A Variational Neural Network Model of Resource-Rational Reward Encoding in Human Planning

Zhuojun Ying (z5ying@ucsd.edu)
Department of Cognitive Science, University of California, San Diego

Frederick Callaway (fredcallaway@nyu.edu)
Department of Psychology, New York University

Roy Fox (royf@uci.edu)
Department of Computer Science, University of California, Irvine

Anastasia Kiyonaga (akiyonaga@ucsd.edu)
Department of Cognitive Science, University of California, San Diego

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

Abstract

Working memory (WM) is essential for planning and decision-making, enabling us to temporarily store and manipulate information about potential future actions and their outcomes. Existing research on WM, however, has primarily considered contexts where stimuli are presented simultaneously and encoded independently. It thus remains unclear how WM dynamically manages information about reward and value during planning, when actions are evaluated sequentially in time and their cumulative values must be integrated to guide choice. To address this gap, we developed an information-theoretic model of WM allocation during planning, implemented using variational recurrent neural networks. In this model, an agent optimizes plan quality while maintaining reward information under WM constraints. To test our model, we designed a task in which participants sequentially observed the rewards available at different future states before executing a sequence of actions, attempting to maximize cumulative rewards. Our results suggest that humans preferentially maintain rewards that are most informative for plan selection, integrating both local and global factors. These findings bridge theories of WM limitations with models of human planning, revealing how cognitive constraints shape decision-making strategies.

Keywords: working memory; planning; information theory; reinforcement learning; variational autoencoder

Introduction

Planning is fundamental to human behavior, requiring us to evaluate potential outcomes of different action sequences to make optimal decisions. Consider planning a road trip: to decide on a route, we might assess each potential route by evaluating features like the cost, travel time, and scenery at each stop, combining these assessments into an overall judgment. While computational models of human planning have provided valuable insights into decision-making processes, they typically assume that people can perfectly encode and maintain information about each evaluated option indefinitely. In reality, planning relies critically on working memory (WM) to temporarily store and manipulate information about different alternatives, and WM's limited capacity constrains how much information can be maintained. This fundamental tension between the computational demands of planning and the limitations of human memory shapes how people approach complex decisions, yet we still lack a comprehensive understanding of how people optimize planning strategies under these constraints.

To understand how working memory supports planning, we must first acknowledge how planning differs from the tasks typically studied in working memory research. The field of working memory has extensively characterized how individuals distribute limited cognitive resources when remembering multiple items (Vogel et al., 2001; Awh et al., 2007; Chunharas et al., 2022). However, these works primarily focus on contexts where people must remember independent pieces of information, such as lists of words or visual features. In these classical WM paradigms, the relevance of each item to the task is independent of other items — remembering one word in a list, for instance, has no bearing on the importance of remembering other words. However, planning presents a fundamentally different challenge: the relevance of any single piece of information depends on its relationship to other encoded information. For example, when planning a road trip, the value of remembering the scenic quality of one stop depends entirely on how it contributes to the overall attractiveness of that route relative to alternatives. This interdependence of information value is not captured by existing models of working memory.

A second critical challenge in understanding how WM supports planning is the temporal nature of information acquisition. Unlike traditional WM tasks where all items are presented simultaneously (e.g., Stocker and Simoncelli, 2006; Sims, 2016; Jakob and Gershman, 2023), planning typically unfolds sequentially. In planning, we evaluate potential actions one at a time. The importance of precisely maintaining earlier actions only becomes apparent as we evaluate later choices. For instance, discovering a highly scenic location later in planning might change how precisely we need to maintain information about earlier stops along that route.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

While prior WM research has demonstrated that people can prioritize goal-relevant information in simple tasks (Ravizza et al., 2021), it remains unclear how they manage these dynamic interdependencies during sequential planning, especially when dealing with continuous rewards that require precise representation.

In this paper, we examine how humans dynamically manage continuous reward information with WM during sequential planning. We developed an information-theoretic model that directly addresses both the interdependence of reward information and the temporal nature of planning, characterizing optimal encoding strategies under WM constraints. To address the interdependence of reward information, we define an information channel that encodes the potential rewards in all candidate plans into a single, integrated representation, as proposed previously (Fox and Tishby, 2012), rather than treating them as independent components. To address the temporal nature of information acquisition in planning, we define a metalevel Markov Decision Process (MDP) where each state represents a mental representation of the possible plans, and the agent learns to optimally map previous reward representations and new action evaluations to updated representations (Callaway et al., 2022).

To test this information-theoretic model of planning, we developed a modified version of the Mouselab task (Callaway et al., 2022), where participants are sequentially presented with rewards at different nodes in a decision tree. By revealing rewards sequentially, the task creates increasing demands on WM as participants must maintain earlier node rewards while processing new information. To derive predictions for the optimal WM allocation policy in this task, we trained a variational recurrent neural network (VRNN). This model combines the ability to compress multivariate continuous information through its variational component with the capacity to track sequential dependencies through its recurrent architecture. This provides a computationally tractable framework for modeling working memory allocation during sequential planning.

Comparing model predictions with behavioral data, we show that people preferentially encode rewards that are most informative for plan selection. Specifically, participants maintained more precise representations when rewards were on higher-valued paths or when competing options were similar in overall value. These findings demonstrate how humans strategically distribute limited WM resources during planning, adapting their encoding precision based on both the interdependence of rewards and the sequential nature of information acquisition.

Results

To investigate how people allocate WM resources during sequential planning, we designed a decision-making task where participants navigated a decision tree. Participants observed the reward at each node in the tree sequentially before selecting the path with the highest cumulative value. After path selection, we assessed the encoding precision of action rewards through participants' recall error for different node rewards. We contrasted participants' choices and recall errors with the predictions of a resource-rational agent that allocates WM resources optimally, implemented using a VRNN.

Behavioral Experiment

Participants We recruited 112 participants with normal or corrected-to-normal vision through UCSD's Sona system. Fourteen subjects were excluded due to low performance, resulting in 98 participants being analyzed.

Stimuli The primary stimulus in this experiment was a game board designed as a 7-node decision tree. The reward at the central node of this board was always set to 0. The rewards at other nodes were visually represented as diamonds with varying orientations, each having a value ranging from -4.5 to 4.5 (Figure 1.a). The reward at each node was randomly sampled from a normal distribution $N(0,5)$. Participants were trained to associate these orientations with their corresponding reward values before starting the main experiment.

Task Each trial consisted of three phases: reward presentation, path selection, and reward recall (Figure 1.b). During reward presentation, participants observed the reward at each node in the decision tree sequentially for 1.5 seconds each without inter-stimulus intervals, following a depth-first manner, exploring one complete path before moving to the next. Subsequently, during path selection, they were asked to control a plane to travel from the central node to one of the leaf nodes, aiming to accumulate the maximum possible rewards in the nodes they had visited. Once the plane moved, it could not return to a visited node. After completing the route, we asked participants to recall the reward at selected nodes in random order. Specifically, they were prompted to match the diamond orientation on a probed node with the corresponding pattern from their memory. They did this by continuously adjusting a slider to control the diamond orientation until it matched their recollection (Figure 1.c). The participants were required to select the best path and recall two selected nodes in the practice trial without errors to proceed to the actual experiment. To ensure that the participants' primary goal was to maximize their points during plan selection rather than perfectly encoding each node, we only probed an average of two nodes per trial.

Procedure Before the main experiment, participants completed a tutorial where they practiced path selection and memory recall separately, followed by a complete practice trial. They then completed up to 60 experimental trials. We incentivized participants' performance in path selection by informing them that the experiment would conclude when 120 points of reward had been accumulated across all trials. This corresponded to the average number of points they would get

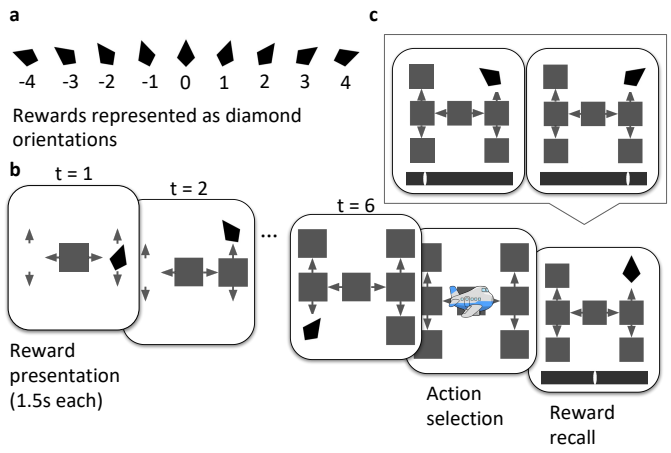

Figure 1: **a:** Diamond orientations corresponding to reward values. The possible reward values $\mathcal{R} \sim N(0,5)$. The orientations representing negative rewards were pointing to the left, and the orientations representing positive rewards were pointing to the right. **b:** Overview of the experimental trial process, with each participant completing a maximum of 60 trials featuring randomly sampled rewards. **c:** Participants reported their reward estimations by adjusting the handle on the slider until the orientation at the probed node matched their recall.

after completing 40 trials perfectly. The reward at each node of the decision tree was randomly sampled across each trial and across participants. We revealed the decision tree with the true node rewards after each trial.

Model

We formalize planning under WM constraints as a sequential decision problem in which an agent evaluates a sequence of actions and their rewards before selecting the optimal plan. At each step, the agent observes the reward associated with a new action and updates its mental representation of the entire space of candidate plans. We assume that storing reward information with WM incurs a cognitive cost that increases with encoding precision. We hypothesize that during planning, people adaptively balance this cost against the benefit of choosing the plan that delivers the best outcome.

This problem is inherently dynamic and recursive. At any given moment, the optimal encoding strategy depends not only on the reward of the current action but also on the rewards of future actions that have yet to be evaluated. For example, a low current reward might initially seem unimportant to encode precisely, but could become critical if subsequent actions in the same path yield high rewards. These dependencies require the agent to consider the potential rewards of the unevaluated nodes and the expected future relevance of each evaluated node in plan selection.

To formalize the problem of reward encoding and maintenance during planning, we adapt the "passive POMDP" framework of Fox and Tishby (2012). At each time step, an agent observes the reward at a pre-specified state, R_t. These rewards are integrated into the agent's representation, or *mental state*, following an encoding/maintenance policy, $p(M_{t+1} \mid M_t, R_t)$. At the end of planning, when all rewards have been observed, the agent executes the sequence of actions that yield the maximal expected value, given the agent's mental state at that moment. The agent's goal is to learn an optimal mental state updating policy that maximizes the expected value of the final decision while minimizing the cognitive cost of maintaining information with WM.

We quantify the WM cost using information theory: each mental state update incurs a cost proportional to $I(M_{t+1}; M_t, R_t)$, that is, the mutual information between the new mental state and the combination of the previous mental state and the observed reward. This term thus captures both how much new information is encoded (R_t) as well as how much old information is maintained (M_t). A precise encoding that preserves small differences in reward distributions requires higher mutual information and thus higher cost. The cost of maintaining information increases with the duration it must be held with WM, making early observations more costly to retain. By adjusting this compression at each step, the agent manages its total WM load during planning.

Implementation with a Variational Recurrent Neural Network We implemented this model using a Variational Recurrent Neural Network (VRNN; Chung et al., 2015). The mental state M was implemented as the hidden state of a recurrent neural network, specifically a 128-dimensional vector. At each time step, the VRNN received a newly observed reward (R_t) and the previous mental state (M_t) and produced a new mental state (M_{t+1}).

The mapping $p(M_{t+1} \mid M_t, R_t)$ was implemented using a variational approach (c.f. Kingma, 2013). The model learned a stochastic encoder that transformed M_t and R_t into parameters (means and standard deviations) of a probability distribution over latent variables Z_t. The specific values z_t were then sampled from this distribution using the reparameterization trick and were passed to a decoder network that produced the new mental state, M_{t+1}. The mutual information term was approximated as the KL divergence between the variational posterior distribution $q(Z_t \mid M_t, R_t)$ and a learned prior $p(Z_t)$ over those parameters. By backpropagating from the final choice outcome and the accumulated WM costs, the VRNN learned to compress reward information in a way that maximized final payoffs while minimizing total memory complexity. This implementation provides a computationally efficient and scalable model of WM-bounded planning.

Encoding Strategies under WM Constraints

We hypothesized that humans integrate reward information across multiple nodes to determine each node's relevance for decision-making. This relevance is determined by two factors:

1. **Path optimality**: The likelihood that the path containing the node is optimal, which increases with greater cumulative values and higher path ranks.

2. **Reward distinctiveness**: The magnitude of reward differences between competitors. Here, competitors are defined as sibling nodes or competing paths in the decision tree.

For our analyses, we define path rank based on cumulative values across the decision tree's four paths, where rank 1 indicates the highest-value path and rank 4 the lowest. We examined how path optimality and reward differences influence recall accuracy using both model simulations and behavioral data. All analyses used repeated measures correlations with Bonferroni correction for multiple comparisons.

Reward estimates improve with reward magnitude and reflect the use of prior information. We observed a significant negative correlation between node reward and recall error, both in simulations ($r = -0.099$, $p < 0.001$, degree of freedom (df) $= 12538$) and behavioral data ($r = -0.120$, $p < 0.001$, df $= 13627$; see Figure 2.a and 2.b). This may suggest that participants preferentially encoded more rewarding actions because they may contribute to a potentially optimal plan.

We also observed a significant positive correlation between the absolute node reward and recall error, both in simulations ($r = 0.523$, $p < 0.001$, df $= 12538$) and behavioral data ($r = 0.287$, $p < 0.001$, df $= 13627$). This suggests that participants may be using a recall strategy where they tend to estimate a reward number closer to the prior mean to reduce the error, since the rewards are drawn from a normal distribution centered at 0.

Preferential encoding of rewards following positive rewards. We observed a significant negative correlation between the middle node reward and leaf node recall error, both in simulations ($r = -0.079$, $p < 0.001$, df $= 40007$) and behavioral data ($r = -0.075$, $p < 0.001$, df $= 6252$; see Figure 2.c and 2.d), even after accounting for leaf node reward (model: $r = -0.049$, $p < 0.001$, df $= 40006$; participants: $r = -0.080$, $p < 0.001$, df $= 6252$; see Figure 2.e). This result shows that leaf node recall precision increases with middle node reward magnitude. This might reflect a strategic allocation of memory resources, where early positive rewards suggest a path is potentially optimal, leading to enhanced encoding of subsequent rewards.

Memory error decreases with relative plan quality. Path rank, defined by the cumulative value of each path, was significantly correlated with leaf node recall error (Figure 2.f and 2.g). Both the model ($r = 0.061$, $p < 0.001$, df $=$ 40007) and behavioral data ($r = 0.152$, $p < 0.001$, df $= 6252$) showed a significant positive correlation between the rank of a path and recall error for the leaf node, even when controlling for leaf node reward (model: $r = 0.092$, $p < 0.001$, df $=$

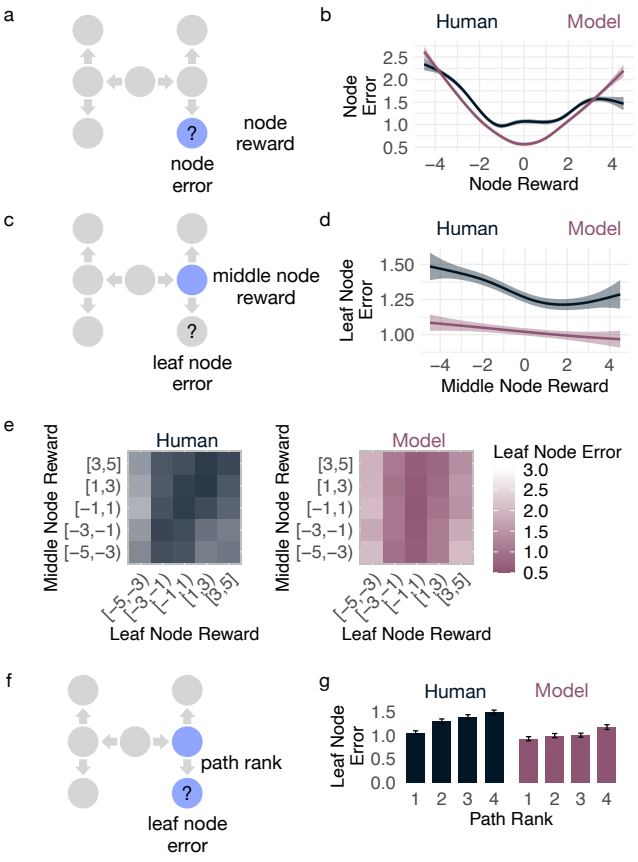

Figure 2: **a**: In this analysis, we examine how node reward influences recall. The decision tree components highlighted in blue represent independent variables, and the component with a question mark represents the dependent variable. **b**: Line plot showing the relationship between node reward versus recall error. **c**: In this analysis, we examine how the middle node reward influences leaf node recall. **d**: Line plot showing the relationship between middle node reward versus leaf node recall error. **e**: Heatmap showing the recall error for different combinations of middle node and leaf node rewards. The x-axis represents the reward value of the leaf node in a path, and the y-axis represents the reward value of the middle node. Color intensity indicates the magnitude of recall error of the leaf node. **f**: In this analysis, we examine how the relative path rank influences leaf node recall. **g**: Leaf node recall error for paths with different ranks.

40006; participants: $r = 0.114$, $p < 0.001$, df $= 6252$). This indicates that rewards in relatively higher-valued paths are encoded with less error, aligning with the goal of maximizing the probability of selecting the best path.

Memory error decreases when alternative actions have similar reward values. We found a significant positive correlation between the reward difference with sibling nodes and recall error in both model simulations ($r = 0.070$, $p <$

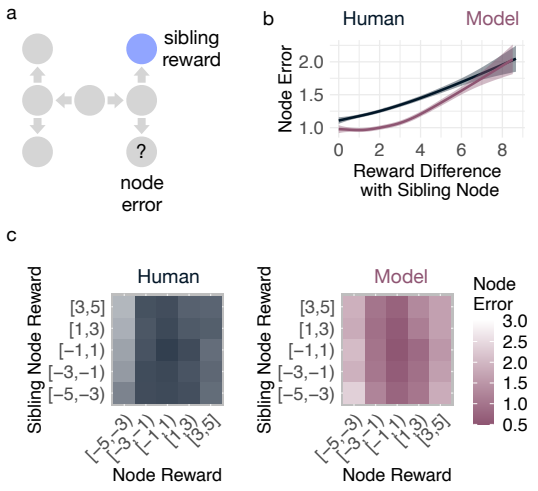

Figure 3: **a**: In this analysis, we examine how the sibling node reward influences recall error. **b**: Line plot showing the relationship between the reward difference between a node and its sibling versus recall error. **c**: Heatmap showing the recall error for different combinations of node and sibling node rewards. The x-axis represents the reward value of the target node, and the y-axis represents the reward value of its sibling node. Color intensity indicates the magnitude of recall error, with darker colors corresponding to lower error values.

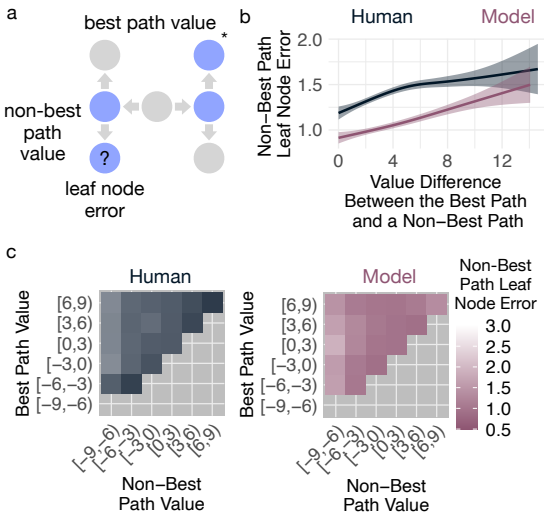

Figure 4: **a**: In this analysis, we examine how the path values of the best path and a non-best path jointly influence recall error for the leaf node of the non-best path. **b**: Line plot showing the relationship between the reward difference between a non-best path and the best path versus the recall error for the leaf node of the non-best path. **c**: Heatmap showing the non-best path leaf node recall error for different combinations of best path and non-best path values. The x-axis represents the non-best path value, and the y-axis represents the best path value. Color intensity indicates the magnitude of recall error of the non-best path leaf node.

$0.001, \text{df} = 40007$) and behavioral data ($r = 0.144$, $p < 0.001, \text{df} = 6252$; see Figure 3.a and 3.b). This correlation remained significant even after controlling for node reward (model: $r = 0.185$, $p < 0.001$, $\text{df} = 40006$; participants: $r = 0.168$, $p < 0.001$, $\text{df} = 4627$; see Figure 3.c). When the sibling node has a similar reward, the agent may need to retain both rewards with less error to differentiate between them.

Memory error for non-optimal plans decreases when the optimal plan has similar values. We assumed the competitor of a non-best path is the best path. For leaf nodes in non-best paths, recall error decreased as the reward difference between the best and the non-best path decreased (see Figure 4.a and 4.b). There was a significant positive correlation between the reward difference and leaf node recall error in both simulations ($r = 0.154$, $p < 0.001$, $\text{df} = 30007$) and behavioral data ($r = 0.099$, $p < 0.001$, $\text{df} = 4627$). This effect persisted after controlling for the non-best path value (model: $r = 0.074$, $p < 0.001$, $\text{df} = 6229$; participants: $r = 0.064$, $p < 0.001$, $\text{df} = 6737$; see Figure 4.c).

Similarly, we assumed the competitor of the best path is the second-best path. For leaf nodes in best paths, in simulation data, recall error significantly decreased as the reward difference between the best and second paths decreased ($r = 0.081$, $p < 0.001$, $\text{df} = 1995$). However, this correlation was not significant in behavioral data ($r = 0.017$, $p = 0.413$, $\text{df} = 2231$). This discrepancy between model predictions and human behavior may suggest that participants preferentially encode the rewards in the best path, regardless of the value of the competitors for the best path.

These findings support our hypothesis that humans strategically allocate WM resources when encoding reward information during planning. The systematic relationships between recall error and both local factors (middle node reward and sibling node differences) and global factors (path rank and competitor differences) suggest that humans adaptively modulate reward encoding precision based on decision-relevant features. This adaptive strategy helps maximize decision quality under WM constraints.

Discussion

In this study, we developed and tested a VRNN model of how people strategically allocate working memory resources during planning. Our modeling results show that an action reward's influence on plan selection depends on the broader context: the rewards of other actions and the values of alternative plans. Our experimental results showed that participants were likewise sensitive to these contextual factors, more accurately recalling rewards that were more contextually relevant.

In particular, we found evidence for contextual effects both *within* and *between* paths in the decision tree. Starting with

within-path effects, we found that leaf nodes in paths with higher total values and better ranks were maintained with more precision, even after controlling for individual node rewards. This finding demonstrates that during planning, WM encoding prioritizes information about the potentially optimal plans to maximize decision quality.

Turning to between-path effects, we found that participants maintained more precise representations when reward differences between sibling nodes were smaller, suggesting that people increase encoding precision in response to closely valued alternative actions that could affect plan selection. Similarly, leaf nodes in non-optimal paths were encoded more precisely when their total value approached that of the optimal path. This pattern reveals that people increase encoding precision when subtle value differences between plans could influence the final choice.

More broadly, these findings contribute to a fuller understanding of how reward influences working memory. Prior research has established that rewards can enhance WM performance (Wallis et al., 2015; Beck et al., 2010; Kennerley and Wallis, 2009), with behavioral (Hu et al., 2016) and neural (Krawczyk et al., 2007; Gazzaley et al., 2005) evidence showing that participants prioritized encoding of goal-relevant information while suppressing encoding of task-irrelevant information. However, most of these studies examined reward effects in relatively simple contexts where the relevance of information was explicitly defined. Our results extend these findings to more complex planning scenarios where information relevance depends on multiple interacting factors. We show that humans can dynamically optimize WM allocation by integrating information across the entire decision tree. This optimization considers not just reward magnitude, but also the relationships between rewards, their position in the decision tree, and their potential impact on plan selection.

Our finding that early positive rewards lead to enhanced encoding of subsequent rewards also aligns with efficient search strategies observed in human planning. Prior work on best-first search and pruning has shown that humans prioritize exploring and evaluating actions in promising paths while reducing consideration of paths following negative outcomes (Huys et al., 2012; Hunt et al., 2021; Callaway et al., 2022). Our results reveal a parallel but distinct mechanism in working memory: beyond just prioritizing *which* actions to evaluate, people also modulate *how precisely* they maintain information about these actions. This may suggest two complementary ways humans optimize planning under cognitive constraints: through selective evaluation of actions, as shown in previous work, and through strategic allocation of memory precision, as demonstrated in our study.

Previous work (Ying et al., 2023) modeled planning under WM constraints using metalevel MDP and showed that path value influences reward encoding in a binary reward planning task. Our current work addresses two limitations of this prior research. First, we use continuous rewards; this puts a greater burden on working memory and allows for more finely graded control of memory precision. Second, we allow the format of the representation to be learned; this allows the model to directly represent integrated reward information, such as the total value of a path. This allowed us to make (and confirm) finer-grained predictions about how people's memory for one reward would depend on other rewards in the decision tree. This more general approach can also accommodate more complex planning scenarios, such as tasks with deeper decision trees or stochastic rewards.

In conclusion, our study shows that humans strategically leverage task-relevance to allocate working memory resources during multi-step planning, preferentially encoding information that is most likely to contribute to optimal decision-making. By dynamically adjusting memory precision based on both local features (such as reward magnitude and sibling similarity) and global properties (such as path ranking and competitor values) of the decision context, people optimize their WM resources to focus on potentially optimal plans. Future work could explore how these mechanisms operate in scenarios that better reflect everyday planning challenges, such as modeling how planning strategies adapt to increased WM load or how reward encoding strategies adjust for planning scenarios where path values are uncertain.

References

Awh, E., Barton, B., and Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological science*, 18(7):622–628.

Beck, S. M., Locke, H. S., Savine, A. C., Jimura, K., and Braver, T. S. (2010). Primary and secondary rewards differentially modulate neural activity dynamics during working memory. *PloS one*, 5(2):e9251.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., and Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, 6(8):1112–1125.

Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., and Bengio, Y. (2015). A recurrent latent variable model for sequential data. *Advances in neural information processing systems*, 28.

Chunharas, C., Rademaker, R. L., Brady, T. F., and Serences, J. T. (2022). An adaptive perspective on visual working memory distortions. *Journal of Experimental Psychology: General*, 151(10):2300.

Fox, R. and Tishby, N. (2012). Bounded planning in passive pomdps. *arXiv preprint arXiv:1206.6405*.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., and D'esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of cognitive neuroscience*, 17(3):507–517.

Hu, Y., Allen, R. J., Baddeley, A. D., and Hitch, G. J. (2016). Executive control of stimulus-driven and goal-directed attention in visual working memory. *Attention, Perception, & Psychophysics*, 78:2164–2175.

Hunt, L., Daw, N., Kaanders, P., MacIver, M., Mugan, U., Procyk, E., Redish, A., Russo, E., Scholl, J., Stachenfeld, K., et al. (2021). Formalizing planning and information search in naturalistic decision-making. *Nature neuroscience*, 24(8):1051–1064.

Huys, Q. J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., and Roiser, J. P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, 8(3):e1002410.

Jakob, A. M. and Gershman, S. J. (2023). Rate-distortion theory of neural coding and its implications for working memory. *Elife*, 12:e79450.

Kennerley, S. W. and Wallis, J. D. (2009). Reward-dependent modulation of working memory in lateral prefrontal cortex. *Journal of Neuroscience*, 29(10):3259–3270.

Kingma, D. P. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Krawczyk, D. C., Gazzaley, A., and D'Esposito, M. (2007). Reward modulation of prefrontal and visual association cortex during an incentive working memory task. *Brain research*, 1141:168–177.

Ravizza, S. M., Pleskac, T. J., and Liu, T. (2021). Working memory prioritization: Goal-driven attention, physical salience, and implicit learning. *Journal of Memory and Language*, 121:104287.

Sims, C. R. (2016). Rate–distortion theory and human perception. *Cognition*, 152:181–198.

Stocker, A. A. and Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585.

Vogel, E. K., Woodman, G. F., and Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of experimental psychology: human perception and performance*, 27(1):92.

Wallis, G., Stokes, M. G., Arnold, C., and Nobre, A. C. (2015). Reward boosts working memory encoding over a brief temporal window. *Visual Cognition*, 23(1-2):291–312.

Ying, Z., Callaway, F., Kiyonaga, A., and Mattar, M. G. (2023). Resource-rational encoding of reward information in planning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46.