Memory mechanisms predict sampling biases in sequential decision tasks

Marcelo G. Mattar (mmattar@princeton.edu)
Princeton Neuroscience Institute, Princeton University
Princeton, NJ 08544, USA

Deborah Talmi (deborah.talmi@manchester.ac.uk)
Division of Neuroscience & Experimental Psychology, University of Manchester
Manchester, UK, M139PL

Nathaniel D. Daw (ndaw@princeton.edu)
Princeton Neuroscience Institute and Department of Psychology, Princeton University
Princeton, NJ 08544, USA

Abstract

Good decisions are informed by past experience. Accordingly, models of memory encoding and retrieval can shed light on the evaluation processes underlying choice. In one classic memory model aimed at explaining biases in free recall, known as the temporal context model (TCM), a drifting temporal context serves as a cue for retrieving previously encoded items. The associations built by this model share a number of similarities to the successor representation (SR) — a particular type of world model used in reinforcement learning to capture the long-run consequences of actions. Here, we show how decision variables may be constructed by retrieval in the TCM, corresponding to drawing samples from the SR. Since the SR and TCM encode long-term sequential relationships, this provides a mechanistic, process level model for evaluating candidate actions in sequential, multi-step tasks, connecting them to the details of memory encoding and retrieval. This framework reveals three ways in which the phenomenology of memory predict novel choice biases that are counterintuitive from a decision perspective: the effects of emotion, of sequential retrieval, and of backward reactivation. The suggestion that the brain employs an efficient sampling algorithm to rapidly compute decision variables offers a normative view on decision biases, explains patterns of memory retrieval during deliberation, and may shed light on psychiatric disorders such as rumination and craving.

Keywords: decision making; reinforcement learning; successor representation; episodic memory; temporal context

Reinforcement learning, in the most general terms, requires drawing on previous experiences to evaluate candidate actions in terms of their anticipated future consequences. This suggests that what we know about how the brain encodes and retrieves memories outside the decision context can shed light on the mechanisms underlying action evaluation and, in particular, of “model-based” learning, i.e. piecing together knowledge from multiple distinct experiences to forecast the results of novel actions.

Here we present a new theoretical framework that combines and builds on two previous advances concerning the relationship between mnemonic and choice processes, offering many new insights and experimental predictions about how memory serves choices. The first line of previous work is the recent advent of episodic sampling models, which posit that the brain constructs decision variables by selectively retrieving a small number of records of the outcomes from previous individual experiences with similar actions (Plonsky, Teodorescu, & Erev, 2015; Gershman & Daw, 2017; Bornstein, Khaw, Shohamy, & Daw, 2017). These models explain a number of detailed aspects of learning behavior, but so far they have been applied only to single-step “bandit” tasks (in which a single choice is presented repeatedly and associated with an outcome, received immediately), and it has been unclear how best to extend them to the sorts of sequential decision tasks (like mazes or games like chess, in which actions occur in series with interdependent outcomes) that particularly exercise integrative memory access for forecasting.

The second foundation for our model is the recent discovery of a relationship between standard memory and choice models; in particular, the temporal context model (TCM) characterizes standard memory experiments like free recall by positing a set of seemingly incidental associations between studied items (by means of a slowly temporal drifting context) due to their temporal proximity (Howard & Kahana, 2002). A simplified form of these associations has recently been shown to coincide with the successor representation (SR), a particular type of world model used in reinforcement learning to capture the long-run consequences of actions (Gershman, Moore, Todd, Norman, & Sederberg, 2012; Dayan, 1993). This representation is useful in planning and choice and has been argued to explain a number of features of human reinforcement learning behavior (Momennejad et al., 2017). The equivalence with the SR is highly suggestive about an adaptive purpose for the encoding phase of the TCM in constructing representations to guide choice, but so far no research has actually delivered on this promise by showing how the retrieval of these learned associations (about which TCM also provide a highly detailed account constrained by extensive experimental data) would actually be useful in constructing decision variables.

Accordingly, in the present work, we demonstrate a relationship between the *retrieval* of memories in temporal context models and the construction of decision variables. In short, we consider the retrieval phase of a simplified version of the TCM to show that retrieved items correspond to samples from the long-run future consequences of a candidate action, drawn from the learned SR. The temporally abstracted form of the SR itself serves to “flatten” the tree-like set of future situations in a sequential task to a set of individual future states, rendering the situation more like bandit problems studied previously and solving the problem of extending sampling models to the sequential case. This skirts issues like depth- vs breadth-first rollouts, goal selection, and pruning, while suggesting a different account for some of the same experimental phenomena that have been previously interpreted in these terms (Cushman & Morris, 2015; Huys et al., 2015). Analogous to episodic sampling models in the bandit domain, this new account predicts the statistics of human choices — but in arbitrary, sequential tasks — as reflecting a small sample from potential outcomes, but (unlike other reinforcement learning models) not necessarily the most recently experienced (Plonsky et al., 2015) or the most imminently expected.

By considering memory retrieval as a dynamic process, as embodied by the TCM in list learning experiments, the current framework also highlights several particulars of human memory that each diverge from a straightforward SR-based sampling model that would be most directly expected on reinforcement learning grounds. Each of these phenomena predicts novel choice biases yet to be tested, and collectively can be understood as promoting value estimators that are biased, but favorable in the small-sample regime.

1. It is well known that emotion tends to enhance memory; for instance, emotionally salient pictures are often preferentially retrieved. Recent work has shown that the detailed pattern of these effects (such as their dependence on the local context of study and the test) can be understood, in temporal context models, as reflecting an elevated learning rate in the initial encoding of items to context (Talmi, Lohnas, & Daw, 2017). In reinforcement learning terms, this corresponds to over-representing such items — particularly high- or low-reward states — in the SR. Extending a recent suggestion from Lieder, Griffiths, and Hsu (2018) to the sequential case, this would tend to focus sampling on rare, relevant outcomes so as to produce a favorable bias-variance trade-off for small-sample evaluations. A similarly biased retrieval could also help explain subject’s preference for attended stimuli on choice (Salomon et al., 2018).

2. The central insight of temporal context models is to capture sequential effects in retrieval (a tendency to successively retrieve items studied nearby in time; the contiguity effect), which in the model results from the learned item-context associations. In the context of reinforcement learning, this means that in contrast to standard sampling models (which draw i.i.d. from the target distribution), retrieved successor states will be (tunably) biased to occur along a path, e.g. each subsequent draw biased toward the successors of the previous draw. (Because they are drawn from a SR, unlike a standard depth-first rollout from a one-step model, these draws can skip over multiple steps at once.) Analogous to eligibility traces in temporal difference learning, we suggest this sequential biasing effect would again serve to reduce variance in the estimation of long-run returns.

3. A key feature of temporal contiguity effects in human memory experiments is that they extend not only forwards but also backwards, i.e. subjects tend to retrieve items studied immediately before as well as immediately after the previously retrieved item (though more often the latter). This is actually not the case for the SR — which by definition represents strictly the *future* consequences of a state or action directionally; and accordingly this aspect of the TCM was simplified away in the derivation of the relationship between TCM and SR by Gershman et al. (2012). We argue that restoring this key feature of the model produces a representation that diverges from the SR but in so doing corrects one of its key deficiencies. In particular, the SR is policy dependent — it predicts the consequences of a new action assuming that, following it, the agent makes choices according to the preferences under which the SR was originally learned. In both machine learning (Lehnert, Tellex, & Littman, 2017) and neuroscience applications (Russek, Momennejad, Botvinick, Gershman, & Daw, 2017), it has been pointed out that this feature results in inflexibility in transfer learning scenarios (i.e., where the agent must draw on its old memories to plan actions in a changed setting) and that a better blend of learning and transfer performance can be obtained by replacing the on-policy SR with one that blends a certain amount of random behavior (Stachenfeld, Botvinick, & Gershman, 2017). In many tasks (e.g., those based in Euclidean space, which corresponds to an undirected graph), this amounts to regularizing a directional policy to include the possibility of backtracking. The current framework thus suggests a quantitative relationship between backwards encoding and retrieval; action evaluations that are biased toward an exploratory policy; and transfer learning performance.

Our theoretical framework establishes a formal relationship between models of episodic retrieval and decision making and has potential impacts in both AI and cognitive neuroscience fields. On the AI front, the relationship suggests a family of sampling algorithms for action evaluation that may be more effective for transfer learning than traditional methods, in particular during the early phases of learning. For cognitive neuroscience, the suggestion that the brain employs an efficient sampling algorithm to rapidly compute decision variables offers a normative view on decision biases, explains patterns of memory retrieval during deliberation, and may shed light on psychiatric disorders such as rumination and craving.

Acknowledgments

We acknowledge support from NIDA through grant R01DA038891, part of the CRCNS program, and Google DeepMind. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

References

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*, 15958.

Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences*, *112*(45), 13817–13822.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, *5*(4), 613–624.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual review of psychology*, *68*, 101–128.

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A., & Sederberg, P. B. (2012). The successor representation and temporal context. *Neural Computation*, *24*(6), 1553–1568.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., . . . Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*(10), 3098–3103.

Lehnert, L., Tellex, S., & Littman, M. L. (2017). Advantages and limitations of using successor features for transfer in reinforcement learning. *arXiv preprint arXiv:1708.00102*.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological review*, *125*(1), 1.

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680.

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological review*, *122*(4), 621.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLOS Computational Biology*, *13*(9), e1005768.

Salomon, T., Botvinik-Nezer, R., Gutentag, T., Gera, R., Iwanir, R., Tamir, M., & Schonberg, T. (2018). The cue-approach task as a general mechanism for long-term non-reinforced behavioral change. *Scientific reports*, *8*(1), 3614.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature neuroscience*, *20*(11), 1643.

Talmi, D., Lohnas, L., & Daw, N. (2017). A retrieved context model of the emotional modulation of memory. *bioRxiv*, 175653.