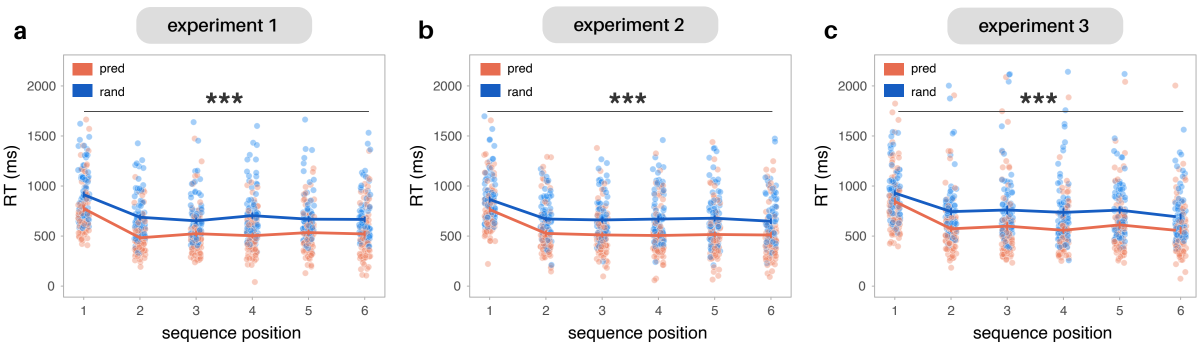
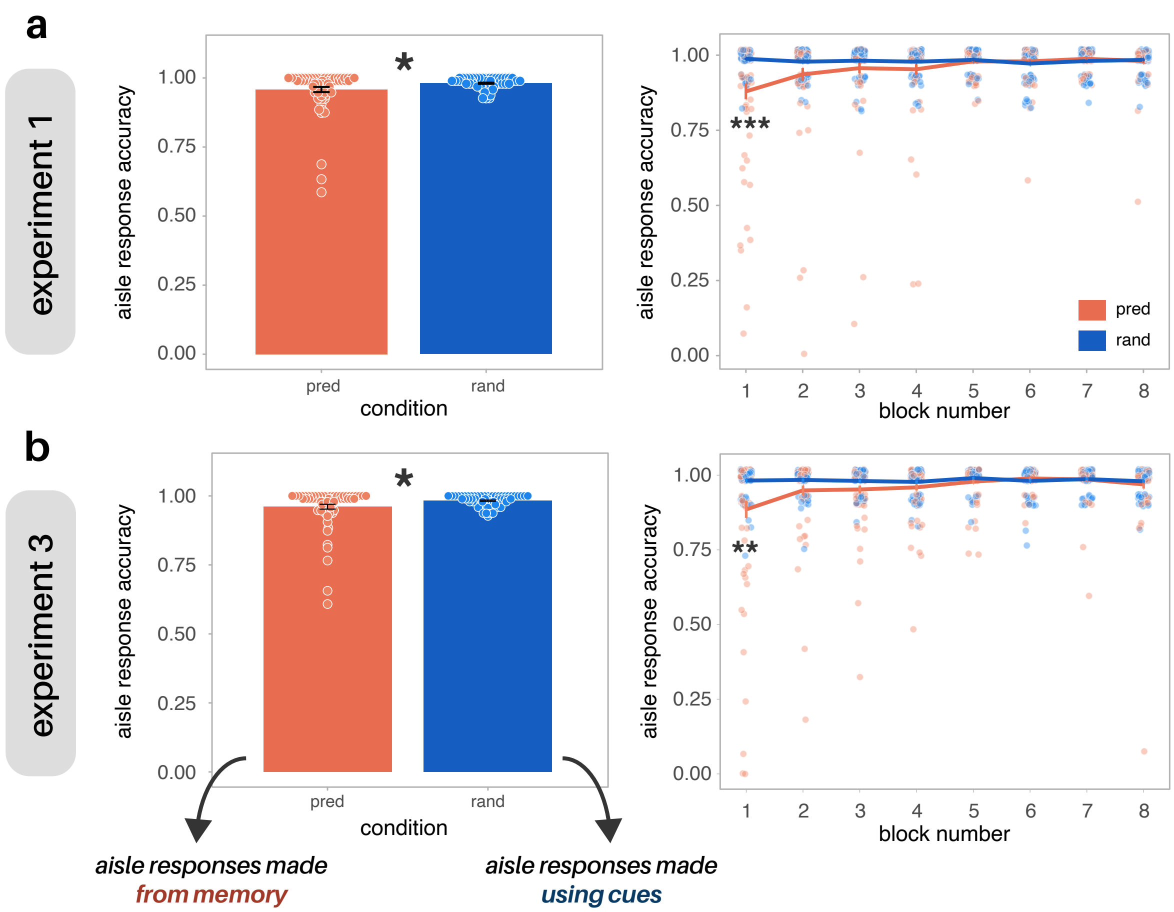
Supplementary Materials

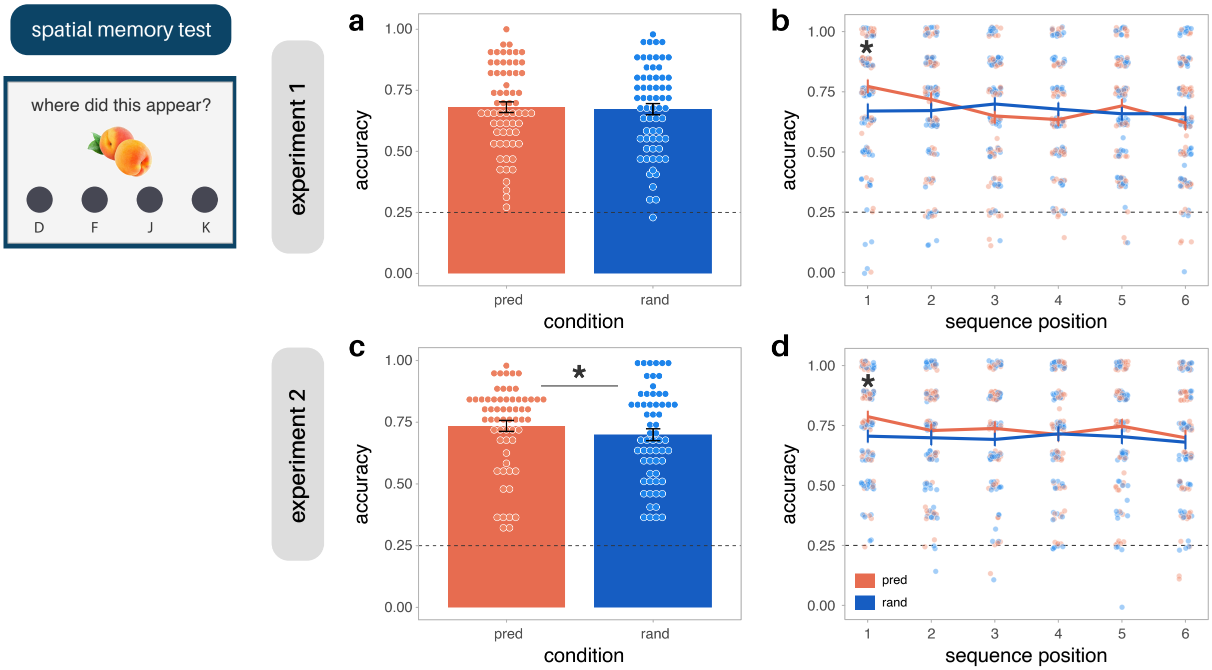
Supplementary Figures



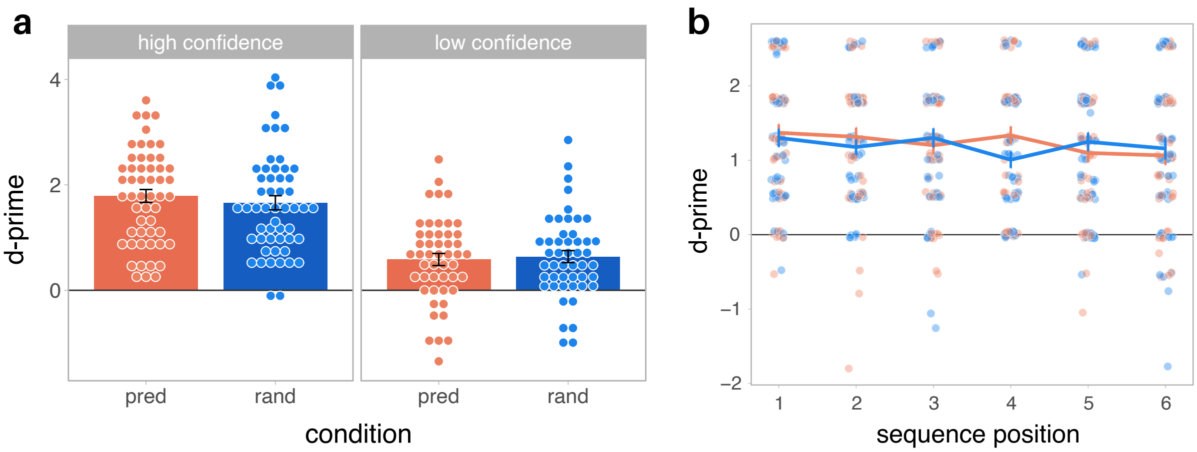
Supplementary Fig. 1 | Reaction times (RTs) to aisle cues during the intermixed pretraining study block. Results from Experiment 1 (a), Experiment 2 (b), and Experiment 3 (c). In all three experiments, participants were consistently faster to respond to aisle cues across all sequence positions in the predictable versus random store. Statistical significance was assessed using Wilcoxon signed-rank tests, FDR-corrected for multiple comparisons. Error bars indicate SEM, and dots represent individual participants. *** $p < 0.001$.



Supplementary Fig. 2 | Accuracy of participants' aisle responses during encoding. Data from Experiment 1 (a) and 3 (b). Left-most plots show the difference in accuracy averaged across all blocks. Right-most plots show accuracy in both conditions as a function of block (with significance values FDR-corrected for multiple comparisons), revealing that performance in these conditions becomes comparable early on in the experiment. Error bars indicate SEM, and dots represent individual participants. $*p < 0.05$, $** p < 0.01$, $*** p < 0.001$.



Supplementary Fig. 3 | Spatial memory test performance. Data from Experiments 1 (top) and 2 (bottom). (a) and (b) show mean accuracy as a function of store condition. (b) and (d) show accuracy in each condition as a function of the sequence position in which an item was encountered during encoding (significance values are FDR-corrected for multiple comparisons). Although both experiments show that participants were more likely to remember the aisle/spatial location of the first item in a predictable versus random event, only Experiment 2 showed a significant effect of store condition on overall spatial memory performance. Error bars indicate SEM, and dots represent individual participants. Dotted lines indicate chance performance (1/4 = 0.25). *p < 0.05.



Supplementary Fig. 4 | Item recognition performance in Experiment 3. D-prime scores broken down by confidence (a) and sequence position (b). (a) In both high and low confidence bins, participants showed no difference in their ability to discriminate between old and new items in the predictable versus random condition. (b) Participants' d-prime scores also did not differ across conditions regardless of an item's sequence position (i.e., when it was encountered within an event/errand). Error bars indicate SEM, and dots represent individual participants.

Supplementary Tables

Supplementary Table 1. Experiment 1: Results of a Bayesian, multilevel logistic regression model examining the effect of condition on order reconstruction memory after controlling for confounds.

	Estimate	*Est. Error*	*l-95% CI*	*u-95% CI*
Fixed Effects				
Intercept	0.478	0.121	0.245	0.717
condition	0.541	0.111	0.322	0.759
seq position	-0.252	0.024	-0.298	-0.206
event position	-0.088	0.040	-0.168	-0.009
block	-0.007	0.025	-0.057	0.044
stim cat	0.352	0.113	0.130	0.571
Random Effects				
sd(Intercept)	0.939	0.097	0.769	1.150
sd(condition)	0.456	0.237	0.027	0.848
sd(seq position)	0.120	0.029	0.064	0.178
sd(event position)	0.222	0.045	0.137	0.314
sd(block)	0.169	0.023	0.129	0.218
sd(stim cat)	0.473	0.235	0.032	0.854

Supplementary Table 2. Experiment 2: Results of a Bayesian, multilevel logistic regression model examining the effect of condition on order reconstruction memory after controlling for confounds.

	Estimate	*Est. Error*	*l-95% CI*	*u-95% CI*
Fixed Effects				
Intercept	0.524	0.141	0.250	0.806
condition	0.283	0.115	0.059	0.509
seq position	-0.253	0.026	-0.305	-0.202
event position	-0.078	0.043	-0.163	0.008
block	-0.060	0.024	-0.107	-0.014
stim cat	0.203	0.115	-0.025	0.428
Random Effects				
sd(Intercept)	1.059	0.111	0.864	1.296
sd(condition)	0.469	0.240	0.027	0.865
sd(seq position)	0.141	0.027	0.091	0.197
sd(event position)	0.247	0.043	0.166	0.337
sd(block)	0.142	0.022	0.101	0.188
sd(stim cat)	0.480	0.240	0.029	0.872

Supplementary Table 3. Experiment 3: Results of a Bayesian, multilevel logistic regression model examining the effect of condition on order reconstruction memory after controlling for confounds.

	Estimate	*Est. Error*	*l-95% CI*	*u-95% CI*
Fixed Effects				
Intercept	1.080	0.149	0.790	1.379
condition	0.291	0.124	0.048	0.536
seq position	-0.216	0.024	-0.263	-0.169
event position	-0.090	0.044	-0.175	-0.003
block	0.045	0.029	-0.013	0.103
stim cat	0.165	0.124	-0.080	0.410
Random Effects				
sd(Intercept)	1.154	0.121	0.941	1.414
sd(condition)	0.521	0.267	0.031	0.962
sd(seq position)	0.106	0.031	0.041	0.164
sd(event position)	0.247	0.046	0.160	0.343
sd(block)	0.200	0.026	0.154	0.256
sd(stim cat)	0.538	0.267	0.034	0.977

Supplementary Table 4. Comparison of order reconstruction memory and spatial memory effects across experiments. Statistics come from independent two-sample t-tests; assumption of equal variance verified using F-tests [2]. *CI* = confidence interval. *d* = Cohen's d.

	t	*df*	*p*	*95% CI*	*d*
Order memory					
Exp. 1 vs. Exp. 2	1.175	122	0.242	[-0.024, 0.094]	0.211
Exp. 1 vs. Exp. 3	1.522	125	0.131	[-0.013, 0.099]	0.270
Exp. 2 vs. Exp. 3	0.290	123	0.773	[-0.048, 0.065]	0.052
Spatial memory					
Exp. 1 vs. Exp. 2	-1.348	122	0.180	[-0.083, 0.016]	0.242

Supplementary Methods

Bayesian multilevel regression models

When exploring the effect of event condition (predictable versus random) on temporal order memory, we also accounted for the influence of other relevant variables likely to interact with memory performance: stimulus category (animals vs. foods), block number (1-8), event position within a list (1-4), and sequence position within an event (1-6). Specifically, we ran a series of Bayesian multilevel logistic regression models (one per experiment), in which binary accuracy on each memory test trial was predicted by condition plus all confound variables. Categorical predictors were effect-coded and continuous predictors were mean-centered prior to analysis. All models included a random intercept for each subject, as well as both fixed and random slopes for all predictor variables. Models were run using the *brms* package (ver. 2.10.0; [1]) in R. All models included uninformative priors for all parameters and were fit using 4 Markov chains, each with 8000 iterations and 1000 warmup iterations.

We used Bayesian regression for this control analysis because it more readily allowed us to estimate complex multilevel models with numerous random effects (whereas comparable Frequentist models produced convergence errors). Critically, when using uninformative priors (as we do here), Frequentist and Bayesian approaches can be treated as conceptually analogous. Although Bayesian models do not yield traditional p-values, we can assess the "significance" of all fixed effects by examining whether the 95% credibility interval (95%-CI) of the corresponding beta coefficient contains zero. A lack of such overlap is taken as evidence that the effect is reliably different from zero.

Supplementary Results

Controlling for confounds in the effect of condition on order memory

For each of our three experiments, we ran a Bayesian logistic regression model (see Supplementary Methods) to examine the effect of condition (predictable vs. random) on trial-level accuracy during the order reconstruction test after controlling for a set of confound variables: stimulus category (animals vs. foods), block number (1-8), event position within a list (1-4), and sequence position within an event (1-6). In each of these three Bayesian models (one for each experiment), controlling for these variables did not change our finding of enhanced order memory for items seen in the predictable versus random store. That is, the effect of condition on memory remained robust (i.e., credibility intervals did not contain zero) in Experiment 1 (b_{cond} = 0.54, *SE* = 0.11, *95%-CI:* [0.32, 0.76]), Experiment 2 (b_{cond} = 0.28, *SE* = 0.11, *95%-CI:* [0.06, 0.51]), and Experiment 3 (b_{cond} = 0.29, *SE* = 0.12, *95%-CI:* [0.05, 0.54]). Full model outputs can be found in Supplementary Tables 1-3.

Accounting for aisle response accuracy during encoding when examining order memory

In Experiments 1 and 3, participants were required to use their memory for the predictable store's aisle sequence to guide their behavior during encoding. In the main text, we showed that participants were generally very accurate in making these responses, reaching a level of performance that was comparable to the random condition (in which cues were provided) after the first block. We also excluded all participants who did not make predictable aisle response with accuracy > 80% from any memory test analyses, ensuring that all participants in our order memory analyses had demonstrated sufficient learning of the predictable action sequence (see Methods). Nevertheless, in relatively rare cases where participants failed to execute the correct aisle response, we cannot be sure that they had access to a representation of the learned action sequence while encountering novel items. To address this concern, we re-examined the difference in order memory (i.e., ordinal accuracy) across conditions including only items in which participants had executed the correct aisle response (prior to viewing the item) during encoding. We found that order memory remained significantly greater for items from the predictable versus random store in both Experiment 1 ($t(62) = 4.55$, $p < 0.001$, $d = 0.49$, [CI: 0.05, 0.14]) and Experiment 3 ($t(64) = 2.67$, $p = 0.010$, $d = 0.27$, [CI: 0.01, 0.09]).

Spatial memory performance and aisle location

In Experiments 1 and 2, we examined how spatial memory performance differed as a function of aisle location. In Experiment 1, a repeated measures ANOVA revealed that participants' spatial memory performance varied across locations ($F(3, 186) = 6.67$, $p < 0.001$, partial $\eta^2 = 0.10$), with memory for items at the leftmost edge of the screen (aisle 1) being better than all other aisles (aisles 2-4) (aisle 1 vs. 2: $t(62) = 2.89$, $p_{FDR} = 0.011$, $d = 0.30$, [CI_{FCR}: 0.01, 0.10]; aisle 1 vs. 3: $t(62) = 3.69$, $p_{FDR} = 0.001$, $d = 0.35$, [CI_{FCR}: 0.02, 0.11]; aisle 1 vs. 4: $t(62)$, $p_{FDR} = 0.001$, $d = 0.32$, [CI_{FCR}: 0.02, 0.10]). A similar effect was present in Experiment 2 ($F(2.45, 149.97) = 13.868$, $p < 0.001$, partial $\eta^2 = 0.19$; Greenhouse-Geisser correction for non-sphericity applied), such that memory was again best for items that appeared in the leftmost aisle (aisle 1 vs. 2: $t(60) = 5.13$, $p_{FDR} < 0.001$, $d = 0.45$, [CI_{FCR}: 0.05, 0.11]; aisle 1 vs. 3: $t(60) = 5.73$, $p_{FDR} < 0.001$, $d = 0.51$, [CI_{FCR}: 0.06, 0.12]; aisle 1 vs. 4: $t(60) = 3.49$, $p_{FDR} = 0.002$, $d = 0.21$, [CI_{FCR}: 0.02, 0.07]). In this sample, spatial memory for the items in the rightmost aisle was also better than those in the middle aisles (aisle 4 vs. 2: $t(60) = 2.11$, $p_{FDR} = 0.047$, $d = 0.18$, [CI_{FCR}: 0.0005, 0.07]; aisle 4 vs. 3: $t(60) = 2.61$, $p_{FDR} = 0.017$, $d = 0.23$, [CI_{FCR}: 0.01, 0.08]). These effects complicate any interpretation of our finding that participants' spatial memory for the first item in an event was enhanced in the predictable versus random store, given that the location of these first items was not balanced across conditions.

Reminder task performance

In Experiments 1 and 3, participants were required to make aisle responses in the predictable store from memory while seeing cues in the random store. Because of this memory demand, we included an explicit reminder of the predictable sequence halfway through the experiment (i.e., after the fourth retrieval block; Fig. 1a). Although the high accuracy of aisle responses throughout encoding (Supplementary Fig. 2) clearly indicates that participants retained memory for the action sequence, we can further confirm this conclusion by checking behavior on the reminder task. Participants began this task by passively observing the learned aisle sequence twice, before then recreating it from memory twice during two consecutive test trials. As expected, performance during these test trials was near ceiling in both Experiment 1 ($M = 0.99$, $SD = 0.05$) and Experiment 3 ($M = 0.99$, $SD = 0.04$).

Supplementary References

1. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
2. Snedecor, G. W. & Cochran, W. G. *Statistical Methods*. (Iowa State University Press, 1989).