Supplementary Methods

Pilot cohort:

Participants
For the initial study, we recruited 125 healthy younger adults using the Washington University in St. Louis SONA research pool. We excluded 24 participants for responding to fewer than 80% of the trials in the decision-making task. We also excluded eight participants because of missing data in their behavioral representation similarity task, rendering certain pairwise comparisons impossible. This left us with an effective sample of 93 younger adults (47 Male, 45 Female, 1 Non-Binary, age range: 18-23 years, mean age: 19.3 years).

Decision-making task
The decision-making task in the pilot study was the same as the experiment described in the main text, except that in the pilot task the number of unique background images per context was 16. In the replication cohort this was shifted to 4 images per context in order to make the memory portion of the task easier. There were no other changes.

Simulation:

Recovery analysis of model parameters
Recovery analysis was performed using simulated agents to ascertain whether the model fitting could recover ground truth parameters. We used a generative version of our model to simulate the behavior of 500 agents. For each of these agents we sampled the true parameters randomly from uniform distributions $\{\alpha, \lambda, \eta, \kappa, w\} \sim U(0,1)$, $\beta \sim U(0,2)$, $\{\pi, \rho\} \sim U(-0.5,0.5)$. Next we used our model-fitting procedure (as described in **Methods: Reinforcement learning model**) to obtain estimated parameters for each simulated agent from its choice behavior. We found correlations between the true and estimated parameters for the model-based weight ($r_{(498)}$ = 0.62, 95% CI = [0.57 0.67], p < 0.001). We also found the following correlations for the other parameters:

Supplementary Table 1: Parameter recovery

	$r_{(498)}$	95% CI	p
learning rate: α	0.61	[0.56, 0.67]	< 0.001
inverse temperature: β	0.54	[0.47, 0.60]	< 0.001
trace decay: λ	0.50	[0.43, 0.56]	< 0.001
stickiness: π	0.70	[0.66, 0.74]	< 0.001
response stickiness: ρ	0.24	[0.15, 0.32]	< 0.001
transition learning rate: η	0.08	[-0.01, 0.16]	0.085

counterfactual transitions: κ	0.02	[-0.07, 0.1]	0.724

Note the inability to recover the η and κ parameters. This is potentially due to the fact that those parameters are only affecting very few trials at the onset of the task and even small values will converge quickly to the true transition matrix. To ascertain whether our inclusion of those parameters altered any of our other parameters we ran a model where both η and κ were hardcoded to 1 (indicating immediate update of transition probabilities). We then assessed the correlations between the other parameters in the model with η and κ as free parameters and the model with η and κ as hardcoded. The results reported in the main paper do not differ when analyzed with the hardcoded model (correlations among all other parameters $r_{(498)}$ = 0.99, CI 95% [0.99, 1.0], p < 0.001).

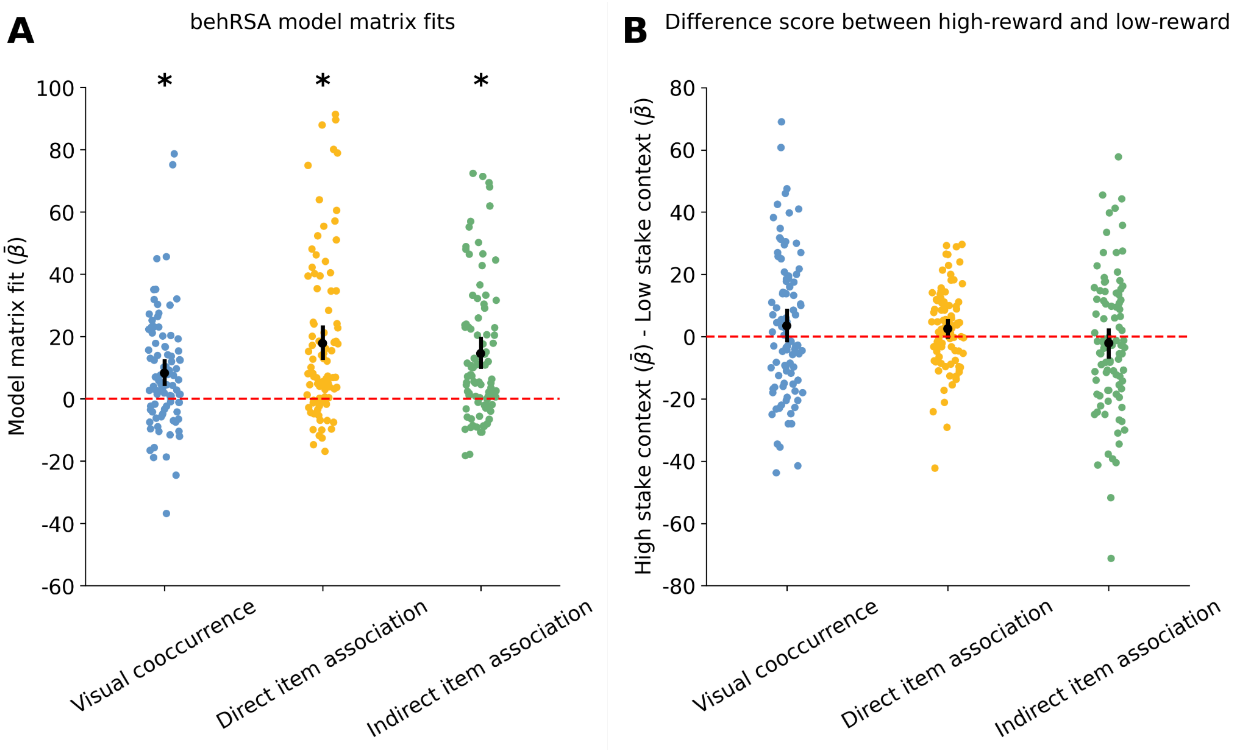
Supplementary Results

Pilot experiment

Participants (n = 93) performed the same relatedness ratings task and two-stage decision-making task as in the main experiment.

Pilot behRSA model fits

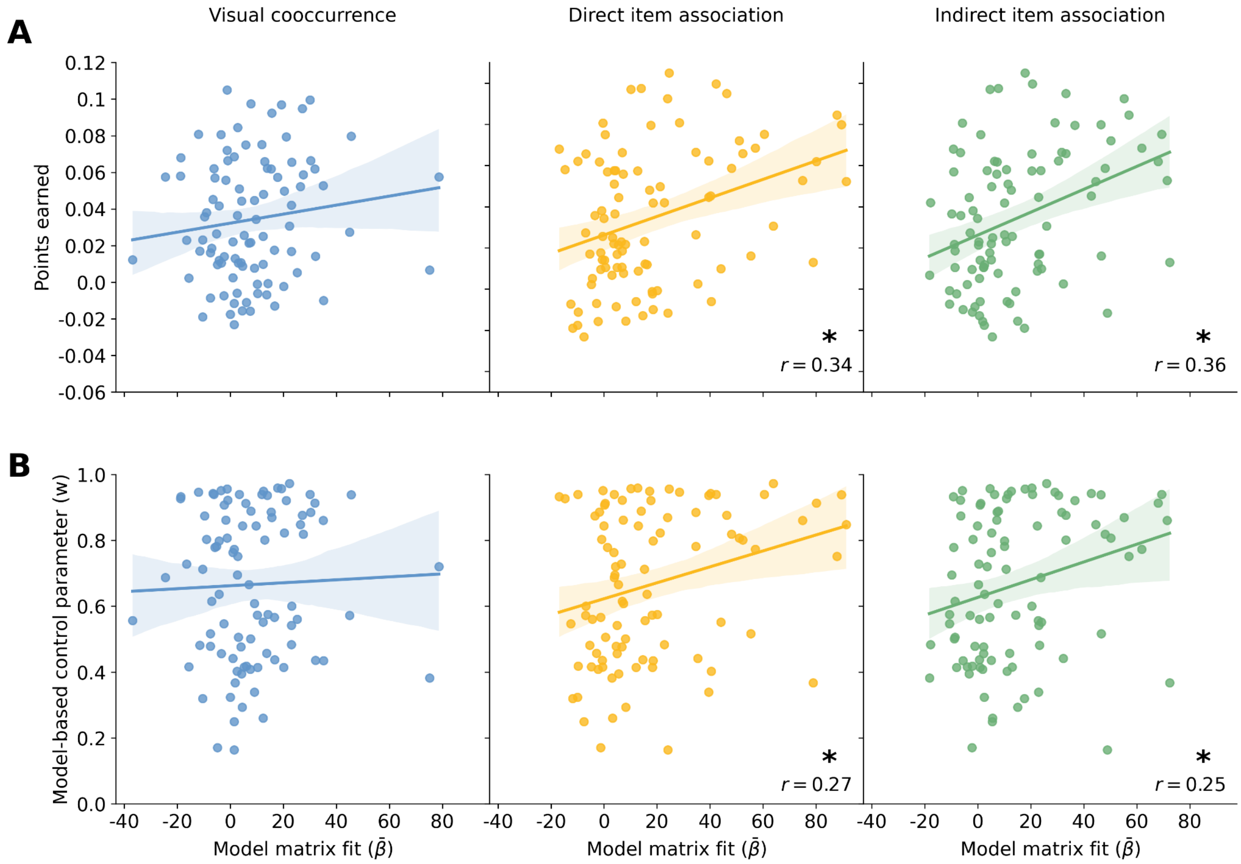
Using the same analysis as described in the main text, we found that each of the hypothesized models was represented in the group (Supplementary Figure 1A). In comparison to the sample reported in the main paper we did not find an effect of high-stake or low-stake contexts on the participants’ representations (Supplementary Figure 1B).



Supplementary Figure 1: Model matrix fits for pilot sample. **A.** Participants' model matrix fits for the three hypothesized models, fit the same way as the primary results. All three of the models are represented within the pilot sample. **B.** Participants on average show no increased effect of cognitive-map-based abstraction for the high-stake vs low-stake context (* represents p < 0.05, error bars are 95% CI)

Pilot decision-making data

Consistent with the main experiment, we observed a correlation between the strength of the direct item association and points earned in the task ($r_{(91)}$ = 0.34, 95% CI = [0.15, 0.51], p < 0.001) and between the strength of the indirect item association and points earned in the task ($r_{(91)}$ = 0.32, 95% CI = [0.17, 0.53], p < 0.001) . There was again no relationship between visual cooccurrence fit and points earned ($r_{(91)}$ = 0.13, 95% CI = [-0.07, 0.33], p = 0.2). After fitting the same reinforcement learning model (see **Methods: Reinforcement learning model**), we found a correlation between model-based control and the strength of the direct item association ($r_{(91)}$ = 0.27, 95% CI = [0.07, 0.45], p = 0.009) and the strength of the indirect item association ($r_{(91)}$ = 0.25, 95% CI = [0.05, 0.43], p = 0.014) We found no correlation between visual cooccurrence and model-based control ($r_{(91)}$ = 0.04, 95% CI = [-0.17, 0.24], p = 0.735).



Supplementary Figure 2: Model-based representations correlate with the decision-making task and reinforcement learning model. **A.** Comparison of behRSA representations of subjects with their performance in the decision-making task. **B.** Reconstruction of task-relevant representations in behRSA also correlate with increased use of model-based control, whereas visual cooccurrence does not. (* represents $p < 0.05$)

Main experiment

Points earned, RT, and w parameters

In line with previous work (Kool et al 2016, 2017) we observed a correlation between the points earned in the task and the model-based weighting parameter w (r = 0.6876, CI = [0.6, 0.76], p < 0.001). This is due to the task being explicitly designed to reward use of model-based control. We further observed correlations for both points earned and participant response time in the first-stage state (r = 0.3919, CI = [0.25, 0.52], p < 0.001) and the model-based weighting parameter and response time (r = 0.3186, CI = [0.17, 0.45], p < 0.001).

ANOVA for 4 w parameter model:

Based on previous work exhibiting an effect of stakes on model-based control[19] we tested whether there was an equivalent stake effect in our experiment. Therefore, we performed a repeated-measures ANOVA to test whether participants exerted more model-based control on the more commonly rewarded first-stage states regardless of the stake multiplier, as well as whether there was an interaction between trial stakes and first-stage state. Replicating prior work, we found that more model-based control was exhibited on high-stake (mean w_{high} = 0.56) compared to low-stakes trials (w_{low} = 0.54) ($F_{(1,160)}$= 7.265, p = 0.008) . We found no effect of first-stage reward context, and no interaction effect (supplementary table 2: Repeated-measures ANOVA).

Supplementary Table 2:
Repeated-measures ANOVA (STAKE x ARM)

	SS	Ddof1	Ddof2	MS	*F*	*p*	partial η^2
arm	0.015	1	160	0.015	1.346	0.248	0.008
stakes	0.091	1	160	0.091	7.265	0.008	0.043
arm * stakes	0.024	1	160	0.024	1.286	0.258	0.008

d' correlations with model-based control parameters

We reasoned that the memory performance in the surprise memory probe would be linked to model-based control but found no significant correlations between the model-based weighting parameter w and any of the *d'* measures. Correlation between w and $d'_{mismatch}$ high-arm (r = -0.01851, CI = [-0.17, 0.14], p = 0.8157). Correlation between *w* fit and $d'_{mismatch}$ low-arm (r = 0.01812, CI = [-0.14, 0.17], p = 0.8196). Correlation between w fit and d'_{lure} high-arm (r = 0.09555, CI = [-0.06, 0.25], p = 0.2280). Correlation between w fit and d'_{lure} low-arm (r = 0.02351, CI = [-0.13, 0.18], p = 0.7672). These results suggest that memory performance was not in fact modulated by model-based control.