Temporally extended decision-making through episodic sampling

Corey Zhou
Department of Cognitive Science
University of California, San Diego
yiz329@ucsd.edu

Deborah Talmi
Department of Psychology
University of Cambridge
dt492@cam.ac.uk

Nathaniel D. Daw
Princeton Neuroscience Institute
Princeton University
ndaw@princeton.edu

Marcelo G. Mattar
Department of Psychology
New York University
marcelo.mattar@nyu.edu

Abstract

A major goal of cognitive science is to characterize how an individual's past experiences guide their present decisions in a sequential task. Various empirical evidence support a process of incremental learning, well-characterized by the framework of reinforcement learning, whereby repeated exposures to similar situations shape decisions. However, in a complex world with sparse data a more sample-efficient process is needed. Prior work has suggested that episodic memory supports decision-making in such settings. Here, we provide novel behavioral evidence that episodic memory supports decision-making in temporally extended settings. We propose that value-based decision-making and episodic memory share common mechanisms to encode and retrieve past events, which in turn shape option evaluation and ultimately choice. In two experiments, we empirically test hypotheses that relate classic dynamics of sequential episodic memory retrieval to response patterns in novel evaluation and decision tasks. We find subjects' reported value estimates are subject to biases analogous to classic episodic memory biases (Experiment 1), and their choices are best captured by an episodic recall-based model (Experiment 2). These results suggest a novel link between value-based decision-making and episodic memory, which could reflect a psychologically plausible mechanism for computing decision variables by Monte Carlo sampling.

Keywords: decision-making; episodic memory; episodic sampling; temporal context; value estimation

Introduction

Past experiences guide behavior by informing the value of different options. We may buy coffee from one store rather than another because of its consistently better quality. We may also pack an extra jacket for a weekend trip because it once snowed out of nowhere. These situations highlight how our memories routinely inform our decisions. Keeping a record of specific actions and their causal link to specific outcomes helps us predict future events and obtain more rewarding experiences, a problem formalized as reinforcement learning (RL). Here, we examine two situations in which episodic memory subserves RL and decision making.

Research in biological RL has described various ways our memory systems may enable different forms of RL. A procedural memory system involving dopamine and prediction errors is most closely linked to the slow and incremental formation of stimulus-response habits, like the choice of your regular coffee shop. A semantic memory system, subserved by cortical learning mechanisms, is alternatively linked to the formation of cognitive maps or models thought to support the deliberative evaluation of candidate actions in goal-directed behavior. Yet, the relatively slow and incremental forms of learning they describe cannot account for the fast and one-shot decisions, such as deciding to pack an extra jacket because it snowed a single time in the past. A promising alternative lies in another extensively studied memory system: episodic memory – autobiographical memory that links different aspects of *individual* events.

Human episodic memory exhibits distinct features during sequential retrieval: we are more likely to recall items observed toward either end of a sequence of events than those observed in the middle (i.e., primacy and recency effects), as well as events close to each other in time (i.e., temporal contiguity effect; see Howard & Kahana, 1999). While the involvement of episodic memory in decision-making has been posited on empirical and theoretical grounds, previous empirical work has, so far, focused only on one-step tasks (Bornstein et al., 2017; Rouhani et al., 2018; Nicholas et al., 2022). As such, they offer a restricted perspective on episodic sampling, while the sequential realm (which we argue is where episodic memory should be most relevant and effective) remains untapped. Yet, if episodic memory also informs decision-making in temporally extended settings, the retrieval process should leave footprints on our choices. Thus, we set out to provide empirical evidence of these footprints.

Formally, we hypothesize that all else equal, events from either end of an episode have a larger weight on memory-based decisions (1A), but such effect may be modulated using temporal contiguity effect (1B). Just like retrieval of temporally adjacent memories is easier than distant events, evaluation of options composed of temporally segregated events should be more challenging (2). These hypotheses bear two important deviations from conventional RL accounts: first, rather than decisions reflecting (incremental averages over) the full trajectory, only samples are used. This is motivated by the observation that even when data is plentiful, people's decisions can still depend on only a handful of individual experiences (Plonsky et al., 2015). In contrast, conventional RL algorithms consider the full trajectory (episode) in evaluating actions and/or state values. Second, the sampling process is psychologically plausible by taking advantage of a shared mechanism with episodic retrieval, unlike previous models that used a stylized memory store (Lengyel & Dayan, 2007; Gershman & Daw, 2017).

To formalize the computational details of this decision-by-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

sampling account, we leverage the temporal context model (TCM; Howard & Kahana, 2002) – first in an evaluation task (Experiment 1), and then in a decision task (Experiment 2). TCM captures all the memory biases in our hypothesis via learning of appropriate associative matrices and retrieval using an abstract, evolving representation of recently experience items. The learned associations closely corresponds to a sampling distribution, allowing past events to be drawn in a manner consistent with episodic retrieval to compute decision variables (Mattar et al., 2019; Zhou et al., 2023).

Our current goal is to empirically test the hypotheses above using two novel experimental paradigms. In Experiment 1, we elicit value estimates (an explicit decision variable and a precursor to choices) from subjects and show they manifest biases outlined in hypotheses 1A and 1B. In Experiment 2, subjects engage in sequential decision-making, whose behavior, as we show, is best predicted by an episodic sampling account. Together, our studies provide novel evidence for a decision-by-sampling mechanism in humans that is subserved by episodic memory.

Experiment 1

We test whether memory-based evaluation weighs items differently based on their absolute and relative temporal positions. To encourage the use of episodic memory, we design a task where the evaluation goal is unknown during encoding.

Methods

Design and Materials The key manipulation involves ad-hoc categories, which are not well established in memory during encoding (Barsalou, 1983) and thus affect memory performance less. Specifically, during each test trial, subjects studied a list of items and were then told to estimate the total value of a "partial" list, a subset of the encoded items that belonged to a category revealed after the full list had been presented (Fig. 1a) Categories were not exclusively colors and also may include things like "packaged items" or "spherical items". All study lists were constructed such that partial list items were spread across serial positions: the first partial list item appeared either as the first or second item, while the last partial list item appeared as one of the last two items.

66 colored images of common grocery store items were collected from various online sources. For each participant, 45 unique items were randomly selected and grouped into 5 study lists. The first list contained 5 items and was used as the practice trial. The rest contained 10 items each. Each 10-item full list corresponded to a 5-item partial list (each corresponding to a random ad-hoc category), and the 5-item full list corresponded to a 3-item partial list. No instruction ever hinted at the size of the partial lists. All items were shown with an integer value between 1 and 12, which was pre-determined according to the national average price.

Two pairs of value estimation and free recall tasks followed list studying. The first pair concerned the partial list and the second was about the full list. The value estimation task prompted the subject to estimate the total price of the corresponding list. Subsequently, the recall task asked for the names of the list items.

A cued recall test was additionally administered during the last trial, where subjects need to recall an object from the list based on a written description (not shown in Fig. 1a). Unbeknownst to the participants, the answer was always the third partial list item. The description was not specific enough to infer the answer from common sense, but could be uniquely determined given the presented items.

Procedure The experiment consisted of one (1) practice trial and four (4) test trials. Feedback was only provided at the end of the practice trial. Subjects had to pass a short quiz with 100% to ensure good understanding of the experiment procedure before moving onto the test trials. During each test trial, an interval of length uniformly drawn between 800ms and 1200ms was inserted between two item presentations.

For the first three test trials, after a 500ms interval following list presentation, subjects completed the two pairs of value estimation and free recall tasks in the order described in the previous section. Each value estimation task had a time limit of 60 seconds. The two free recall tasks had a time limit of 30 seconds (partial list) and 40 seconds (full list) respectively.

On the last test trial, subjects completed a cued recall test before proceeding with the value estimation and free recall tasks. This test was not timed.

Participants 200 subjects were recruited through Prolific, 66 of which were excluded from analysis due to excessive low effort responses (e.g., zero recall for >50% of trials) or responses that indicate gross over- or underestimation of values (i.e., N = 134). Specifically, overestimation is defined as an estimate larger than the maximum possible item value times the number of recalled items, and underestimation is defined as an estimate smaller than the minimum possible item value times the number of recalled items. Neither criterion depends on how accurately individual values were remembered.

Results

Serial Order Effects Free recall of episodic memory exhibits primacy and recency effects. Here, consistent with the classic findings in the free recall literature, subjects recalled more items more often from either end of the partial list (Fig. 1b; position 1 vs. position 2: $t(133) = 3.28, p < 0.001$; 1 vs. 3: $t(133) = 2.97, p = 0.003$; 1 vs. 4: $t(133) = 3.73, p < 0.001$; 5 vs. 2: $t(133) = 3.74, p < 0.001$; 5 vs. 3: $t(133) = 3.46, p < 0.001$; 5 vs. 4: $t(133) = 4.21, p < 0.001$). Thus, these results suggest primacy and recency are also observed in partial list free recall when the retrieval criteria (ad-hoc categories) are unknown during encoding.

Value Estimation To quantify the effect of serial list position and episodic memory retrieval on value estimation, we fit a mixed-effect regression model of the form on data from the first three blocks

$$\mathbf{V}_{reported} = \mathbf{V}\beta + \mathbf{Z}\mathbf{u} + \varepsilon. \quad (1)$$

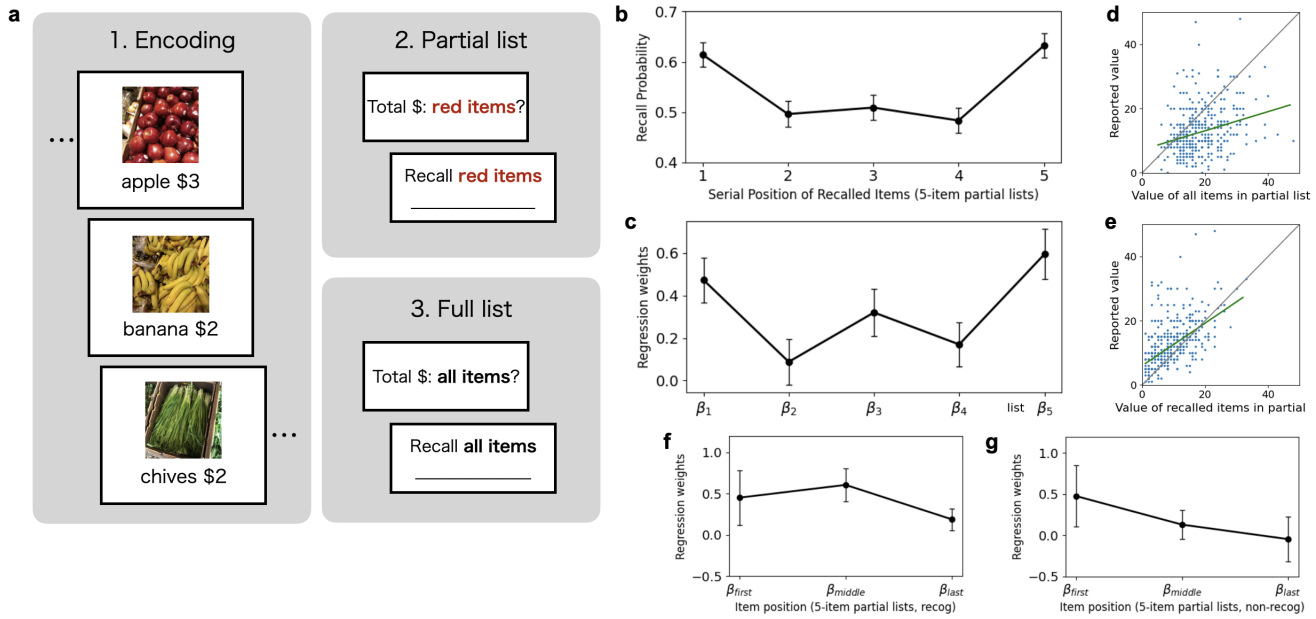

Figure 1: **(a)** Experiment procedure for an example trial (test trials have 10 items). **(b)** Average probability of recall as a function of an item's position in the 5-item partial list. **(c)** Fitted fixed effect regression weights on subjects' reported partial list value. β_i's corresponds to position i in the 5-item partial list. **(d)** Reported partial list value against the true partial list value. **(e)** Reported partial list value against the true value of recalled partial list item(s). Each dot represents one list from one subject. i.e. a subject may contribute to multiple data points. **(f)** Fitted mixed effect regression weights on reported partial list value by subjects who correctly answered the cued recall test. β_{first} corresponds to the first item, β_{last} corresponds to the last item, and β_{middle} corresponds to the collection of middle items. **(g)** As in (f) for subjects who answered the cued recall test incorrectly.

The fixed effects $\beta = (\beta_0, \beta_1, \ldots, \beta_m, \beta_n)'$ represent the average effect of different partial-list positions on the overall value estimate, and the random effect $\mathbf{u}$ captures effects specific to individual subjects. $\mathbf{V}_{reported}$ is a vector of reported partial list values with length L equal to the total number of trials. Each row of the design matrix $\mathbf{V}$ contains the individual item prices in a partial list, plus the total number of partial list items recalled. $\mathbf{Z}$ contains subject ids. The inclusion of the number of recalls helps to explain an additional 12.2% fixed effect variance and 5.0% overall model variance. A model comparison also shows that including the number of recalled partial list items significantly improves the model ($\chi^2_1 = 57.4, p < 0.001$).

The first and the last items have a significantly higher weight on the value estimate compared to the three middle items (β_{first} vs. β_{middle}: $F(1,133) = 5.24, p = 0.023$; β_{last} vs. β_4: $F(1,133) = 10.24, p = 0.001$). Comparison of each pair of partial list positions reveals a similar trend, where the first and last items receive a larger weight in general (Fig. 1c; β_1 vs. β_2: $F(1,133) = 6.27, p = 0.013$; β_1 vs. β_4: $F(1,133) = 3.67, p = 0.057$; β_5 vs. β_2: $F(1,133) = 10.1, p = 0.002$; β_5 vs. β_4: $F(1,133) = 6.98, p = 0.009$), consistent with subjects' serial recall patterns.

The parallel between the recall probability and regression weights in this task agrees with our hypothesis. To further test whether episodic memory recall predicts evaluation, we regress subjects' reported partial list value on two quantities separately using Huber loss: (i) the true partial list value, and (ii) the true total value of only the recalled partial list items. The reported partial list values turn out to strongly correlate with the true value of recalled items (Fig. 1e; $\beta_r = 0.640, R^2 = 0.383$), more so than the true partial list value v_t (Fig. 1d; $\beta_t = 0.291, R^2 = 0.103$). Note the latter corresponds to an alternative account that people keep track of an aggregated statistic (i.e., independent of the specific items they can remember), akin to a model-free reinforcement learning agent. Regressing the reported values on (i) and (ii) together also show that (ii) was more predictive than (i).

Temporal Contiguity The cued recall test was inserted in the last trial in an attempt to update the temporal context of the subject before recall and change subsequent recall dynamics. Because the correct answer is always the item in position 3 of the partial list, the temporal contiguity effect would predict an enhanced probability of recalling the items at position 2 and/or 4, while attenuating both the primacy effect and the recency effect (i.e., worse recall accuracy of item 1 and 5).

Subjects who correctly recognized the cued item show attenuated primacy and recency effects (Fig. 1f). On average, the middle items have an increased influence on the value estimate compared to the previous trials, as no pairs of the regression weights are significantly different. Those who failed to recognize the cued item do not show any significant primacy

or recency effect either (Fig. 1g), likely because the (failed) memory retrieval also changed their internal context to some extent. Thus the results suggest a temporal contiguity effect even in subjects who did not respond correctly to the cue.

Discussion

Results from Experiment 1 suggest people rely on episodic memory to adaptively compute values in our task, supporting an account that memory plays an important role in general decision-making. Participants memorized lists of everyday grocery items. They were then asked about the total price of a subset of items based on an ad-hoc category – a task meant to mimic the computation of a decision variable – and finally, recalled these items. Regression analysis shows that the first and last items in the subset weighs more heavily in value estimates, suggesting primacy and recency effects were at play. Furthermore, when an additional cued recall test was introduced to shift subjects' temporal context to the middle of the list, previous signs of primacy and recency effects are reduced, suggesting temporal contiguity at play. Together, these results are consistent with our hypothesis, providing novel evidence for a psychologically plausible mechanism to compute decision variables using episodic memory samples.

Experiment 2

Building onto the link between episodic retrieval and evaluation in Experiment 1, we next seek to establish a more direct connection between episodic retrieval and decisions: whether people's choice can be well predicted by what their average recall rate at decision time. Again, to encourage the use of episodic memory, decision-relevant information is unknown during encoding. Additionally, we take advantage of the temporal contiguity effect to manipulate subjects' recall rate.

Methods

Design and Materials The temporal contiguity effect suggests that items encoded close together in time are more likely to be retrieved together, and more so in the same order as presented (Howard & Kahana, 1999). This implies that discontiguously encoded items - that is, items presented at relatively distant temporal steps - would have a worse average recall rate. At the same time, the spatial distance between items should be controlled to minimize the spatial contiguity effect (Miller, Lazarus, Polyn, & Kahana, 2013). We hence designed a gridworld which subjects explored in multiple runs. They then decided between two subsets of items that were encoded with different levels of temporal contiguity.

123 black-and-white cartoon battle items were collected from an online resource (`https://game-icons.net/`). For each participant, 108 items were randomly selected to form 9 gridworlds with 12 unique items each, all of which needed to be fully explored (Fig. 2a). The items were hidden behind the grey squares. To fully explore a gridworld, they took four zigzag routes composed of gray squares by starting from the top-center location (marked by $\star$) every time and pressing either the left or the right arrow key to move to an (diagonally) adjacent grid (e.g., $\star \to \square \to A \to B \to C \to D$). Each zigzag path therefore contained 4 items. Black squares were inaccessible and the steps were irreversible. The top row did not contain any item and was solely for navigation purposes.

The image of each item was only shown once when the participant first navigated to its location by pressing on a keyboard. Each item had two attributes - one attack value and one defense value, both of which were shown along the image. All future visits to the location showed a black cross, and the subject could visit a grid twice only if doing so was the only way to access another novel item. For example, to access item F after first taking the path A(item) $\to$ B(item) $\to$ C(item) $\to$ D(item), the participant may take the path A(X) $\to$ E(item) $\to$ C(X) $\to$ F(item). None of the locations on the right half may be accessed thereafter.

A decision task appeared after exploration, where the participant chose between two "paths" to maximize either the total attack or total defense points. They were not told which attribute was relevant until this point. Both paths in the decision task were taken from the four zigzag paths the subject actually took and were paired so they did not overlap with one another. Crucially, exactly one of them was encoded *contiguously*, meaning the subject saw all the items on the path without encountering black crosses and the temporal distance between spatially adjacent items is one (e.g., ABCD in the example above), while the other was encoded *discontiguously*, such that two of the locations were crossed out at exploration and the temporal distance between spatially adjacent items may be three or more (e.g., AECF). We refer to the two types of paths as "full" and "partial" paths respectively.

Two recall tasks then followed the decision, asking the participant to write down names of the items in each path option.

Procedure The experiment consisted of three (3) practice trials and nine (9) test trials. Participants first familiarized themselves with the keyboard control to navigate through the maze efficiently during practice, and completed a trial similar to the test trials except with a smaller gridworld (6 hidden items). They had to pass a short quiz with 100% correctness to ensure good understanding of the experiment procedure before moving onto the test trials.

During each test trial, participants were given 75 seconds to uncover all hidden items. Each item was shown no more than once for 4 seconds when the navigation reached its location for the first time. If a gridworld was not fully explored, the current trial would be skipped and the participant was warned that the trial had timed out.

All decision and recall tasks had a time limit of 60 seconds. The specific attribute queried was random for each trial, and was not revealed until the decision task. Path highlight colors were random and did not indicate full/partial types. The trial was excluded if no choice was made.

Feedback was provided at the end of each trial after the last recall task. Participants could view the full gridworld with items shown, as well as the values of the path options.

Participants 100 subjects were recruited through SONA, 16 of which were excluded from analysis due to excessive low effort responses or responses that indicate note-taking (e.g., perfect recall with exact order as presented in all trials), resulting in 733 trials with at least one recall.

Results

Temporal Contiguity The key manipulation of this study is the construction of partial paths: items along the partial paths are experienced in a different order than full paths – specifically, the second item and the fourth item were observed one *temporal* step apart, even though they are two *spatial* steps apart. Since the temporal contiguity effect implies that temporally adjacent items are more likely to be retrieved in succession, we hypothesize that two-step transitions (i.e., relative lag = ± 2, where lag is defined as the number of key presses) in free recall are more probable in the partial path condition.

Indeed, free recall of partial path items shows a clear disruption of the temporal contiguity effect when relative distance between items is defined spatially as opposed to temporally (Fig. 2b). While subjects still tend to recall in the forward direction (i.e., recalling items in the order they were observed), the recalled item that immediately follows a previous recall is as likely to be from one (spatial) step ahead (+1) as two (spatial) steps ahead (+2; $t(428) = -0.095$, $p = 0.92$). In contrast, serial recalls of full path items display classic temporal contiguity, such that subsequent recalls are much more likely to come from one step ahead.

Serial recall Temporal contiguity effect indicates that episodic retrieval of items is most likely to ensue in close temporal proximity with respect to the order that items are experienced and encoded. It is less likely for people to subsequently recall an item seen a few temporal steps away than something that directly follows a just recalled item during exploration. Since the four items in each partial path are experienced in two separate runs, we expect more difficulty in retrieving all four items and thus lower recall accuracy.

On average, subjects recalled 1.94 items from a queried path, regardless of the path type (full/partial). Across trials, an average of 2 items were recalled from each full path (s.e.m. = 0.05), while an average of 1.88 items were recalled from each partial path (s.e.m. = 0.05), which is significantly lower ($t(732) = 2.87$; $p = 0.004$) and is consistent with our hypothesis (Fig. 2c). There was no difference in the number of recall intrusions between full and partial paths ($t(732) = -1.32$; $p = 0.19$, suggesting that the reduced recall accuracy is unlikely due to incorrectly recalling items from the full path overlapping with the queried partial path.

Choice Prediction We first inspected subjects' performance on the test trials. Despite the perceived task difficulty, they made the optimal choice on 69.09% of the trials. On the subject level, choice accuracy is significantly better than chance ($t(83) = 10.43$, $p < 0.001$), with no obvious prior bias towards either option ($t(83) = 1.35$, $p = 0.180$).

To test the hypothesis that episodic recalls are better predictors of people's adaptive choice, we compare different computational accounts of the decision process to see which one predicts actual trial-wise choices more accurately. Specifically, we consider three candidate models:

1. Model free (MF) - the agent accumulates values for both attack and defense along each path. This gives perfect value estimation for the full paths. Since individual item values are similarly distributed regardless of the grid location ($F(3,729) = 1.46$; $p = 0.22$), the accumulated value of partial paths is in expectation half of its actual value. Thus the agent estimates the partial path value as twice the accumulated number, as it has no information about the individual items. It chooses greedily based on the two value estimates and has no free parameters.
2. Perfect memory (PM) - the agent is assumed to remember perfectly the individual items and their attributes. It always chooses optimally by adding up the individual values and greedily picking the option with the higher value. It has no free parameters.
3. Recall-based (RB) - the agent draws episodic samples of individual items and adds up the attack/defense values of sampled items as needed. We assume it encodes and retrieves values perfectly. Additionally, it uses a single value as the expected value for any item it fails to recall. Since the total number of items within each option is obvious (unlike Experiment 1), this quantity acts as a reasonable placeholder to make up for anything forgotten. For instance, if it recalls two out of four items with attack value 1 and 4 respectively, with $\mathbb{E}[\text{value of unrecalled item}] = 2$, the path value estimate is 1 + 4 + 2 * 2 = 9. It uses a greedy decision policy and has one free parameter $\mathbb{E}[\text{value of unrecalled item}]$.

To predict choices, we fit an RB model to each subject – that is, each RB model shares the same subject-level recall probabilities as the corresponding subject (e.g., if the subject recalls the first item of a full path 70% of the time, the model will successfully sample the value of the first full item 70% of the time; see Fig. 2d for group level probabilities), and a subject-specific value filler (i.e., $\mathbb{E}[\text{value of unrecalled item}]$) is fitted to maximize choice prediction accuracy. Value fillers are found using a grid search with a grid size of 0.1.

Among the three candidate models, the recall-based model predicts subjects' actual choices most accurately, with a trial-wise average of 75.85% (Fig. 2e), much better than both the model-free model (57.91%) and the perfect memory account (61.95%). All choice prediction accuracies are evaluated using leave-one-out cross validation. We have tried ε-greedy and softmax policies as well, but the additional parameters (ε, temperature) of the best-fitting models indicated near-random choice, so we leave them out of the current analysis.

Discussion

Experiment 2 offers additional insight into how episodic retrieval guides choice when value computation is delayed till decision time. In a novel task, participants explored grid-

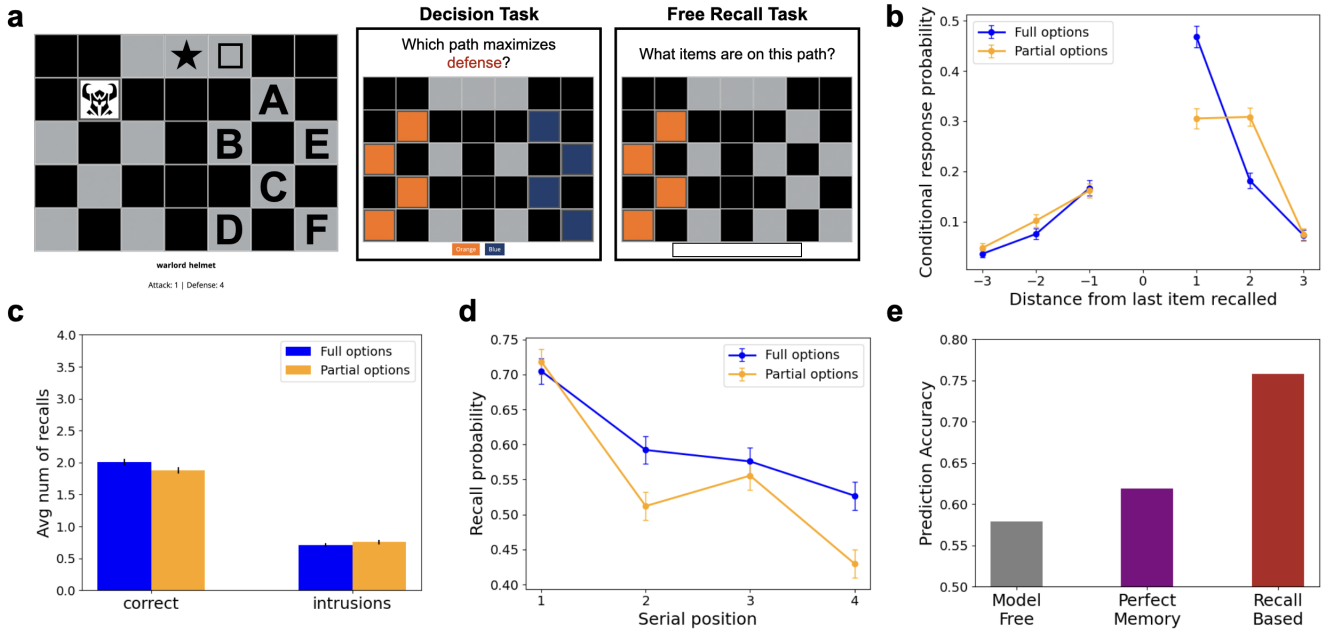

Figure 2: **(a)** Example gridworld and tasks during a trial. **(b)** Probability of successively recalled items as a function of their relative spatial distance (number of key presses). Results are averaged across trials, and error bars indicate standard errors. **(c)** Trial-wise verage number of correctly recalled items and intrusions. **(d)** Trial-wise overall recall probability of an item as a function of its serial position along the queried path. **(e)** Trial-wise prediction accuracy of the three candidate models (see text for details) with respect to subjects' actual choices.

worlds to encode game items. They then chose between two subsets of items encoded with different levels of temporal contiguity based on an attribute selected randomly ad hoc. Finally, they performed free recall of the items within each subset. A comparison of people's responses with different decision models reveals that the recall-based account best predicts actual choice, followed by a perfect memory model and a model-free account. This suggests that subjects' choices are most accurately captured by what they recall at decision time (plus a placeholder for items that they *know* they cannot recall). In contrast, the model-free strategy does not retain any item-specific information to be recalled, and explains human decisions the poorest, even though it solves the task.

It is worth noting that the disagreement between the perfect memory account (which always makes optimal decision) and the recall-based model highlights the way memory biases decision – people rely on episodic retrieval to compute action values, so when fewer episodic samples are drawn successfully (e.g., partial paths), their decisions are *predictably* suboptimal. This finding supports the decision-by-sampling hypothesis, suggesting that people's choices are best predicted by their average recall rates at decision time.

General Discussion

Through a series of experiments, we find that both value estimates (Experiment 1) and adaptive choices (Experiment 2) show footprints of episodic memory biases and can be predicted from episodic recall patterns. Consistent with our hypothesis, memory-based decisions weigh events differently based on their serial position analogous to the primacy and recency effects in episodic memory. Such effect can be modulated by the temporal contiguity effect. Moreover, people's choices are better predicted by their average recall pattern than aggregated statistics or a purely model-based strategy. The results suggest that an episodic sampling mechanism underlies adaptive decision-making in humans, such that encoded information may be integrated for decisions ad hoc.

Two features distinguish our paradigms in comparison to previous recall experiments. First, we extend classic word list learning tasks to investigate the role of episodic retrieval in evaluation and action. Second, we ensure that encoding strategies could not influence the answer so as to minimize its confounding effect. By design, our paradigm forces subjects to retrieve information and compute decision variables only after encoding has been completed.

The rich connection between experiences and choice has attracted interest from various subfields of cognitive science; here, we highlight an under-explored account of decision by episodic sampling using two novel behavioral tasks that suggest recruitment of episodic memory in evaluation and choice. These findings provide empirical support for computational accounts that combine sample-based decision-making with episodic retrival models such as TCM, which we hope to expand in future work.

References

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211-227.

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*, 15958. doi: 10.1038/ncomms15958

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.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *25*, 923-941.

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

Lengyel, M., & Dayan, P. (2007). Hippocampal contributions to control: The third way. In *NeurIPS'20.*

Mattar, M. G., Talmi, D., & Daw, N. D. (2019). Memory mechanisms predict sampling biases in sequential decision tasks. In *The 4th Multi-disciplinary Conference on Reinforcement Learning and Decision Making.*

Miller, J., Lazarus, E., Polyn, S., & Kahana, M. (2013). Spatial clustering during memory search. *Journal of experimental psychology: Learning, Memory, and Cognition*, *39*(3), 773–781. doi: 10.1037/a0029684

Nicholas, J., Daw, N. D., & Shohamy, D. (2022). Uncertainty alters the balance between incremental learning and episodic memory. *Elife*, *11*, e81679.

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

Rouhani, N., Norman, K., & Niv, Y. (2018, 03). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*. doi: 10.1037/xlm0000518

Zhou, C. Y., Talmi, D., Daw, N. D., & Mattar, M. G. (2023). *Episodic retrieval for model-based evaluation in sequential decision tasks.*