & Trabasso 1963, Daw & Courville 2008, Gallistel et al. 2004). This type of insight learning has been observed not only in simple associative learning studies, but also in experiments in which humans inferred hidden task structure (Schuck et al. 2015). Insight learning through sampling is also consistent with eye-tracking studies of multidimensional learning, in which humans appear to abruptly switch between different task representations (Leong, Radulescu et al. 2017, Rehder & Hoffman 2005). These findings suggest a mechanism by which humans sample different hypotheses about task structure and use them as a stand-in for the true structure of the task.

One recent study has explicitly modeled selective attention measured with eye-tracking as a sequential sampling algorithm that performs approximate statistical inference about task structure (Radulescu et al. 2019b). Under this view, humans build task representations by testing one hypothesis at a time about which features of the world are relevant for representing states. The hypothesis is updated by sampling new hypotheses in proportion to how consistently they predict recent observations stored in memory. The current hypothesis, in turn, directs attention to particular aspects of multidimensional stimuli (Radulescu et al. 2019a). Sequential sampling can thus provide a coherent mechanistic account for how attention and memory processes support representation learning, grounding it in principles of statistical inference (Daw & Courville 2008, Radulescu et al. 2019b). This view is consistent with empirical work showing that the focus of attention depends on what aspects of past experience are remembered (Goldfarb et al. 2016, Günseli & Aly 2020, Hutchinson & Turk-Browne 2012, Myers et al. 2017).

It is possible that sequential sampling algorithms could accommodate problems of increasing complexity, as those proposed by Shepard et al. (1961), by gradually adding more than one hypothesis or by considering alternative ways to parametrize the proposal distribution for new particles (Ballard et al. 2018, Goodman et al. 2008, Radulescu et al. 2019a, Song et al. 2020, Wilson & Niv 2012). Hypotheses themselves could take the form of task rules or causal links between observations and latent causes (Radulescu et al. 2019a). While additional work is needed to disambiguate between these different implementations, sequential sampling algorithms are a useful general way to think about how people test hypotheses about which features are most relevant for the task at hand (Ballard et al. 2018, Radulescu et al. 2019b, Sanborn et al. 2010, Song et al. 2020, Wilson & Niv 2012).

But how do humans know how long to sample in order to approximate the utility of a task representation well enough to make good decisions? Resource-rational theory provides a principled answer to this question by framing the problem in terms of the computational cost of sampling additional hypotheses (Hay et al. 2014, Russell & Wefald 1992). To be resource rational, an agent should only sample for as long as the cost of sampling is smaller than the additional value gained by expending computational resources. This idea has successfully been applied to modeling how humans make decisions in consumer choice. Analyzing gaze data in a ternary choice experiment, Callaway & Griffiths (2019) have shown that people's fixation sequences are consistent with an agent that optimally solves the trade-off between

performance and cost of computation. This approach has also been leveraged to model more complex computations, including planning and task decomposition (Callaway et al. 2018, Correa et al. 2020). In conjunction with sampling-based approaches to representation learning, resource rationality may therefore provide a set of mechanistic principles by which good-enough representations can flexibly be learned for a variety of tasks.

4.   FATHOMING THE UNFATHOMABLE MINDS OF OTHERS

Arguably the most complex part of our environment consists of other people, whose key features such as current beliefs (Baker et al. 2017), mental states, and traits (Tamir & Thornton 2018) are rarely observable. As a result, we often infer these latent states from observable actions, physical characteristics of the target person (e.g., formal attire), and situational factors (e.g., job interview) with the overall goal of making accurate predictions about behaviors of others (Vélez & Gweon 2020).

The interactive-POMDP framework by Gmytrasiewicz & Doshi (2005) extended POMDPs to situations where multiple agents interact, each representing the other agents in their own POMDP. A later Bayesian theory of mind (BToM) model (Baker et al. 2017) further assumed that an observer uses POMDPs to represent the relationship between the inherently hidden minds of others (e.g., beliefs and desires) and observable variables such as actions and environments. In this study, human observers watched an animated agent (henceforth, a target) attempting to navigate to different goals (food trucks) with its view partially blocked by obstacles. The BToM model correctly predicted observers' answers to questions regarding beliefs (e.g., does the target believe a Korean food truck is on the other side of the parking lot?) and desires (e.g., does the target like Korean food better than Lebanese food?) given the observation (e.g., the target goes past the Lebanese food truck to where they can look beyond an obstacle), appropriately capturing an observer's inference about the target's beliefs and desires given the observation.

Beyond predicting other people's actions, inferring other people's hidden cognitive states (e.g., knowledge and intention) is useful when we need to learn from feedback given by other people. For instance, we can make inferences about someone's knowledge from observed actions and adjust how much we learn from them (for a review, see Shafto et al. 2012). Similarly, if we receive evaluative feedback (reward) from someone who we believe intends to communicate a rule to us, we can use their feedback to figure out the rule rather than trying to maximize the reward (Ho et al. 2019). On top of knowledgeability and intention to teach, people consider whether the teacher is trustworthy or has bad intentions (Landrum et al. 2015). Children and adults take into account niceness and honesty (Mascaro & Sperber 2009) as much as the smartness of a teacher when deciding whom to ask for information, suggesting that we draw inferences from traits, a deeper layer of hidden states.

Trait inference is beneficial in that it provides a stable basis for generalizing behaviors across different situations. People spontaneously form mental representations of other people's traits (Winter & Uleman 1984), even when the immediate task does not require trait inference and could be performed better when ignoring such information. For instance, Hackel et al. (2015) had participants make a series of decisions about a partner in a game

where the other player divided a pool of points between the two players. Importantly, immediate rewards (the number of points the target shared with the participant on a trial) and generosity (the proportion shared from the pool) were orthogonalized, and participants could maximize their share by just focusing on the immediate rewards, ignoring the trait generosity. Supporting the idea that spontaneous trait inference guides decisions, participants used both the reward magnitude and learned generosity information in making their decisions, with the ventral striatum responding to both reward and trait prediction errors. These results suggest the importance of including a mental model of other people in the state space, as a basis for guiding decisions. In addition to traits such as generosity, people seem to spontaneously represent how powerful are people around them, relative to themselves (Kumaran et al. 2016), with the hippocampus representing information regarding other people's power and affiliation, creating a social map that can help humans navigate the social world (Tavares et al. 2015).

We also often build models of entire social groups—namely, stereotypes (Hilton & von Hippel 1996). For example, people associate gender with different professions, and even after learning counter-stereotypic information (e.g., "Elizabeth is a doctor, and Jonathan is a nurse."), the stereotypic association lingers and slows down responses that are counter-stereotypic (Cao & Banaji 2016). Such associations emerge early in development and bias inferences, for instance about racial group membership when information about characteristics such as wealth is provided (Olson et al. 2012). It is worth noting that, as in nonsocial decision-making, features that are relevant to social decision-making can vary depending on the problem at hand. Therefore, the same questions about representation learning, and the involvement of attention and memory processes, are likely to be relevant —and even exacerbated—in the social setting. For instance, when people are evaluating a group of people based on interactions with group members, experiences with outlying group members weigh more heavily in the overall evaluation of the group, as predicted by models that assume representation learning and latent-cause inference. In particular, these models infer that outlier experiences come from a separate latent cause or group, and as a result overweigh their contribution to overall assessment (Shin & Niv, in press). Considering the hierarchical nature of groups composed of individuals therefore suggests another open question: How do the inferred characteristics of social groups inform the state inference of stable traits of an individual group member and, in turn, the transient mental state of the individual?

In the real world, we often generate spontaneous thoughts by representing the mental models of others (Mildner & Tamir 2018). And we effortlessly copy others' strategies to solve complex tasks (Rendell et al. 2010). Studying how we build models of the highly rich and partially observable social world is therefore valuable not only because it will help understand social decision-making, but also because it will provide deeper insights into how we represent the world as a whole (Kampis & Southgate 2020). The field of social psychology has generated a wealth of research on topics that parallel questions in representation learning, for instance, how social groups organize memories about group members (Sherman et al 2002), how memory about a person and group-level representations interact with one another to guide predictions about individual behavior (Brewer et al. 1995), and how attention is allocated to information that fits our stereotypes (Plaks et al. 2001;

Bastian & Haslam 2007). It would be useful to draw upon this existing body of research to further our understanding of inference processes and representation learning, especially given that our partially observable world so often involves making decisions about other people.

5. ISSUES AND FUTURE DIRECTIONS

We have reviewed work on how attention and memory processes contribute to representation learning, describing how these processes can be formalized within the framework of approximate Bayesian statistical inference within partially observable Markov decision processes, and arguing that people often organize experiences around their inferences about the social world. We now outline some open questions and possible future directions.

5.1. Determining the Balance Between Elements and Configurations

Echoing the debate between elemental and configural theories discussed in Section 2.2.1, several studies on representation learning have employed trial-by-trial model fitting to ask under which conditions humans adopt feature- versus object-based representations (Ballard et al. 2018; Farashahi et al. 2017, 2020; Mack et al. 2016; Markovi et al. 2015). An open question concerns the dynamics of the balance between the two and their neural implementation.

Humans and animals are able to encode unique conjunctions on the fly (Melchers et al. 2008), with the hippocampus arbitrating the degree to which individual memories generalize into a coherent whole (Ballard et al. 2019, Duncan et al. 2018, Kumaran & McClelland 2012). Some recent evidence suggests the possibility that feature- and object-based representations may be reinforced in parallel (Ballard et al. 2019) and could be subserved by different specialized neural systems, all of which project to parts of the striatum. For instance, while the hippocampus may support object-based representations via projections to the ventral striatum (Haber & Knutson 2010), visual areas could convey feature-based representations via projections to the caudate tail (Seger 2013). This specialized organization in corticostriatal synapses appears to be maintained downstream in nigrostriatal dopaminergic synapses, potentially allowing updates of both elemental and configural state features in parallel (Engelhard et al. 2019, Hebart et al. 2018, Lee et al. 2020).

5.2. Hypothesis Spaces for State Representation

All algorithms for representation learning discussed in this review assume a low-dimensional hypothesis space over which inference can occur. In most experiments, this space is assumed to consist of separable dimensions such as color or shape. A pressing question remains regarding what are the building-block dimensions that define the axes of hypotheses in the real world (Boroditsky & Ramscar 2001, Medin et al. 1993, Navarro 2006). One possibility is that these dimensions are themselves learned via an unsupervised statistical inference process that extracts natural variation in the environment (e.g., covariance and hierarchical structure) (Kemp et al. 2007, Sanborn et al. 2009). For instance, Sanborn et al. (2009) showed that human-like dimensional biases emerge when

training latent-cause inference algorithms (see section 2.2.1) on perceptual inputs that match those encountered by children.

But there remains a gap between the computational-level description of such models and their algorithmic implementation and neural substrates. Future work on Bayesian approaches to representation learning stands to benefit from understanding whether the assumptions such models make about representational primitives (e.g., simple features, conjunctions/ compositions, cluster assignments, and motor programs) are borne out in neural data (Mack et al. 2016, Park et al. 2020, Tomov et al. 2018).

Another way to bridge this gap is to ask whether the kinds of structure that Bayesian models capture can be learned and represented in biologically plausible artificial neural networks (ANNs). Research at the intersection of human and artificial intelligence is focused on endowing ANNs with the kinds of dimensional biases humans display, for example, by tailoring the training regime to data constraints present in human development (Feinman & Lake 2018, Orhan et al. 2020, Smith et al. 2011) or by teaching agents to implicitly represent structure (e.g., objects and scenes) in an unsupervised manner (van den Oord et al. 2018, Sitzmann et al. 2019).

Various data-driven approaches have also led to the insight that conceptual representations are much richer than previously thought (Battleday, Peterson, & Griffiths, 2020; Hebart, Zheng, Pereira, & Baker, 2020; Hornsby, Evans, Riefer, Prior, & Love, 2020). For example, Hornsby et al. (2020) applied a Bayesian nonparametric model to consumer rating data to challenge a long-standing assumption of laboratory studies. They found that rather than organizing concepts only by their intrinsic dimensions and features (e.g., bananas are long and yellow), people organize concepts around what they can do with them (e.g., bananas go well in smoothies). Such work demonstrates the special status of actions in conceptual organization and points to a possible role for affordances as a key element of hypothesis spaces that guide representation learning (Khetarpal et al. 2020).

5.3. Toward Domain Generality in Representation Learning

Several recent studies suggests that human task representations differ across domains. For instance, Wu et al. (2020) found significant differences in exploration behavior depending on whether the task was defined in a spatial or conceptual domain, and the extent to which humans adopt elemental or configural representations depends on whether task stimuli are abstract or naturalistic (Farashahi et al. 2020). These results point to the possibility that the so-called psychological space (Shepard 1987) may be meaningfully decomposed into domains that have vastly different, yet relevant, underlying structures (e.g., perceptual features, physical properties, abstract conceptual properties, social graphs, and time).

Whether there exist learning mechanisms that can capture structure across different domains remains an open question. One approach that may move us closer to an answer is to complement typical laboratory experiments with naturalistic data from task environments that more closely match those that humans are likely to encounter in real life (Nastase et al. 2020, Smith et al. 2011). Allowing humans to be embodied and interact with naturalistic environments would allow us to study the full gamut of inductive biases that they use

to solve in the real world even those tasks that have long been considered canonical lab experiments, such as visual search (Radulescu et al. 2020).

In the social realm, it is plausible, for example, that people reason by learning the structure of discrete graphs (Lau et al. 2020, Parkinson & Du 2020, Parkinson et al. 2017, Wu et al. 2020) and build a cognitive map from piecemeal learning (Park et al. 2020). Integrating this knowledge with models of other agents (Baker et al. 2017) may support a range of functions, from cooperation to imitation learning (Kleiman-Weiner et al. 2016, Rendell et al. 2010), and facilitate the transmission of knowledge about task structure.

In short, systematically testing, in naturalistic domains, the theories of representation learning reviewed here is a critical step forward in building a richer picture of how humans learn to represent a wide range of tasks.

ACKNOWLEDGMENTS

LITERATURE CITED

Anderson JR. 1991. The adaptive nature of human categorization. Psychol. Rev 98(3):409–29

Anderson JR, Milson R. 1989. Human memory: an adaptive perspective. Psychol. Rev 96(4):703

Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB. 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat. Hum. Behav 1(4):1–10

Ballard I, Miller EM, Piantadosi ST, Goodman ND, McClure SM. 2018. Beyond reward prediction errors: Human striatum updates rule values during learning. Cereb. Cortex 28(11):3965–75 [PubMed: 29040494]

Ballard I, Wagner AD, McClure SM. 2019. Hippocampal pattern separation supports reinforcement learning. Nat. Commun 10(1):1073 [PubMed: 30842581]

Barron HC, Dolan RJ, Behrens TEJ. 2013. Online evaluation of novel choices by simultaneous representation of multiple memories. Nat. Neurosci 16(10):1492–98 [PubMed: 24013592]

Barto AG. 1995. Adaptive critics and the basal ganglia. In Models of Information Processing in the Basal Ganglia, Vol. 11, ed. Houk JC, Davis J, Beiser D, pp. 215–32. Cambridge, MA: MIT Press

Bastian B, & Haslam N (2007). Psychological essentialism and attention allocation: Preferences for stereotype-consistent versus stereotype-inconsistent information. The Journal of social psychology, 147(5), 531–541 [PubMed: 18225833]

Battleday RM, Peterson JC, & Griffiths TL (2020). Capturing human categorization of natural images by combining deep networks and cognitive models. Nature communications, 11(1), 1–14

Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, et al. 2018. What is a cognitive map? Organizing knowledge for flexible behavior. Neuron 100(2):490–509 [PubMed: 30359611]

Bellman R 1957. A Markovian decision process. J. Math. Mech 6(5):679–84

Bengio Y, Courville A, Vincent P. 2013. Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell 35(8):1798–828 [PubMed: 23787338]

Berg EA (1948). A simple objective technique for measuring flexibility in thinking. The Journal of general psychology, 39(1), 15–22. [PubMed: 18889466]

Biderman N, Bakkour A, Shohamy D. 2020. What are memories for? The hippocampus bridges past experience with future decisions. Trends Cogn. Sci 24(7):542–56 [PubMed: 32513572]

Birrell JM, & Brown VJ (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. Journal of Neuroscience, 20(11), 4320–4324 [PubMed: 10818167]

Bishara AJ, Kruschke JK, Stout JC, Bechara A, McCabe DP, & Busemeyer JR (2010). Sequential learning models for the Wisconsin card sort task: Assessing processes in substance dependent individuals. Journal of mathematical psychology, 54(1), 5–13. [PubMed: 20495607]

Bishop CM. 2006. Pattern Recognition and Machine Learning. New York: Springer

Bornstein AM, Khaw MW, Shohamy D, Daw ND. 2017. Reminders of past choices bias decisions for reward in humans. Nat. Commun 8(1):15958 [PubMed: 28653668]

Bornstein AM, Norman KA. 2017. Reinstated episodic context guides sampling-based decisions for reward. Nat. Neurosci 20(7):997–1003 [PubMed: 28581478]

Boroditsky L, Ramscar M. 2001. First, we assume a spherical cow… Behav. Brain Sci 24(4):656–57

Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, Hassabis D. 2019. Reinforcement learning, fast and slow. Trends Cogn. Sci 23(5):408–22 [PubMed: 31003893]

Bower G, Trabasso T. 1963. Reversals prior to solution in concept identification. J. Exp. Psychol 66(4):409–18 [PubMed: 14053258]

Brewer MB, Weber JG, Carini B. 1995. Person memory in intergroup contexts: categorization versus individuation. J. Personal. Soc. Psychol 69(1):29–40

Brunec IK, Bellana B, Ozubko JD, Man V, Robin J, Liu ZX, … & Moscovitch M (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. Current Biology, 28(13), 2129–2135 [PubMed: 29937352]

Callaway F, Rangel A & Griffiths T (2019). Fixation patterns in simple choice reflect optimal information sampling. PsyArXiv. https://psyarxiv.com/57v6k/ doi:10.31234/osf.io/57v6k

Callaway F, Lieder F, Das P, Gul S, Krueger PM, Griffiths T. 2018. A resource-rational analysis of human planning. In Proceedings of the 40th Annual Conference of the Cognitive Science Society 2018. Eds. Kalish C, Rau M, Zhu J, Rogers T.

Cao J, Banaji MR. 2016. The base rate principle and the fairness principle in social judgment. PNAS 113(27):7475–80 [PubMed: 27325760]

Cohen JD, Dunbar K, McClelland JL. 1990. On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychol. Rev 97(3):332–61 [PubMed: 2200075]

Collins AGE, Brown JK, Gold JM, Waltz JA, Frank MJ. 2014. Working memory contributions to reinforcement learning impairments in schizophrenia. J. Neurosci 34(41):13747–56 [PubMed: 25297101]

Collins AGE, Frank MJ. 2012. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur. J. Neurosci 35(7):1024–35 [PubMed: 22487033]

Collins AGE, Frank MJ. 2013. Cognitive control over learning: creating, clustering, and generalizing task-set structure. Psychol. Rev 120(1):190–229 [PubMed: 23356780]

Collins AGE, Frank MJ. 2014. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol. Rev 121(3):337–66 [PubMed: 25090423]

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci 3(3):201–15 [PubMed: 11994752]

Correa CG, Ho MK, Callaway F, Griffiths TL. 2020. Resource-rational task decomposition to minimize planning costs. arXiv:2007.13862 [cs.AI]

Courville AC, Daw ND, Touretzky DS. 2006. Bayesian theories of conditioning in a changing world. Trends Cogn. Sci 10(7):294–300 [PubMed: 16793323]

Craiu RV, Rosenthal JS. 2014. Bayesian computation via Markov chain Monte Carlo. Annu. Rev. Stat. Appl 1:179–201

Dayan P, Kakade S, & Montague PR (2000). Learning and selective attention. Nature neuroscience, 3(11), 1218–1223 [PubMed: 11127841]

Daw N 2011. Trial by trial data analysis using computational models. In Decision Making, Affect, and Learning: Attention and Performance XXIII, ed. Delgado MR, Phelps EA, Robbins TW, pp. 3–38. Oxford, UK: Oxford Univ. Press

Daw N, Courville A. 2008. The pigeon as particle filter. Adv. Neural Inf. Process. Syst 20:369–76

Daw N, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci 8(12):1704–11 [PubMed: 16286932]

Doucet A, Johansen AM. 2009. A tutorial on particle filtering and smoothing: fifteen years later. In The Oxford Handbook of Nonlinear Filtering, ed. Crisan D, Rozovskiĭ R, pp. 656–704. New York: Oxford Univ. Press

Duncan K, Doll BB, Daw ND, Shohamy D. 2018. More than the sum of its parts: a role for the hippocampus in configural reinforcement learning. Neuron 98(3):645–57 [PubMed: 29681530]

Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, et al. 2019. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature 570(7762):509–13 [PubMed: 31142844]

Farashahi S, Rowe K, Aslami Z, Lee D. 2017. Feature-based learning improves adaptability without compromising precision. Nat. Commun 8:1768 [PubMed: 29170381]

Farashahi S, Xu J, Wu S-W, Soltani A. 2020. Learning arbitrary stimulus-reward associations for naturalistic stimuli involves transition from learning about features to learning about objects. Cognition 205:104425 [PubMed: 32958287]

Feinman R, Lake BM. 2018. Learning inductive biases with simple neural networks. arXiv:1802.02745 [cs.CL]

Fiser J, Lengyel G. 2019. A common probabilistic framework for perceptual and statistical learning. Curr. Opin. Neurobiol 58:218–28 [PubMed: 31669722]

Franklin NT, Norman KA, Ranganath C, Zacks JM, & Gershman SJ (2020). Structured Event Memory: A neuro-symbolic model of event cognition. Psychological Review, 127(3), 327. [PubMed: 32223284]

Gallistel CR, Fairhurst S, Balsam P. 2004. The learning curve: implications of a quantitative analysis. PNAS 101(36):13124–31 [PubMed: 15331782]

Garvert MM, Dolan RJ, & Behrens TE (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. Elife, 6, e17086. [PubMed: 28448253]

Gershman SJ, Blei DM. 2012. A tutorial on Bayesian nonparametric models. J. Math. Psychol 56(1):1–12

Gershman SJ, Blei DM, Niv Y. 2010. Context, learning, and extinction. Psychol. Rev 117(1):197–209 [PubMed: 20063968]

Gershman SJ, Monfils M-H, Norman KA, Niv Y. 2017. The computational nature of memory modification. eLife 6:e23763 [PubMed: 28294944]

Gershman SJ, Norman KA, Niv Y. 2015. Discovering latent causes in reinforcement learning. Curr. Opin. Behav. Sci 5:43–50

Gmytrasiewicz PJ, & Doshi P (2005). A framework for sequential planning in multi-agent settings. Journal of Artificial Intelligence Research, 24, 49–79.

Goldfarb EV, Chun MM, Phelps EA. 2016. Memory-guided attention: independent contributions of the hippocampus and striatum. Neuron 89(2):317–24 [PubMed: 26777274]

Goodman ND, Tenenbaum JB, Feldman J, Griffiths TL. 2008. A rational analysis of rule-based concept learning. Cogn. Sci 32(1):108–54 [PubMed: 21635333]

Gottlieb J 2012. Perspective attention, learning, and the value of information. Neuron 76(2):281–95 [PubMed: 23083732]

Griffiths TL, Lieder F, Goodman ND. 2015. Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. Top. Cogn. Sci 7(2):217–29 [PubMed: 25898807]

Grossberg S (1982). Processing of expected and unexpected events during conditioning and attention: a psychophysiological theory. Psychological review, 89(5), 529. [PubMed: 7178332]

Günseli E, Aly M. 2020. Preparation for upcoming attentional states in the hippocampus and medial prefrontal cortex. eLife 9:e53191 [PubMed: 32255423]

Haber SN, Knutson B. 2010. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35(1):4–26 [PubMed: 19812543]

Hackel LM, Doll BB, Amodio DM. 2015. Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. Nat. Neurosci 18(9):1233–35 [PubMed: 26237363]

Haefner RM, Berkes P, Fiser J. 2016. Perceptual decision-making as probabilistic inference by neural sampling. Neuron 90(3):649–60 [PubMed: 27146267]

Hamrick JB (2019). Analogues of mental simulation and imagination in deep learning. Current Opinion in Behavioral Sciences, 29, 8–16

Harlow HF. 1949. The formation of learning sets. Psychol. Rev 56(1):51–65 [PubMed: 18124807]

Hay N, Russell S, Tolpin D, Shimony SE. 2014. Selecting computations: theory and applications. arXiv:1207.5879 [cs.AI]

Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM. 2018. The representational dynamics of task and object processing in humans. eLife 7:e32816 [PubMed: 29384473]

Hebart M, Zheng CY, Pereira F, Baker C. 2020. Revealing the multidimensional mental representations of natural objects underlying human similarity judgments. Nat. Hum. Behav 4:1173–85 [PubMed: 33046861]

Hilton JL, von Hippel W. 1996. Stereotypes. Annu. Rev. Psychol 47:237–71 [PubMed: 15012482]

Ho MK, Cushman F, Littman ML, Austerweil JL. 2019. People teach with rewards and punishments as communication, not reinforcements. J. Exp. Psychol 148(3):520–49

Hornsby AN, Evans T, Riefer PS, Prior R, Love BC. 2020. Conceptual organization is revealed by consumer activity patterns. Comput. Brain Behav 3:162–73 [PubMed: 32455337]

Hoskin AN, Bornstein AM, Norman KA, Cohen JD. 2019. Refresh my memory: Episodic memory reinstatements intrude on working memory maintenance. Cogn. Affect. Behav. Neurosci 19(2):338–54 [PubMed: 30515644]

Hutchinson JB, Turk-Browne NB. 2012. Memory-guided attention: control from multiple memory systems. Trends Cogn. Sci 16(12):576–79 [PubMed: 23141429]

Joel D, Niv Y, Ruppin E. 2002. Actor–critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw. 15(4–6):535–47 [PubMed: 12371510]

Jones M, Cañas F. 2010. Integrating reinforcement learning with models of representation learning. In Proceedings of the 32nd Annual Meeting of the Cognitive Science Society 2010, Vol. 1, ed. Ohlsson S, Catrambone R, pp. 1258–63. Red Hook, NY: Curan

Kaelbling LP, Littman ML, Cassandra AR. 1998. Planning and acting in partially observable stochastic domains. Artif. Intell 101(1–2):99–134

Kampis D, Southgate V. 2020. Altercentric cognition: how others influence our cognitive processing. Trends Cogn. Sci 24(11):945–59 [PubMed: 32981846]

Kemp C, Perfors A, Tenenbaum JB. 2007. Learning overhypotheses with hierarchical Bayesian models. Dev. Sci 10(3):307–21 [PubMed: 17444972]

Khetarpal K, Ahmed Z, Comanici G, Abel D, Precup D. 2020. What can I do here? A theory of affordances in reinforcement learning. arXiv:2006.15085 [cs.LG]

Kleiman-Weiner M, Ho MK, Austerweil JL, Littman ML, Tenenbaum JB. 2016. Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction. In Proceedings of the 38th Annual Conference of the Cognitive Science Society 2016, Eds. Papafragou A, Grodner D, Mirman D, Trueswell JC. Austin, TX: Cognitive Science Society

Kober J, Bagnell JA, Peters J. 2013. Reinforcement learning in robotics: a survey. Int. J. Robot. Res 32(11):1238–74

Kruschke JK. 1992. ALCOVE: an exemplar-based connectionist model of category learning. Psychol. Rev 99(1):22–44 [PubMed: 1546117]

Kumaran D, Banino A, Blundell C, Hassabis D, Dayan P. 2016. Computations underlying social hierarchy learning: distinct neural mechanisms for updating and representing self-relevant information. Neuron 92(5):1135–47 [PubMed: 27930904]

Kumaran D, McClelland JL. 2012. Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. Psychol. Rev 119(3):573–616 [PubMed: 22775499]

Lake BM, Salakhutdinov R, Tenenbaum JB. 2015. Human-level concept learning through probabilistic program induction. Science 350(6266):1332–38 [PubMed: 26659050]

Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. 2017. Building machines that learn and think like people. Behav. Brain Sci 40:e253 [PubMed: 27881212]

Landrum AR, Eaves BS, Shafto P. 2015. Learning to trust and trusting to learn: a theoretical framework. Trends Cogn. Sci 19(3):109–11 [PubMed: 25563822]

Lau T, Gershman SJ, Cikara M. 2020. Social structure learning in human anterior insula. eLife 9:e53162 [PubMed: 32067635]

Lee J, Wang W, Sabatini BL. 2020. Anatomically segregated basal ganglia pathways allow parallel behavioral modulation. Nat. Neurosci 23:1388–98 [PubMed: 32989293]

Leong YC, Radulescu A, Daniel R, Dewoskin V, Niv Y, et al. 2017. Dynamic interaction between reinforcement learning and attention in multidimensional environments. Neuron 93(2):451–63 [PubMed: 28103483]

Lieder F, Griffiths TL. 2017. Strategy selection as rational metareasoning. Psychol. Rev 124(6):762–94 [PubMed: 29106268]

Lindsay GW. 2020. Attention in psychology, neuroscience, and machine learning. Front. Comput. Neurosci 14:29 [PubMed: 32372937]

Luck SJ, Vogel EK. 1997. The capacity of visual working memory for features and conjunctions. Nature 390(6657):279–81 [PubMed: 9384378]

Mackintosh NJ (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. Psychological review, 82(4), 276

Mack ML, Love BC, Preston AR. 2016. Dynamic updating of hippocampal object representations reflects new conceptual knowledge. PNAS 113(46):13203–8 [PubMed: 27803320]

Mark S, Moran R, Parr T, Kennerley SW, & Behrens TE (2020). Transferring structural knowledge across cognitive maps in humans and models. Nature communications, 11(1), 1–12.

Markovi D, Gläscher J, Bossaerts P, O'Doherty J, Kiebel SJ. 2015. Modeling the evolution of beliefs using an attentional focus mechanism. PLOS Comput. Biol 11(10):e1004558 [PubMed: 26495984]

Mascaro O, & Sperber D (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. Cognition, 112(3), 367–380 [PubMed: 19540473]

McCallum R 1997. Reinforcement learning with selective perception and hidden state. PhD thesis, Univ. Rochester, Rochester, NY

Medin DL, Goldstone RL, Gentner D. 1993. Respects for similarity. Psychol. Rev 100(2):254–78

Melchers KG, Shanks DR, Lachnit H. 2008. Stimulus coding in human associative learning: flexible representations of parts and wholes. Behav. Process 77(3):413–27

Mildner J, Tamir D. 2018. The people around you are inside your head: social context shapes spontaneous thought. PsyArXiv. 10.31234/osf.io/xmzh7

Milner B (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. Archives of neurology, 9(1), 90–100

Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, et al. 2015. Human-level control through deep reinforcement learning. Nature 518(7540):529–33 [PubMed: 25719670]

Montague PR, Dayan P, Sejnowski TJ. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci 16(5):1936–47 [PubMed: 8774460]

Myers NE, Stokes MG, Nobre AC. 2017. Prioritizing information during working memory: beyond sustained internal attention. Trends Cogn. Sci 21(6):449–61 [PubMed: 28454719]

Nastase SA, Goldstein A, Hasson U. 2020. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. NeuroImage 222:117254 [PubMed: 32800992]

Navarro D 2006. From natural kinds to complex categories. In Proceedings of the 28th Annual Conference of the Cognitive Science Society 2006. Distributed by Lawrence Erlbaum Associates, Inc, Mahwah, NJ

Niv Y 2009. Reinforcement learning in the brain. J. Math. Psychol 53(3):139–54

Niv Y 2019. Learning task-state representations. Nat. Neurosci 22(10):1544–53 [PubMed: 31551597]

Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, et al. 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. J. Neurosci 35(21):8145–57 [PubMed: 26019331]

Nosofsky RM, Gluck MA, Palmeri TJ, McKinley SC, Glauthier P. 1994. Comparing modes of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). Mem. Cogn 22(3):352–69

Olson KR, Shutts K, Kinzler KD, Weisman KG. 2012. Children associate racial groups with wealth: evidence from South Africa. Child Dev. 83(6):1884–99 [PubMed: 22860510]

Orhan AE, Gupta VV, Lake BM. 2020. Self-supervised learning through the eyes of a child. arXiv:2007.16189 [cs.CV]

Park SA, Miller DS, Nili H, Ranganath C, Boorman ED. 2020. Map making: constructing, combining, and inferring on abstract cognitive maps. Neuron 107(6):1226–38.e8 [PubMed: 32702288]

Parkinson C, Du M. 2020. How does the brain infer hidden social structures? Trends Cogn. Sci 24(7):497–98 [PubMed: 32451240]

Parkinson C, Kleinbaum AM, Wheatley T. 2017. Spontaneous neural encoding of social network position. Nat. Hum. Behav 1(5):0072

Pearce JM, & Hall G (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological review, 87(6), 532 [PubMed: 7443916]

Plaks JE, Stroessner SJ, Dweck CS, Sherman JW. 2001. Person theories and attention allocation: preferences for stereotypic versus counterstereotypic information. J. Personal. Soc. Psychol 80(6):876–93

Radulescu A, Niv Y, Ballard I. 2019a. Holistic reinforcement learning: the role of structure and attention. Trends Cogn. Sci 23(4):278–92 [PubMed: 30824227]

Radulescu A, Niv Y, Daw N. 2019b. A particle filtering account of selective attention during learning. Poster presented at the 2019 Conference on Cognitive Computational Neuroscience, Berlin, Sept. 14

Radulescu A, van Opheusden B, Callaway F, Griffiths T, Hillis J. 2020. From heuristic to optimal models in naturalistic visual search. Paper presented at the Eighth International Conference on Learning Representations, April 26

Rehder B, Hoffman AB. 2005. Eyetracking and selective attention in category learning. Cogn. Psychol 51(1):1–41 [PubMed: 16039934]

Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, et al. 2010. Why copy others? Insights from the social learning strategies tournament. Science 328(5975):208–13 [PubMed: 20378813]

Roelfsema PR, van Ooyen A. 2005. Attention-gated reinforcement learning of internal representations for classification. Neural Comput. 17(10):2176–214 [PubMed: 16105222]

Rouhani N, Norman KA, Niv Y. 2018. Dissociable effects of surprising rewards on learning and memory. J. Exp. Psychol. Learn. Mem. Cogn 44(9):1430–43 [PubMed: 29553767]

Rouhani N, Norman KA, Niv Y, Bornstein AM. 2020. Reward prediction errors create event boundaries in memory. Cognition 203:104269 [PubMed: 32563083]

Roy N, Gordon G, Thrun S. 2005. Finding approximate POMDP solutions through belief compression. J. Artif. Intell. Res 23:1–40

Russell S, Norvig P. 2002. Artificial Intelligence: A Modern Approach. Upper Saddle River, NJ: Prentice Hall

Russell S, Wefald E. 1992. 'Principles of metareasoning. Artif. Intell 49(1–3):361–95

Sanborn A, Chater N. 2016. Bayesian brains without probabilities. Trends Cogn. Sci 20(12):883–93 [PubMed: 28327290]

Sanborn A, Chater N, Heller KA. 2009. Hierarchical learning of dimensional biases in human categorization. In Advances in Neural Information Processing Systems 22, ed. Bengio Y, Schuurmans D, Lafferty J, Williams C, Culotta A, pp. 727–735. San Diego, CA: NeurIPS

Sanborn A, Griffiths TL, Navarro DJ. 2010. Rational approximations to rational models: alternative algorithms for category learning. Psychol. Rev 117(4):1144–67 [PubMed: 21038975]

Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, & Botvinick MM (2013). Neural representations of events arise from temporal community structure. Nature neuroscience, 16(4), 486–492 [PubMed: 23416451]

Schneegans S, Taylor R, Bays PM. 2020. Stochastic sampling provides a unifying account of visual working memory limits. PNAS 117(34):20959–68 [PubMed: 32788373]

Schuck NW, Gaschler R, Wenke D, Heinzle J, Frensch PA, et al. 2015. Medial prefrontal cortex predicts internally driven strategy shifts. Neuron 86(1):331–40 [PubMed: 25819613]

Schultz W 2004. Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. Curr. Opin. Neurobiol 14(2):139–47 [PubMed: 15082317]

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. Science 275(5306):1593–99 [PubMed: 9054347]

Scolari M, Seidl-Rathkopf KN, Kastner S. 2015. Functions of the human frontoparietal attention network: evidence from neuroimaging. Curr. Opin. Behav. Sci 1:32–39 [PubMed: 27398396]

Seger CA. 2013. The visual corticostriatal loop through the tail of the caudate: circuitry and function. Front. Syst. Neurosci 7:104 [PubMed: 24367300]

Shadlen MN, Shohamy D. 2016. Decision making and sequential sampling from memory. Neuron 90(5):927–39 [PubMed: 27253447]

Shafto P, Eaves B, Navarro DJ, Perfors A. 2012. Epistemic trust: modeling children's reasoning about others' knowledge and intent. Dev. Sci 15(3):436–47 [PubMed: 22490183]

Shepard R 1987. Toward a universal law of generalization for psychological science. Science 237(4820):1317–23 [PubMed: 3629243]

Shepard R, Hovland C, Jenkins H. 1961. Learning and memorization of classifications. Psychol. Monogr. Gen. Appl 75(13):1–42

Sherman SJ, Castelli L, & Hamilton DL (2002). The spontaneous use of a group typology as an organizing principle in memory. Journal of Personality and Social Psychology, 82(3), 328–342. 10.1037/0022-3514.82.3.328 [PubMed: 11902620]

Shin YS, & DuBrow S (2020). Structuring memory through inference-based event segmentation. Topics in Cognitive Science.

Shin YS & Niv Y (in press). Biased evaluations emerge from inferring hidden causes. Nature Human Behavior

Sitzmann V, Zollhöfer M, Wetzstein G. 2019. Scene representation networks: continuous 3D-structure-aware neural scene representations. In Advances in Neural Information Processing Systems 32, ed. Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, pp. 1121–32. San Diego, CA: NeurIPS

Smith LB, Yu C, Pereira AF. 2011. Not your mother's view: the dynamics of toddler visual experience. Dev. Sci 14(1):9–17 [PubMed: 21159083]

Song M, Niv Y, Cai MB. 2020. Learning what is relevant for rewards via value-based serial hypothesis testing. Paper presented at the 42nd Annual Meeting of the Cognitive Science Society, July 29

Soto FA, Gershman SJ, Niv Y. 2014. Explaining compound generalization in associative and causal learning through rational principles of dimensional generalization. Psychol. Rev 121(3):526–58 [PubMed: 25090430]

Speekenbrink M 2016. A tutorial on particle filters. J. Math. Psychol 73:140–52

Steinke A, Lange F, Seer C, & Kopp B (2018). Toward a computational cognitive neuropsychology of Wisconsin card sorts: a showcase study in Parkinson's disease. Computational Brain & Behavior, 1(2), 137–150.

Sutton RS. 1988. Learning to predict by the methods of temporal differences. Mach. Learn 3(1):9–44

Sutton RS, Barto AG. 2018. Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press

Tamir DI, Thornton MA. 2018. Modeling the predictive social mind. Trends Cogn. Sci 22(3):201–12 [PubMed: 29361382]

Tavares RM, Mendelsohn A, Grossman Y, Williams CH, Shapiro M, et al. 2015. A map for social navigation in the human brain. Neuron 87(1):231–43 [PubMed: 26139376]

Todd MT, Niv Y, Cohen JD. 2009. Learning to use working memory in partially observable environments through dopaminergic reinforcement. In Advances in Neural Information Processing Systems, ed. Koller D, Schuurmans D, Bengio Y, Bottou L, pp. 1689–96. San Diego, CA: NeurIPS

Tomov MS, Dorfman HM, Gershman SJ. 2018. Neural computations underlying causal structure learning. J. Neurosci 38(32):7143–57 [PubMed: 29959234]

Ungerleider LG, Kastner S. 2000. Mechanisms of visual attention in the human cortex. Annu. Rev. Neurosci 23:315–41 [PubMed: 10845067]

van den Oord A, Li Y, Vinyals O. 2018. Representation learning with contrastive predictive coding. arXiv:1807.03748 [cs.LG]

Vélez N, Gweon H. 2020. Learning from other minds: an optimistic critique of reinforcement learning models of social learning. PsyArXiv. 10.31234/osf.io/q4bxr

Vinyals O, Babuschkin I, Czarnecki WM, Mathieu M, Dudzik A, et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature 575(7782):350–54 [PubMed: 31666705]

Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, et al. 2018. Prefrontal cortex as a meta-reinforcement learning system. Nat. Neurosci 21(6):860–68 [PubMed: 29760527]

Whittington JC, Muller TH, Mark S, Chen G, Barry C, Burgess N, & Behrens TE (2020). The Tolman-Eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. Cell, 183(5), 1249–1263 [PubMed: 33181068]

Wilson RC, Collins AGE. 2019. Ten simple rules for the computational modeling of behavioral data. eLife 8:e49547 [PubMed: 31769410]

Wilson RC, Niv Y. 2012. Inferring relevance in a changing world. Front. Hum. Neurosci 5:189 [PubMed: 22291631]

Winter L, & Uleman JS (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. Journal of personality and social psychology, 47(2), 237. [PubMed: 6481615]

Wu CM, Schulz E, Garvert MM, Meder B, Schuck NW. 2020. Similarities and differences in spatial and non-spatial cognitive maps. PLOS Comput. Biol 16(9):e1008149 [PubMed: 32903264]

SUMMARY POINTS

1. Humans learn compact representations of tasks that help them learn faster in the future.

2. Such representation learning relies on selective attention and memory to extract the most relevant features of both past and present experiences.

3. Approximate Bayesian inference provides an integrative framework for modeling how experiences are converted into task representations.

4. The ubiquity of this framework becomes clear when we consider how people use abstract world models to interpret social cues and predict others' behavior.

FUTURE ISSUES

1. How do attention and memory interact to give rise to task representations?

2. Are elemental features and configural objects represented and reinforced in parallel?

3. What are the so-called primitives that constrain inference of task representations, and are they also learned from experience?

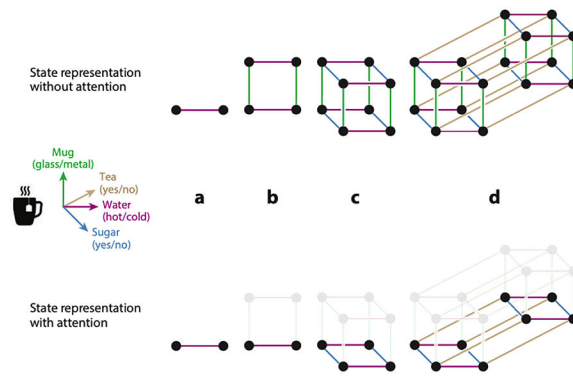4. Do inference algorithms for representation learning extend to naturalistic settings?

Figure 1.
Selective attention for state representation. Consider a representational space for the task of making tea, defined along four dimensions: Water temperature (hot/cold), Mug material (glass/metal), Sugar (yes/no), and Tea (yes/no), each involving, for simplicity, a binary feature. The top row depicts how the number of states grows as we add dimensions. Considering only one dimension (Water) results in two possible states (*a*). Two dimensions would mean four states (*b*), three dimensions eight states (*c*), and four dimensions would mean 16 unique states (*d*). This exponential increase in the number of states as the number of dimensions grows is known as the curse of dimensionality. The bottom row shows how selective attention can solve this problem: ignoring even one dimension (Mug) reduces the size of the state space by a factor of two.
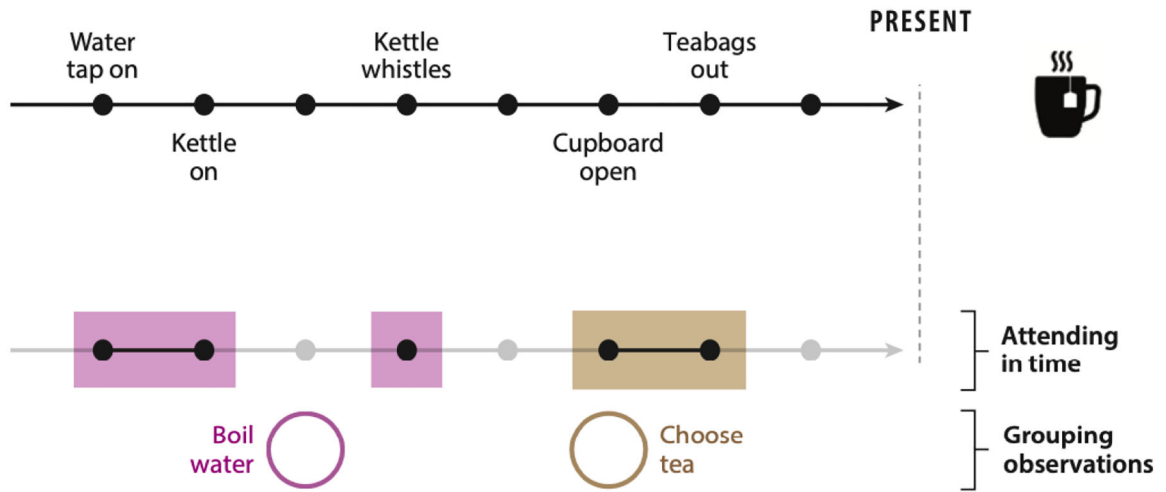
Figure 2:
Selectively remembering and organizing past observations. For any given task, time can be viewed as an axis of the representational space (Figure 1). Consider a chain of observations one could make during the task of making tea. Only some of those past observations have relevance for the task. The question of which observations to remember can thus be reframed as attention learning in time. The agent has to selectively attend not only to different aspects of present sensory data but also to different past observations that should be included in the state. Once stored in memory, observations can be grouped in such a way as to facilitate retrieval when they become relevant for the task.