A Metacognitive Model of Memory Encoding Modulated by Rewards

Si Ma (si.ma@rutgers.edu)
Department of Psychology
Rutgers University–New Brunswick
Piscataway, NJ 08854 USA

Qiong Zhang (qiong.z@rutgers.edu)
Department of Psychology, Department of Computer Science, Rutgers Center for Cognitive Science
Rutgers University–New Brunswick
Piscataway, NJ 08854 USA

Abstract

Despite robust empirical evidence supporting the role of reward in enhancing memory, the relationship between reward and memory shows complex patterns. We present a novel computational model that considers how people optimally allocate limited cognitive resources during memory encoding. Unlike previous accounts, which assume that higher rewards directly lead to stronger memory encoding, we allow our model to adaptively decide how much to encode for each item based on the overall reward environment and one's limited cognitive resources. Our model's predictions align closely with human recall patterns across three experiments. It successfully explains why high-reward items are better remembered than low-reward items only when they are presented together but not separately (Exp 1). Analyzing an existing dataset (Exp 2), our model accounts for how memory is modulated not only by the current reward but also by the rewards of preceding (but not future) items. To further test our proposed model, we collected new data (Exp 3) demonstrating that this insensitivity to rewards of future items can be reversed when participants can anticipate upcoming reward values. These findings provide evidence that memory encoding is an active process involving meta-level control, where cognitive resources are strategically allocated to maximize overall rewards, rather than a passive response to individual reward values.

Keywords: reward, memory encoding, metacognition, rational analysis

Instructions

The ability to selectively encode information is an adaptive feature of episodic memory. It is rational to prioritize resources and effort in encoding information associated with higher rewards than those with lower rewards. Previous studies found that people better remember information if it will help them gain reward (Bowen et al., 2020; Gong & Li, 2014; Grandoit et al., 2024; Manga et al., 2020; Middlebrooks et al., 2017; Talmi et al., 2021). Multiple accounts have been proposed to explain this phenomenon of reward-enhanced memory. For instance, high-reward items may capture greater attention than low-reward items (Allen & Ueno, 2018; Sandry & Ricker, 2020), be more deeply encoded (Castel, 2007; Cohen et al., 2014), or receive more selective study (Castel et al., 2013). These accounts are consistent with a computational implementation in which memory encoding strength increases when reward increases (Talmi et al., 2021; Zhou et al., 2023). Despite robust empirical evidence supporting the role of reward in enhancing memory, this relationship does not always hold and the influence of reward on memory formation can exhibit more complex patterns. The goal of the present work is to propose a novel account of reward's effect on memory that can reconcile a range of empirical findings.

While high-reward items are generally better remembered than low-reward items, Talmi et al. (2021) discovered an intriguing exception. This memory advantage only exists when high- and low-reward items are mixed within the same list (e.g., "LHLLHLL"; mixed-list condition) but not when they are presented in separate lists (e.g., "LLLLLLL" or "HHHHHHH"; pure-list condition). These effects have been replicated in several follow-up studies (Hellerstedt & Talmi, 2022; Hellerstedt et al., 2023). This raises an important question: if people are rational in selectively remembering more valuable information, why would they recall an equal amount from a high-reward list compared to a low-reward list? Past research has shown that memory behavior reflects not only how people adapt to environmental goals but also the computational limitations imposed by human cognitive architecture (Anderson, 1990; Callaway et al., 2024; Lieder & Griffiths, 2020; Lu et al., 2024; Van den Berg & Ma, 2018; Xu et al., 2024; Zhang et al., 2023). We hypothesize that better memory for high-reward items results from the *adaptive* allocation of *limited cognitive resources* to these items. High-reward items are better remembered than low-reward items only when they are presented together, as more cognitive resources are allocated to high-reward items when they are directly competing with low-reward items.

To formally test our hypothesis, we build a novel metacognitive model of memory encoding and validate our model against three sets of experimental results, including the difference between pure-list and mixed-list conditions (Talmi et al., 2021). We propose that metacognitive mechanisms play a key role in memory encoding. We draw inspiration from early theoretical developments in the metamemory literature (Nelson, 1990), which posit that our memory system includes an additional "meta-level" component. This meta-level component adaptively *controls* ongoing memory processes by *monitoring* the current state of memory (at the object-level) with the goal of maximizing performance (see Figure 1). While long-term memory storage is vast (Brady et al., 2008), our ability to recall information is constrained by moment-to-moment fluctuations during encoding (Noh et al., 2014; Paller & Wagner, 2002; Sundby et al., 2019). Recent theoretical advances in computational modeling have linked memory performance with the amount of cognitive re-

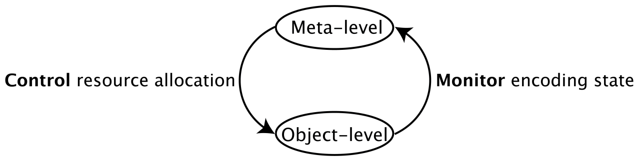

Figure 1: Our proposed framework. Meta-level monitors the state of memory encoding (e.g., available resources, reward, and serial position) and controls how to allocate cognitive resources to optimize overall rewards.

sources available when encoding items into long-term memory (Popov & Reder, 2020; Reder et al., 2000, 2007). The assumption of a finite amount of cognitive resources that needs time to recover after depletion (i.e., resource-depletion-and-recovery assumption) can account for numerous behavioral (Kowialiewski et al., 2021; Mızrak & Oberauer, 2021; Oberauer, 2022; Popov & Reder, 2020; Popov et al., 2019, 2022) and neural findings (Ma et al., 2024). Given these constraints at memory encoding, we hypothesize that a "meta-level" process continuously monitors the memory state and strategically determines how to allocate cognitive resources.

Critically, when different rewards are present, our hypothesis differs from accounts where the strength of memory encoding is directly related to the reward magnitude of a given item (Talmi et al., 2021). Instead, we posit that the strength of memory encoding is a consequence of optimizing total reward gain through the adaptive allocation of limited cognitive resources. Our proposed model is supported by three sets of experimental results:

- In Experiment I, we simulated our model based on the same reward structures as those used by Talmi et al. (2021). Under the constraints of limited cognitive resources, we obtained the optimal policy to allocate these resources to each item in order to maximize total rewards. The optimal behavior aligns with observations from Talmi et al. (2021) where high-reward items were better remembered than low-reward items in mixed lists but not in pure lists.
- In Experiment II, we further examined how items with varying rewards within a list compete for allocation of limited cognitive resources, analyzing a publicly shared dataset with rewards ranging from 1 to 10 (Middlebrooks et al., 2017). We showed that an item's memory performance is not only modulated by its associated reward but also by the rewards of preceding (but not future) items. These observations align with the model predictions, where resources allocated to each item depend not only on its associated reward but also on the overall reward environment.
- In Experiment III, we collected a new dataset to provide a strong test for our proposed model. Although there was no effect of future reward in Experiment II, our model predicts that if people are rational in allocating limited cognitive resources, they should reserve them when anticipating high-reward items in the near future.

Metacognitive model

In this section, we specify the details of the proposed metacognitive model. The model aims to capture the encoding processes of a list-learning paradigm, where participants first study a list of items, and then attempt to recall as many items as they can from the list in any order (free recall task; Murdock Jr, 1962; Roberts, 1972; Standing, 1973). The model consists of two processes: an object-level process responsible for memory encoding, and a meta-level process that monitors and adaptively controls the encoding process (Figure 1). The interaction between the object-level and the meta-level is a continuous process. As each item in a study list is presented to the model one after another, the meta-level determines the amount of resources to allocate based on information from the object-level and the object-level uses the allocated resources to strengthen the memory and moves on to the next word. This process repeats until the end of the list and the model is asked to recall all the items from the list.

Object-level model

At the object-level, we follow the model implementation from prior studies in characterizing how items are encoded into episodic memory under the resource-depletion-and-recovery assumption (Ma et al., 2024; Popov & Reder, 2020; Reder et al., 2007). The assumption has been supported by numerous behavioral (Kowialiewski et al., 2021; Mızrak & Oberauer, 2021; Oberauer, 2022; Popov & Reder, 2020; Popov et al., 2019, 2022) and neural findings (Ma et al., 2024).

At the beginning of studying a list of items, the amount of resources available is at its maximum ($W_{max} = 1$). Upon studying an item at position k, a proportion (τ) of currently available resources, W_k, is depleted for semantically processing the item, expressed as:

$$W_{sem,k} = \tau W_k \tag{1}$$

Meantime, a proportion (δ) of currently available resources ($W_k - W_{sem,k}$) is allocated to encode the item into episodic memory, expressed as:

$$W_{epi,k} = \delta(W_k - W_{sem,k}) \tag{2}$$

The memory strength of the encoded item ($B_{epi,k}$) depends on the amount of resources allocated to it:

$$B_{epi,k} = \sqrt{W_{epi,k}} \tag{3}$$

The amount of resources is finite, and once depleted it needs time to recover. Resources recover linearly at a rate of r per second until it reaches W_{max}. Therefore, the amount of resources at the beginning of encoding the next item at position $k+1$ follows:

$$W_{k+1} = min(W_{max}, W_k - W_{sem,k} - W_{epi,k} + rt_k) \tag{4}$$

where t_k is the time between item presentations.

The larger the memory strength $B_{epi,k}$, the higher the probability of recalling the item at position k, expressed as:

$$p_k = \Phi\left(\frac{B_{epi,k} - \theta_{epi}}{\sigma_{epi}}\right) \tag{5}$$

where Φ represents the cumulative distribution function of the standard normal distribution, with θ_{epi} and σ_{epi} as the mean and the standard deviation used to standardize $B_{epi,k}$.

Meta-level model

While resource-based memory models have traditionally assumed passive allocation of resources with a fixed value of δ (Popov & Reder, 2020; Reder et al., 2000, 2007), our proposed model will examine the adaptive allocation of these resources through metacognitive control. The meta-level process addresses the problem of how much cognitive resources to allocate by varying δ for each item to maximize overall reward gain. To obtain the optimal policy for this behavior, we use a reinforcement learning framework (Mnih et al., 2015; Sutton, 2018), where the agent at the meta-level receives the state information S and the reward information R from the object level (in a monitoring process) and determines what action A to take (in a control process) to maximize cumulative rewards over time.

- **States (S):** In our model, s_k contains the information needed to characterize the current state of memory when encoding an item at position k, including 1) the amount of resources available W_k, 2) the property (i.e., reward r_k) of the current item being encoded, and 3) how far has it been in the list (i.e., k).

- **Actions (A):** A is the action the agent takes. In our model, a_k is the proportion (δ) of currently available resources allocated to the item at position k, ranging from 0 to 1.

- **Reward (R):** R_k is the reward received by the agent after taking an action a_k at item position k:

$$R_k = \begin{cases} \sum_{i=1}^{L}(p_i \cdot r_i), & \text{if } k = L \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

  where p_i is the probability of the item at position i being recalled later (Equation (5)), r_i is the reward assigned to the item, and L is the list length. Neither participants nor the model has access to item-specific recall probabilities at the time of encoding the item. A reward was determined and delivered only after the presentation of the full list (length L). Participants are informed of their rewards for a given list after their recall attempts, and the model uses the above equation to approximate the rewards.

- **Policy (π):** π is a policy that decides which action to take at a given state s. At each position k, the agent aims to maximize the expected discounted return from the current position k to the end of the list:

$$G_k = \sum_{t=0}^{L-k} \gamma^t R_{k+t} \tag{7}$$

  where γ is the discount factor that controls the relative importance of future versus immediate rewards. We set $\gamma = 0.95$. The agent's objective is to learn an optimal policy π^* that maximizes the expected return:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi}[G_k] \tag{8}$$

  We used RecurrentPPO (Raffin et al., 2021), an instance of policy-gradient methods (Schulman et al., 2017), to find the optimal behavior of metacognitive control π^*.

Model setup and simulations

To generate model predictions for specific experiments, we set up the list and reward structure to be identical to the experiment that we aim to model. Parameters at the object-level describe constraints of cognitive resources, which we inherited from previous implementations of resource-based memory models ($\tau = 0.072$, $r = 0.08$, $\theta_{epi} = 0.367$, $\sigma_{epi} = 0.256$; Ma et al., 2024). We then obtained how the model optimally allocates cognitive resources to each item (δ) when faced with different list and reward structures in a given experimental design. Importantly, the model has never been directly fitted to the empirical patterns it seeks to explain. Instead, resource constraint parameters are fixed based on previous studies, and resource allocation behavior is derived through optimization. As a result, the model's outcomes serve as *model predictions* generated by the proposed theoretical framework.

Experiment I

Many studies support that high-reward items are better remembered than low-reward ones (Gong & Li, 2014; Grandoit et al., 2024; Middlebrooks et al., 2017; Talmi et al., 2021). However, this memory advantage disappears when high-reward items and low-reward items are studied in separate lists (Talmi et al., 2021). In Experiment I, we sought to explain this effect by proposing that participants adaptively allocate limited cognitive resources during memory encoding in a metacognitive process. Our proposed model was set up based on the same stimuli, rewards, and trial structure as in Talmi et al. (2021) and was trained to maximize its overall reward gain. We then compared the recall patterns predicted by the model with those of the experiment.

Talmi et al. (2021)

29 participants were recruited (aged 18-21; Experiment 1 in Talmi et al., 2021) and were asked to complete a free recall task. In each trial, participants were presented with a list of 16 pictures, each of which lasted for 2 seconds followed by a randomized interval (of approximately 4 seconds) of white screen. Immediately following the list presentation, participants completed a distractor task for 60 seconds after which they were given 3 minutes to describe the pictures they remembered in any order. Some pictures were framed and recalling those would give participants a high reward of £1, while unframed pictures had a lower reward of 10 pence. Each participant completed six trials: two trials

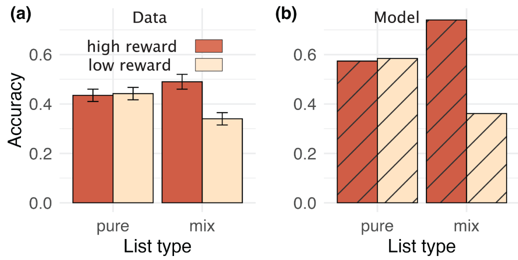

Figure 2: High-reward items are better remembered than low-reward items only in mixed lists (e.g., LHLLHL) but not in pure lists (e.g., LLLLLL, HHHHHH), shown for both (a) behavioral data reproduced from Figure 2A in Talmi et al. (2021) and (b) model predictions.

of high-reward pure lists, two trials of low-reward pure lists, and two trials of mixed lists (half low-reward and half high-reward pictures).

Results and Discussions

We observed an alignment between model predictions (Figure 2b) and human recall patterns (Figure 2a) in pure lists versus mixed lists. Specifically, high-reward items were better recalled than low-reward items in mixed lists but not in pure lists. Our model captures these effects: In mixed lists, high-reward items and low-reward items compete directly for limited cognitive resources, with high-reward items prioritized over low-reward items. In contrast, in pure lists, competition occurs between items with the same rewards, so there is no clear advantage in allocating more resources to some items over others. As the total amount of available resources remains the same across both high-reward and low-reward pure lists, items within each list receive comparable resources during encoding. The alignment between model predictions and human recall patterns supports our hypothesis that participants adaptively allocate limited cognitive resources during memory encoding based on reward levels.

Experiment II

Our model can explain recall patterns in Experiment I through competitions between items for limited resources during encoding controlled by a meta-level process. The goal of Experiment II is to further investigate these mechanisms by examining how memory for a specific item is affected by other list items across more fine-grained reward magnitudes. We analyzed a publicly shared dataset with fine-grained reward magnitudes ranging from 1 to 10 points (Middlebrooks et al., 2017). If items compete for limited cognitive resources during encoding, as our model proposes, we would expect an item's reward value to influence not only its own recall performance but also that of temporally adjacent items in the study list.

Middlebrooks et al. (2017)

72 participants were recruited and asked to complete a free recall task. Participants were presented with six trials each containing a list of 20 words with a presentation rate of 3 seconds. Each word was given a reward value ranging from 1 to 10 points, with two words allocated to each point value per list, analogous to the mixed list condition in Experiment I. Immediately after the list presentation, participants were asked to recall the words in any order and maximize their total points. While participants showed varying degrees of sensitivity to reward values, we focused our analysis on the half of participants who demonstrated stronger reward sensitivity, as measured by the slope of a linear mixed-effects model fitted to their recall performance.

Results and Discussions

Our model predictions align well with human data, revealing key patterns in the adaptive control of cognitive resources during encoding. Using a dataset with fine-grained reward values ranging from 1 to 10 points, we first examine how an item's reward value affects its own recall performance. Specifically, we analyzed the influence of reward at position k (r_k) on the recall probability of the word at the same position (m_k; Figure 3a). As shown in Figure 3b, participants had an increasingly better recall performance as the reward value increased (linear mixed-effects model: $\beta = 0.063, SE = 0.003, t(323) = 24.286, p < 0.01$), consistent with our model predictions (Figure 3c). This result reproduces the same effects in the mixed list condition in Experiment 1 (Figure 2). Our model successfully captures these effects regardless of the granularity of reward levels.

While alternative accounts could explain why memory performance increases as a function of reward for the item at the same position, our model makes unique predictions on how reward affects the memory performance of adjacent items in a study list. Specifically, our model predicts that while a higher reward r_k for an item increases the recall probability for that item m_k (Figure 3c), it decreases the recall probability for the subsequent item m_{k+1} (Figure 3f). This prediction results from the limited cognitive resources during encoding: once resources are depleted, they require time to recover. When more cognitive resources are depleted in encoding an item with a higher reward at position k, fewer resources remain available to encode the subsequent item at position $k+1$, resulting in lower recall probability at position $k+1$. Furthermore, although reward at position k affects the recall probability of the subsequent item at position $k+1$ (forward direction; Figure 3f), it leaves the recall probability of the preceding item at position $k-1$ unchanged (backward direction; Figure 3i). No reward effect was shown in the backward direction because cognitive resources that have already been allocated to the item at position $k-1$ cannot be reallocated or released, even if a high-reward item is subsequently encountered at position k. Recall patterns in Middlebrooks et al. (2017) align with model predictions, with an effect of r_k on m_{k+1} in the forward direction (Figure 3e; $\beta = -0.012, SE = 0.003, t(323) = -4.64, p < 0.01$), but no effect of r_k on m_{k-1} in the backward direction (Figure 3h; $\beta = -0.005, SE = 0.003, t(323) = -1.836, p = 0.067$). While ex-

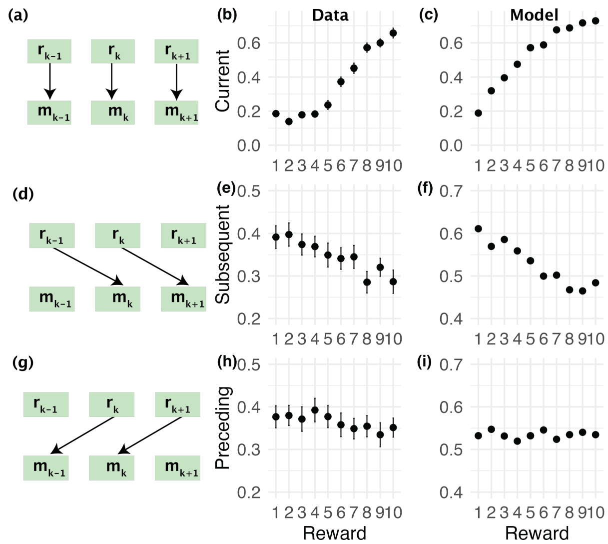

Figure 3: The effect of reward magnitude (from 1 to 10) on the recall probability of the current (a-c), subsequent (d-f), and preceding item (g-i) shown for both behavioral data (middle column) and model predictions (right column).

isting accounts of rewards can explain the enhanced memory for high-reward versus low-reward items, the effect of reward in the forward direction and a lack of effect in the backward direction, along with evidence from Experiment I, are unique evidence supporting our proposed model. These results support our hypothesis that limited cognitive resources and specific reward environments jointly shape memory behavior.

Experiment III

In this experiment, we introduced a novel manipulation and collected a new dataset to provide a strong test for our proposed metacognitive model. Previously, we have shown that rewards affect temporally adjacent items but only in the forward direction. This is because of the constraint that after resources are allocated to an item at position $k-1$, they cannot be reallocated or released even upon seeing a high-reward item at position k. While the reward of an item was not known to participants in Experiments I and II until the item was presented, we designed an experiment where participants were informed of the reward structure of an entire list ahead of its presentation. We hypothesize that with the knowledge of upcoming rewards when encoding an item, participants can adaptively adjust their resource allocation at the current position according to rewards at future positions. This would provide a strong test for our proposed account of reward effects on memory. As participants have limited time to learn the reward values for an entire list, we simplified the reward structure for participants to learn and used short lists consisting of six words per list with at most one switch in reward values mid-list, “HHHHHH”, “HHHLLL”, “LLLHHH”, and “LLLLLL”. Under this simplified reward structure, we then examined how reward values of the first or second half of the list influence the recall performance of the current half, the preceding half, and the subsequent half of the list. The experiment, analyses, and model predictions were pre-registered (https://aspredicted.org/jyx6-mdqh.pdf).

Methods

Participants 125 participants (aged 18-40) were recruited on the Prolific platform. They were all fluent English speakers and consented to participate in the study. Following our preregistered exclusion criteria, we excluded participants who failed the attention check, quit halfway, or reported using external help or indicating a lack of diligence in completing the task. 71 participants remained in our analysis.

Stimuli All word trials were randomly selected from a prior word pool of 326 words (Polyn et al., 2011). Each trial contained six words, each belonging to a different semantic category (Polyn et al., 2011). Some words were displayed with a gray rectangular frame around them, indicating a high-reward value (3 points), while unframed words carried a low-reward value (1 point). The experiment included four different types of reward structures across lists: “HHHHHH”, “HHHLLL”, “LLLHHH”, and “LLLLLL”. Participants were informed of the reward structure before each list presentation. Each participant completed 17 trials in total, including one practice trial at the beginning and 16 experimental trials (four trials for each reward structure). The order of experimental trials was randomized for each participant.

Procedure Participants completed a free recall task. In each trial, they were presented with six words, displayed sequentially. Each word appeared on the screen for two seconds, followed by a 0.5-second blank screen. Words were either framed or unframed, with framed words awarding high rewards (3 points) and unframed words awarding low rewards (1 point). At the beginning of each trial, participants were informed of the reward structure for the upcoming list. As each

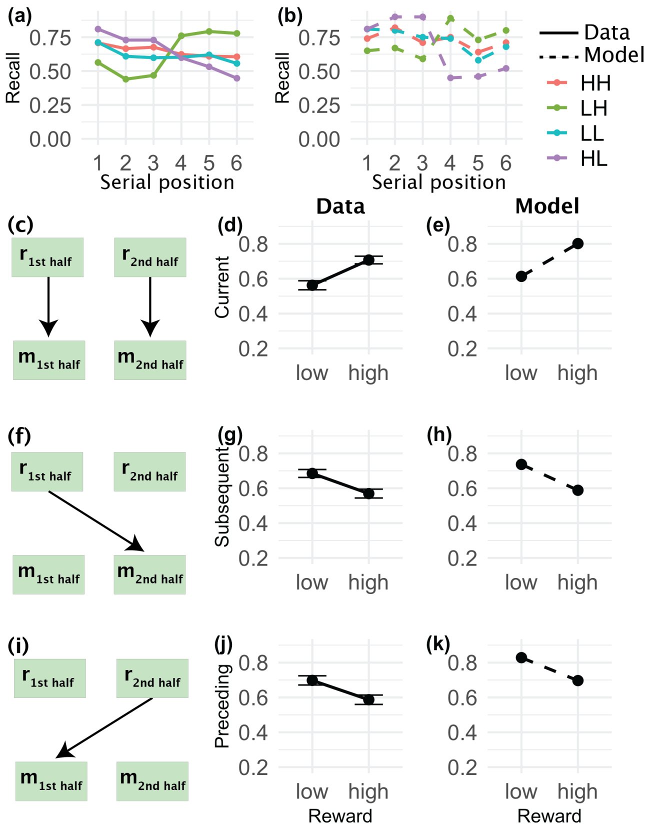

Figure 4: The effect of reward on recall probability across serial positions for behavioral data (a) and model predictions (b) for each list type. The effect of reward on the recall probability of the current (c-e), subsequent (f-h), and preceding items (i-k), shown for both the behavioral data (middle column) and model predictions (right column).

word was presented, they performed a size judgment task by pressing "Q" if the word was smaller than a shoebox or "P" if it was larger. To receive any points for a list, participants had to correctly classify more than half of the words in the list. This task was designed as an attention check. Following the list presentation, participants completed a 12-second distractor task, during which they solved three math problems in the form of A + B + C = ?. Bonuses were awarded based on their performance. After the distractor task, participants had 15 seconds to recall the words from the just-presented list in any order. Their objective was to maximize their total reward points which influenced their final payment. The experiment was implemented using PsiTurk and Heroku and lasted approximately 25 minutes.

Results and Discussions

Figure 4a and 4b compare human recall patterns and model predictions for the recall probability across different serial positions for each list type. We analyzed the effect of reward values on current, subsequent, and preceding items following a similar approach to the analyses conducted in Experiment II. Recall probabilities were averaged for each half of the list.

Consistent with the results of Experiment II, when comparing rewards and recall probabilities in the same or different halves of a list (Figure 4c, 4f, and 4i), we found that while high-reward items were better remembered than low-reward items (one-tailed Wilcoxon signed-rank test; $V = 252.5$, $P < 0.01$; Figure 4d), subsequent items following high-reward items were associated with lower memory performance compared with those following low-reward items ($V = 1854$, $P < 0.01$; Figure 4g). These results align with model predictions (Figure 4e and 4h), supporting our hypothesis that participants adaptively allocate limited cognitive resources during memory encoding based on reward levels.

While rewards affect current and subsequent items (Figure 4d and 4g) in a manner comparable to Experiment II (Figure 3b and 3e), there are differences in how rewards affect preceding items between these two experiments. Experiment II did not exhibit an effect of reward on preceding memory performance; this is because resources cannot be reallocated to a high-reward item at position k once they have already been used to encode another item at position $k-1$ (a key assumption in the object-level model). However, the above result could be reversed if one decides not to allocate resources at position $k-1$ in the first place. A strong test for the optimal model is to examine if participants reserve resources at position $k-1$ if they anticipate high-reward items in the upcoming future. In the current experiment, participants were informed of (so was the model) the reward structure of a list prior to the list presentation. We analyzed how rewards of the second halves influence memory performance of the first halves (Figure 4i). Consistent with our model predictions (Figure 4k), items preceding high-reward items were associated with lower memory performance compared with those preceding low-reward items ($V = 1842.5$, $P < 0.01$; Figure 4j). This is in contrast to those in Experiment II where there was no effect of reward on the preceding items (Figure 3h). These results provide strong evidence that the meta-level process adaptively reserves and allocates cognitive resources to encode items given the resource constraints at the object-level.

Conclusion

In this work, we presented an optimal model of metacognitive control for memory encoding to understand the role of reward in memory. The model is constrained by limited cognitive resources available at encoding, where resource depletion increases memory strength but requires time for recovery. We validated our model across three experiments, which explain why high-reward items do not always show memory advantages, how they can affect the memory of subsequent but not preceding items, and how reward anticipation can reverse the insensitivity of the preceding items to rewards. Together, our results provide strong evidence that reward-modulated memory encoding is an adaptive process involving meta-level control, rather than a passive response to individual reward values.

References

Allen, R. J., & Ueno, T. (2018). Multiple high-reward items can be prioritized in working memory but with greater vulnerability to interference. *Attention, Perception, & Psychophysics*, *80*, 1731–1743.

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Bowen, H. J., Gallant, S. N., & Moon, D. H. (2020). Influence of reward motivation on directed forgetting in younger and older adults. *Frontiers in Psychology*, *11*, 1764.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

Callaway, F., Griffiths, T. L., Norman, K. A., & Zhang, Q. (2024). Optimal metacognitive control of memory recall. *Psychological Review*, *131*(3), 781.

Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. *Psychology of Learning and Motivation*, *48*, 225–270.

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging*, *28*(1), 232.

Cohen, M. S., Rissman, J., Suthana, N. A., Castel, A. D., & Knowlton, B. J. (2014). Value-based modulation of memory encoding involves strategic engagement of fronto-temporal semantic processing regions. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 578–592.

Gong, M., & Li, S. (2014). Learned reward association improves visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 841.

Grandoit, E., Cohen, M. S., & Reber, P. J. (2024). Reward enhancement of item-location associative memory spreads to similar items within a category. *Cognition and Emotion*, 1–16.

Hellerstedt, R., Bekinschtein, T., & Talmi, D. (2023). Can neural correlates of encoding explain the context dependence of reward-enhanced memory? *Psychophysiology*, *60*(9), e14322.

Hellerstedt, R., & Talmi, D. (2022). Reward does not modulate forgetting in free recall tests. *Learning & Memory*, *29*(12), 430–434.

Kowialiewski, B., Lemaire, B., & Portrat, S. (2021). How does semantic knowledge impact working memory maintenance? computational and behavioral investigations. *Journal of Memory and Language*, *117*, 104208.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, *43*, e1.

Lu, Q., Nguyen, T. T., Zhang, Q., Hasson, U., Griffiths, T. L., Zacks, J. M., Gershman, S. J., & Norman, K. A. (2024). Reconciling shared versus context-specific information in a neural network model of latent causes. *Scientific reports*, *14*(1), 16782.

Ma, S., Popov, V., & Zhang, Q. (2024). A neural index reflecting the amount of cognitive resources available during memory encoding: A model-based approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Manga, A., Vakli, P., & Vidnyánszky, Z. (2020). The influence of anticipated monetary incentives on visual working memory performance in healthy younger and older adults. *Scientific Reports*, *10*(1), 8817.

Middlebrooks, C. D., Kerr, T., & Castel, A. D. (2017). Selectively distracted: Divided attention and memory for important information. *Psychological Science*, *28*(8), 1103–1115.

Mızrak, E., & Oberauer, K. (2021). What is time good for in working memory? *Psychological Science*, *32*(8), 1325–1337.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482.

Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation* (pp. 125–173, Vol. 26). Elsevier.

Noh, E., Herzmann, G., Curran, T., & de Sa, V. R. (2014). Using single-trial eeg to predict and analyze subsequent memory. *NeuroImage*, *84*, 712–723.

Oberauer, K. (2022). When does working memory get better with longer time? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(12), 1754.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*(2), 93–102.

Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2011). Semantic cuing and the scale insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 766.

Popov, V., Marevic, I., Rummel, J., & Reder, L. M. (2019). Forgetting is a feature, not a bug: Intentionally forgetting some things helps us remember others by freeing up working memory resources. *Psychological Science*, *30*(9), 1303–1317.

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review*, *127*(1), 1–46.

Popov, V., So, M., & Reder, L. M. (2022). Memory resources recover gradually over time: The effects of word frequency, presentation rate, and list composition on binding errors and mnemonic precision in source memory. *Journal of Ex-*

perimental Psychology: Learning, Memory, and Cognition, *48*(9), 1263.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, *22*(268), 1–8.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 294–320.

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: A computational model of the encoding/retrieval trade-off with familiarity. *Psychology of Learning and Motivation*, *48*, 271–312.

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, *92*(3), 365.

Sandry, J., & Ricker, T. J. (2020). Prioritization within visual working memory reflects a flexible focus of attention. *Attention, Perception, & Psychophysics*, *82*, 2985–3004.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222.

Sundby, C. S., Woodman, G. F., & Fukuda, K. (2019). Electrophysiological and behavioral evidence for attentional up-regulation, but not down-regulation, when encoding pictures into long-term memory. *Memory & Cognition*, *47*, 351–364.

Sutton, R. S. (2018). Reinforcement learning: An introduction. *A Bradford Book*.

Talmi, D., Kavaliauskaite, D., & Daw, N. D. (2021). In for a penny, in for a pound: Examining motivated memory through the lens of retrieved context models. *Learning & Memory*, *28*(12), 445–456.

Van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *ELife*, *7*, e34963.

Xu, Z., Hemmer, P., & Zhang, Q. (2024). Towards a generalized bayesian model of reconstructive memory. *Computational Brain & Behavior*, 1–13.

Zhang, Q., Griffiths, T. L., & Norman, K. A. (2023). Optimal policies for free recall. *Psychological Review*, *130*(4), 1104.

Zhou, Z., Kahana, M. J., & Schapiro, A. C. (2023). A unifying account of replay as context-driven memory reactivation. *bioRxiv*, 2023–03.