AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 10, 2588–2604
https://doi.org/10.1037/xge0001638

First Impressions or Good Endings? Preferences Depend on When You Ask

Alyssa H. Sinclair[1, 2], Yuxi C. Wang[1], and R. Alison Adcock[1, 3]

[1] Department of Psychology and Neuroscience, Duke University
[2] Center for Science, Sustainability, and the Media, University of Pennsylvania
[3] Department of Psychiatry and Behavioral Sciences, Duke University

Rewards often unfold over time; we must summarize events in memory to guide future choices. Do first impressions matter most, or is it better to end on a good note? Across nine studies ($N = 569$), we tested these competing intuitions and found that preferences depend on when rewards occur and when we are asked to evaluate an experience. In our "garage sale" task, participants opened boxes containing sequences of objects with values. All boxes were equally valuable, but rewards were either evenly distributed or clustered at the beginning, middle, or end of the sequence. First, we tested preferences and valuation shortly after learning; we consistently found that boxes with rewards at the beginning were strongly preferred and overvalued. Object-value associative memory was impaired in boxes with early rewards, suggesting that value information was linked to the box rather than the objects. However, when tested after an overnight delay, participants equally preferred boxes with any cluster of rewards, whether at the beginning, middle, or end of the experience. Finally, we demonstrated that evaluating shortly after an experience led to lasting preferences for early rewards. Overall, we show that people summarize rewarding experiences in a nonlinear and time-dependent way, unifying prior work on affect, memory, and decision making. We propose that short-term preferences are biased by first impressions. However, when we wait and evaluate an experience after a delay, we summarize rewarding events in memory to inform adaptive longer term preferences. Preferences depend on *when rewards occur* and *when we first evaluate an experience*.

Public Significance Statement
These studies demonstrate that our preferences are influenced by how positive or negative experiences unfold over time, as well as when we are first asked to express our preferences. Shortly after an experience (as in rating a restaurant or judging a competition), preferences are driven by first impressions. However, when we wait and evaluate after a longer delay (as in returning to a restaurant or selecting a job candidate), preferences instead reflect *clusters* of rewards that occurred at any time during the experience. Our findings offer broader implications for understanding everyday experiences, including consumer choices, comparative decisions, and social interactions.

Keywords: reward, preferences, memory, valuation, decision making

Supplemental materials: https://doi.org/10.1037/xge0001638.supp

How do we summarize rewarding experiences to guide future choices? In daily life, rewards often unfold over time. For example, a restaurant outing may feature an excellent appetizer followed by a mediocre entrée and dessert. Alternatively, the meal could begin with a mediocre appetizer and entrée but end with a delicious dessert. Immediately afterward, one may leave the restaurant a positive or negative review. Later, one may recall this experience to decide whether or not to revisit the restaurant. The temporal distribution of reward may influence our preferences, evaluations, and memories. We also integrate rewards or evidence over time in other situations,

This article was published Online First September 9, 2024.
Kimberly Fenn served as action editor.
Alyssa H. Sinclair https://orcid.org/0000-0003-0447-3959
Findings from the studies reported in this article were previously presented as posters at the annual meetings of the Psychonomic Society (2022) and the Cognitive Neuroscience Society (2022). Data and code are provided in a public repository hosted by the Open Science Framework (https://doi.org/10.17605/OSF.IO/C846S). The studies were funded by a Duke University Duke Health Scholars Award to R. Alison Adcock.

Alyssa H. Sinclair played a lead role in conceptualization, formal analysis, investigation, methodology, project administration, visualization, writing–original draft, and writing–review and editing. Yuxi C. Wang played a supporting role in conceptualization, investigation, methodology, and writing–review and editing. R. Alison Adcock played a lead role in funding acquisition and supervision, a supporting role in investigation, and an equal role in writing–review and editing.

Correspondence concerning this article should be addressed to Alyssa H. Sinclair, Department of Psychology and Neuroscience, Duke University, 202 South 36th Street, Philadelphia, PA 19104, United States, or R. Alison Adcock, Department of Psychology and Neuroscience, Duke University, 308 Research Drive, Durham, NC 27710, United States. Email: sinclair.allie@gmail.com or raa@duke.edu

such as product reviews, job interviews, competition rankings, social interactions, or jury decisions. Do first impressions matter most, or is it more important to end on a good note? Diverse evidence from studies of brain and behavior set up competing predictions for this key question.

The *primacy effect* describes how the first items in a sequence tend to be preferred and remembered better. This phenomenon has been demonstrated in ballot voting behavior (van Erkel & Thijssen, 2016), belief revision (Peterson & DuCharme, 1967), and impression formation (N. H. Anderson & Norman, 1964; Sullivan, 2019). Other studies have shown enhanced memory for the first items in a sequence, such as in free recall of word lists (Murdock, 1962) and memory for Super Bowl commercials (Li, 2010). The first items in a sequence may be better encoded because of lower memory load and interference (Sederberg et al., 2006). If the same item is encountered repeatedly with minor changes, memory is biased toward the first version of the stimulus; first impressions can exert a disproportionate impact on memory, at the expense of subsequent information (Digirolamo & Hintzman, 1997).

There is also an extensive literature emphasizing the importance of endings. The *recency effect* describes how memory and preferences are driven by the last items in a sequence. The recency effect often diminishes after a delay but can endure under some conditions (Baddeley & Hitch, 1977). The recency effect has been demonstrated in many contexts, including free recall of word lists (Murdock, 1962), semantic and episodic memory (Healy et al., 2000; Neath, 2010), juror decisions (Furnham, 1986; Schweitzer & Nuñez, 2021), and category learning (Jones & Sieck, 2003).

In the emotional memory domain, the *peak–end rule* describes how retrospective evaluations of affective experiences are defined by specific moments (Fredrickson, 2000; Kahneman, 1999; Kahneman et al., 1997). When we recall and reflect on emotional events, we are biased by the strongest affective moment and the affect experienced at the end. Individuals prefer videos and commercials that have more positive affect at the peaks and ends (Baumgartner et al., 1997; Fredrickson & Kahneman, 1993). Likewise, aversive experiences (e.g., distressing videos, cold pressor tests, and painful medical procedures) are retrospectively perceived as more negative if the peaks and ends were worse (Fredrickson & Kahneman, 1993; Kahneman et al., 1993; Redelmeier & Kahneman, 1996). Extending an experience to end on a positive or negative note biases subsequent evaluations (Do et al., 2008; Kahneman et al., 1993, 1997; Redelmeier & Kahneman, 1996).

Neural and behavioral studies have shown that rewards capture attention, enhance memory encoding, and trigger the release of dopamine to support memory formation in the hippocampus (Dickerson & Adcock, 2018; Murty & Dickerson, 2016; Shohamy & Adcock, 2010). Importantly, rewards can bias preferences for items that were indirectly associated with reward, even in the absence of explicit memory (Wimmer & Shohamy, 2012). Rewards can also influence the organization of items in memory, structuring free recall by reward categories rather than temporal order (Horwath et al., 2023). Larger, more salient rewards elicit greater phasic dopamine responses and memory enhancements (Miendlarzewska et al., 2016). Importantly, reward can induce memory benefits that extend to stimuli encountered *before* or *after* the reward. Therefore, encountering salient rewards at any point during an episode may enhance subsequent memory. In some cases, the effects of reward on memory only emerge after a delay (e.g., Murayama & Kitagami, 2014; Patil et al., 2017). Reward can enhance hippocampal replay and memory consolidation during wakeful rest or sleep (Ambrose et al., 2016; Atherton et al., 2015; Murayama & Kitagami, 2014). Taken together, these prior findings suggest that salient rewards may bias memory and preferences, regardless of whether the rewards occurred at the beginning, middle, or end of an episode. However, prior studies have not tested whether the *same amount* of reward can exert different effects on memory, depending on whether those rewards are evenly distributed or presented in dense clusters.

The separate literatures described above generate distinct hypotheses about preferences over time. The peak–end rule predicts that late rewards should influence preferences regardless of the delay between the initial experience and the retrospective evaluation (Fredrickson, 2000; Fredrickson & Kahneman, 1993; Redelmeier & Kahneman, 1996). The primacy effect predicts that early rewards should drive preferences, both immediately and after a delay (Mantonakis et al., 2009; Rey et al., 2020). The recency effect predicts a bias toward late rewards, likely most apparent immediately (Baddeley & Hitch, 1977). The memory consolidation literature predicts that salient rewards at any point within an episode should bias memory, but effects may not emerge until after an overnight delay (Miendlarzewska et al., 2016).

Across nine studies, we tested these competing predictions by varying the temporal distribution of reward in equally valuable episodes and then assessing preferences, valuation, and memory. We conceptualize preferences as choices that are informed by our memories of past experiences, though preferences may be distinct from explicit memory for value or other details of those experiences. To foreshadow, we found that early rewards strongly biased short-term preferences and value estimation. However, when participants were tested after a delay that permitted consolidation, *clusters* of reward (in the beginning, middle, or end of an experience) determined longer term preferences. Crucially, preferences depended on when participants were *first* asked to evaluate the episodes—expressing preferences shortly after a learning experience led to a lasting bias in favor of early rewards.

Study 1

Method

Participants

We recruited participants from Prolific, an online platform for recruiting paid participants. We chose to recruit from Prolific for several reasons: (a) online testing offered a safe and efficient option during the pandemic, (b) the Prolific participant pool is more diverse than other online testing platforms, (c) participants are vetted to prevent the use of automated bots, and (d) Prolific participants perform better on attention checks, comprehension checks, and measures of honesty and effort (Peer et al., 2017, 2021).

Inclusion criteria were as follows: fluent in English, currently residing in the United States, normal or corrected-to-normal vision, and no psychiatric or neurological diagnoses. Demographic information was obtained from Prolific prescreening data; in these prescreening questions, participants were asked to report their age in years, sex (as recorded on legal documents), and race (simplified categories specified by Prolific). These inclusion criteria and demographic prescreening questions apply to Studies 1–9.

In Study 1, the mean age of participants was 25.48 years ($SD = 7.70$). The sample consisted of 50% women and 50% men. The racial distribution was as follows: 67% White, 22% Black, 6% Asian, and 5% mixed/other. Participants were compensated with a base rate of $5.80 for completing the study, which took approximately 35 min. In addition, participants received a bonus payment of $5. All studies were approved by the Duke University Campus Institutional Review Board (Protocol No. 2022-0031).

The target sample size was determined by a power analysis conducted with G*Power software (Faul et al., 2007); we estimated that a sample of 42 participants would yield 95% power to detect a medium-sized effect in an analysis of variance (ANOVA) with three within-subjects conditions. We anticipated that some participants may be excluded for failing attention checks, so we aimed to recruit 50 participants for each study. Prolific occasionally recruits additional subjects beyond the target sample if some subjects do not finish the task in the expected time frame. Therefore, some studies included more than 50 participants.

We determined a priori that all data would be excluded from participants who failed two or more attention checks during the task; one attention check occurred during each of the six runs of the task. In Study 1, we excluded all data from one participant who failed two attention checks. The final sample for Study 1 consisted of 52 participants.

Procedure

Garage Sale Task. The task was programed with PsychoPy Version 2021.2.3 (Peirce et al., 2019) and hosted by Pavlovia, an online server for running experiments. The task was described as a "garage sale game" in which participants opened boxes in their garage in search of high-value items that they could sell at a garage sale.

After providing informed consent, participants read instructions that stated:

> Imagine that you are searching through boxes to find objects that you can sell at a garage sale. On each round, choose to open either the left box or the right box. You will view a sequence of objects from inside the box. You will also see the sell value of each object. Some objects will be very cheap because they are in poor condition. Other objects will be worth more money. Choosing boxes with high-value objects earns you money! The value of the objects will become a real cash bonus added to your payment at the end of the study.

At the beginning of each run of the garage sale task, participants viewed two pictures of distinctive boxes and chose one box to open (Figure 1A, left). A randomized sequence of 20 trial-unique objects was then displayed on the screen (Figure 1A, middle). Each object was labeled with value information below the image. Objects and values were presented for 3.5 s each, with a 1-s fixation cross between each stimulus. An icon depicting the current box was presented at the top of the screen throughout the entire sequence. After finishing the sequence, participants were informed that there were no more items in the box. Participants then began the next run by viewing a new pair of boxes; the process was repeated for six runs in total. All box and object images were presented in a randomized order for each participant.

Importantly, although we gave participants the illusion of agency, choices did not actually influence earnings. Unbeknownst to participants, each box was worth $5 in total, but we manipulated the temporal distribution of reward in each box (Figure 1A, right). Each box contained five high-value objects (each worth approximately $0.80) and 15 low-value objects (each worth approximately $0.10). Although the distinction between high-value and low-value objects is essentially binary, we included minor variations in the values (up to $0.05 above or below the high-value and low-value anchors) to make it difficult for participants to sum the values of the objects within each box.

We used six predetermined reward sequences: two sequences featured a cluster of five high-value objects at the beginning (*Early boxes*), two sequences featured a cluster of five high-value objects at the end (*Late boxes*), and two sequences evenly distributed the five high-value objects throughout (*Even boxes*). The order of these predetermined reward sequences was randomized across the six runs of the task for each participant. Although the value sequences within each box type were predetermined, the box-value associations and object-value association values were fully randomized. Note that Figure 1A also depicts a box with a cluster of high-value objects in the middle of the sequence (*Middle boxes*); this box type was not included in Study 1, but was included in subsequent studies.

We also included an attention check during each run. Participants were instructed to look for a dog who would occasionally enter the garage to interfere with the search. The dog appeared once during each box sequence, between trials. Participants had a 3-s window to make a keyboard response when the dog appeared. If a participant did not respond in time, they failed the attention check and the task paused to display a warning and a reminder about paying attention. Resuming the task after the warning required a keyboard response. To ensure that the attention check would not disrupt a cluster of high-value objects, the dog appeared on a random trial between objects 7 and 14 in each sequence.

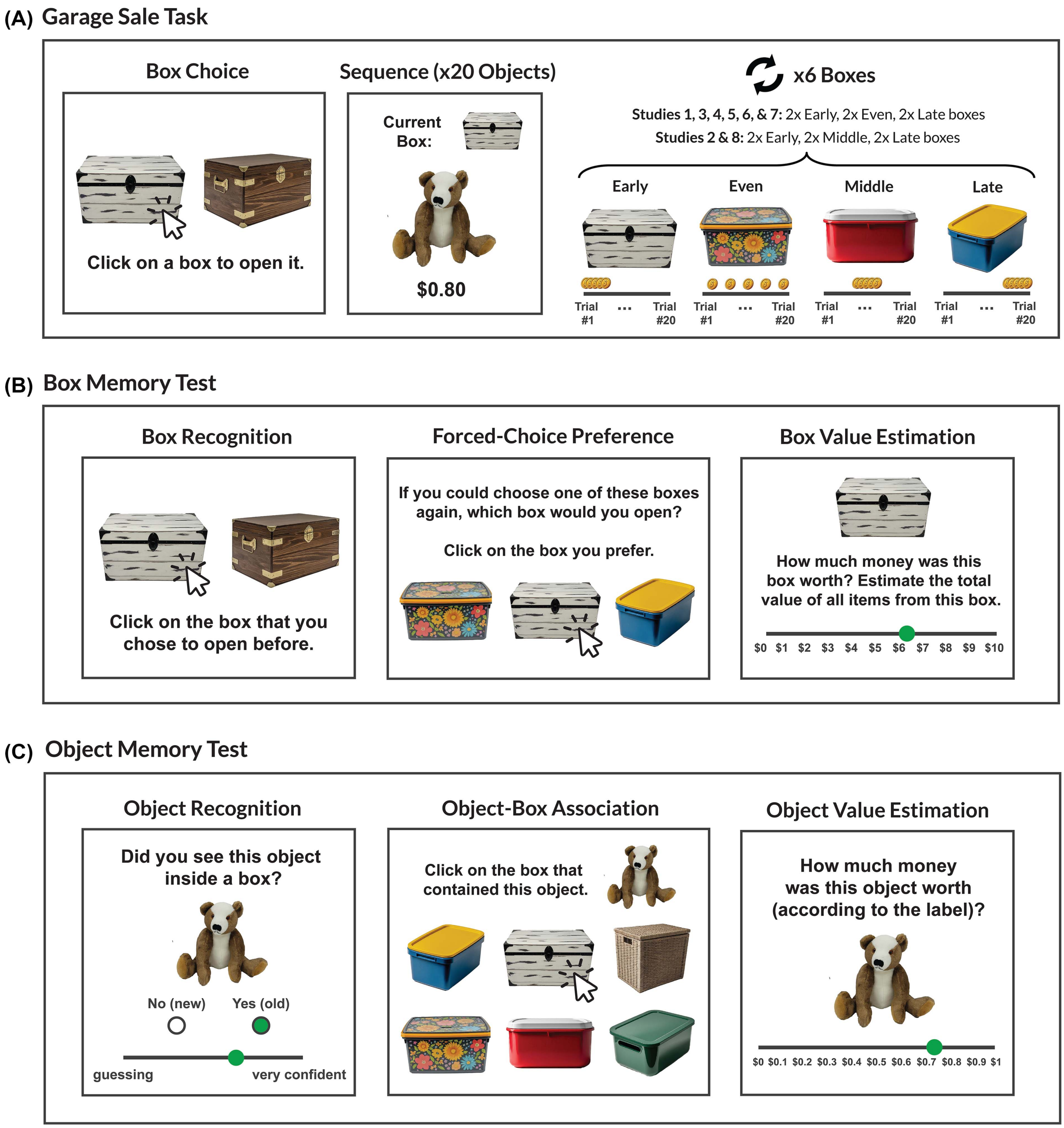
Test Phase. After completing the six runs of the garage sale task, participants were asked to answer questions about the boxes. First, participants viewed pairs of boxes (the same pairs shown at the beginning of each run) and identified the box that they had chosen during the garage sale task (Figure 1B, left). We then assessed forced-choice preference by displaying three images of previously chosen boxes (an Early box, an Even box, and a Late box). Participants were asked to choose the box that they would prefer to open again (Figure 1B, middle). We then repeated the forced-choice preference question with the other three chosen boxes. Finally, participants used a sliding scale (ranging from $0 to $10) to estimate the total value of all objects from inside each box (Figure 1B, right).

The second portion of the test assessed memory for the objects. On each trial, participants viewed one object image, reported whether it was old or new, and rated their confidence on a sliding scale from 0 (*guessing*) to 1 (*very confident*; Figure 1C, left). If a participant reported that an object was old, we then asked them to report which box had been associated with the object (Figure 1C, middle) and recall the value of the object (Figure 1C, right). Participants were instructed to recall the value of the object that had been shown during the garage sale task, but not to estimate the real-world value of the object.

Stimuli

The garage sale task stimuli consisted of 12 images of distinctive boxes and 120 images of household objects (e.g., a lamp, a calculator, a teddy bear). All images were sourced online with Google Images, edited to remove backgrounds, and resized/cropped to ensure consistent dimensions. The test stimuli included 80 novel images of household objects, in addition to all the stimuli used during the garage sale task.

Figure 1
Overview of Paradigm



Note. (A) At the start of each run of the garage sale task, participants chose one of two distinctive boxes to open. They then viewed a sequence of 20 trial-unique objects inside the box, each labeled with a sell value. The process was repeated until participants had opened six boxes in total. Unbeknownst to participants, all boxes were worth $5 in total, but the temporal distribution of high-value objects varied. Two of the six boxes had rewards clustered at the beginning (Early boxes), two boxes had rewards evenly distributed throughout (Even boxes), and two boxes had rewards clustered at the end (Late boxes). The reward sequences were predetermined, but the order of the box types and objects was randomized. (B) During the box memory test (either immediately after the Garage Sale Task or after a 1-day delay), participants recognized boxes, chose their preferred boxes, and estimated the value of each box. (C) During the object memory test, participants recognized objects, recalled object-box associations, and estimated the value of each object. See the online article for the color version of this figure.

Each object image belonged to a unique subordinate category, but some images shared a basic-level category (e.g., a basketball and a beach ball).

Statistical Analysis

The following information about statistical analysis applies to Studies 1–9. All analyses were conducted with R (v4.1.1) in RStudio (v2021.09.0). Mixed-effects analyses were conducted with the *lme4* package (Bates et al., 2014). Effect sizes were obtained with the packages *effectsize* (Ben-Shachar et al., 2020) and *rcompanion* for chi-squared tests (Mangiafico, 2022). Follow-up pairwise comparisons were conducted with the packages *emmeans* (Lenth, 2021) and *RVAideMemoire* for chi-squared tests (Hervé, 2022). Plots were produced with the packages *ggplot2* (Wickham, 2016) and *sjPlot* (Lüdecke, 2021). In accordance with current best practices for mixed-effects modeling, we included random intercepts for subjects and random slopes for all fixed effects when appropriate (Bates et al., 2014; Matuschek et al., 2017). Random slopes that explained very little variance were removed from the models in cases of convergence failure or overfitting (singular fit). Information about the random effects included in each mixed-effects model is provided in the corresponding results tables within the Supplemental Material. Follow-up pairwise comparisons (e.g., among the three box types) were corrected for multiple comparisons with Tukey's honestly significant difference (HSD) method.

Transparency and Openness

Data and analysis code associated with all studies are provided in a permanent, public repository hosted by the Open Science Framework (https://osf.io/c846s/; Sinclair, 2023). The studies were not preregistered. Additional materials (e.g., stimulus images and task program) are available from the first author by reasonable request.

Results

Box Memory and Valuation

Average recognition memory for chosen boxes was very accurate (M = 92.95% correct), confirming that participants attended to the task. Box recognition accuracy did not differ across box types (Early, Even, or Late) when compared with a repeated-measures ANOVA, $F(2, 102) = 0.05$, $p = .952$, $\eta_p^2 < 0.001$, 95% CI [0.00, 0.01].

We then tested whether preferences depended on the box type. Using a chi-squared goodness-of-fit test, we compared the observed responses to the forced-choice preference test with the null hypothesis that preferences should be evenly distributed across the three box types. We found that participants were biased to prefer Early boxes over Even and Late boxes (Figure 2A), with a medium effect size, $\chi^2(2) = 11.85$, $p = .003$, Cramér's $V = 0.24$, 95% CI [0.11, 0.38].

We then compared value estimation across box types. We calculated *value estimation error* scores by subtracting $5 (the true value of each box) from each value estimate. Using a repeated-measures ANOVA, we compared estimates across box types. There was a significant effect of box type on value estimation error, with a medium effect size, $F(2, 102) = 9.84$, $p < .001$, $\eta_p^2 = 0.16$, 95% CI [0.05, 0.29]. We conducted follow-up pairwise comparisons, adjusted for multiple comparisons with Tukey's HSD, and found that the estimated value of Early boxes was significantly greater than the estimated value of Even boxes ($t = 3.39$, $p = .003$, Cohen's $d = 0.67$, 95% CI [0.27, 1.07]) and Late boxes ($t = 4.17$, $p < .001$, Cohen's $d = 0.83$, 95% CI [0.42, 1.23]); there was no difference between Even and Late boxes ($t = 0.78$, $p = .71$, Cohen's $d = 0.16$, 95% CI [−0.23, 0.54]). Participants falsely believed that Early boxes were more valuable than Even and Late boxes. We also used post hoc analyses to test the accuracy of value estimation for each box type. Participants significantly overestimated the value of Early boxes; on average, participants estimated that Early boxes were worth $5.48 ($t = 2.02$, $p = .047$, Cohen's $d = 0.40$, 95% CI [0.02, 0.79]). In contrast, participants accurately estimated the value of Even and Late boxes; value estimation error scores did not differ from zero (Even: $t = 0.28$, $p = .779$, Cohen's $d = 0.06$, 95% CI [−0.33, 0.44]; Late: $t = -0.03$, $p = .904$, Cohen's $d = 0.02$, 95% CI [−0.36, 0.41]). Overall, we found that participants overestimated the value of Early boxes and falsely believed that Early boxes were more valuable than Even and Late boxes, despite all boxes being equally valuable.

Object Memory

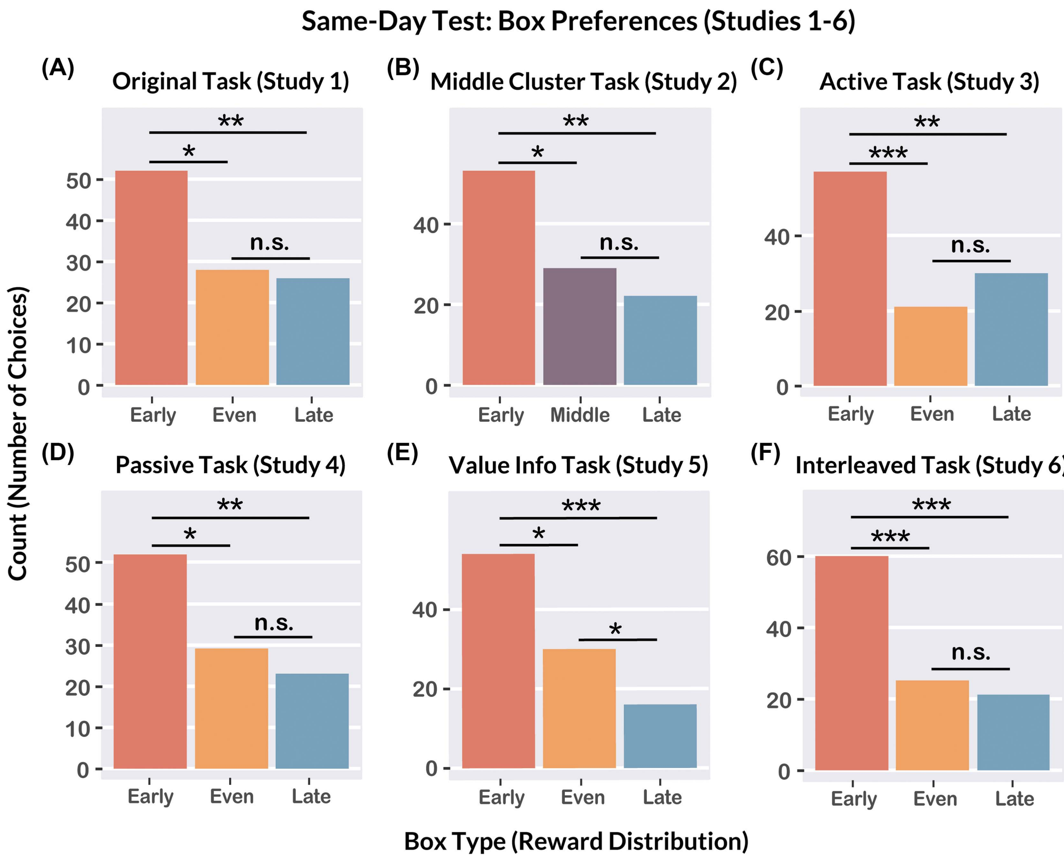
The average accuracy on the object memory test was 75.74%, substantially above the chance performance level of 50%, $t(51) = 19.37$, $p < .001$, $d = 2.69$, 95% CI [2.12, 3.30]. We used a binomial generalized mixed-effects model to predict trial-by-trial *recognition accuracy* from the variables *box type* (Early, Even, or Late), *object value* (low or high), and the interaction term. The model also included a covariate of no interest for the *learning trial number* (to account for fatigue effects). There was a significant effect of object value on memory; high-value objects were more likely to be remembered ($\beta = 0.34$, 95% CI [0.16, 0.51], $z = 3.76$, $p < .001$, Cohen's $d = 0.36$). There was no effect of box type, nor did box type interact with object value, on object recognition memory. Additional model statistics are reported in Supplemental Table 1.

Next, we investigated memory for object-value associations. We defined "high-value hits" as trials in which the participant correctly identified that a high-value object was high-value (estimated value ≥ $0.70, capturing all estimates close to the true value of high-value objects). In this analysis, we treat the object value estimates as a binary variable because the true object values were binary (with minor variations to make it difficult for participants to sum the values during the task). The distributions of true object values and participants' value estimates are both bimodal, as shown in Supplemental Figure 1.

Using a binomial generalized mixed-effects model, we predicted high-value hits/misses (1 or 0) from the box type and learning trial number (covariate of no interest). There was a significant effect of box type on high-value hits, $\chi^2(2) = 15.55$, $p < .001$ (Figure 3C). Follow-up pairwise comparisons (adjusted for multiple comparisons with Tukey's HSD) indicated that object-value association memory was significantly worse for Early boxes relative to Late boxes ($\beta = -0.70$, 95% CI [−1.12, −0.27], $z = -3.40$, $p < .001$, Cohen's $d = -0.70$) and Even boxes ($\beta = -0.46$, 95% CI [−0.88, −0.04], $z = -2.59$, $p = .026$, Cohen's $d = -0.46$). Object-value association memory did not significantly differ between Even and Late boxes ($\beta = -0.23$, 95% CI [−0.66, 0.19], $z = -1.30$, $p = .398$, Cohen's $d = -0.23$). Additional model statistics are reported in Supplemental Table 2.

Finally, we tested memory for object-box associations. Overall, object-box association memory performance was poor (M = 24.29%; chance level was 16.67%). Using a binomial generalized mixed-effects model, we predicted trial-by-trial object-box association accuracy

Figure 2
Same-Day Test: Box Preferences (Studies 1–6)



Note. The figure compares preferences for different box types across Studies 1–6. We consistently found that participants disproportionately preferred Early boxes over Even/Middle and Late boxes, across all studies. Black horizontal lines indicate pairwise comparisons conducted after the global chi-squared goodness-of-fit test for each study, adjusted by the false discovery rate to account for multiple comparisons. n.s. = not significant. See the online article for the color version of this figure.
* $p < .05$. ** $p < .01$. *** $p < .001$.

from the variables *box type*, *object value*, the interaction term, and the learning trial number (covariate of no interest). Note that this analysis necessarily only includes trials in which participants correctly recognized old objects. There were no significant effects of interest; statistics for all model terms are reported in Supplemental Table 3.

Overall, we found that participants preferred and overestimated the value of Early boxes. Despite this bias, they were less likely to successfully remember which objects had been high-value in the Early boxes. Together, these findings suggest that at the beginning of an episode, value information may be associated with the broader context (the box) rather than specific details (the objects). In subsequent studies, we sought to replicate these findings and investigate potential boundary conditions.

Studies 2–6

Method

Participants

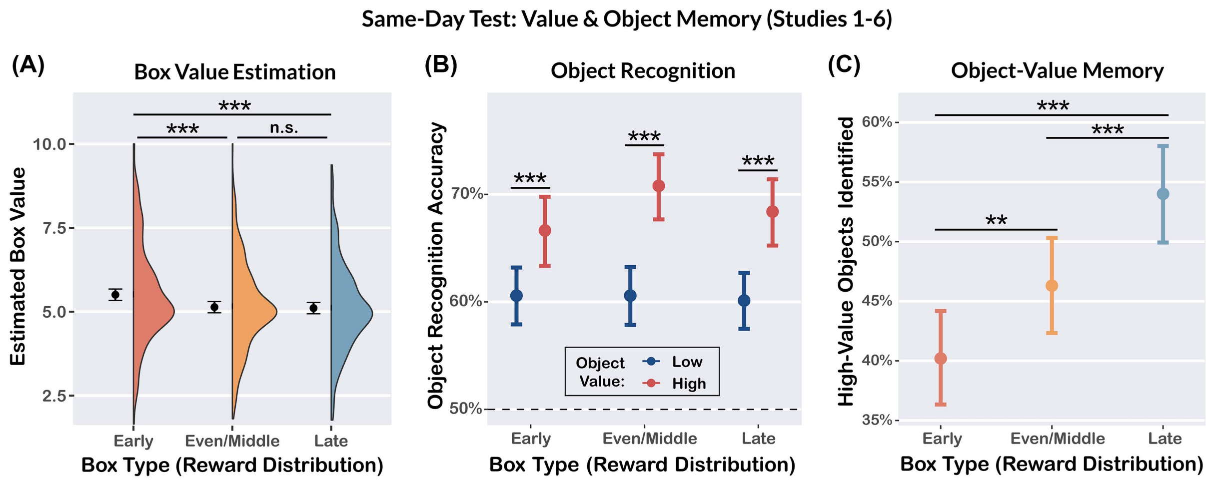
Recruitment and inclusion criteria were consistent with Study 1. Across Studies 2–6, the mean age of participants was 30.83 years (*SD* = 11.63). The sample consisted of 53% women and 47% men. The racial distribution was as follows: 71% White, 9% Black, 6% Asian, 8% mixed race, and 6% other.

We aimed to collect approximately 50 participants per study, as per the power analysis described previously. Using the same exclusion criteria as in Study 1, we excluded all data from participants who failed two or more attention checks (zero participants in Study 2, six participants in Study 3, three participants in Study 4, three participants in Study 5, and two participants in Study 6). Additionally, we excluded one participant in Study 3 who aborted and then restarted the task before finishing. After exclusions, the final sample sizes were as follows: Study 2 $N = 51$, Study 3 $N = 48$, Study 4 $N = 49$, Study 5 $N = 47$, and Study 6 $N = 51$.

Procedure

Studies 2–6 each tested minor variations on the paradigm described in Study 1. We aimed to replicate our key findings, investigate possible boundary conditions, and rule out potential confounding variables. To this end, we compared each sample with the original Study 1 sample. A summary of all task versions is provided in Table 1.

Figure 3
Same-Day Test: Value and Object Memory (Studies 1–6)



Note. The figure depicts box and object memory results combined across Studies 1–6. Error bars indicate 95% confidence intervals. All pairwise contrasts were corrected for multiple comparisons with Tukey's honestly significant difference method. (A) Average estimated value of boxes. Participants significantly overestimated the value of Early boxes (within-subjects effect). Black points represent estimated means from a linear mixed-effects regression model. Colored shapes depict the distribution of the raw data. (B) Predicted trial-wise object recognition accuracy from a mixed-effects logistic regression model. Across all box types, participants were more likely to recognize high-value objects than low-value objects. Recognition accuracy did not differ across box types. (C) Predicted trial-wise object-value association memory from a mixed-effects logistic regression model. Participants were less successful at recalling that high-value objects from Early boxes had been high value. n.s. = not significant. See the online article for the color version of this figure.
** $p < .01$. *** $p < .001$.

In Study 2 (Middle Cluster task), we replaced the Even boxes (high-value objects evenly distributed throughout the sequence) with Middle boxes (a cluster of high-value objects in the middle of the sequence). The purpose of this version was to test the effect of a reward peak in the middle of an experience; this version of the task isolates the effect of reward placement while controlling for reward density across all conditions.

In Study 3 (Active task), we used the same paradigm as in Study 1 but modified the instructions shown prior to the task. We emphasized agency and future utility in the instructions, encouraging participants to choose rewarding boxes and implying that chosen boxes would be relevant later for future choices. The purpose of this version was to test whether expectations about agency and future choices influenced preferences.

In Study 4 (Passive task), we modified the Study 1 paradigm to remove the illusion of agency. At the start of each run, participants viewed only one box and were instructed to click on it (instead of choosing between two boxes). The purpose of this version was to test whether our effects depended on choice or would generalize to a passive viewing experience.

In Study 5 (Value Information task), we used the same paradigm as in Study 1 but modified the instructions shown prior to the task. In

Table 1
Summary of Task Versions Across the Eight Studies

Study no.	Task version	Task description	*N*
1	Original	Paradigm described in detail in the Study 1 Method section	52
2	Middle Cluster	Replaced Even boxes with Middle boxes (reward cluster in the middle of sequence)	51
3	Active	Instructions emphasized choice, agency, and future utility of learning about boxes	48
4	Passive	Removed choices; passive viewing of boxes/objects	49
5	Value Information	Set reward expectations by specifying the average value of low- and high-value objects	47
6	Interleaved Value Estimation	Box value estimation immediately after each box, and all other test questions at the end of the session	51
7	Next Day Test	Same as Study 1, but with a 1-day delay-to-test	48
8	Next Day Test, Middle Cluster	Same as Study 2, but with a 1-day delay-to-test	51
9	Same-Day + Next-Day Test	Same as Study 7, but with the three boxes tested same-day and three boxes tested next-day	172/158

Note. *N*s for Study 9 represent same-day/next-day tests, respectively.

Early boxes, participants may experience high-value objects before learning that low-value objects are worth much less. Therefore, we explicitly informed participants about the range of approximate values of the low-value and high-value objects that would appear in the upcoming task. We modified the instructions to state, "Many objects will be very cheap (about 10 cents) because they are in poor condition. A few special objects will be worth more money (about 80 cents)." The purpose of this version was to test whether reward expectations influence preferences.

Finally, in Study 6 (Interleaved Value Estimation task), we modified the Study 1 paradigm to have participants estimate the value of each box immediately after viewing the sequence of objects inside. The box preference test and object memory test still occurred at the end of the session after viewing all six boxes. The purpose of this version was to test whether participants would still overestimate the value of the Early boxes if asked immediately after the experience, rather than when retrospectively comparing all six boxes.

Studies 1–6 all tested preferences, valuation, and memory on the same day as the learning experience.

Results

Study 1 yielded three key findings: (1) participants preferred to choose Early boxes again, (2) participants overestimated the value of Early boxes, and (3) participants were less successful at identifying the high-value objects from Early boxes. Here, we focus on these three key analyses in each study. Additional analyses of object recognition memory and object-box association memory are reported in Supplemental Tables 5 and 6.

Box Preferences and Memory

Consistent with Study 1, box recognition accuracy was high across all studies (Study 2 $M = 92.48\%$, Study 3 $M = 91.13\%$, Study 5 $M = 94.33\%$, and Study 6 $M = 92.16\%$). High performance on box recognition confirmed that participants were attending to the task. Note that in Study 4 (Passive task), participants were not asked to identify their chosen boxes because the box choice component of the task was removed.

We found that the bias in favor of Early boxes was remarkably robust (Figure 2). The forced-choice preference test revealed that participants significantly preferred Early boxes over Late and Even/Middle boxes in every study, Study 2: $\chi^2(2) = 13.47$, $p = .001$, Cramér's $V = 0.26$, 95% CI [0.12, 0.40]; Study 3: $\chi^2(2) = 19.50$, $p < .001$, Cramér's $V = 0.30$, 95% CI [0.17, 0.43]; Study 4: $\chi^2(2) = 13.52$, $p = .001$, Cramér's $V = 0.25$, 95% CI [0.12, 0.40]; Study 5: $\chi^2(2) = 22.16$, $p < .001$, Cramér's $V = 0.33$, 95% CI [0.21, 0.46]; Study 6: $\chi^2(2) = 26.06$, $p < .001$, Cramér's $V = 0.35$, 95% CI [0.32, 0.49]. Overall, we strongly replicated this key finding five times. Crucially, we demonstrated that the bias to prefer boxes with Early rewards did not depend on the density of reward in the middle of the sequence (Study 2), agency and beliefs about future utility (Studies 3 and 4), expectations about reward values (Study 5), or the delay period before estimating values (Study 6).

Next, we compared box value estimation error across studies (Figure 3A), using a mixed ANOVA with box type as a within-subjects factor and task version as a between-subjects factor. To test whether each manipulation produced results that were significantly different from those observed in Study 1, we included the task version (Study 1, 2, 3, 4, 5, or 6) as a factor variable with Study 1 as the reference level. A significant interaction between task version and box type would indicate that the effect of box type on value estimation differed across studies. In this analysis, Middle boxes (Study 2) were grouped with Even boxes (Studies 1, 3, 4, 5, and 6).

Across all studies, we replicated the effect of box type on value estimation error, $F(2, 582) = 29.66$, $p < .001$, $\eta_p^2 = 0.09$, 95% CI [0.05, 0.14]. Follow-up comparisons (corrected with Tukey's HSD) indicated that participants estimated that Early boxes were more valuable than Even/Middle boxes ($t = 6.29$, $p < .001$, Cohen's $d = 0.52$, 95% CI [0.36, 0.69]) and Late boxes ($t = 7.05$, $p < .001$, Cohen's $d = 0.58$, 95% CI [0.42, 0.75]). Estimated value did not differ between Even/Middle and Late boxes ($t = 0.76$, $p = .729$, Cohen's $d = 0.06$, 95% CI [−0.10 0.23]). There was no significant main effect of task version, $F(4, 291) = 0.37$, $p = .869$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.01], nor an interaction between task version and box type, $F(8, 582) = 0.75$, $p = .676$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.02]. We consistently found that participants falsely believed that Early boxes were more valuable than Even and Late boxes. Additional statistics comparing value estimation across box types within each study separately are provided in Supplemental Table 4.

We also used post hoc analyses to test the accuracy of value estimation (comparing the estimated marginal mean for each box type to zero). In the combined data from Studies 1 to 6, we again found that participants overestimated the value of Early boxes ($t = 6.47$, $p < .0001$, Cohen's $d = 0.54$, 95% CI [0.37, 0.70]). On average, participants estimated that Early boxes were worth $5.53, more than the true value of $5. Participants also slightly overestimated the value of Even boxes by $0.16 ($t = 1.98$, $p = .048$, Cohen's $d = 0.16$, 95% CI [0.00, 0.33]), but they did not overestimate the value of Late boxes ($t = 1.44$, $p = .151$, Cohen's $d = 0.12$, 95% CI [−0.04, 0.28]).

In sum, we found that across all task versions, participants overestimated the value of Early boxes and preferred Early boxes. These effects were not influenced by any of the potential boundary conditions that we investigated.

Object Memory

Consistent with Study 1, object recognition accuracy was substantially above chance performance (50%) in all studies (Study 2 $M = 72.22\%$, Study 3 $M = 72.65\%$, Study 4 $M = 74.15\%$, Study 5 $M = 74.28\%$, Study 6 $M = 72.57\%$). Object recognition performance in Studies 1–6 is visualized in Figure 3B. Additional analyses pertaining to object recognition memory and object-box association memory are reported in Supplemental Tables 5 and 6.

Next, we tested memory for object-value associations across studies (Figure 3C). As in Study 1, we used a binomial generalized mixed-effects model to predict high-value hits (correctly identifying high-value objects as high-value) from the variables *box type* (Early, Even/Middle, or Late), *task version* (Study 1, 2, 3, 4, 5, or 6), the interaction term, and learning trial number. Across all studies, we found a main effect of box type on object-value association memory, $\chi^2(2) = 50.76$, $p < .001$. Follow-up pairwise comparisons (adjusted for multiple comparisons with Tukey's HSD) indicated that object-value association memory was significantly worse for Early boxes relative to Late boxes ($\beta = -0.56$, 95% CI [−0.74, −0.37], $z = -7.12$, $p < .001$, Cohen's $d = -0.56$, 95% CI [−0.71, −0.40]), and Even/Middle boxes ($\beta = -0.25$, 95% CI [−0.43, −0.07], $z = -3.28$, $p = .003$, Cohen's $d = -0.25$, 95% CI [−0.40, −0.10]). Object-value association memory was significantly worse for Even/Middle boxes than Late boxes ($\beta = -0.31$, 95% CI [−0.49, −0.13], $z = -4.02$, $p <$ .001, Cohen's $d = -0.31$, 95% CI [−0.46, −0.16]). There was no significant main effect of task version, $\chi^2(5) = 8.15$, $p = .148$, nor an interaction between task version and box type, $\chi^2(10) = 13.49$, $p =$ .197. Additional model statistics are provided in Supplemental Table 7. In an exploratory analysis, we also investigated whether object-value association memory accuracy was related to the overestimation of value in Early boxes (i.e., suggesting a within-subjects trade-off between these two memory measures); there were no significant effects in any task versions (Supplemental Material, *Object-Value and Box-Value Associations*).

Overall, we observed that object-value association accuracy was lowest for Early boxes and highest for Late boxes, with Even/Middle boxes at an intermediate level. The effects observed in Study 1 were consistent across task versions, with no significant differences among the studies.

Interim Discussion

In Studies 1–6, participants consistently preferred and overestimated the value of Early boxes; these effects did not differ across task versions. Memory for object-value associations in Early boxes was impaired, suggesting that value information was associated with the broader context (the box) instead of specific events or details (the objects). Overall, our results did not differ depending on the density of reward in the middle of the sequence (Study 2), the active versus passive nature of the task (Studies 3 and 4), expectations about reward values (Study 5), or whether value estimation was interleaved or retrospective (Study 6).

Prior studies have shown that salient rewards can enhance memory for temporally adjacent stimuli, but only after a delay that allows for memory replay and consolidation (Ambrose et al., 2016; Cowan et al., 2021; Miendlarzewska et al., 2016; Murayama & Kitagami, 2014; Patil et al., 2017; Shohamy & Adcock, 2010). Although these prior studies have not directly contrasted dense and spaced rewards, we expected that dense clusters of reward may be especially salient, potentially eliciting a greater phasic dopamine response within a short time window and thus increasing the likelihood that rewards would enhance memory and bias future choices. We predicted that after a delay, participants would prefer episodes that included a *cluster* of rewards, regardless of whether a cluster occurred at the beginning, middle, or end of an experience. To test this prediction, in Studies 7 and 8, we investigated preferences after an overnight delay.

Studies 7 and 8

Method

Participants

Recruitment and inclusion criteria were consistent with previous studies. The average age of participants was 39.93 years ($SD = 13.57$). The sample consisted of 56% women and 44% men. The racial distribution was as follows: 86% White, 3% Black, 5% Asian, 5% mixed race, and 3% other. As in the previous studies, we aimed to collect approximately 50 participants per study. In Study 7, 10 participants did not return for Session 2, and two participants were excluded for failing two or more attention checks. Therefore, we also collected a top-up sample with an additional 10 participants to compensate for high attrition. In Study 8, one participant did not return for Session 2, and no participants failed two or more attention checks. After exclusions, the final sample sizes were 48 participants for Study 7 and 51 participants for Study 8. Participants were compensated with $1.67 for completing Session 1 (approximately 10 min duration) and $5.13 for completing Session 2 (approximately 25 min duration). The $5 bonus payment was awarded after completing Session 2.

Procedure

In Studies 7 and 8, we investigated preferences after an overnight delay. We introduced a delay to permit memory consolidation by separating the paradigm into two sessions; instead of immediately after the garage sale task, the test phase occurred after a 1-day delay. In Study 7 (Next Day Test task), we used the same paradigm as in Study 1 (Original task), but with a 1-day delay between the garage sale task and the memory test. We expected that after a delay, dense *clusters* of reward would drive preferences; we predicted that Early boxes would still be preferred over Even boxes, but Late boxes would also be preferred over Even boxes after a delay. In other words, when asked to evaluate after a delay, participants would no longer exclusively prefer Early boxes.

In Study 8 (Next Day Test, Middle Cluster task), we introduced a delay to the paradigm used in Study 2 (Middle Cluster task), in which we replaced Even boxes with Middle boxes. We expected that if *reward density* is the key factor that drives preferences after a delay, then there should be no difference among the Early, Middle, and Late boxes, as all three of these box types feature a dense cluster of rewards at some point in the sequence.

Results

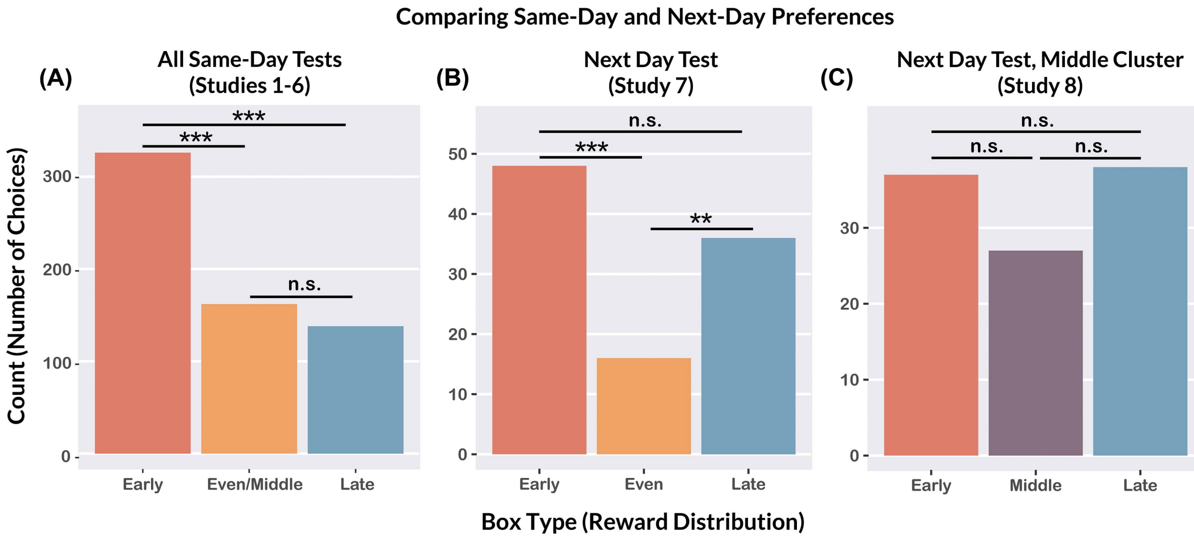
Box Preferences

Due to the 1-day delay to test, box recognition memory accuracy was slightly lower in Study 7 (87.50%) and Study 8 (88.56%), relative to previous studies. However, accuracy was still substantially above chance (50%) in both Study 7, $t(47) = 13.46$, $p < .001$, Cohen's $d = 1.94$, 95% CI [1.47, 2.45], and Study 8, $t(50) = 16.05$, $p < .001$, Cohen's $d = 2.25$, 95% CI [1.74, 2.79], indicating that participants still remembered their chosen boxes the following day.

Next, we compared preferences across box types (Figure 4). In Study 7, there was a significant difference in proportions among the three box types, $\chi^2(2) = 15.68$, $p < .001$, Cramér's $V = 0.28$, 95% CI [0.17, 0.42]. Crucially, however, the distribution across box types was distinct from the pattern observed in Studies 1–6. In all six prior studies, participants strongly preferred Early boxes; Even and Late boxes were equally unfavored (Figure 4A). In Study 7, we found that both Early and Late boxes were significantly preferred over Even boxes (Early > Even: $p < .001$, Late > Even: $p = .006$). There was no significant difference in proportions for Early and Late boxes ($p = .191$). In other words, Late boxes were initially unfavored, but a preference for Late boxes emerged only after a 24-hr delay that allowed for consolidation (Figure 4B). Participants showed a preference for Early boxes immediately, and this preference also persisted after a delay (relative to Even boxes). Directly comparing box preferences in Study 1 and Study 7 confirmed that the distributions

Figure 4

Comparing Same-Day and Next-Day Preferences



Note. Figure depicts the distribution of responses to the forced-choice preference test in Studies 1–6 (A), Study 7 (B), and Study 8 (C). After a 1-day delay, both Early and Late boxes were preferred over Even boxes (B). Preferences for Middle boxes also increased after a delay (C), indicating that clusters of reward drive preferences after a delay that permits consolidation. The patterns shown in Panels B and C were significantly different from one another ($p = .007$), and both were significantly different from the pattern shown in Panel A (A vs. B: $p = .009$, A vs. C: $p = .006$). Black horizontal lines indicate pairwise comparisons conducted after the global chi-squared goodness-of-fit test for each study, adjusted by the false discovery rate to account for multiple comparisons. n.s. = not significant. See the online article for the color version of this figure.
$^{**}p < .01$. $^{***}p < .001$.

were significantly different, $\chi^2(2) = 9.49$, $p = .009$, Cramér's $V = 0.22$, 95% CI [0.09, 0.37].

We predicted that *clusters* of reward would prioritize information in memory. With time, replay, and consolidation, these reward-related memories can be strengthened adaptively. In Study 8, we tested whether reward *density* (clustering) or *temporal placement* was the crucial factor for this consolidation-dependent change in preferences. Therefore, we replaced the Even boxes with Middle boxes in a paradigm with a 24-hr delay to test. In contrast to the previous seven studies, there was no significant difference in proportions across box types, $\chi^2(2) = 2.18$, $p = .337$, Cramér's $V = 0.10$, 95% CI [0.03, 0.24]. After a delay, preferences for Middle boxes were also boosted nearly to the level of Early and Late boxes (Figure 4C). Directly comparing box preferences in Study 7 and Study 8, we found that the distributions were significantly different, $\chi^2(2) = 9.96$, $p = .007$, Cramér's $V = 0.22$, 95% CI [0.08, 0.39]. Likewise, the pattern of box preferences in Study 8 was significantly different from the patterns observed in Study 1, $\chi^2(2) = 10.13$, $p = .006$, Cramér's $V = 0.22$, 95% CI [0.10, 0.39], and in Study 2, $\chi^2(2) = 15.62$, $p < .001$, Cramér's $V = 0.28$, 95% CI [0.14, 0.44]. Taken together, the results from Studies 7 and 8 demonstrate that clusters of reward (regardless of temporal placement) biased preferences, but only when participants were asked to evaluate after a delay that permitted memory consolidation.

Box Value Estimation

We first compared box value estimation in Study 1 (Original task) and Study 7 (Next Day task), using a mixed ANOVA with box type as a within-subjects factor and task version as a between-subjects factor. As in all previous studies, there was a significant effect of box type on value estimation error, $F(2, 198) = 9.68$, $p < .001$, $\eta_p^2 = 0.09$, 95% CI [0.02, 0.17]. There was no significant main effect of task version, $F(1, 99) = 0.37$, $p = .544$, $\eta_p^2 = 0.003$, 95% CI [0.00, 0.06], nor an interaction between task version and box type, $F(2, 198) = 0.48$, $p = .618$, $\eta_p^2 = 0.005$, 95% CI [0.00, 0.03]. Although there was no significant interaction and the pattern of results observed in Study 7 was generally consistent with Study 1, we did note that the effects in Study 7 were qualitatively weaker (Early > Late: $t = 2.18$, $p = .077$; Early > Even: $t = 1.88$, $p = .146$; Even > Late: $t = 0.30$, $p = .952$, pairwise contrasts corrected with Tukey's HSD).

Next, we conducted the same analysis but compared Study 2 (Middle Cluster task) with Study 8 (Next Day Test, Middle Cluster task). In contrast to previous studies, there was no significant main effect of box type on value estimation, $F(2, 200) = 0.70$, $p = .500$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.04]. Instead, there was a significant interaction between box type and task version, $F(2, 200) = 3.93$, $p = .021$, $\eta_p^2 = 0.04$, 95% CI [0.00, 0.10], indicating that the effect of box type was only evident in Study 2 (Middle Cluster task). There was no significant main effect of task version, $F(1, 100) = 0.00$, $p = .983$, $\eta_p^2 = 0.00$, 95% CI [0.00, 0.00].

Overall, we found that after a delay, the overestimation of Early boxes was weakened. When boxes included a cluster of rewards (as opposed to evenly distributed rewards), they were equally favored after a delay (Figure 4). Clusters of reward in Early and Late boxes were both preferred over Even boxes. Interestingly, these findings demonstrate a dissociation between explicit value estimation and revealed preferences.

Object Memory

As expected given the 1-day delay before the memory test, object recognition memory was lower in Study 7 (60.54%) and Study 8 (62.67%) than in previous studies. However, object recognition accuracy was still significantly better than chance (50%) in both Study 7, $t(47) = 8.25$, $p < .001$, Cohen's $d = 1.19$, 95% CI [0.82, 1.57], and Study 8, $t(50) = 9.02$, $p < .001$, Cohen's $d = 1.26$, 95% CI [0.90, 1.64]. Additional analyses for object recognition memory and object-box association memory are reported in Supplemental Tables 8 and 9.

Finally, we tested memory for object-value associations. Using binomial generalized mixed-effects models, we predicted high-value hits (correctly identifying high-value objects as high-value) from the variable *box type* (Early, Even/Middle, or Late), testing Study 7 and Study 8 in separate models. The models included a covariate of no interest for the learning trial number. Unlike in Studies 1–6, there were no significant effects of box type in Study 7, $\chi^2(2) = 1.22$, $p = .543$, or Study 8, $\chi^2(2) = 1.38$, $p = .501$. Additional model statistics are provided in Supplemental Table 10. Overall, we found that the effect of Early rewards impairing object-value association memory was only evident on a same-day test. After a delay, object-value association memory did not differ across box types.

Interim Discussion

In Studies 1–6, we found that participants consistently preferred and overestimated the value of Early boxes. This bias in favor of Early boxes was remarkably robust across task versions. In Studies 7 and 8, we investigated whether preferences would differ after a delay. We found that when participants were prompted to evaluate the boxes after an overnight delay, they no longer showed an exclusive bias in favor of Early boxes. In Study 7, Early boxes were still preferred over Even boxes. However, Late boxes were also preferred over Even boxes, and there was no significant difference in preferences for Early versus Late boxes. We thus reasoned that *clusters* of dense rewards, regardless of where the cluster occurred during an episode, may drive preferences after a delay. In Study 8, we tested this prediction by replacing Even boxes with Middle boxes. As predicted, in Study 8, there were no significant differences in preferences among Early, Middle, and Late boxes, as all three of these box types featured clusters of reward.

Studies 7 and 8 demonstrated that preferences differ depending on how much time elapses before an experience is recalled and evaluated. When participants were asked to evaluate their experiences immediately, first impressions drove preferences. In contrast, when participants were asked to evaluate after an overnight delay, reward density drove preferences. Importantly, however, these studies did not test whether preferences *change* over time, with repeated testing. In a final study, we investigated this question with a within-subjects manipulation of the delay-to-test.

Study 9

Method

Participants

Recruitment and inclusion criteria were consistent with previous studies. We recruited more participants than in previous studies because we planned to test preferences and valuation of boxes at two time points, including only half of the test items on each test. A power analysis indicated that a sample of 142 participants would yield 95% power to detect the effect of box type on preferences. In anticipation of attrition and exclusions, we recruited 175 participants.

The average age of participants was 43.53 years ($SD = 12.99$). The sample consisted of 49.7% women and 50.3% men. The racial distribution was as follows: 70.3% White, 15.4% Black, 7.4% Asian, 5.1% mixed race, and 1.7% other. Fourteen participants did not return for Session 2. Three participants were excluded for failing two or more attention checks. The final sample included 172 participants with same-day test data and 158 participants with next-day test data. Participants were compensated with $2.40 for completing Session 1 (approximately 12 min duration) and $4.60 for completing Session 2 (approximately 23 min duration). The $5 bonus payment was awarded after completing Session 2.

Procedure

In Study 9, we tested participants at two time points. During the same-day test, completed shortly after the garage sale task, participants first completed a recognition memory test including all chosen and unchosen boxes. We then assessed preferences for the boxes, testing only three of the six boxes. As in previous studies, participants were shown a trio consisting of one Early box, one Even box, and one Late box in a randomized order and asked to select their preferred box. Participants were also asked to estimate the total value of all objects from within each of those three boxes. The remaining three boxes were not tested during this same-day test session. Due to a technical error, box value estimation data from the same-day test were not saved and cannot be analyzed.

The following day, participants were invited to return for Session 2. The next-day test session began with an assessment of box preferences and value estimation for the three boxes that had not yet been tested. Participants then completed the object memory test, as in previous studies. All other aspects of the task were consistent with the original paradigm used in Study 1.

We expected that only Early boxes would be preferred on the same-day test (comparable to Studies 1–6), but both Early and Late boxes would be preferred over Even boxes on the next-day test (comparable to Study 7). An alternative prediction, however, was that Early boxes would be favored at both time points; the process of evaluating and expressing preferences during the same-day test could reinforce the primacy bias and make it endure over time.

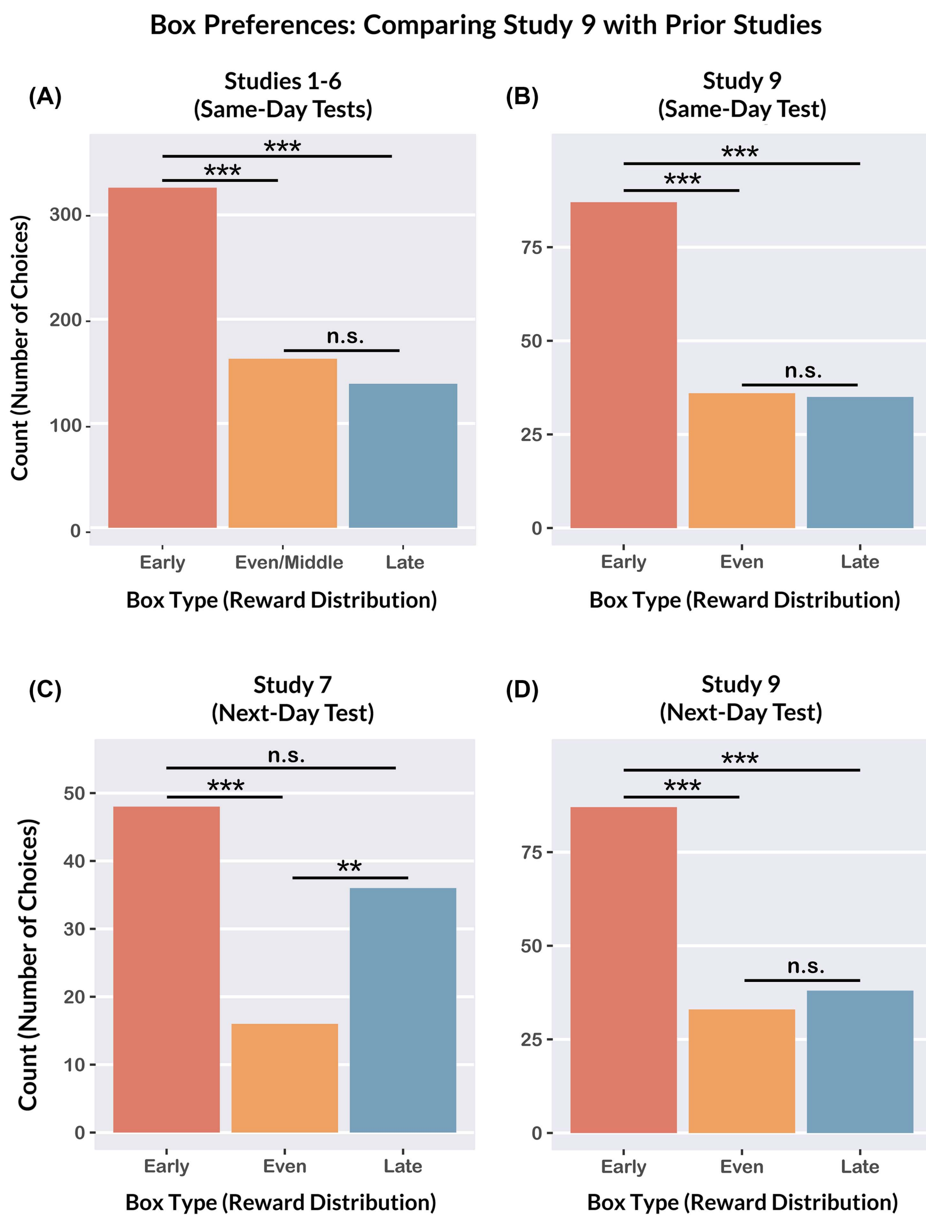
Results

Box Memory and Preferences

Box recognition accuracy was comparable to previous studies ($M = 88.7\%$). Same-day box preferences were comparable to Studies 1–6 (Figure 5A and 5B). As expected, box preferences were not evenly distributed, $\chi^2(2) = 33.58$, $p < .0001$, Cramér's $V = 0.33$, 95% CI [0.22, 0.44]. As in Studies 1–6, Early boxes were strongly favored over Even boxes ($p < .0001$) and Late boxes ($p < .0001$), and there was no difference between Even and Late boxes ($p = .91$).

Next, we investigated box preferences on the next-day test. If both Early and Late boxes were preferred over Even boxes (comparable

Figure 5
Box Preferences: Comparing Study 9 With Prior Studies



Note. The figure depicts the distribution of responses to the forced-choice preference test in the same-day tests in Studies 1–6 (A), the same-day test in Study 9 (B), the next-day test in Study 7 (C), and the next-day test in Study 9 (D). Same-day test results in Study 9 (B) were comparable to Studies 1–6 (A), replicating the bias in favor of Early boxes. Next-day test results in Study 9 (D) diverged from the next-day test results observed in Study 7 (C). In Study 9, the exclusive preference for Early boxes remained when participants were tested again after an overnight delay. Results suggest that expressing preferences immediately may lead to a lasting primacy bias (D), preventing the delayed preference for Late boxes (relative to Even boxes) observed in Study 7 (C). Black horizontal lines indicate pairwise comparisons conducted after the global chi-squared goodness-of-fit test for each study, adjusted by the false discovery rate to account for multiple comparisons. n.s. = not significant. See the online article for the color version of this figure.
** $p < .01$. *** $p < .001$.

to Study 7, see Figure 5C), this would indicate that within-subjects preferences changed from the same-day test to the next-day test. In contrast, if the exclusive preference for Early boxes remained (comparable to Study 1), this would suggest that the process of evaluating shortly after an experience influences next-day preferences, perhaps reinforcing the bias in favor of Early boxes.

Supporting this second idea, we found that the exclusive preference for Early boxes remained strong on the next-day test, $\chi^2(2) = 33.81$,

$p < .0001$, Cramér's $V = 0.33$, 95% CI [0.21, 0.45]. Participants in Study 9 continued to favor Early boxes over Even boxes ($p < .0001$) and Late boxes ($p < .0001$), and there was no difference between Even and Late boxes ($p = .55$; Figure 5D). This exclusive preference for Early boxes was consistent with the pattern observed in the previous day, $\chi^2(2) = 0.51$, $p = .776$, Cramér's $V = 0.04$, 95% CI [0.01, 0.17], and consistent with the pattern observed in Study 1, $\chi^2(2) = 3.01$, $p = .223$, Cramér's $V = 0.10$, 95% CI [0.03, 0.22]. In contrast, this pattern was significantly different from the pattern observed in Study 7, $\chi^2(2) = 10.27$, $p = .006$, Cramér's $V = 0.18$, 95% CI [0.09, 0.28]; Figure 5C.

Box Value Estimation

In Study 9, participants estimated the value of three boxes during the same-day test and then estimated the value of the remaining three boxes during the next-day test. However, due to a technical error, responses from the same-day box value estimation test were not recorded. On the next-day test, there were no differences in value estimates across box types, $F(2, 318) = 0.11$, $p = .893$, $\eta_p^2 = 0.001$, 95% CI [0.00, 0.01]. This result contrasts with Study 7, in which we observed a pattern of box value estimation that was similar to Study 1 (Early > Even > Late). However, the effect in Study 7 was weak; our studies may not be sufficiently powered to reliably detect next-day differences in box value estimation. Importantly, results from both Study 7 and Study 9 demonstrate a clear dissociation between explicit value estimation and revealed preferences when participants are tested after a delay. Despite no longer overestimating the value of Early boxes, participants strongly preferred to choose them again.

Object Memory

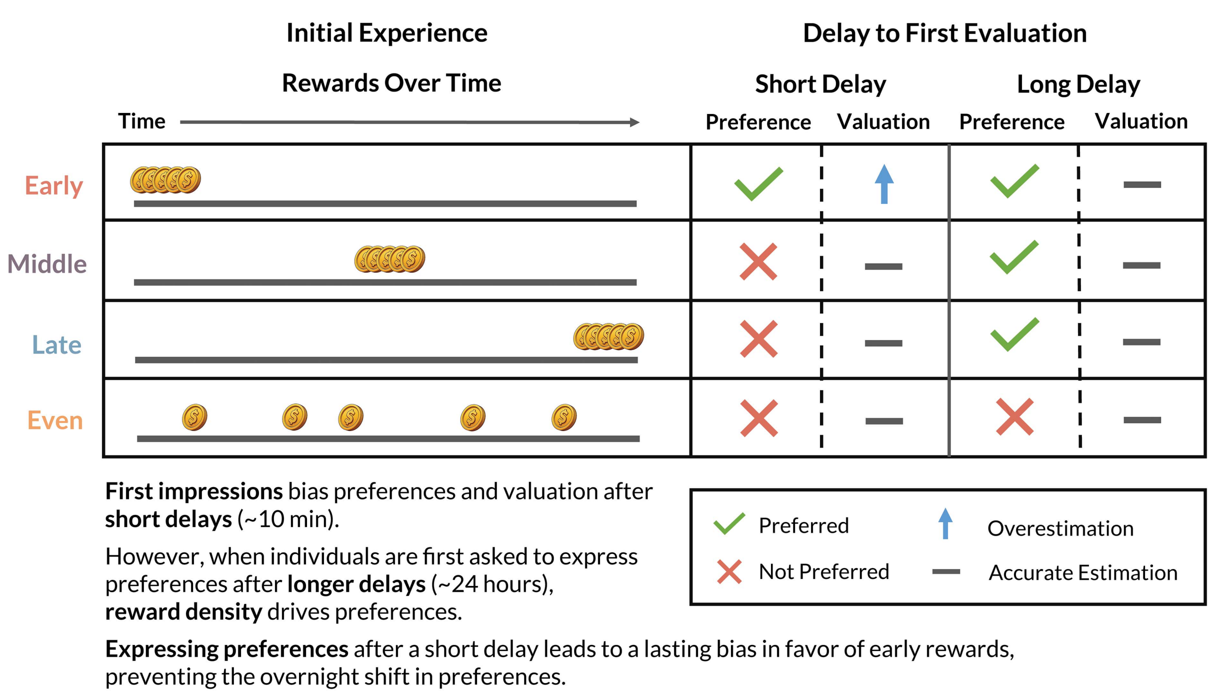
Object recognition memory in Study 9 was comparable to Studies 7 and 8 (62.00%) and significantly greater than chance performance, $t(157) = 15.18$, $p < .0001$, $d = 1.21$, 95% CI [1.00, 1.41]. Additional analyses reporting object recognition memory and object-box association memory are reported in Supplemental Tables 11 and 12.

As in previous studies, we tested memory for object-value associations using a binomial generalized mixed-effects model. We predicted *high-value hits* (correctly identifying high-value objects as high-value) from the variable *box type* (Early, Even/Middle, or Late). The model included a covariate of no interest for the learning trial number. Consistent with the other next-day tests (Studies 7 and 8), there were no significant differences in object-value association memory across box types, $\chi^2(2) = 3.71$, $p = .156$.

General Discussion

Here, we tested whether first impressions or good endings biased preferences and valuation. Crucially, we found that preferences depend on both *when rewards occur* and when you ask (Figure 6). In our

Figure 6
Summary of Findings



Note. Rows compare results for the four different box types tested (Early, Middle, Late, and Even) across Studies 1–9. Coins represent the temporal distribution of reward, though all box types have the same total reward value. After a short delay (~10 min), participants strongly preferred Early boxes over all other box types. Participants also overestimated the value of Early boxes. In contrast, when participants were first asked to express preferences after a long delay (~24 hr), they equally preferred Early, Middle, and Late boxes (all featuring clusters of reward) over Even boxes. After a delay, revealed preferences were dissociated from explicit value estimation. Importantly, expressing preferences after a short delay led to a lasting bias in favor of Early boxes, preventing the overnight shift in preferences when participants again reported preferences after a longer delay. See the online article for the color version of this figure.

"garage sale" task, participants opened boxes and viewed sequences of objects with values and then later evaluated the boxes. Across Studies 1–6, we consistently found that participants were strongly biased by first impressions; participants preferred and overvalued boxes with early rewards. In Studies 7 and 8, participants were asked to evaluate the episodes after an overnight delay. After a delay, preferences were shaped by reward *density* rather than primacy: Participants equally preferred boxes with clusters of reward at the beginning, middle, or end of the sequence (relative to boxes with evenly distributed rewards). Finally, in Study 9, we tested whether evaluating shortly after an experience influences next-day preferences. In Study 9, participants again showed a strong bias in favor of Early boxes, both immediately and after an overnight delay.

We conclude that first impressions strongly bias preferences and valuation shortly after an experience. After a delay, preferences are driven by reward *density* rather than primacy or recency. However, expressing preferences immediately after an experience immediately may lead to a lasting primacy bias, preventing this adaptive shift in preferences. We propose that clusters of reward may prioritize memories for replay and consolidation, summarizing experiences in memory and guiding rational choice.

Early Rewards Bias Evaluations Shortly After an Experience

In Study 1, we found a primacy effect whereby participants preferred Early boxes and overestimated the value of these boxes. Importantly, participants could not stop sampling objects from the box during the sequence; therefore, receiving rewards earlier did not offer any advantage. Despite the preference for Early boxes and the overall effect of reward-enhancing object recognition, participants were less accurate at recalling which objects from Early boxes were high-value. We propose that at the beginning of an episode, reward information is associated with the broader context (the box) rather than individual events or details (the objects). This value-context association may help form adaptive memories that guide future choices.

In Studies 2–6, we systematically tested possible constraints on these phenomena. The bias in favor of Early boxes was very reliable; it did not depend on the density of reward in the middle of the experience (Study 2), the active versus passive nature of the task (Studies 3 and 4), expectations about object values (Study 5), or whether value estimation was interleaved or retrospective (Study 6). In Study 2, Early boxes were still favored over boxes with a peak cluster of reward in the middle of the sequence. Studies 3 and 4 showed that results did not depend on whether participants had agency to choose among boxes. In Study 5, we calibrated reward expectations (to reduce surprise experienced during the first box) by explicitly informing participants about the approximate values of low-value ($0.10) and high-value objects ($0.80) in the instructions. Value expectations did not influence preferences or memory. In Study 6, participants estimated the value immediately after each box; participants still preferred and overvalued Early boxes.

Overall, we found that early rewards consistently biased preferences and valuation when participants were asked to evaluate shortly after an experience. Challenging several theoretical predictions, we found no evidence that late rewards drove preferences. The peak–ends effect predicts that endings are particularly important, especially when an experience is linked to a goal or future choices (Fredrickson, 2000). The recency effect predicts a bias toward late rewards, particularly when tested immediately after learning (Baddeley & Hitch, 1977; Mantonakis et al., 2009; Murdock, 1962). Similarly, in computational models of reward learning, current reward estimates are driven by recent experiences (Momennejad et al., 2017; Sutton & Barto, 1998). Contrary to these predictions, upon immediate evaluation, Late boxes were never preferred, even when we emphasized agency and future utility (Study 3) or probed value estimation immediately after each box (Study 6).

Clusters of Reward Drive Preferences After a Delay

Salient rewards enhance memory for associated stimuli, but this effect can depend on a delay between encoding and test (Miendlarzewska et al., 2016). Neurally, rewards enhance hippocampal replay and memory consolidation (Atherton et al., 2015; Cowan et al., 2021; Murayama & Kitagami, 2014; Patil et al., 2017; Shohamy & Adcock, 2010), perhaps by "tagging" memories to prioritize consolidation of reward-related information (Ballarini et al., 2009; Moncada et al., 2015). Importantly, the process of associating reward with related items in memory can bias subsequent preferences and choices (Wimmer & Shohamy, 2012). Although prior studies have not directly compared the effects of dense and spaced rewards (of equal total value), prior evidence indicates that larger, more salient rewards elicit greater phasic dopamine responses and larger memory enhancements (Aberg et al., 2020; Miendlarzewska et al., 2016). Therefore, we expected that dense clusters of reward may be especially salient, exerting greater effects on memory and preferences than evenly spaced rewards. We predicted that *clusters* of reward would influence memory and preferences after a delay that permits consolidation, regardless of whether the clusters occurred at the beginning or end of an experience.

Study 7 affirmed this prediction by demonstrating that after a 1-day delay, both Early and Late boxes were equally preferred over Even boxes. Interestingly, participants did not show a corresponding increase in the perceived value of Late boxes, demonstrating a dissociation between explicit value estimation and revealed preferences. This dissociation parallels prior findings: Neural signatures of subjective valuation and choice can be dissociated (Liu et al., 2012), and implicit and explicit evaluations are correlated but distinct (Nosek, 2005).

In Study 8, we expanded on findings from Study 7 by testing whether a cluster of rewards in the middle of a sequence would also drive preferences after a delay. We expected that after a delay, there would be no differences in preferences among boxes with Early, Middle, and Late clusters, as all three box types featured dense clusters of reward. As predicted, participants equally preferred Early, Middle, and Late boxes, supporting the idea that reward *density*, rather than primacy, determines preferences after a delay.

Overall, Studies 7 and 8 showed that clusters of reward (in the beginning, middle, or end of an experience) drove preferences after an overnight delay. Short-term preferences were dominated by an exclusive preference for Early boxes, but preferences were more rational when participants were not asked to express preferences until the following day.

Evaluating Immediately Leads to a Lasting Primacy Bias

Finally, in Study 9, we investigated whether expressing preferences shortly after an experience would influence subsequent preferences after a delay. In this study, participants were tested on half of the boxes immediately after the garage sale task and then tested on the other half of the boxes after an overnight delay. Results from the same-day test were consistent with Studies 1–6; participants strongly preferred Early boxes over Even and Late boxes.

However, results from the next-day test contrasted with the pattern observed in the other studies with next-day tests. In Studies 7 and 8, we found that participants no longer showed an exclusive preference for Early boxes when tested after a delay. In Study 9, we found that participants continued to strongly prefer Early boxes over Even and Late boxes when tested after an overnight delay. Importantly, participants were not tested on the same boxes twice— they demonstrated an exclusive preference for Early boxes after both short and long delays, evaluating different boxes on each test. As in Study 7, we found that next-day preferences were decoupled from explicit value estimation.

This finding suggests that the process of recalling and evaluating an experience influences subsequent preferences, even for related but untested items. In other words, reporting preferences shortly after an experience led to a lasting primacy bias, diverging from the reward density-dependent preferences that we observed in Studies 7 and 8. This interpretation aligns with prior evidence from studies on *retrieval-induced forgetting*; retrieving a subset of items can inhibit the activation of related but unretrieved items, impairing memory for the untested items and biasing decision making (M. C. Anderson et al., 2000; Iglesias-Parro & Gómez-Ariza, 2006; Murayama et al., 2014). Relatedly, other studies have shown that assigning a value (positive or negative) to a subset of stimuli can retrospectively link that value to related stimuli; these behavioral effects are evident after a delay, but not immediately (Dunsmoor et al., 2015; Patil et al., 2017). This process is hypothesized to occur through *synaptic tagging* that influences which memories are strengthened or pruned during consolidation. In our paradigm, this effect may be driven by the process of explicitly stating preferences, the box recognition probes, or both mechanisms.

Synthesis: Adaptive Preferences

Synthesizing results from Studies 1 to 9, we propose the following: When individuals are prompted to recall and evaluate an experience shortly afterward, preferences are strongly influenced by early rewards. This primacy bias is evident immediately and does not depend on consolidation. Clusters of reward that occur at any point during an episode (beginning, middle, or end) tag memories for consolidation, but this process must unfold over a delay. Thus, when individuals do not evaluate until after a delay, preferences are driven by reward *density* rather than primacy. Importantly, however, recalling and evaluating experiences immediately—even if only a subset of items—cements the primacy bias, preventing this overnight shift in preferences. We speculate that the process of recalling and evaluating an experience (i.e., assigning value) may strengthen memory for Early boxes and inhibit memories for Late and Even boxes, such as through retrieval-induced forgetting. Future research could explore these ideas further by applying methods from the behavioral tagging literature (e.g., testing if the primacy bias influences preferences for categorically related but unrewarded items).

A primacy bias can be adaptive, particularly in the short term. In situations where we are able to choose when to end an experience and explore other options, it can be adaptive to make decisions based on early rewards (or lack thereof). For example, in daily life, we might skim the first few pages of a book while standing in the library or watch the first few minutes of a movie at home. Early evidence can help us decide whether or not to continue an experience before we allocate too much time and effort. In our paradigm, however, a primacy bias was not rational or adaptive. By choosing to open a box, participants committed to viewing the full sequence of objects and earning the total value of all the objects. Receiving rewards early did not offer any advantage, and all boxes were equally valuable. In daily life, we also experience situations like this; for example, few people would abandon a restaurant after a bad appetizer or leave a movie theater if they did not enjoy the beginning of a film. After a delay, when we no longer have the option to choose whether to extend or end an experience, it is more adaptive to evaluate past experiences based on reward density rather than primacy.

Limitations and Future Directions

Future studies could make changes to our paradigm to address limitations. Memory for objects and object-value associations was poor in Studies 7–9 (next-day tests), and memory for object-box associations was generally poor. Using different stimuli that are more memorable (e.g., famous faces) could improve performance and thus test for additional effects that depend on box type and consolidation. Additionally, participants did not have the opportunity to repeatedly sample the same boxes. A variant of this paradigm that is more similar to a foraging task could offer insight into repeated choices over time.

In Study 7, we observed that after an overnight delay, participants preferred both Early and Late boxes over Even boxes. It is possible that participants in Study 7 would also have reported a preference for Late boxes if tested immediately (i.e., the distinct preferences could be explained by an idiosyncratic feature of this sample, rather than the overnight delay). However, we consider this explanation unlikely given the remarkably robust evidence observed in Studies 1–6 and Study 9, demonstrating an exclusive bias in favor of Early boxes when participants are tested immediately. Furthermore, Study 8 again demonstrated that this exclusive preference for Early boxes is absent after a delay, demonstrating that this effect was not unique to the sample of participants in Study 7.

It is possible that the design of our task inflated the accuracy of box value estimates, as the true value was the midpoint of the scale. Although we did observe that participants used the full range of the scale (Supplemental Figure 2), future studies could probe open-ended value estimates without providing a range of possible values. We also found that next-day preferences were decoupled from explicit value estimation; future research could investigate other factors that may drive preferences after a delay. Finally, future studies could also extend this paradigm to other contexts, such as consumer choices (e.g., integrating across product reviews) or social dynamics (e.g., determining trustworthiness after multiple interactions).

Conclusion

Overall, both first impressions and good endings can influence valuation and preferences, but at different time points after an experience. Returning to the restaurant example, while a diner's review immediately after a meal may be biased by an excellent appetizer, returning to the restaurant in the future could be motivated by excellent dishes from any time during the meal. However, leaving a review immediately after the meal might lead to an enduring primacy bias. Our findings offer theoretical insight as well as broader implications for understanding consumer choices, social interactions, and comparative decisions.

Our results both challenge and unify prior research on reward learning, decision making, episodic memories, and affective experiences. We showed that humans summarize rewarding experiences in a nonlinear and time-dependent way, impacting future choices. Contrary to predictions from recency effects and reinforcement learning, ending with a cluster of rewards did *not* bias short-term preferences or valuation. Instead, first impressions reliably determined short-term evaluations. In daily life, we often make evaluations shortly after an experience, such as when rating a restaurant, judging a competition, or making a purchase after reading product reviews. Our results imply that these immediate evaluations are driven by first impressions.

However, preferences differed when individuals waited and evaluated their experiences after a delay. We found that after an overnight delay, preferences were driven by reward *density* rather than temporal placement. This shift implies that consolidation permits the adaptive integration of rewards or evidence over an entire episode. Important decisions are also made after longer delays, such as when deciding whether to return to a restaurant, selecting a job candidate, or considering a second date. We propose that consolidation changes how rewarding episodes are summarized in memory, reinforcing the wisdom of sleeping on important decisions. Importantly, however, evaluating experiences immediately solidified the short-term primacy bias, preventing this adaptive shift in preferences over time.

Both immediate and delayed evaluations are important in everyday life. We show that preferences derived from rewarding experiences depend on *when* those experiences are first recalled and evaluated. We propose that, like memories, preferences are transformed over time to support adaptive future choices.

Constraints on Generality

The present studies included online samples of adults currently residing in the United States. The distributions of gender and race in our sample were comparable to the population-level distributions in the United States. Although we did not limit our sample to a specific age range, more younger adults participated than older adults due to the makeup of the Prolific participant pool. Our results may not generalize to other demographic groups or cultural contexts.

References

Aberg, K. C., Kramer, E. E., & Schwartz, S. (2020). Interplay between midbrain and dorsal anterior cingulate regions arbitrates lingering reward effects on memory encoding. *Nature Communications*, *11*(1), Article 1829. https://doi.org/10.1038/s41467-020-15542-z

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron*, *91*(5), 1124–1136. https://doi.org/10.1016/j.neuron.2016.07.047

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review*, *7*(3), 522–530. https://doi.org/10.3758/BF03214366

Anderson, N. H., & Norman, A. (1964). Order effects in impression formation in four classes of stimuli. *The Journal of Abnormal and Social Psychology*, *69*(5), 467–471. https://doi.org/10.1037/h0047472

Atherton, L. A., Dupret, D., & Mellor, J. R. (2015). Memory trace replay: The shaping of memory consolidation by neuromodulation. *Trends in Neurosciences*, *38*(9), 560–570. https://doi.org/10.1016/j.tins.2015.07.004

Baddeley, A. D., & Hitch, G. J. (1977). Recency reexamined. In S. Dornic (Ed.), *Attention and performance* (Vol. 6, pp. 647–667). Routledge.

Ballarini, F., Moncada, D., Martinez, M. C., Alen, N., & Viola, H. (2009). Behavioral tagging is a general mechanism of long-term memory formation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(34), 14599–14604. https://doi.org/10.1073/pnas.0907078106

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. https://arxiv.org/abs/1406.5823

Baumgartner, H., Sujan, M., & Padgett, D. (1997). Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research*, *34*(2), 219–232. https://doi.org/10.1177/002224379703400203

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), Article 2815. https://doi.org/10.21105/joss.02815

Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., & Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bulletin & Review*, *28*(6), 1796–1810. https://doi.org/10.3758/s13423-021-01978-x

Dickerson, K. C., & Adcock, R. A. (2018). Motivation and memory. In E. A. Phelps & L. Davachi (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, learning and memory* (pp. 215–242). Wiley. https://doi.org/10.1002/9781119170174.epcn107

Digirolamo, G. J., & Hintzman, D. L. (1997). First impressions are lasting impressions: A primacy effect in memory for repetitions. *Psychonomic Bulletin & Review*, *4*(1), 121–124. https://doi.org/10.3758/BF03210784

Do, A. M., Rupert, A. V., & Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, *15*(1), 96–98. https://doi.org/10.3758/PBR.15.1.96

Dunsmoor, J. E., Murty, V. P., Davachi, L., & Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature*, *520*(7547), 345–348. https://doi.org/10.1038/nature14106

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, *14*(4), 577–606. https://doi.org/10.1080/026999300402808

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55. https://doi.org/10.1037/0022-3514.65.1.45

Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. *The Journal of General Psychology*, *113*(4), 351–357. https://doi.org/10.1080/00221309.1986.9711045

Healy, A. F., Havas, D. A., & Parker, J. T. (2000). Comparing serial position effects in semantic and episodic memory using reconstruction of order tasks. *Journal of Memory and Language*, *42*(2), 147–167. https://doi.org/10.1006/jmla.1999.2671

Hervé, M. (2022). *RVAideMemoire: Testing and plotting procedures for biostatistics*. https://CRAN.R-project.org/package=RVAideMemoire

Horwath, E. A., Rouhani, N., DuBrow, S., & Murty, V. P. (2023). Value restructures the organization of free recall. *Cognition*, *231*, Article 105315. https://doi.org/10.1016/j.cognition.2022.105315

Iglesias-Parro, S., & Gómez-Ariza, C. J. (2006). Biasing decision making by means of retrieval practice. *European Journal of Cognitive Psychology*, *18*(6), 899–908. https://doi.org/10.1080/09541440500501189

Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *29*(4), 626–640. https://doi.org/10.1037/0278-7393.29.4.626

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 3–25). Sage Publications.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, *112*(2), 375–406. https://doi.org/10.1162/003355397555235

Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.6.0) [R; Computer Software]. https://CRAN.R-project.org/package=emmeans

Li, C. (2010). Primacy effect or recency effect? A long-term memory test of Super Bowl commercials. *Journal of Consumer Behaviour*, *9*(1), 32–44. https://doi.org/10.1002/cb.291

Liu, L., Feng, T., Wang, J., & Li, H. (2012). The neural dissociation of subjective valuation from choice processes in intertemporal choice. *Behavioural Brain Research*, *231*(1), 40–47. https://doi.org/10.1016/j.bbr.2012.02.045

Lüdecke, D. (2021). *sjPlot: Data visualization for statistics in social science*. https://CRAN.R-project.org/package=sjPlot

Mangiafico, S. (2022). *rcompanion: Functions to support extension education program evaluation*. https://CRAN.R-project.org/package=rcompanion

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, *20*(11), 1309–1312. https://doi.org/10.1111/j.1467-9280.2009.02453.x

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Miendlarzewska, E. A., Bavelier, D., & Schwartz, S. (2016). Influence of reward motivation on human declarative memory. *Neuroscience and Biobehavioral Reviews*, *61*, 156–176. https://doi.org/10.1016/j.neubiorev.2015.11.015

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680–692. https://doi.org/10.1038/s41562-017-0180-8

Moncada, D., Ballarini, F., & Viola, H. (2015). Behavioral tagging: A translation of the synaptic tagging and capture hypothesis. *Neural Plasticity*, *2015*, Article 650780. https://doi.org/10.1155/2015/650780

Murayama, K., & Kitagami, S. (2014). Consolidation power of extrinsic rewards: Reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology: General*, *143*(1), 15–20. https://doi.org/10.1037/a0031992

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *140*(5), 1383–1409. https://doi.org/10.1037/a0037505

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. https://doi.org/10.1037/h0045106

Murty, V. P., & Dickerson, K. C. (2016). Motivational influences on memory. In S. Kim, J. Reeve, & M. Bong (Eds.), *Recent developments in neuroscience research on human motivation* (Vol. 19, pp. 203–227). Emerald Group Publishing Limited. https://doi.org/10.1108/S0749-742320160000019019

Neath, I. (2010). Evidence for similar principles in episodic and semantic memory: The presidential serial position function. *Memory & Cognition*, *38*(5), 659–666. https://doi.org/10.3758/MC.38.5.659

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*(4), 565–584. https://doi.org/10.1037/0096-3445.134.4.565

Patil, A., Murty, V. P., Dunsmoor, J. E., Phelps, E. A., & Davachi, L. (2017). Reward retroactively enhances memory consolidation for related items. *Learning & Memory*, *24*(1), 65–69. https://doi.org/10.1101/lm.042978.116

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peterson, C. R., & DuCharme, W. M. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology*, *73*(1), 61–65. https://doi.org/10.1037/h0024139

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*(1), 3–8. https://doi.org/10.1016/0304-3959(96)02994-6

Rey, A., Le Goff, K., Abadie, M., & Courrieu, P. (2020). The primacy order effect in complex decision making. *Psychological Research*, *84*(6), 1739–1748. https://doi.org/10.1007/s00426-019-01178-2

Schweitzer, K., & Nuñez, N. (2021). The effect of evidence order on jurors' verdicts: Primacy and recency effects with strongly and weakly probative evidence. *Applied Cognitive Psychology*, *35*(6), 1510–1522. https://doi.org/10.1002/acp.3884

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage*, *32*(3), 1422–1431. https://doi.org/10.1016/j.neuroimage.2006.04.223

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, *14*(10), 464–472. https://doi.org/10.1016/j.tics.2010.08.002

Sinclair, A. H. (2023). *First impressions or good endings? Preferences depend on when you ask* [Data set]. https://doi.org/10.17605/OSF.IO/C846S

Sullivan, J. (2019). The primacy effect in impression formation: Some replications and extensions. *Social Psychological and Personality Science*, *10*(4), 432–439. https://doi.org/10.1177/1948550618771003

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

van Erkel, P. F. A., & Thijssen, P. (2016). The first one wins: Distilling the primacy effect. *Electoral Studies*, *44*, 245–254. https://doi.org/10.1016/j.electstud.2016.09.002

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273. https://doi.org/10.1126/science.1223252

Received June 15, 2023
Revision received June 4, 2024
Accepted June 12, 2024 ■