Methods

Participants

All participants were right handed, native English speakers recruited from New York University and the neighboring communities. Before participating, informed consent was obtained as approved by the University Committee on Activities Involving Human Subjects. In addition to being paid $10/hour for their participation, participants were told that they could earn up to $21 in performance-based bonuses.

Thirty-one adults (mean age = 23.10, age range = 18-34, 22 females) participated in Experiment 1. Data from seven participants were excluded from the analyses: one for not following instructions, two for not attending to low reward trials during the encoding phase, two for attempting to recall images before the memory test, and two for expecting a memory test. Due to a technical error, encoding data for one participant was missing. This participant is still included in the analysis; excluding their data for recognition test did not change the pattern of results. Thus, our final analyses included twenty-four participants (mean age = 22.17, age range = 18-34, 18 females).

Twenty-nine adults (mean age = 23.83, age range: 19-30, 24 females) participated in Experiment 2. Data from four participants were excluded from the analyses: one for not attending to low reward trials during the encoding phase, two for attempting to recall images before the memory test, and one for expecting a memory test. Thus, our analyses included twenty-five participants (mean age = 24.08, age-range = 19-30, 18 females). They were recruited, consented, and compensated as in Experiment 1.

Stimuli

A total of 360 colored images of animals and tools (adapted from Dunsmoor et al. 2014) were used in the experiments and were randomly selected to be in each condition across participants. They were displayed on a gray background on a computer screen.

Procedure

The experiment consisted of a pre-reward phase, reward phase, and memory test.

Pre-reward phase: Participants performed a delayed match to sample task. On each trial, they were presented with an image of either an animal or a tool for 2 seconds. After a five-second delay, they were presented with the same image along with a trial-unique, foil image from the same image category. Participants were instructed to indicate which image matched the preceding image (i.e., the target) within 600 milliseconds (“1” for the image on the left, “2” for the image on the right). An intertrial interval of 1.5 seconds included a fixation cross on the center of a blank screen. Images from the same category were not presented more than 3 times in a row. The location of the target image (left, right) was randomized across trials. 60 images from each image category were used, with half of them presented as foil images. This phase consisted of 2 blocks, 30 trials each, and lasted for around 9 minutes.

Accuracy on each trial of the delayed match to sample task (during pre-reward and reward phase) was calculated based on their key response as well as their reaction time. If they

responded correctly to the target within 600 milliseconds, they were given a score of '1'. If they responded incorrectly to the target or took > 600 milliseconds to make a response they were given a score of '0'.

Reward phase: The reward phase task was identical to the pre-reward task except reward could be earned by correct performance indicated by performance feedback in the form of green or white stars for 1 second. A star indicated correct performance (see above) and a lack of a star indicated incorrect performance. The color of the star was always associated with a single image category (animals, tools); however, participants were not instructed of these relationships. Before the reward phase, participants were notified that a performance of 90% or more on the "green star" trials would result in a $20 bonus, whereas a performance of 90% or more on the "white star" trials would result in a $1 bonus. Assignment of the low and high reward to the image categories was counterbalanced across participants. Participants were not given feedback about their performance-based bonus until after the end of the experiment. The conditioning phase also consisted of 2 blocks, 30 trials each, and lasted for 10 minutes.

Memory test: A recognition memory test was administered either immediately (Experiment 2) or ~24 hours (Experiment 1) after the encoding session. It included 120 images of animals and 120 images of tools (4 blocks of 60 trials each), half of which were target images from the pre-reward and reward phase and half of which were novel foil images. Foil images from the delayed match-to-sample task were not included in this test. On each trial, participants had 6 second to indicate whether they had seen the displayed image in the experiment before on a confidence rating scale from 1 to 4 ("Definitely yes", "Maybe yes", "Maybe no", and "Definitely no"). Memory was analyzed as corrected recognition (high confidence hits > high confidence false alarms). These scores were calculated separately for images that belonged to the high reward image category and were presented in the pre-reward phase (Pre-reward R+), images that belonged to the low reward image category and were presented in the pre-reward phase (Pre-reward R-), images that belonged to the high reward image category but were presented in the reward phase (Reward R+), images that belonged to the low reward image category and were presented in the reward phase (Reward R-).

After the memory test, participants were asked if they had expected a memory test and if they had attempted to recall images. To ensure that we were investigating incidental encoding and avoiding active rehearsal post-encoding, data from seven participants who said yes to either question were excluded from the analyses.

Supplemental Results

Control analyses for Delayed Match-To-Sample RT and Accuracy:

Our results indicated that reward's influence on memory only emerged after a delay (i.e., larger influences of reward on memory in the delayed versus immediate group). We ran the following control analyses to assure that these findings were not driven by differences in performance on the delayed match-to-sample task across groups.

First, we ran a three-factor ANOVA characterizing the effects of phase (pre-reward, reward), category (high reward, low reward), and delay (immediate, 24-hour) on delayed match-to-sample performance. For accuracy, there a main effect of phase (i.e., higher accuracy for reward versus pre-reward phase; $F(1,46)=26.72$, $p<0.001$), and no other significant main effects

or interactions. Critically, there was no performance*group interactions, suggesting that differences in accuracy across groups could account for the different patterns of reward memory.

Next, we ran a three-factor ANOVA on RT with the same factors described above. For RT, there was a main effect of phase (i.e., faster RT during the reward versus pre-reward phase; $F(1,46)=10.75$, $p=0.002$), a delay*category interaction (i.e., faster RTs in the reward versus non-reward condition in the delayed versus immediate group, $F(1,46)=4.40$, $p=0.04$), and a phase*category interaction (i.e., faster RT to high versus low reward items in the reward versus pre-reward phase, $F(1,46)=5.64$, $p=0.02$). Critically, differences in RTs for rewarded versus non-rewarded trials in the delayed versus immediate group could have influenced our recognition memory results. To directly test if our memory results could be confounded by RT differences in performance during the delayed match-to-sample task across Experiments 1 and 2, we tested whether RT during the delayed match-to-sample task predicted recognition memory within subjects. Specifically, we used a generalized mixed-effects model (*glmer* function of the lme4 function in *R* (Bates, Maechler, Bolker, & Walker, 2014) and included RT during delayed match-to-sample as within-subject predictors both as fixed effects and random slopes varying over participant, and found that RT did not reliably predict recognition memory ($\beta=-0.53$, $z=-1.13$, $p=0.259$). These findings suggest that differences in RT during the delayed match-to-sample task does not account for the pattern of our recognition memory results.

Cited References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7).

Supplemental Tables

Table S1A:
Experiment 1: Accuracy and RT for delayed match-to-sample task

Stimulus Category	Pre-reward Phase		Reward Phase	
	Accuracy	RT	Accuracy	RT
Animals	0.929 (0.054)	0.403 (0.016)	0.970 (0.036)	0.393 (0.015)
Tools	0.945 (0.048)	0.399 (0.016)	0.975 (0.032)	0.387 (0.014)

*Reported values indicate mean (standard error).

Table S1B:
Experiment 1: Accuracy and RT for delayed match-to-sample task

Stimulus Category	Pre-reward Phase		Reward Phase	
	Accuracy	RT	Accuracy	RT
Animals	0.957 (0.040)	0.409 (0.016)	0.98 (0.028)	0.399 (0.014)
Tools	0.948 (0.044)	0.413 (0.018)	0.975 (0.031)	0.394 (0.015)

*Reported values indicate mean (standard error).