Supplemental materials for "Putting value in context: A role for context memory in decisions for reward".

Aaron M. Bornstein [1*], Kenneth A. Norman [1,2]

1. Neuroscience Institute, Princeton University, Princeton, NJ, USA.

2. Department of Psychology, Princeton University, Princeton, NJ, USA.

* To whom correspondence should be addressed: aaronmb@princeton.edu.

Supplemental Methods

Context-aware sampling model

To investigate how single-trial and context reward trade off with each other as the number of past episodes sampled increases, we simulated the task as performed by a context-aware episode sampling model. In this simulation, all choices are made using episode sampling alone (no influence of model-free values), so as to clearly isolate the influence of changing the number of samples. In episode sampling, option values are estimated using the values encountered on one or more past episodes, with the likelihood of sampling a given episode diminishing exponentially with its recency. The context-aware episode sampling model augments this idea, by positing that additional samples after the first are (with some probability) selected uniformly from the same context as the preceding sample.

The model maintains a cache of episodes representing each experienced trial. When a subject responds correctly to a valid memory probes, the model with some probability "reinstates" the episode by copying the reminded trial to the front of the cache–thus, making it more likely to be drawn when evaluating options during the next choice. If the subject's correct response to the memory probe is of high confidence, then the context of the probed trial is included in the episode copied to the front of the cache.

We used this model to simulate subjects who used different numbers of samples from episodic memory to make decisions. The model had four parameters which were fixed across all simulations: α_{direct}, or the decay rate on temporal recency; α_{evoked}, or the probability of reinstating evoked trials as a result of memory probes; β, the softmax temperature; β_p, the choice perseveration term; and π, the likelihood of drawing sample k from the same context as sample $k-1$ (as opposed to based on temporal recency).

A final parameter, the number of samples drawn to make a decision, was varied between 1 and 15. For each fixed number of samples, we simulated 1,600 subjects each performing an instantiation of the task. The simulated subject's parameters α_{direct}, β, and β_p were set to those fit to the real subjects with the number of samples equal to 1, and π was set to 1.

For simplicity, in the first simulation α_{evoked} was set to 1. This assumption was relaxed for our second set of simulations, during which we varied α_{evoked} between 0 and 1 to illustrate the impact of changing this parameter. The subjects were programmed to make, on average, the same proportion of low confidence and incorrect responses as did the subjects from Experiments 1 and 2.

Fitting incremental learning models to our data

To compare our context reward model to the others, we generated the timeseries of reinstated context reward values that would be learned according to three different specifications of model-free RL.

The first variant followed the traditional method, learning the value of each card deck without regard to changes in room context. Specifically, at each step, the value for the chosen card deck, B_t, was updated according to the reward received on that trial, R_t, and the learning rate α:

$$Q_B = Q_B + \alpha * (R_t - Q_B) \tag{S1}$$

We refer to this as the *standard* model.

The second variant reset the value of each card deck when the room changed, giving separate values $Q_{B,C}$ for each room-context, following Equation S2.

$$Q_{B,C} = Q_{B,C} + \alpha * (R - Q_{B,C}) \tag{S2}$$

We refer to this as the *room-reset* model.

A third variant also reset action values, but this time it did so at a variable trial number within each room (e.g. 1, 2, 3, or more trials after room change). We designed this model to account for the possibility that participants took note of payoff reversals and discounted information prior to the switch [1]. We refer to this model as the *reversal* model.

The three incremental learning models fit to choices in rooms one through six, using maximum likelihood estimation of parameters. To compare the model likelihoods we first transformed them using the Bayesian Information Criterion (BIC; Schwarz 2), which penalized the third model for its additional parameter.

The best-fitting model was used to generate a replacement for the context-reward regressor as specified in main text, *Methods* subsection *Regression analysis*. The resulting regression model weights were compared to the original (shown in Figure 5 in the main text).

Supplemental Results

Sampling model simulations

We demonstrated the context-aware sampling model (described in *Supplemental Methods*, subsection *Context-aware sampling model*) predicts that the effect of reminded context

should be greater than the effect of reminded trials (as in Figure 5 of the main text). To show this we simulated the context-aware sampling model 1,600 times, and split the sampled subjects into 50 populations of 32 each. (Figure S2 shows the results for one simulated population of 32 subjects each sampling 12 episodes in the time between probe and choice.) As the number of samples to make a decision increased, so did the ratio of the effect of context reward to that of the reminded trial (correlation between context and single-trial effects: $R = -0.9436$, $p = 1.3125e - 07$; Figure S1).
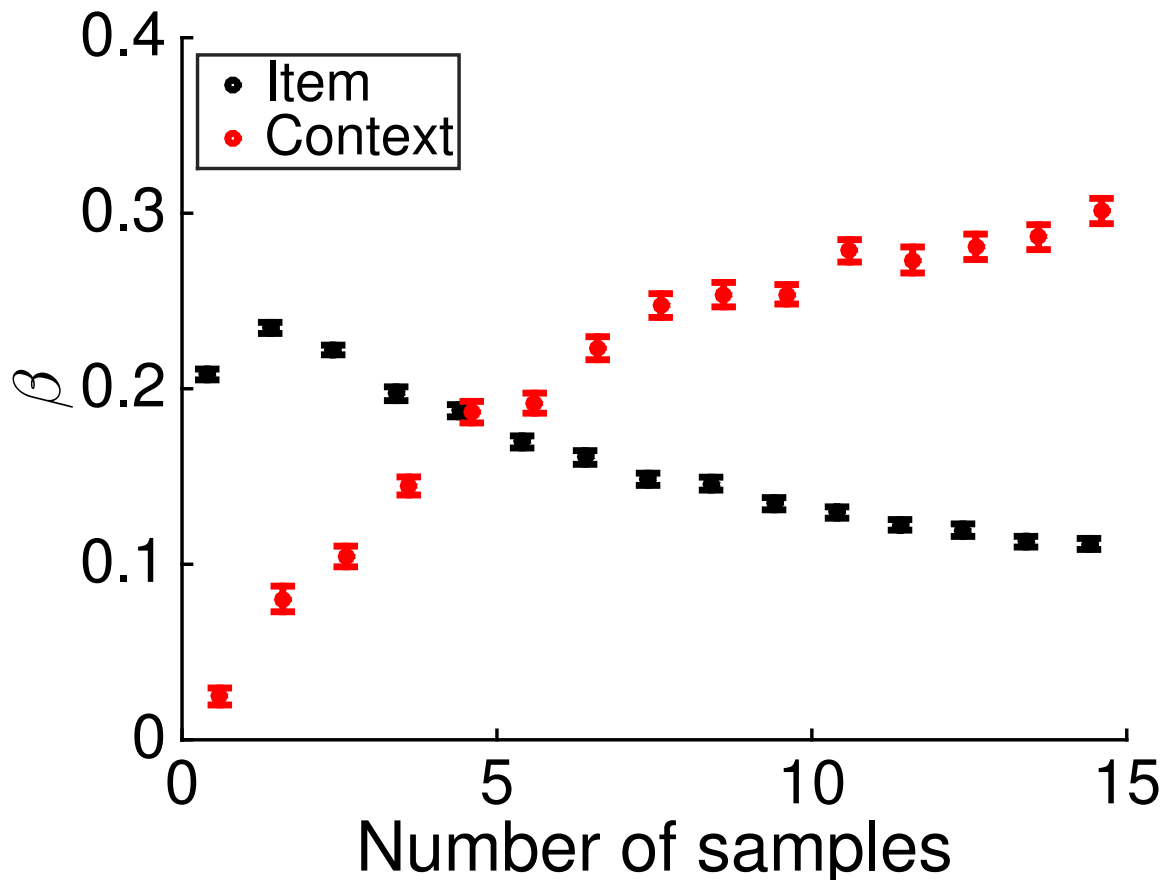

Figure S1: **Simulations show that the influence of context reward should increase with greater numbers of episode samples.** We simulated the context-aware sampling model, holding fixed all parameters except the number of samples drawn in support of each decision. We then performed the regression analysis shown in Figure 5 of the main text on this simulated data, and plot here the regression weights for single-trial (item) rewards and context rewards. As the number of samples drawn increases, the effect of context reward increases, while the effect of item reward decreases.

The effects of both reminded trials and reminded context are lower in Experiment 2 than Experiment 1, relative to the effect of recent reinforcement. In Experiment 1, the
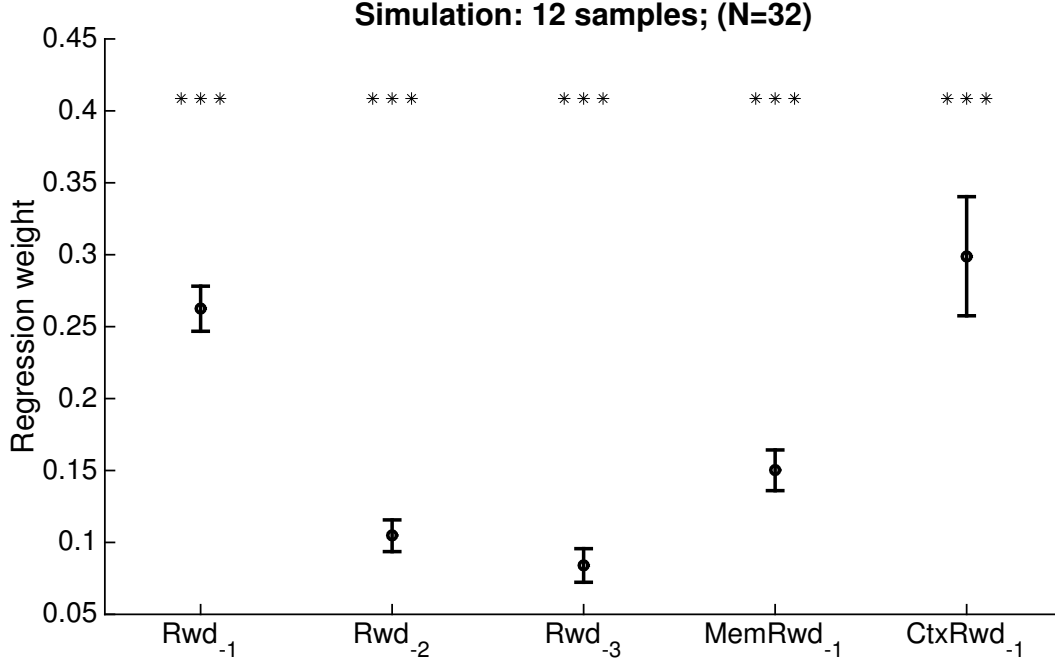
Figure S2: Full regression model fit to a simulated population of 32 subjects that sampled 12 episodes between each probe and the ensuing choice.

context reward effect is not significantly different from the effect of the most recent reward ($p = 0.6601$) nor the second most recent ($p = 0.2449$), but is significantly larger the effect of the third most recent reward ($p = 0.0128$). In Experiment 2, the context reward effect is lower than the effect of the most recent reward ($p = 0.0027$), but not distinguishable from the second most recent ($p = 0.3148$) nor the third most recent reward ($p = 0.8191$). In other words, subjects in Experiment 2 showed a lower influence of rewards from probe-evoked information, and thus were relatively more likely to make decisions on the basis of recent reinforcement.

We used the simulation to illustrate how our mechanism can eliminate the effect of reminded trials while sparing an effect of recent rewards and context. We followed the simulation procedure above, this time generating populations of 1,600 subjects, with each population using a different value of α_{evoked}. All other model parameters were fixed. Figure S3 shows the results for one population of 32 simulated subjects with α_{evoked} set to 0.3.

Figure S4 shows how, as the value of α_{evoked} decreases, so do both memory effects. Crucially, though, because the reminded context has a larger impact on behavior, relative to the reminded trial (for reasons discussed above), it remains impactful, while the item effect drops to nearly zero.
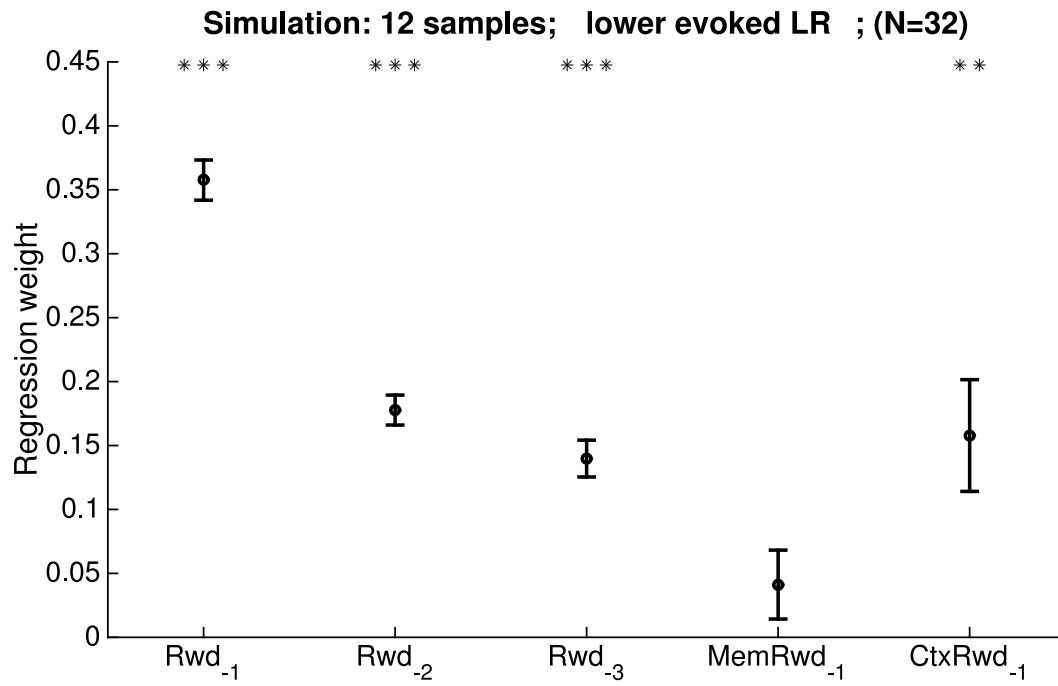
Figure S3: Full regression model fit to a simulated population of 32 subjects that sampled 12 episodes before each choice, with α_{evoked} set to 0.3.
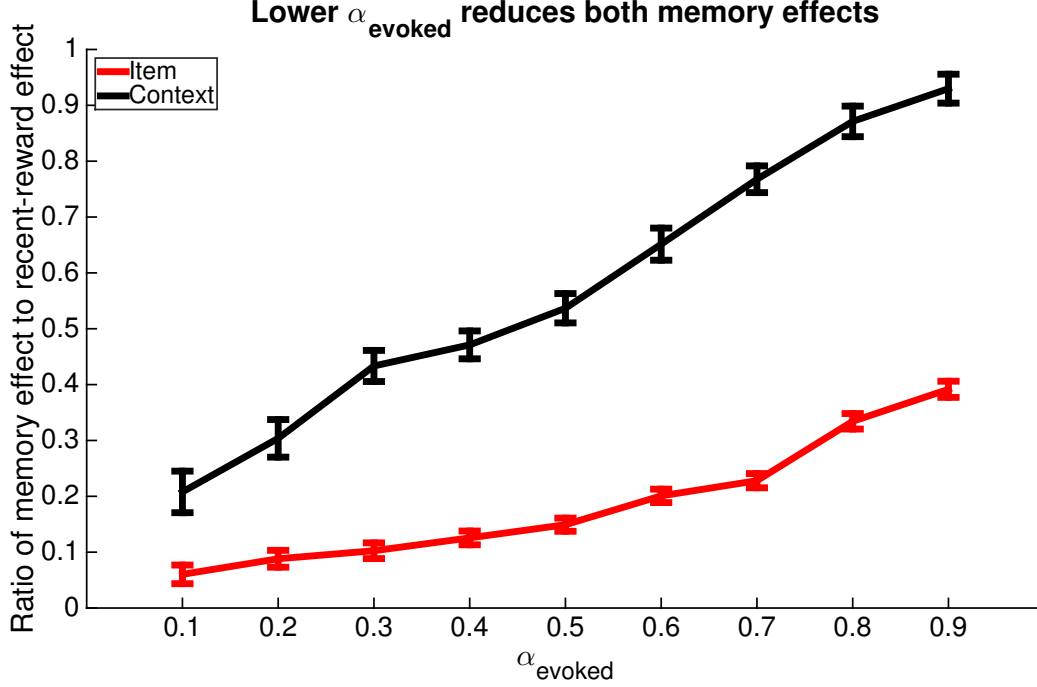
Figure S4: As the probability of reinstating reward information decreases, so does the contribution of both item and context memory to decisions. We ran the experiment for 9 populations of 1,600 simulated subjects. Each population used a different value of the α_{evoked} parameter, the probability that a reminder probe would result in reward information being reinstated from the reminded trial and context. We then measured the regression weights for the reminded trial and the reminded context, plotted here by their ratio to the regression weight for reward experienced one trial ago. As α_{evoked} decreased, so did the ratio of each type of memory-based effect to the effect of recent rewards. While the effect of reminded context is preserved at even very small values for α_{evoked}, the effect of the reminded trial drops to near zero. This pattern matches the reduced effect of both reminded trial and context in Experiment 2.

Fitting incremental learning models to our data

First, we compared the fit to behavior of three alternative RL models. Each model learned cached action values for the three card decks; two of these models reset those values when context changed–in one model, the context shifts/value resets were at the time that the room changed, in the other model, the context shifts were at a variable trial after the start of each room (to account for context boundaries being drawn at the time the payoffs changed).

The model that reset action values when the room changed was the best fit to behavior. By BIC versus the second-best model, in Experiment 1, the room-reset model was superior for 18/20 subjects, mean difference in BIC: 5.6981; In Experiment 2, 26/32 subjects, 4.6406. The fit parameters for the room-reset model–learning rate α, softmax temperature β,

choice stickiness β_p–were: Expt. 1 α mean 0.4802 SEM 0.0617, β mean 0.2333 SEM 0.2829, β_p mean 0.4348 SEM 0.1392; Expt. 2 α mean 0.5738 SEM 0.0375, β mean 0.4623 SEM 0.0313, β_p mean 0.2231 SEM 0.0738. These values were consistent across the two experiments (by two-sample t-test: α $p = 0.1745$; β $p = 0.3156$, β_p $p = 0.1480$).

We then used this room-reset model to generate a replacement context reward regressor, used in place of context reward the final action-values learned for each card deck in each room. The mean correlation between this new regressor and the corresponding original regressor was $R = 0.0564$ ($p = 0.3926$) in Expt. 1 and $R = 0.1714$ ($p = 5.4728e - 06$) in Expt. 2.

Context reward effects are specifically modulated by scene evidence in PPA

To confirm that context reward was specifically modulated by scene reinstatement, we repeated the quartiles analysis of Figure 5 in the main text to look for an effect of univariate activity in PPA (as opposed to scene evidence). Across quartiles and subjects, neither reminded trial (mean slope= -0.0422, SEM 0.0397, $p = 0.3154$), nor reminded context (mean slope= -0.0162, SEM 0.0184, $p = 0.3875$) showed a relationship with PPA activity, nor did we observe an interaction between these relationships ($p = 0.7559$).

We next defined a region of interest that was differentially responsive to the "scrambled" scenes that were used as a control in our localizer task (section *Methods*, subsection *ROI definition*). Across quartiles and subjects, neither reminded trial (mean slope= 0.0194, SEM 0.0479, $p = 0.6956$), nor reminded context (mean slope= -0.0214, SEM 0.0212, $p = 0.3210$) showed a relationship with univariate activity levels in this ROI, nor did we observe an interaction between these relationships ($p = 0.5047$). We next trained a classifier (analogously to how we trained the scene evidence classifier in the main analysis) to decode both "scrambled scene" evidence and scene evidence from this ROI, and explored how these classifier evidence values related to trial and context effects in our regression model. Across quartiles and subjects, neither reminded trial (mean slope= -0.0033, SEM 0.0160, $p = 0.8391$), nor reminded context (mean slope= 0.0314, SEM 0.0200, $p = 0.1264$) showed a relationship with "scrambled scene" evidence in this ROI, nor did we observe an interaction between these effects ($p = 0.1467$). Across quartiles and subjects, neither reminded trial (mean slope= 0.0041, SEM 0.0132, $p = 0.7602$), nor reminded context (mean slope= 0.0185, SEM 0.0158, $p = 0.2518$) showed a relationship with scene evidence in this ROI, nor did we observe an interaction between these effects ($p = 0.5046$).

As described in the main text, we also tested whether each regressor of interest was modulated by scene evidence, by repeating the quartiles analysis of Figure 5 for each regressor. We found that scene evidence did not reliably modulate any of the other regressors of interest (all $p > 0.14$; Figure S5).
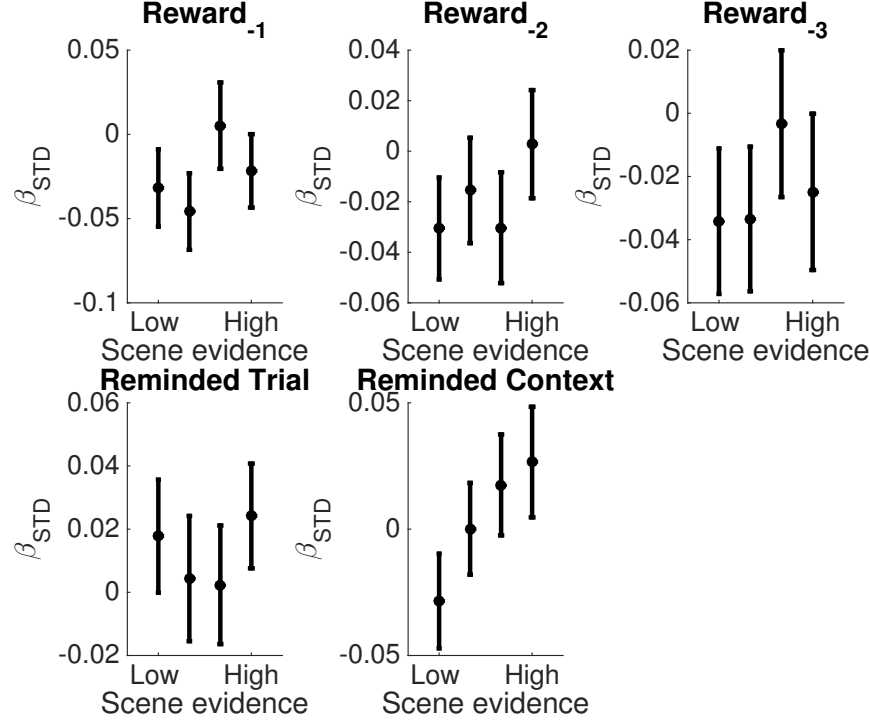
Figure S5: Each regressor of interest plotted as a function of scene reinstatement evidence.

Hippocampal activity during decision-making

The hippocampus has long been understood to be critical to episodic memory encoding and retrieval. We investigated whether our regression weights were correlated with univariate activity levels, or classifier evidence for scenes, in this structure (Figure S6).

First, using the FreeSurfer atlas, we defined for each subject an anatomical mask of bilateral hippocampus proper.

Next, at each correct, valid probe trial, we took the average activity level in hippocampus across our timepoints of interest (Figure 6b of the main text). We then divided trials into quartiles based on this measure, and recomputed the regression in Figure 6c of the main text for each bin. We repeated the regression analysis within each of these quartiles.

While the relationship between hippocampal activity levels and single-trial reward effect was numerically negative (average slope $= -0.0231$ SEM 0.0288, $p = 0.4444$), and the context reward effect was numerically positive (average slope $= 0.0277$ SEM 0.0285, $p = 0.3422$), neither were statistically significant, not was there an interaction between the effects ($p = 0.2928$).

We next tested for a relationship between scene reinstatement evidence in hippocampus and the regression effects. For this analysis, we trained a scene classifier on multi-voxel patterns in the hippocampus, following a procedure analogous to that used for PPA in the

main text. We then divided trials into quartiles based on this measure, and recomputed the regression analysis (Figure 6c of the main text) for each quartile. We did not observe a reliable relationship between hippocampal scene evidence and the effect of either reminded trial or reminded context (both $p > 0.12$).

These results show that, as a practical matter, it is easier to detect these scene context effects in PPA than in hippocampus; this does not mean that hippocampus is uninvolved (we do not want to over-interpret a null result) but it does fit with other work showing that applying a classifier to PPA activity patterns is a reliable way of tracking scene recollection [3–5].
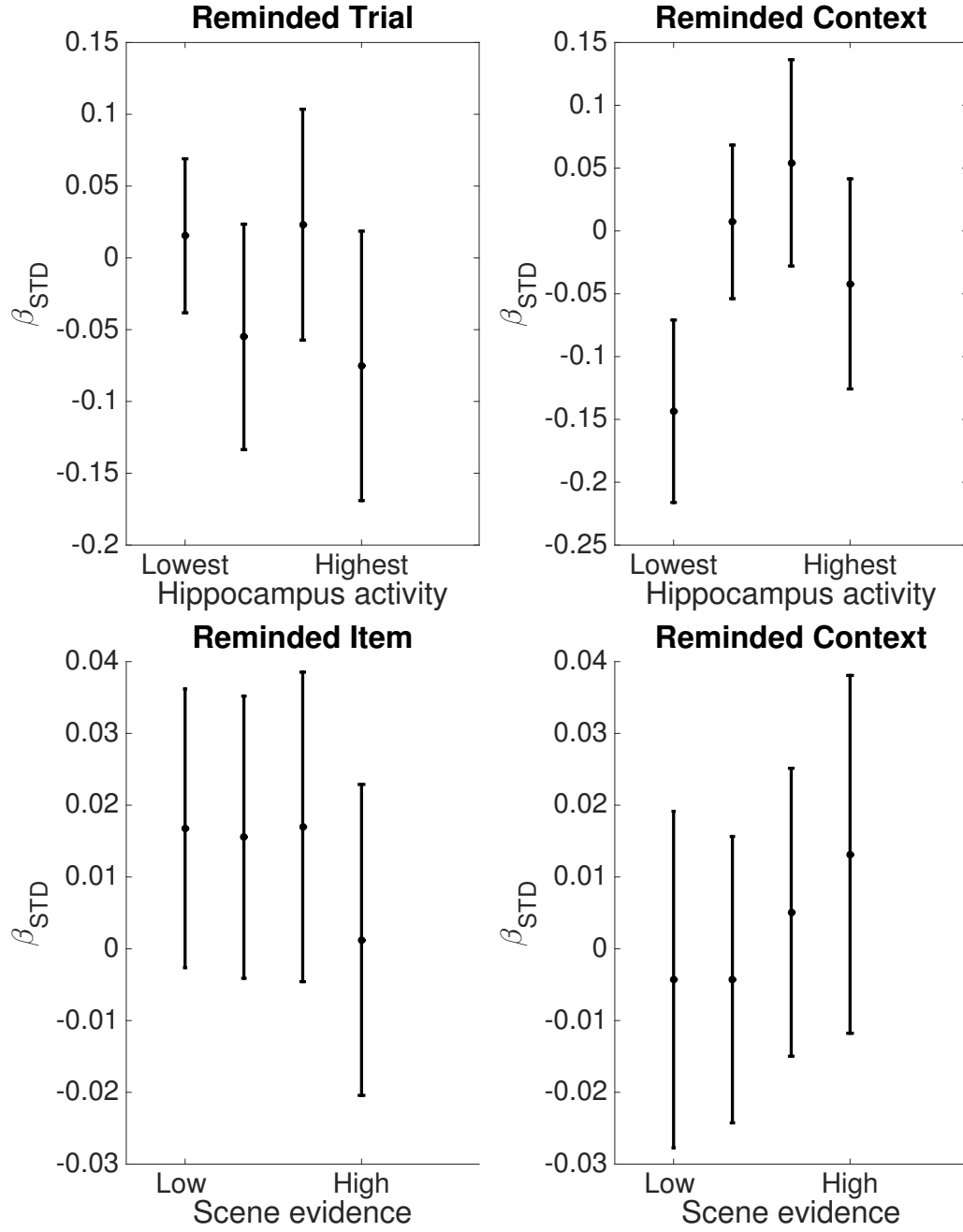
Figure S6: The effect of reminded trial and of reminded context reward as a function of univariate activity and scene evidence in hippocampus.

References

[1] Timothy E J Behrens, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9):1214–21, sep 2007. ISSN 1097-6256. doi: 10.1038/nn1954. URL http://www.ncbi.nlm.nih.gov/pubmed/17676057.

[2] Gideon Schwarz. Estimating the Dimension of a Model. *Annals of Statistics*, 6(2):461–464, 1978.

[3] Samuel J Gershman, Anna C Schapiro, Almut Hupbach, and Kenneth A Norman. Neural context reinstatement predicts memory misattribution. *The Journal of Neuroscience*, 33 (20):8590–5, may 2013. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.0096-13.2013. URL http://www.ncbi.nlm.nih.gov/pubmed/23678104.

[4] Greg J. Detre, Annamalai Natarajan, Samuel J. Gershman, and Kenneth A. Norman. Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia*, 51(12):2371–2388, 2013. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2013.02.017. URL http://dx.doi.org/10.1016/j.neuropsychologia.2013.02.017.

[5] Jeremy R Manning, Justin C Hulbert, Jamal Williams, Luis Piloto, Lili Sahakyan, and Kenneth A Norman. A neural signature of contextually mediated intentional forgetting. *Psychonomic Bulletin & Review*, 2016. ISSN 1531-5320. doi: 10.3758/s13423-016-1024-7. URL http://www.ncbi.nlm.nih.gov/pubmed/27150815.